(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 994 064**

(51) Int. Cl.:

C12N 15/113 (2010.01)
C12N 15/10 (2006.01)
C12N 9/22 (2006.01)
C12N 15/82 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **17.06.2016 E 19171732 (1)**

(97) Fecha y número de publicación de la concesión europea: **01.05.2024 EP 3604532**

(54) Título: **Nuevas enzimas y sistemas CRISPR**

(30) Prioridad:

18.06.2015 US 201562181739 P
16.07.2015 US 201562193507 P
05.08.2015 US 201562201542 P
16.08.2015 US 201562205733 P
24.09.2015 US 201562232067 P
18.12.2015 US 201514975085
07.01.2016 EP 16150428

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
16.01.2025

(73) Titular/es:

THE BROAD INSTITUTE, INC. (33.33%)
415 Main Street
Cambridge, MA 02142, US;
MASSACHUSETTS INSTITUTE OF TECHNOLOGY (33.33%) y
PRESIDENT AND FELLOWS OF HARVARD COLLEGE (33.33%)

(72) Inventor/es:

ZHANG, FENG;
ZETSCHE, BERND;
GOOTENBERG, JONATHAN, S.;
ABUDAYYEH, OMAR, O. y
SLAYMAKER, IAN

(74) Agente/Representante:

LEHMANN NOVO, María Isabel

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Nuevas enzimas y sistemas CRISPR

CAMPO DE LA INVENCIÓN

La presente invención se relaciona en general con sistemas, métodos y composiciones usados en el control de la expresión genética que comprende direccionamiento de secuencias, tal como perturbación de transcriptos de genes o edición de ácidos nucleicos, que pueden emplear sistemas de vectores relacionados con los grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR) y componentes de las mismas.

ANTECEDENTES DE LA INVENCIÓN

Los avances recientes en las técnicas de secuenciación genómica y en los métodos de análisis han acelerado significativamente la capacidad para clasificar y mapear los factores genéticos asociados con una amplia gama de funciones biológicas y enfermedades. Se necesitan tecnologías de direccionamiento a genomas precisas para posibilitar una ingeniería inversa sistemática de las variaciones genéticas causales mediante una perturbación selectiva de elementos genéticos individuales, así como para lograr avances en las aplicaciones biotecnológicas y médicas y en la biología sintética. Si bien se cuenta con técnicas de edición de genomas, tales como dedos de zinc de diseño, efectores tipo activadores de la transcripción (TALE) o meganucleasas dirigidas [*homing*], para producir perturbaciones dirigidas al genoma, persiste la necesidad de nuevas tecnologías de manipulación de genomas que empleen nuevas estrategias y mecanismos moleculares y que sean factibles, fáciles de configurar, se puedan llevar a escala y pasibles de un direccionamiento a múltiples posiciones dentro del genoma eucariota. Esto puede proporcionar un recurso importante para aplicaciones nuevas en la ingeniería genética de genomas y la biotecnología.

Los sistemas CRISPR-Cas de inmunidad adaptativa bacteriana y de Archaea presentan una amplia diversidad en cuanto a composición de proteínas y arquitectura genómica de loci. Los loci del sistema CRISPR-Cas comprenden más de 50 familias de genes y no hay genes estrictamente universales indicativos de una evolución rápida y una diversidad extrema de la arquitectura de los loci. Hasta ahora, se ha logrado una identificación general de los genes *cas* de 395 perfiles para 93 proteínas Cas, adoptando un abordaje multidimensional. La clasificación incluye perfiles de patrones de genes además de patrones de la arquitectura del locus. Se propone una nueva clasificación de los sistemas CRISPR-Cas según la cual estos sistemas se dividen en general en dos clases, la Clase 1, con complejos efectores de múltiples subunidades, y la Clase 2, con módulos efectores de una sola subunidad, ésta última ejemplificada por la proteína Cas9. Se pueden desarrollar nuevas proteínas efectoras asociadas con los sistemas CRISPR-Cas de Clase 2 como herramientas poderosas para la manipulación del genoma mediante ingeniería genética y para la predicción de nuevas proteínas efectoras putativas y es importante su ingeniería y optimización.

La cita o identificación de cualquier documento en esta solicitud no es un reconocimiento que dicho documento constituya un tema del arte anterior para la presente invención.

SUMARIO DE LA INVENCIÓN

Existe una necesidad apremiante de sistemas y técnicas alternativas y robustas para el direccionamiento de ácidos nucleicos o polinucleótidos (por ejemplo, ADN o ARN o cualquier híbrido o derivado de los mismos) con una gran variedad de aplicaciones. Esta invención está dirigida a esta necesidad y ofrece ventajas relacionadas. La incorporación de nuevos sistemas de direccionamiento de ADN o de ARN de la presente solicitud en el repertorio de tecnologías de direccionamiento genómico y epigenómico puede transformar el estudio y la perturbación o edición de sitios diana específico por medio de una detección, análisis y manipulación directos. Para poder utilizar eficazmente los sistemas de direccionamiento de ADN o de ARN de la presente solicitud en el direccionamiento genómico o epigenómico sin efectos perjudiciales, resulta crítico comprender los aspectos de la ingeniería y optimización de estas herramientas de direccionamiento de ADN o ARN.

La invención proporciona un método *in vitro* o *ex vivo* de direccionamiento de un polinucleótido, que comprende: poner en contacto una muestra que puede comprender el polinucleótido con un complejo de CRISPR-Cas que comprende (a) una proteína Cas de tipo V que comprende un dominio de nucleasa RuvC pero que no comprende un dominio HNH y (b) una guía manipulada que es capaz de dirigir la unión específica de secuencia del complejo a una secuencia diana del polinucleótido, dirigiéndose así o detectando dicho polinucleótido. Similarmente, la invención proporciona un método de direccionamiento de un polinucleótido, que comprende: poner en contacto una muestra que puede comprender el polinucleótido con un complejo de CRISPR-Cas que comprende (a) una proteína Cas de tipo V que comprende un dominio de nucleasa RuvC pero que no comprende un dominio HNH y (b) una guía manipulada que es capaz de dirigir la unión específica de secuencia del complejo a una secuencia diana del polinucleótido, dirigiéndose así al polinucleótido, en donde el método no es un método de modificación de la identidad genética de la línea germinal de seres humanos. En una forma de realización preferida, el método comprende además la introducción de una rotura de cadena, inactivación génica específica del sitio, edición de base específica del sitio o edición de genoma específico del sitio. En realizaciones particulares, se modifica un producto génico o se modifica la cantidad o expresión de un producto génico. En realizaciones particulares, el método comprende además detectar la unión del complejo a la secuencia diana. En realizaciones particulares, la proteína Cas está asociada con uno o más dominios funcionales o comprende al menos una mutación. En los métodos de la invención, la secuencia diana puede ser ARN o ADN. En

realizaciones particulares, la secuencia diana está asociada a la enfermedad.

La solicitud proporciona además la composición que comprende: una proteína Cas de tipo V que comprende un dominio de nucleasa RuvC pero que no comprende un dominio HNH o un polinucleótido que codifica la proteína, y un ARN guía manipulado que es capaz de formar un complejo CRISPR-Cas con la proteína Cas y dirigir la unión específica de secuencia del complejo CRISPR-Cas a una secuencia diana o un polinucleótido que codifica el ARN guía. En realizaciones particulares, la proteína Cas está asociada con uno o más dominios funcionales o al menos una mutación, y para su uso en una célula eucariota.

La invención proporciona además un método *in vitro* o *ex vivo* de direccionamiento de un polinucleótido, que comprende: poner en contacto una muestra que puede comprender el polinucleótido con la composición de la invención, en donde el ARN guía manipulado es capaz de formar un complejo CRISPR-Cas con la proteína Cas y dirigir la unión específica de secuencia del complejo CRISPR-Cas a una secuencia diana del polinucleótido. En realizaciones particulares de estos métodos, la proteína Cas está asociada con uno o más dominios funcionales o al menos una mutación, y para su uso en una célula eucariota. La invención proporciona similarmente un método de direccionamiento de un polinucleótido, que comprende: poner en contacto una muestra que puede comprender el polinucleótido con la composición de la invención, en donde el ARN guía manipulado es capaz de formar un complejo CRISPR-Cas con la proteína Cas y dirigir la unión específica de secuencia del complejo CRISPR-Cas a una secuencia diana del polinucleótido, en donde el método no es un método de modificación de la identidad genética de la línea germinal de seres humanos.

Se podrá apreciar que los términos enzima Cas, enzima CRISPR, proteína CRISPR proteína Cas y CRISPR Cas en general se usan indistintamente y todos los puntos de referencia en la presente se refieren por analogía a las nuevas proteínas efectoras CRISPR que también se describen en esta solicitud, a menos que sea evidente de otra manera, tal como por referencia específica a Cas9. Las proteínas efectoras CRISPR que se describen en la presente son preferiblemente proteínas efectoras Cpf1.

Aspectos de la invención se relacionan con composiciones tal como se definen en la presente, en donde, en realizaciones particulares la proteína efectora de tipo V es Cpf1, teniendo los complejos de proteína efectora Cpf1 uno o más componentes de ácidos nucleicos no naturales o manipulados o modificados u optimizados. En una forma de realización preferida, el componente de ácido nucleico del complejo puede comprender una secuencia guía unida a una secuencia de repetición directa, en donde dicha secuencia de repetición directa comprende uno o más bucles-tallos o estructuras secundarias optimizadas. En una forma de realización preferida, la repetición directa tiene una longitud mínima de 16 nts y un solo bucle-tallo. En formas de realización adicionales, la repetición directa tiene una longitud mayor que 16 nts, preferiblemente de más de 17 nts, y tiene más de un bucle-tallo o estructuras secundarias optimizadas. En una forma de realización preferida, la repetición directa se puede modificar para comprender uno o más aptámeros de ARN de unión a proteínas. En una forma de realización preferida, se puede incluir uno o más aptámeros, tal como formando parte de una estructura secundaria optimizada. Dichos aptámeros se pueden unir a una proteína de la envoltura de bacteriófagos. La proteína de la envoltura de bacteriófagos se puede seleccionar del grupo que comprende Qβ, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, φCb5, φCb8r, φCb12r, φCb23r, 7s y PRR1. En una forma de realización preferida, la proteína de la envoltura de bacteriófagos es MS2. La invención también provee un componente de ácido nucleico del complejo de 30 o más, 40 o más o 50 o más nucleótidos de longitud.

En los métodos reivindicados en la presente, el locus diana de interés puede estar comprendido en una molécula de ADN *in vitro*. En una forma de realización preferida, la molécula de ADN es un plásmido.

En los métodos reivindicados en el presente documento, el locus diana que comprende la secuencia diana de interés puede estar comprendido en una molécula de ADN dentro de una célula.

La célula puede ser una célula procariota o una célula eucariota. La célula puede ser una célula de mamífero. La célula de mamífero puede ser de un mamífero no humano, por ejemplo, de primate, bovinos, ovinos, porcinos, caninos, roedores, Leporidae, por ejemplo, una célula de mono, vaca, ovejas, cerdos, perros, conejos, ratas o ratón. La célula puede ser una célula eucariota que no es de mamífero tal como una célula de aves de corral (por ejemplo, de pollo), de peces vertebrados (por ejemplo, de salmón) o mariscos (por ejemplo, ostras, almejas, langostas, camarones). La célula también puede ser una célula vegetal. La célula vegetal puede ser de una monocotiledónea o dicotiledónea o de una planta de cultivos o de granos, tal como de mandioca, maíz, sorgo, soja, sorgo, avena o arroz. La célula vegetal también puede ser de algas, árboles o plantas productoras, de frutas o verduras (por ejemplo, árboles tales como árboles de cítricos, por ejemplo, árboles de naranja, pomelo o limón; durazno o nectarina; árboles de manzano o peras; árboles de frutos secos, tales como árboles de almendras o nueces o pistacho; plantas de sombra nocturna; plantas del género *Brassica*; plantas del género *Lactuca*; plantas del género *Spinacia*; plantas del género *Capsicum*; algodón, tabaco, espárrago, zanahoria, repollo, brócoli, coliflor, tomate, berenjena, pimiento, lechuga, espinaca, frutilla, arándano azul, frambuesa, zarzamora, uvas, café, cacao, etc.).

En los métodos descritos reivindicados en la presente, el locus diana de interés puede ser un locus genómico o epigenómico de interés. En cualquiera de los métodos descritos, el complejo se puede suministrar con múltiples guías para su uso multiplexado. En cualquiera de los métodos descritos, se puede usar más de una proteína.

En formas de realización preferidas de la invención, el clivaje bioquímico o *in vitro* o *in vivo* de las secuencias asociadas con un locus diana de interés, o presentes en el mismo, resulta en la ausencia de una secuencia del ARNcr transactivador putativo (ARNtracr), por ejemplo, un clivaje mediante una proteína efectora FnCpf1. En otras formas de realización de la invención, el clivaje puede dar como resultado la presencia de una secuencia de ARNcr transactivador putativo (ARNtracr), por ejemplo, un clivaje por otras proteínas efectoras de la familia CRISPR; sin embargo, después de evaluar el locus FnCpfI, los Solicitantes concluyeron que clivaje de un ADN diana mediante un complejo de proteína efectora CpfI no requiere de un ARNtracr. Los Solicitantes determinaron que los complejos de la proteína efectora CpfI que solamente comprenden una proteína efectora CpfI y un ARNcr (ARN guía que comprende una secuencia de repetición directa y una secuencia guía) eran suficientes para clivar el ADN diana. Por lo tanto, la invención provee métodos para modificar un locus diana de interés que se describió previamente en la presente, en donde la proteína efectora es una proteína CpfI y la proteína efectora forma complejos con la secuencia diana sin la presencia de un tracr.

En cualquiera de los métodos descritos, la proteína efectora (es decir, CpfI) y los componentes de ácidos nucleicos tal como se definen previamente se pueden proveer mediante una o más moléculas de polinucleótidos que codifican la proteína y/o uno o más componentes de ácidos nucleicos, y en donde dichas una o más moléculas de polinucleótidos están configuradas operativamente para expresar la proteína y/o dichos uno o más componentes de ácidos nucleicos. Dichas una o más moléculas de polinucleótidos pueden comprender uno o más elementos reguladores configurados operativamente para expresar la proteína y/o dichos uno o más componentes de ácidos nucleicos. Dichas una o más moléculas de polinucleótidos pueden estar comprendidas en uno o más vectores. La invención proporciona dichas moléculas de polinucleótidos, por ejemplo, moléculas de polinucleótidos configuradas operativamente para expresar la proteína y/o los componentes de ácidos nucleicos, así como dichos vectores.

En cualquiera de los métodos descritos, la rotura de hebra puede ser una rotura de hebra simple o una rotura de hebra doble.

Los elementos reguladores pueden comprender promotores inducibles. Los sistemas de polinucleótidos y/o de vectores pueden comprender sistemas inducibles.

En cualquiera de los métodos descritos, dichas una o más moléculas de polinucleótidos pueden estar comprendidas en un sistema de suministro, o dichos uno o más vectores pueden estar comprendidos en un sistema de suministro.

En cualquiera de los métodos descritos, la composición no natural o manipulada se puede suministrar mediante liposomas, partículas (por ejemplo, nanopartículas), exosomas, microvesículas, una pistola de genes o uno o más vectores, por ejemplo, vectores virales o moléculas de ácidos nucleicos.

La invención también provee una composición no natural o manipulada que es una composición que tiene las características descritas en la presente o definidas en cualquiera de los métodos descritos en la presente.

También se describe, pero no se reivindica específicamente en la presente, un sistema de vectores que comprende uno o más vectores, donde dichos uno o más vectores comprenden una o más moléculas de polinucleótidos que codifican los componentes de una composición no natural o manipulada que es una composición que tiene las características descritas en la presente o definidas en cualquiera de los métodos descritos en la presente.

También se describe, pero no se reivindica específicamente en la presente, un sistema de suministro que comprende uno o más vectores o una o más moléculas de polinucleótidos, dichos uno o más vectores o moléculas de polinucleótidos que comprenden una o más moléculas de polinucleótidos que codifican componentes de una composición no natural o manipulada que es una composición que tiene las características que se describen en la presente o que se definen en cualquiera de los métodos descritos en la presente.

La descripción también detalla métodos y algoritmos computacionales para predecir nuevos sistemas CRISPR-Cas de Clase 2 y para identificar los componentes de los mismos.

La invención también provee métodos y composiciones como se reivindica en la presente en donde se puede modificar uno o más residuos de aminoácidos de la proteína efectora, por ejemplo, una proteína efectora o CpfI manipulada o no natural. En una forma de realización, la modificación puede comprender la mutación de uno o más residuos de aminoácidos de la proteína efectora. Dichas una o más mutaciones pueden ser en uno o más dominios catalíticamente activos de la proteína efectora. La proteína efectora puede tener una actividad nucleasa reducida o anulada en comparación con una proteína efectora sin dichas una o más mutaciones. Es posible que la proteína efectora no dirija el clivaje de una u otra hebra de ADN o ARN en el locus diana de interés. Es posible que la proteína efectora no dirija el clivaje de ninguna de las hebras de ADN o ARN en el locus diana de interés. En una forma de realización preferida, dichas una o más mutaciones pueden comprender dos mutaciones. En una forma de realización preferida, se modifican dichos uno o más residuos de aminoácidos en una proteína efectora CpfI, por ejemplo, una proteína efectora o CpfI manipulada o no natural. En una forma de realización preferida, la proteína efectora CpfI es una proteína efectora FnCpfI. En una forma de realización preferida, dichos uno o más residuos de aminoácidos modificados o mutados son D917A, E1006A o D1255A con referencia a la numeración de posiciones de aminoácidos de la proteína efectora FnCpfI. En formas de realización preferidas adicionales, dichos uno o más residuos de aminoácidos mutados son D908A, E993A, D1263A con referencia a las posiciones de aminoácidos en AsCpfI o LbD832A, E925A, D947A o

D1180A con referencia a las posiciones de aminoácidos en LbCpfl.

La invención también provee, en los métodos y composiciones como se reivindica en la presente, una o más mutaciones o dichas dos o más mutaciones en un dominio catalíticamente activo de la proteína efectora que comprende un dominio RuvC. En algunas formas de realización de la invención, el dominio RuvC puede comprender un dominio RuvCI, RuvCII o RuvCIII o un dominio catalíticamente activo que es homólogo de un dominio RuvCI, RuvCII o RuvCIII, etc., o cualquier dominio relevante descrito en cualquiera de los métodos que se describen en la presente. La proteína efectora puede comprender uno o más dominios funcionales heterólogos. Dichos uno o más dominios funcionales heterólogos pueden comprender uno o más dominios de señales de localización nuclear (NLS). Dichos uno o más dominios funcionales heterólogos pueden comprender por lo menos dos o más dominios NLS. Dichos uno o más dominios NLS se pueden ubicar en o cerca o próximos a un extremo terminal de la proteína efectora (por ejemplo, Cpfl) y si hay dos o más NLS, cada uno de los dos se pueden ubicar en o cerca o próximos a un extremo terminal de la proteína efectora (por ejemplo, Cpfl). Dichos uno o más dominios funcionales heterólogos pueden comprender uno o más dominios de activación de la transcripción. En una forma de realización preferida, el dominio de activación de la transcripción puede comprender VP64. Dichos uno o más dominios funcionales heterólogos pueden comprender uno o más dominios de represión de la transcripción. En una forma de realización preferida, el dominio de represión de la transcripción comprende un dominio KRAB o un dominio SID (por ejemplo, SID4X). Dichos uno o más dominios funcionales heterólogos pueden comprender uno o más dominios de nucleasas. En una forma de realización preferida, un dominio de nucleasas comprende Fok1.

La invención también provee, en los métodos y composiciones reivindicados en la presente, uno o más dominios funcionales heterólogos con una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad de factor de liberación de la transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de clivaje de un ARN de hebra doble, actividad de clivaje de ARN de hebra doble, actividad de clivaje de ADN de hebra simple, actividad de clivaje de ADN de hebra doble y actividad de unión a ácidos nucleicos. Por lo menos uno o más dominios funcionales heterólogos pueden estar o cerca del extremo amino terminal de la proteína efectora y/o en donde por lo menos uno o más dominios funcionales heterólogos están en o cerca del extremo carboxilo terminal de la proteína efectora. Dichos uno o más dominios funcionales heterólogos se pueden fusionar a la proteína efectora. Dichos uno o más dominios funcionales heterólogos se pueden unir a la proteína efectora. Dichos uno o más dominios funcionales heterólogos se pueden conectar a la proteína efectora mediante una porción conectora.

La invención también provee, para los métodos y composiciones como se reivindica en la presente, la proteína efectora (por ejemplo, una Cpfl) que comprende una proteína efectora (por ejemplo, una Cpfl) de un organismo de un género que comprende *Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Metanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium* o *Acidaminococcus.*

La invención también provee, para los métodos y composiciones como se reivindica en la presente, la proteína efectora (por ejemplo, una Cpfl) que comprende una proteína efectora (por ejemplo, una Cpfl) de un organismo de *S. mutans, S. agalactiae, S. equisimilis, S. sanguinis, S. pneumonia; C. jejuni, C. coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C. tetani, C. sordellii.*

La proteína efectora puede comprender una proteína efectora quimérica que comprende un primer fragmento de un primer ortólogo de la proteína efectora (por ejemplo, una Cpfl) y un segundo fragmento de un segundo ortólogo de la proteína efectora (por ejemplo, una Cpfl), y en donde los ortólogos de la primera y segunda proteínas efectoras son diferentes. Por lo menos una entre los ortólogos de la primera y segunda proteína efectora (por ejemplo, una Cpfl) puede comprender una proteína efectora (por ejemplo, una Cpfl) de un organismo que comprende *Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Metanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium* o *Acidaminococcus*; por ejemplo, una proteína efectora quimérica que comprende un primer fragmento y un segundo fragmento, en donde cada uno de dichos primer y segundo fragmentos se selecciona entre una Cpfl de un organismo que comprende: *Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Metanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium* o *Acidaminococcus;* en donde el primer y segundo fragmentos no provienen de la misma bacteria; por ejemplo, una proteína efectora quimérica que comprende un primer fragmento y un segundo fragmento en donde cada uno de dichos primer y segundo fragmentos se selecciona entre una Cpfl de *S. mutans, S. agalactiae, S. equisimilis, S.*

sanguinis, S. pneumonia; C. jejuni, C. coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C. tetani, C. sordellii; Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Metanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens y *Porphyromonas macacae*, en donde el primer y segundo fragmentos no son de la misma bacteria.

En realizaciones preferidas, la proteína efectora deriva de un locus Cpf1 (en la presente dichas proteínas efectoras también se denominan "Cpf1p"), por ejemplo, una proteína Cpf1 (y dicha proteína efectora o proteína Cpf1 o proteína derivada de un locus Cpf1 también se conoce como "enzima CRISPR"). Los loci de Cpf1 incluyen, pero en un sentido no taxativo, los loci Cpf1 de las especies bacterianas enumeradas en la Figura 64. En una forma de realización más preferida, la Cpf1p deriva de una especie bacteriana seleccionada entre *Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Metanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens* y *Porphyromonas macacae.* En determinadas formas de realización, la Cpf1p deriva de una especie bacteriana seleccionada entre *Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020.* En determinadas formas de realización, la proteína efectora deriva de una subespecie de *Francisella tularensis 1* que incluye, pero en un sentido no taxativo, *Francisella tularensis subsp. Novicida.*

En formas de realización adicionales de la invención, un motivo adyacente del protoespaciador (PAM) o un motivo tipo PAM dirige la unión del complejo de la proteína efectora al locus diana de interés. En una forma de realización preferida de la invención, el PAM es 5' TTN, donde N es A/C/G o T y la proteína efectora es FnCpf1p. En otra forma de realización preferida de la invención, el PAM es 5' TTTV, donde V es A/C o G y la proteína efectora es AsCpf1, LbCpf1 o PaCpf1p. En determinadas formas de realización, el PAM es 5' TTN, donde N es A/C/G o T, la proteína efectora es FnCpf1p, y el PAM está ubicado corriente arriba con respecto al extremo 5' del protoespaciador. En determinadas formas de realización de la invención, el PAM es 5' CTA, donde la proteína efectora es FnCpf1p, y el PAM está ubicado corriente arriba con respecto al extremo 5' del protoespaciador o del locus diana. En formas de realización preferidas, la invención provee un rango de direccionamiento extendido para la edición de genomas guiada por ARN de nucleasas, en donde los PAM ricos en T de la familia Cpf1 permite el direccionamiento y la edición de genomas ricos en AT.

En determinadas formas de realización, la enzima CRISPR está modificada y puede comprender una o más mutaciones que reducen o eliminan una actividad nucleasa. Las posiciones de aminoácidos en el dominio RuvC FnCpf1p incluyen, pero en un sentido no taxativo, D917A, E1006A, E1028A, D1227A, D1255A, N1257A, D917A, E1006A, E1028A, D1227A, D1255A y N1257A. Los Solicitantes también han identificado un segundo dominio nucleasa putativo que es más similar a la superfamilia de nucleasas PD-(D/E)XK y tipo endonucleasa HincII. Las mutaciones puntuales que se generarán en este dominio nucleasa putativo para reducir sustancialmente la actividad nucleasa incluyen, pero en un sentido no taxativo, N580A, N584A, T587A, W609A, D610A, K613A, E614A, D616A, K624A, D625A, K627A y Y629A. En una forma de realización preferida, la mutación en el dominio RuvC FnCpf1p es D917A o E1006A, en donde dicha mutación D917A o E1006A inactivará por completo la actividad de clivaje de ADN de la proteína efectora FnCpf1. En otra forma de realización, la mutación en el dominio RuvC FnCpf1p es D1255A, en donde la proteína efectora FnCpf1 mutada presenta una actividad nucleolítica significativamente reducida.

Las posiciones de aminoácidos en el dominio RuvC AsCpf1p incluyen, pero en un sentido no taxativo, 908, 993 y 1263. En una forma de realización preferida, la mutación en el dominio RuvC AsCpf1p es D908A, E993A y D1263A, en donde las mutaciones D908A, E993A y D1263A inactivan por completo la actividad de clivaje de ADN de la proteína efectora AsCpf1. Las posiciones de aminoácidos en el dominio RuvC LbCpf1p incluyen, pero en un sentido no taxativo, 832, 947 o 1180. En una forma de realización preferida, la mutación en el dominio RuvC LbCpf1p es LbD832A, E925A, D947A o D1180A, en donde dichas mutaciones LbD832A E925A, D947A o D1180A inactivan por completo la actividad de clivaje de ADN de la proteína efectora LbCpf1.

También se pueden efectuar mutaciones en residuos vecinos, por ejemplo, en los aminoácidos cercanos a los indicados previamente que participan en la actividad nucleasa. En algunas formas de realización, solamente se inactiva el dominio RuvC y en otras formas de realización, se inactiva otro dominio nucleasa putativo, en donde el complejo de la proteína efectora funciona como una nickasa y solamente cliva una hebra de ADN. En una forma de realización preferida, el otro dominio nucleasa putativo es un dominio endonucleasa tipo HincII. En algunas formas de realización, se usan dos variantes de FnCpf1, AsCpf1 o LbCpf1 (cada una es una nickasa diferente) para aumentar la especificidad, dos variantes de nickasa se usan para clivar ADN en una diana (donde ambas nickasas clivan la hebra de ADN, minimizando o eliminando al mismo tiempo las modificaciones fuera de la diana donde solamente se cliva una hebra de ADN y a continuación se repara). En formas de realización preferidas, la proteína efectora Cpf1 cliva las secuencias asociadas con un locus diana de interés, o presentes en el mismo, como un homodímero que comprende dos moléculas de la proteína efectora Cpf1. En una forma de realización preferida, el homodímero puede comprender dos moléculas de la proteína efectora Cpf1 que comprenden una mutación diferente en sus respectivos dominios RuvC.

La invención, como se define por los métodos y composiciones reivindicados, contempla métodos para utilizar dos o

más nickasas, en particular un abordaje de nickasas dual o doble. En algunos aspectos y formas de realización, se puede suministrar un solo tipo de nickasa FnCpfl, AsCpfl o LbCpfl, por ejemplo, una FnCpfl, AsCpfl o LbCpfl modificada o una nickasa FnCpfl, AsCpfl o LbCpfl modificada como se describe en la presente. Esto da como resultado la unión del ADN diana por dos nickasas FnCpfl. Además, también se contempla el uso de diferentes ortólogos, por ejemplo, una nickasa FnCpfl, AsCpfl o LbCpfl en una hebra (por ejemplo, la hebra codificante) del ADN y un ortólogo en la hebra de ADN no codificante u opuesta. El ortólogo puede ser, pero en un sentido no taxativo, una nickasa Cas9 tal como una nickasa SaCas9 o una nickasa SpCas9. Puede ser ventajoso emplear dos ortólogos diferentes que requieren PAM diferentes y también pueden tener requerimientos de guía diferentes, permitiendo de esa manera un control mayor por parte del usuario. En determinadas formas de realización, el clivaje de ADN comprenderá por lo menos cuatro tipos de nickasas, en donde cada tipo es guiado a una secuencia de ADN diana diferente, en donde cada par introduce un primer corte monocatenario en una hebra de ADN y el segundo introduce un corte monocatenario en la segunda hebra de ADN. En dichos métodos, se introducen por lo menos dos pares de roturas de hebra simple en el ADN diana, en donde tras la introducción del primer y segundo pares de roturas de hebra simple, se recortan las secuencias diana entre el primer y segundo pares de roturas de hebra simple. En determinadas formas de realización, uno o ambos ortólogos son controlables, es decir, son inducibles.

En determinadas formas de realización de la invención, el ARN guía o el ARNcr maduro comprende, consiste esencialmente en o consiste en una secuencia de repetición directa y una secuencia guía o una secuencia espaciadora. En determinadas formas de realización, el ARN guía o el ARNcr maduro comprende, consiste esencialmente en o consiste en una secuencia de repetición directa unida a una secuencia guía o una secuencia espaciadora. En determinadas formas de realización, el ARN guía o el ARNcr maduro comprende 19 nts de una repetición directa parcial seguido por 20-30 nt de una secuencia guía o una secuencia espaciadora, ventajosamente de aproximadamente 20 nt, 23-25 nt o 24 nt. En determinadas formas de realización, la proteína efectora es una proteína efectora FnCpfl, AsCpfl o LbCpfl y requiere por lo menos 16 nt de secuencia guía para lograr un clivaje de ADN detectable y un mínimo de 17 nt de secuencia guía para lograr un clivaje de ADN eficiente *in vitro*. En determinadas formas de realización, la secuencia de repetición directa está ubicada corriente arriba (es decir, 5') con respecto a la secuencia guía o la secuencia espaciadora. En una forma de realización preferida, la secuencia semilla (es decir la secuencia crítica esencial para el reconocimiento y/o la hibridación con la secuencia en el locus diana) del ARN guía de la FnCpfl, AsCpfl o LbCpfl se encuentra aproximadamente dentro de los primeros 5 nt por el extremo 5' de la secuencia guía o la secuencia espaciadora.

En formas de realización preferidas de la invención, el ARNcr maduro comprende un bucle-tallo o una estructura de bucle-tallo optimizada o una estructura secundaria optimizada. En formas de realización preferidas, el ARNcr maduro comprende un bucle-tallo o una estructura de bucle-tallo optimizada en la secuencia de repetición directa, en donde dicho bucle-tallo o estructura de bucle-tallo optimizada es importante para la actividad de clivaje. En determinadas formas de realización, el ARNcr maduro preferiblemente comprende un solo bucle-tallo. En determinadas formas de realización, la secuencia de repetición directa preferiblemente comprende un solo bucle-tallo. En determinadas formas de realización, la actividad de clivaje del complejo de la proteína efectora se modifica mediante la introducción de mutaciones que afectan la estructura del dúplex de ARN de bucle-tallo. En formas de realización preferidas, se pueden introducir mutaciones que conservan el dúplex de ARN del bucle-tallo, con lo cual se conserva la actividad de clivaje del complejo de la proteína efectora. En otras formas de realización preferidas, se pueden introducir mutaciones que alteran la estructura del dúplex de ARN del bucle-tallo, con lo cual se anula por completo la actividad de clivaje del complejo de la proteína efectora.

La invención también provee la secuencia de nucleótidos que codifica la proteína efectora de codones optimizados para su expresión en un eucariota o en una célula eucariota en cualquiera de los métodos o las composiciones que se describen en la presente. En una forma de realización de la invención, la secuencia de nucleótidos de codones optimizados de la proteína efectora es FnCpfl. AsCpfl o LbCpfl y es de codones optimizados para su operabilidad en una célula u organismo eucariota, por ejemplo, dicha célula u organismo mencionado en otra parte en la presente, por ejemplo, en un sentido no taxativo, una célula de levadura o una célula de mamífero o un organismo mamífero, incluyendo una célula de ratón, una célula de rata y una célula humana o un organismo eucariota no humano, por ejemplo, una planta.

En determinadas formas de realización de la invención, se une por lo menos una señal de localización nuclear (NLS) a las secuencias de ácidos nucleicos que codifican las proteínas efectoras Cpfl. En formas de realización preferidas, se unen por lo menos una o más NLS C-terminales o N-terminales (y por ende las moléculas de ácidos nucleicos que codifican la proteína efectora Cpfl pueden incluir la codificación de NLS de modo que el producto expresado comprende una o más NLS unidas o conectadas). En una forma de realización preferida, de une una NLS C-terminal para una expresión óptima y el direccionamiento nuclear en células eucariotas, preferiblemente células humanas. En una forma de realización preferida, la proteína efectora de codones optimizados es FnCpflp, AsCpfl o LbCpfl y la longitud del espaciador del ARN guía es de entre 15 y 35 nt. En determinadas formas de realización, la longitud del espaciador del ARN guía es de por lo menos 16 nucleótidos, tal como de por lo menos 17 nucleótidos. En determinadas formas de realización, la longitud del espaciador es de entre 15 y 17 nt, entre 17 y 20 nt, entre 20 y 24 nt, por ejemplo, de 20, 21, 22, 23 o 24 nt, entre 23 y 25 nt, por ejemplo, de 23, 24 o 25 nt, entre 24 y 27 nt, de 27-30 nt, de 30-35 nt o de 35 nt o más. En determinadas formas de realización de la invención, la proteína efectora de codones optimizados es FnCpflp y la longitud de la repetición directa del ARN guía es de por lo menos 16 nucleótidos. En determinadas formas de realización, la proteína efectora de codones optimizados es FnCpflp y la longitud de la repetición directa del ARN guía

es de entre 16 y 20 nt, por ejemplo, de 16, 17, 18, 19 o 20 nucleótidos. En determinadas formas de realización preferidas, la longitud de la repetición directa del ARN guía es de 19 nucleótidos.

La invención también abarca métodos para suministrar múltiples componentes de ácidos nucleicos, en donde cada componente de ácido nucleico es específico de un locus diana de interés diferente con lo cual se modifican múltiples loci blancos de interés. El componente de ácido nucleico del complejo puede comprender uno o más aptámeros de ARN de unión a proteínas. Dichos uno o más aptámeros se pueden unir a una proteína de la envoltura de bacteriófagos. La proteína de la envoltura de bacteriófagos se puede seleccionar del grupo que comprende Qβ, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, φCb5, φCb8r, φCb12r, φCb23r, 7s y PRR1. En una forma de realización preferida, la proteína de la envoltura de bacteriófagos es MS2. La invención también provee un componente de ácido nucleico del complejo de 30 o más, 40 o más o 50 o más nucleótidos de longitud.

También se describe, pero no se reivindica específicamente en la presente, células, componentes y/o sistemas que comprenden cantidades traza de los cationes presentes en las células, los componentes y/o los sistemas. Ventajosamente, el catión es magnesio, tal como Mg^{2+}. El catión puede estar presente en cantidades traza. Un rango preferido puede comprender entre aproximadamente 1 mM y aproximadamente 15 mM del catión, que ventajosamente es Mg^{2+}. Una concentración preferida puede ser de aproximadamente 1 mM para células, componentes y/o sistemas de base humana y de entre aproximadamente 10 mM y aproximadamente 15 mM en el caso de células, componentes y/o sistemas basados en bacterias. Véase, por ejemplo, Gasiunas y col., *PNAS*, publicada en línea el 4 de septiembre, 2012, www.pnas.org/cgi/doi/10.1073/pnas.1208507109.

Por lo tanto, un objeto de la invención comprende no abarcar en la misma ningún producto, ni un proceso para elaborar el producto o un método de uso del producto conocidos previamente, de modo tal que los Solicitantes se reservan el derecho y por la presente divulgan la exclusión de cualquier producto, proceso o método conocidos previamente. Se hace notar además que la invención no pretende abarcar dentro del alcance de la misma ningún producto, proceso o elaboración del producto o método de uso del producto, que no cumpla con la descripción escrita y los requerimientos de habilitación de USPTO (35 U.S.C. §112, primer párrafo) o EPO (Artículo 83 de EPC), de modo tal que los Solicitantes se reservan el derecho y por la presente divulgan la exclusión de cualquier producto, proceso de elaboración de dicho producto o método de uso del producto conocidos previamente. Será ventajoso para la práctica de la invención cumplir con el Artículo 53(c) de EPC y con las Reglas 28(b) y (c) de EPC. En la presente, nada debe considerarse como una promesa.

Cabe destacar que en esta divulgación, y en particular en las reivindicaciones y/o párrafos, los términos tales como "comprende", "comprendía", "que comprende" y semejantes tendrán el significado atribuido a los mismos en la Ley de Patentes de los EE.UU.; por ejemplo, pueden significar "incluye", "incluido", "que incluye" y semejantes; y aquellos términos tales como "que consiste esencialmente en" y "consiste esencialmente en" tienen el significado que se les adjudica en la Ley de Patentes de los EE.UU.

Estas y otras formas de realización se divulgan o resultan evidentes y están comprendidas en la siguiente Descripción detallada.

DESCRIPCIÓN BREVE DE LAS FIGURAS

Las nuevas características de la invención se establecen en forma particular en las reivindicaciones adjuntas. Se obtendrá una mejor comprensión de las características y ventajas de la presente invención en referencia a la siguiente descripción detallada que establece formas de realización ilustrativas, en donde se utilizan los principios de la invención, y las figuras acompañantes en las cuales:

Las Figuras 1A-1B describen una nueva clasificación de los sistemas CRISPR-Cas. La Clase 2 incluye complejos de efectores y ARNcr de subunidades múltiples (Cascada) y la clase 2 incluye complejos de efectores y ARNcr de subunidad individual (tipo Cas9).

La Figura 2 provee una organización molecular de CRISPR-Cas.

Las Figuras 3A-3D proveen estructuras de complejos de efectores de Tipo I y III: arquitectura común/linaje común, aunque con gran divergencia de secuencia.

La Figura 4 muestra a CRISPR-Cas como un sistema centrado en un motivo de reconocimiento de ARN (RRM).

Las Figuras 5A-5D muestran la filogenia de Cas1 en donde la recombinación de módulos de adaptación y ARNcr-efector muestra un aspecto principal de la evolución de CRISPR-Cas.

La Figura 6 muestra un censo de CRISPR-Cas, específicamente una distribución de tipos/subtipos de CRISPR-Cas entre *archaea* y *bacteria*.

La Figura 7 muestra un flujo de trabajo para identificar candidatos de Cas.

Las Figuras 8A-8D representan una organización de sistemas de loci completos de Clase 2.

Las Figuras 9A-9B representan vecindades de C2c1.

Las Figuras 10A-10C representan un árbol de Cas1.

Las Figuras 11A-11B representan una organización de dominios de familias de clase 2.

Las Figuras 12A-12B representan regiones de homología con TnpB en proteínas de Clase 2 (SEQ ID NOS 246-428, respectivamente, en orden de aparición).

Las Figuras 13A-13B representan vecindades C2c2.

Las Figuras 14A-14E representan el motivo HEPN RxxxxH en la familia C2c2 (SEQ ID NOS 429-1032, respectivamente, en orden de aparición).

La Figura 15 representa a C2C1: 1. Alicyclobacillus acidoterrestris ATCC 49025 (SEQ ID NOS 1034-1037, respectivamente, en orden de aparición).

La Figura 16 representa a C2C1: 4. Desulfonatronum thiodismutans cepa MLF-1 (SEQ ID NOS 1038-1041, respectivamente, en orden de aparición).

La Figura 17 representa a C2C1: 5. Opitutaceae bacterium TAV5 (SEQ ID NOS 1042-1045, respectivamente, en orden de aparición).

La Figura 18 representa a C2C1: 7. Bacillus thermoamylovorans cepa B4166 (SEQ ID NOS 1046-1049, respectivamente, en orden de aparición).

La Figura 19 representa a C2C1: 9. Bacillus sp. NSP2.1 (SEQ ID NOS 1050-1053, respectivamente, en orden de aparición).

La Figura 20 representa a C2C2: 1. Lachnospiraceae bacterium MA2020 (SEQ ID NOS 1054-1057, respectivamente, en orden de aparición).

La Figura 21 representa a C2C2: 2. Lachnospiraceae bacterium NK4A179 (SEQ ID NOS 1058-1064, respectivamente, en orden de aparición).

La Figura 22 representa a C2C2: 3. [Clostridium] aminophilum DSM 10710 (SEQ ID NOS 1065-1068, respectivamente, en orden de aparición).

La Figura 23 representa a C2C2: 4. Lachnospiraceae bacterium NK4A144 (SEQ ID NOS 1069 y 1070, respectivamente, en orden de aparición).

La Figura 24 representa a C2C2: 5. Carnobacterium gallinarum DSM 4847 (SEQ ID NOS 1071-1074, respectivamente, en orden de aparición).

La Figura 25 representa a C2C2: 6. Carnobacterium gallinarum DSM 4847 (SEQ ID NOS 1075-1081, respectivamente, en orden de aparición).

La Figura 26 representa a C2C2: 7. Paludibacter propionicigenes WB4 (SEQ ID NO: 1082).

La Figura 27 representa a C2C2: 8. Listeria seeligeri serovariedad 1/2b (SEQ ID NOS 1083-1086, respectivamente, en orden de aparición).

La Figura 28 representa a C2C2: 9. Listeria weihenstephanensis FSL R9-0317 (SEQ ID NO: 1087).

La Figura 29 representa a C2C2: 10. Listeria bacterium FSL M6-0635 (SEQ ID NOS 1088 y 1091, respectivamente, en orden de aparición).

La Figura 30 representa a C2C2: 11. Leptotrichia wadei F0279 (SEQ ID NO: 1092).

La Figura 31 representa a C2C2: 12. Leptotrichia wadei F0279 (SEQ ID NOS 1093-1099, respectivamente, en orden de aparición).

La Figura 32 representa a C2C2: 14. Leptotrichia shahii DSM 19757 (SEQ ID NOS 1100-1103, respectivamente, en orden de aparición).

La Figura 33 representa a C2C2: 15. Rhodobacter capsulatus SB 1003 (SEQ ID NOS 1104 y 1105, respectivamente, en orden de aparición).

La Figura 34 representa a C2C2: 16. Rhodobacter capsulatus R121 (SEQ ID NOS 1106 y 1107, respectivamente, en orden de aparición).

La Figura 35 representa a C2C2: 17. Rhodobacter capsulatus DE442 (SEQ ID NOS 1108 y 1109, respectivamente,

en orden de aparición).

La Figura 36 representa un árbol de DR

La Figura 37 representa un árbol de C2C2s

Las Figuras 38A-38BB muestran el alineamiento de secuencias de ortólogos de Cas-Cpf1 (SEQ ID NOS 1033 y 1110-1166, respectivamente, en orden de aparición).

Las Figuras 39A-39B muestran el resumen del alineamiento de loci de Cpfl

Las Figuras 40A-40X muestran la construcción de vector PACYC184 FnCpfl (PY001) (SEQ ID NO: 1167 y SEQ ID NOS 1168-1189, respectivamente, en orden de aparición).

Las Figuras 41A-41I muestran la secuencia de PaCpfl humanizado, con la secuencia nucleotídica de SEQ ID NO: 1190 y la secuencia proteica de SEQ ID NO: 1191.

La Figura 42 representa un ensayo de desafío de PAM

La Figura 43 representa un esquema de un locus FnCpfl endógeno. pY0001 es un esqueleto de pACY184 (de NEB) con un locus parcial FnCpfl. El locus FnCpfl se amplificó por PCR en tres partes y se clonó en pACYC184 cortado con Xba1 y Hind3 usando ensamblaje de Gibson. El PY0001 contiene el locus endógeno FnCpfl entre las 255 pb de la secuencia 3' de la acetiltransferasa hasta la cuarta secuencia espaciadora. Solamente el espaciador 1-3 es potencialmente activo debido a que el espacio 4 ya no está flanqueado por repeticiones directas.

La Figura 44 representa bibliotecas PAM, las que divulgan a las SEQ ID NOS 1192-1195, respectivamente, en orden de aparición. Ambas bibliotecas PAM (izquierda y derecha) están en pUC19. La complejidad de la biblioteca PAM izquierda es de 48 ~ 65k y la complejidad de la biblioteca PAM derecha es de 47 ~ 16k. Ambas bibliotecas se prepararon con una representación > 500.

La Figura 45A-4E representa Análisis Computacional de Cribado de PAM FnCpfl. Después de la secuenciación del ADN cribado, se extrajeron las regiones correspondientes al PAM izquierdo o bien al PAM derecho. Para cada muestra, se comparó el número de PAM presentes en la biblioteca secuenciada con el número de PAM esperados en la biblioteca (4^8 para la biblioteca izquierda, 4^7 para la derecha). La Figura 44A muestra que la biblioteca izquierda mostró agotamiento de PAM. Para cuantificar este agotamiento, se calculó una proporción de enriquecimiento. Para ambas condiciones (pACYC control o FnCpfl conteniendo pACYC) se calculó la proporción para cada PAM de la biblioteca como

$$\text{proporción} = -\log_2 \frac{\text{muestra} + 0{,}01}{\text{biblioteca inicial} + 0{,}01}$$

El gráfico de la distribución muestra poco enriquecimiento en la muestra control y enriquecimiento en ambas réplicas biológicas. Las Figuras 44B-44D representan distribuciones de proporciones de PAM. La Figura 44E que se recolectaron todos los PAM por arriba de una proporción de 8, y se graficaron las distribuciones de frecuencia, revelando un PAM 5' YYN.

La Figura 46 representa un análisis de ARNseq del locus Cpfl de tolerancia a Francisella, que muestra que el locus CRISPR se expresa activamente. Además de los genes Cpfl y Cas genes, dos transcriptos no codificantes pequeños se transcriben en alto nivel, los que podrían ser ARNtracr putativos. El arreglo de CRISPR también se expresa. Tanto los ARNtracr putativos como el arreglo CRISPR se transcriben en la misma dirección que los genes Cpfl y Cas. Aquí, todos los transcriptos de ARN identificados a través de experimento de ARNseq se mapean contra el locus. Después de una evaluación adicional del locus FnCpfl, los Solicitantes concluyeron que la escisión de ADN diana mediante un complejo de proteína efectora Cpfl no requiere un ARNtracr. Los Solicitantes determinaron que los complejos de proteína efectora Cpfl que comprenden solo una proteína efectora Cpfl y un ARNcr (ARN guía que comprende una secuencia repetitiva directa y una secuencia guía) fueron suficientes para escindir el ADN diana.

La Figura 47 representa un acercamiento en el arreglo de CRISPR Cpfl. Se pueden identificar muchos transcriptos cortos diferentes. En este gráfico, todos los transcriptos de ARN identificados se mapean contra el locus Cpfl.

La Figura 48 representa la identificación de dos ARNtracr putativos después de seleccionar los transcriptos que tienen menos de 85 nucleótidos de longitud

La Figura 49 representa un acercamiento en el ARNtracr putativo 1 (SEQ ID NO: 1196) y el arreglo de CRISPR

La Figura 50 representa un acercamiento en el ARNtracr putativo 2 que divulga a las SEQ ID NOS 1197-1203, respectivamente, en orden de aparición.

La Figura 51 representa secuencias de ARNcr putativos (repetición en azul, espaciador en negro) (SEQ ID NOS 1205 y 1206, respectivamente, en orden de aparición).

La Figura 52 muestra un esquema del ensayo para confirmar el PAM FnCpfl predicho in vivo.

La Figura 53 muestra células que portan el locus FnCpfl y células control transformadas con pUC19 que codifica para espaciador endógeno 1 con 5' TTN PAM.

La Figura 54 muestra un esquema que indica las posiciones de secuencia de ARNtracr putativo en el locus FnCpfl, el ARNcr (SEQ ID NO: 1207) y el vector protoespaciador pUC.

La Figura 55 es un gel que muestra el fragmento de PCR con TTa PAM y secuencia protoespaciadora 1 incubados en lisado celular.

La Figura 56 es un gel que muestra el pUC-espaciador1 con diferentes PAM incubados en lisado celular.

La Figura 57 es un gel que muestra la digestión con BasI después de incubar en lisado celular.

La Figura 58 es un gel que muestra resultados de digestión para tres secuencias de ARNcr putativos (SEQ ID NO: 1208).

La Figura 59 es un gel que muestra el ensayo de diferentes longitudes de espaciador contra una porción de ADN diana que contiene el sitio diana: 5'-TTAgagaagtcatttaataaggccactgttaaaa-3' (SEQ ID NO: 1209). Los resultados muestran que los ARNcr 1-7 mediaron una escisión exitosa del ADN diana in vitro con FnCpfl. Los ARNcr 8-13 no facilitaron la escisión del ADN diana. SEQ ID NOS 1210-1248 se divulgan, respectivamente, en orden de aparición.

La Figura 60 es un esquema que indica el locus FnCpfl mínimo.

La Figura 61 es un esquema que indica la guía Cpfl mínima (SEQ ID NO: 1249).

La Figura 62A-62E representa un Análisis Computacional de Cribado de PAM PaCpfl. Después de la secuenciación del ADN cribado, se extrajeron las regiones correspondientes al PAM izquierdo o bien al PAM derecho. Para cada muestra, se comparó el número de PAM presentes en la biblioteca secuenciada con el número de PAM esperados en la biblioteca (4^7). (Figura 62A) La biblioteca izquierda mostró un agotamiento muy leve de PAM. Para cuantificar este agotamiento, se calculó una proporción de enriquecimiento. Para ambas condiciones (pACYC control o PaCpfl conteniendo pACYC) se calculó la proporción para cada PAM de la biblioteca como

$$\text{proporción} = -\log_2 \frac{\text{muestra} + 0{,}01}{\text{biblioteca inicial} + 0{,}01}$$

El gráfico de la distribución muestra poco enriquecimiento en la muestra control y enriquecimiento en ambas réplicas biológicas. Las Figuras 62 B-D representan distribuciones de proporciones de PAM. La Figura 62E muestra que se recolectaron todos los PAM por arriba de una proporción de 4,5, y se graficaron las distribuciones de frecuencia, relevando un PAM 5' TTTV, en donde V es A o C o G.

La Figura 63 muestra un mapa de vector de la secuencia de PaCpfl optimizada para codones humanos que se muestra como CBh-NLS-huPaCpfl-NLS-3xHA-pA.

Las Figuras 64A-64B muestran un árbol filogenético de 51 loci Cpfl en diferentes bacterias. Las cajas resaltadas indican las Referencias de Gen #: 1-17. Los ortólogos en caja/numerados se ensayaron para ver la actividad de escisión in vitro con ARNcr maduro predicho; los ortólogos con cajas alrededor de sus números mostraron actividad en el ensayo in vitro.

Las Figuras 65A-65H muestran los detalles de secuencia optimizada para codones humanos para Lachnospiraceae bacterium MC2017 1 Cpfl con una longitud génica de 3849 nucleótidos (Ref #3 en la Figura 64). Figura 65A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 65B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 65C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 65D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 65E: Secuencias Repetitivas Eliminadas. Figura 65F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3849, % de GC de 54,70) (SEQ ID NO: 1250). Figura 65H: Secuencia de Proteína (SEQ ID NO: 1251).

Las Figuras 66A-66H muestran los detalles de secuencia optimizada para codones humanos para Cpfl de Butyrivibrio proteoclasticus con una longitud génica de 3873 nucleótidos (Ref #4 en la Figura 64). Figura 66A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 66B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100

para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 66C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 66D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 66E: Secuencias Repetitivas Eliminadas. Figura 66F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3873, % de GC de 54,05) (SEQ ID NO: 1252). Figura 66H: Secuencia de Proteína (SEQ ID NO: 1253).

Las Figuras 67A-67H muestran los detalles de secuencia optimizada para codones humanos para CpfI de Peregrinibacteria bacterium GW2011_GWA2_33_10 con una longitud génica de 4581 nucleótidos (Ref #5 en la Figura 64). Figura 67A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 67B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 67C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 67D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 67E: Secuencias Repetitivas Eliminadas. Figura 67F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 4581, % de GC de 50,81) (SEQ ID NO: 1254). Figura 67H: Secuencia de Proteína (SEQ ID NO: 1255).

Las Figuras 68A-68H muestran los detalles de secuencia optimizada para codones humanos para CpfI de Parcubacteria bacterium GW2011_GWC2_44_17 con una longitud génica de 4206 nucleótidos (Ref #6 en la Figura 64). Figura 68A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 68B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 68C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 68D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 68E: Secuencias Repetitivas Eliminadas. Figura 68F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 4206, % de GC de 52,17) (SEQ ID NO: 1256). Figura 68H: Secuencia de Proteína (SEQ ID NO: 1257).

Las Figuras 69A-69H muestran los detalles de secuencia optimizada para codones humanos para CpfI de Smithella sp. SCADC con una longitud génica de 3900 nucleótidos (Ref #7 en la Figura 64). Figura 69A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 69B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 69C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 69D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 69E: Secuencias Repetitivas Eliminadas. Figura 69F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3900, % de GC de 51,56) (SEQ ID NO: 1258). Figura 69H: Secuencia de Proteína (SEQ ID NO: 1259).

Las Figuras 70A-70H muestran los detalles de secuencia optimizada para codones humanos para CpfI de Acidaminococcus sp. BV3L6 con una longitud génica de 4071 nucleótidos (Ref #8 en la Figura 64). Figura 70A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 70B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 70C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 70D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 70E: Secuencias Repetitivas Eliminadas. Figura 70F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 4071, % de GC de 54,89) (SEQ ID NO: 1260). Figura 70H: Secuencia de Proteína (SEQ ID NO: 1261).

Las Figuras 71A-71H muestran los detalles de secuencia optimizada para codones humanos para CpfI de Lachnospiraceae bacterium MA2020 con una longitud génica de 3768 nucleótidos (Ref #9 en la Figura 64). Figura 71A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 71B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 71C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es

entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 71D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 71E: Secuencias Repetitivas Eliminadas. Figura 71F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3768, % de GC de 51,53) (SEQ ID NO: 1262). Figura 71H: Secuencia de Proteína (SEQ ID NO: 1263).

Las Figuras 72A-72H muestran los detalles de secuencia optimizada para codones humanos para Cpfl de Candidatus Methanoplasma termitum con una longitud génica de 3864 nucleótidos (Ref #10 en la Figura 64). Figura 72A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 72B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 72C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 72D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 72E: Secuencias Repetitivas Eliminadas. Figura 72F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3864, % de GC de 52,67) (SEQ ID NO: 1264). Figura 72H: Secuencia de Proteína (SEQ ID NO: 1265).

Las Figuras 73A-73H muestran los detalles de secuencia optimizada para codones humanos para Cpfl de Eubacterium eligens con una longitud génica de 3996 nucleótidos (Ref #11 en la Figura 64). Figura 73A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 73B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 73C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 73D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 73E: Secuencias Repetitivas Eliminadas. Figura 73F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3996, % de GC de 50,52) (SEQ ID NO: 1266). Figura 73H: Secuencia de Proteína (SEQ ID NO: 1267).

Las Figuras 74A-74H muestran los detalles de secuencia optimizada para codones humanos para Cpfl de Moraxella bovoculi 237 con una longitud génica de 4269 nucleótidos (Ref #12 en la Figura 64). Figura 74A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 74B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 74C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 74D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 74E: Secuencias Repetitivas Eliminadas. Figura 74F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 4269, % de GC de 53,58) (SEQ ID NO: 1268). Figura 74H: Secuencia de Proteína (SEQ ID NO: 1269).

Las Figuras 75A-75H muestran los detalles de secuencia optimizada para codones humanos para Cpfl de Leptospira inadai con una longitud génica de 3939 nucleótidos (Ref #13 en la Figura 64). Figura 75A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 75B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 75C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 75D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 75E: Secuencias Repetitivas Eliminadas. Figura 75F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3939, % de GC de 51,30) (SEQ ID NO: 1270). Figura 75H: Secuencia de Proteína (SEQ ID NO: 1271).

Las Figuras 76A-76H muestran los detalles de secuencia optimizada para codones humanos para Cpfl de Lachnospiraceae bacterium ND2006 con una longitud génica de 3834 nucleótidos (Ref #14 en la Figura 64). Figura 76A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 76B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 76C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 76D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 76E: Secuencias Repetitivas Eliminadas. Figura 76F-G:

Secuencia Optimizada (Longitud de Secuencia Optimizada: 3834, % de GC de 51,06) (SEQ ID NO: 1272). Figura 76H: Secuencia de Proteína (SEQ ID NO: 1273).

Las Figuras 77A-77H muestran los detalles de secuencia optimizada para codones humanos para Cpf1 de Porphyromonas crevioricanis 3 con una longitud génica de 3930 nucleótidos (Ref #15 en la Figura 64). Figura 77A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 77B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 77C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 77D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 77E: Secuencias Repetitivas Eliminadas. Figura 77F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3930, % de GC de 54,42) (SEQ ID NO: 1274). Figura 77H: Secuencia de Proteína (SEQ ID NO: 1275).

Las Figuras 78A-78H muestran los detalles de secuencia optimizada para codones humanos para Cpf1 de Prevotella disiens con una longitud génica de 4119 nucleótidos (Ref #16 en la Figura 64). Figura 78A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 78B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 78C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 78D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 78E: Secuencias Repetitivas Eliminadas. Figura 78F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 4119, % de GC de 51,88) (SEQ ID NO: 1276). Figura 78H: Secuencia de Proteína (SEQ ID NO: 1277).

Las Figuras 79A-79H muestra los detalles de secuencia optimizada para codones humanos para Cpf1 de Porphyromonas macacae con una longitud génica de 3888 nucleótidos (Ref #17 en la Figura 64). Figura 79A: Índice de Adaptación de Codones (CAI). Distribución de frecuencia de uso de codones a lo largo de toda la longitud de la secuencia génica. Un CAI de 1,0 se considera como perfecto en el organismo de expresión deseado, y un CAI de > 0,8 se considera como bueno, en términos de alto nivel de expresión génica. Figura 79B: Frecuencia de Codones Óptimos (FOP). Porcentaje de distribución de codones en grupos de calidad de codones computados. Se configura un valor de 100 para el codón con la frecuencia de uso más alta para un aminoácido dado en el organismo de expresión deseado. Figura 79C: Ajuste de Contenido de GC. El porcentaje ideal de rango de contenido de GC es entre 30 y 70%. Los picos de % de contenido de GC en una ventana de 60 pb se han eliminado. Figura 79D: Enzimas de Restricción y Elementos que Actúan en CIS. Figura 79E: Secuencias Repetitivas Eliminadas. Figura 79F-G: Secuencia Optimizada (Longitud de Secuencia Optimizada: 3888, % de GC de 53,26) (SEQ ID NO: 1278). Figura 79H: Secuencia de Proteína (SEQ ID NO: 1279).

La Figura 80A-80I muestra las secuencias repetitivas directa (DR) para cada ortólogo (en referencia a la numeración Ref # 3-17 en la Figura 64) y su estructura plegada predicha. Las SEQ ID NOS 1280-1313, respectivamente, se divulgan en orden de aparición.

La Figura 81 muestra la escisión de un amplicón de PCR del locus humano Emx1. Las SEQ ID NOS 1314-1318, respectivamente, se divulgan en orden de aparición.

La Figura 82A-82B muestra el efecto de truncación de 5' DR sobre la actividad de escisión. La Figura 82A muestra un gel en donde se indican los resultados de la escisión con 5 truncaciones de DR. La Figura 82B muestra un diagrama en donde el ADNcr deltaDR5 alteró el tallo y bucle del extremo 5'. Esto indica que el tallo y bucle del extremo 5' es esencial para la actividad de escisión. Las SEQ ID NOS 1319-1324, respectivamente, se divulgan en orden de aparición.

La Figura 83 muestra el efecto de la no coincidencia de ARNcr y ADN diana sobre la eficacia de escisión. Las SEQ ID NOS 1325-1335, respectivamente, se divulgan en orden de aparición.

La Figura 84 muestra la escisión de ADN usando Cpf1 purificado de Francisella y Prevotella. Se divulga la SEQ ID NO: 1336.

Las Figuras 85A-85B muestran diagramas de estructuras secundarias de DR. La Figura 85ª muestra la estructura secundaria de DR de FnCpf1 (SEQ ID NO: 1337) (tallo y bucle resaltado). La Figura 85B muestra una estructura secundaria de DR de PaCpf1 (SEQ ID NO: 1338) (tallo y bucle resaltado, idénticos excepto por una diferencia de base individual en la región del bucle).

La Figura 86 muestra una representación adicional de un análisis de ARNseq del locus FnCp1.

Las Figuras 87A-87B muestran esquemas de secuencias de ARNcr maduro. La Figura 87A muestra las secuencias de ARNcr maduro para FnCpf1. La Figura 87B muestra las secuencias de ARNcr maduro para PaCpfl. Las SEQ ID NOS 1339-1342, respectivamente, se divulgan en orden de aparición.

La Figura 88 muestra la escisión de ADN usando FnCpf1 de Francisella novicida optimizado para codones humanos. La banda superior corresponde al fragmento de longitud completa sin escindir (606 pb). Los tamaños de los productos esperados de escisión de -345 pb y -261 pb se indican con triángulos.

La Figura 89 muestra un ensayo de ortólogo in vitro que demuestra la escisión por los ortólogos de Cpfl.

Las Figuras 90A-90C muestran PAM obtenidos por computadora a partir de ensayo de corte in vitro.

La Figura 91 muestra el corte de Cpfl en forma escalonada con sobreextensiones 5'. Las SEQ ID NOS 1343-1345, respectivamente, se divulgan en orden de aparición.

La Figura 92 muestra el efecto de la longitud del espaciador sobre el corte. Las SEQ ID NOS 1346-1352, respectivamente, se divulgan en orden de aparición.

La Figura 93 muestra los datos de SURVEYOR para indeles mediados por FnCpfl en células HEK293T.

Las Figuras 94A-94F muestran el procesamiento de transcriptos cuando se eliminan secciones del locus FnCpfl en comparación con el procesamiento de transcriptos en un locus de FnCpfl de tipo salvaje. Las Figuras 95B, 95D y 95F son un acercamiento del espaciador procesado. Las SEQ ID NOS 1353-1401, respectivamente, se divulgan en orden de aparición.

Las Figuras 95A-95E muestran que el locus Cpfl CRISPR de Francisella tularensis subsp. novicida U112 provee inmunidad contra la transformación de plásmidos que contienen protoespaciadors flanqueados por un PAM 5'-TTN. La Figura 95A muestra la organización de dos loci CRISPR hallados en Francisella tularensis subsp. novicida U112 (NC_008601). Se compara la organización de dominios de FnCas9 y FnCpfl. La Figura 95B provee una ilustración esquemática del ensayo de agotamiento plasmídico para descubrir la identidad y posición de PAM. Se transformaron células E. coli competentes que portan el plásmido con FnCpfl de locus heterólogo (pFnCpfl) o el vector control vacío con una biblioteca de plásmidos que contienen el protoespaciador correspondiente flanqueado por secuencias randomizadas de PAM 5' o 3' y se seleccionaron con antibióticos para agotar los plásmidos que portan PAM blancos exitosos. Los plásmidos de las colonias sobrevivientes se extrajeron y secuenciaron para determinar las secuencias de PAM agotados. Las Figuras 95C-95D muestran logos de secuencia para el PAM de FnCpfl determinado por el ensayo de agotamiento plasmídico. La altura de la letra en cada posición está determinada por el contenido de información; las barras de error muestran el intervalo de confianza Bayesiano de 95%. La Figura 95E muestra que las E. coli que portan pFnCpfl demuestran una interferencia robusta contra los plásmidos que portan PAM 5'-TTN (n = 3, las barras de error representan la media ± S.E.M.).

Las Figuras 96A-96C muestran que la expresión heteróloga de arreglo CRISPR de FnCpfl en E. coli es suficiente para mediar la interferencia del ADN plasmídico y la maduración del ARNcr. La ARNseq de ARN pequeños de Francisella tularensis subsp. novicida U112 (Figura 96A) revela la transcripción y el procesamiento del arreglo de CRISPR de FnCpfl. El ARNcr maduro comienza con una repetición directa parcial de 19 nt seguida por una secuencia protoespaciadora de entre 23 y 25 nt. La ARNseq de ARN pequeños de E. coli transformada con un plásmido que porta un arreglo de FnCpfl y CRISPR conducido por un promotor sintético (Figura 96B) muestra el procesamiento de ARNcr independiente de los genes Cas y otros elementos de secuencia en el locus FnCpfl. La Figura 96C representa a E. coli portando diferentes truncaciones del locus FnCpfl CRISPR y muestra que solo se requieren FnCpfl y el arreglo de CRISPR para la interferencia del ADN plasmídico (n = 3, las barras de error muestran la media ± S.E.M.). Se divulga la SEQ ID NO: 1580.

Las Figuras 97A-97E muestran que FnCpfl es diana de ARNcr para escindir ADN in vitro. La Figura 97A es un esquema del complejo de direccionamiento de ARNcr-ADN de FnCpfl. Los sitios de escisión están indicados por flechas rojas (las SEQ ID NOS 1402 y 1403, respectivamente, se divulgan en orden de aparición). FnCpfl y ARNcr solos mediaron la escisión guiada por ARN del ADN diana en una forma dependiente de ARNcr y Mg2+ (Figura 97B). La Figura 97C muestra que FnCpfl escinde ADN tanto lineal como superenrollado. La Figura 97D muestra trazas de secuenciación de Sanger de diana digerido con FnCpfl con sobreextensiones escalonadas (las SEQ ID NOS 1404 y 1406, respectivamente, se divulgan en orden de aparición). El agregado sin molde de una alanina adicional, indicada como N, es un artefacto de la polimerasa que se usó en la secuenciación. La lectura con el cebador reverso se representó como complemento reverso para ayudar a la visualización. La Figura 97E muestra que la escisión es dependiente del apareamiento de bases en 5' PAM. FnCpfl solo puede reconocer el PAM con ADN apareado correctamente según las reglas de Watson-Crick.

Las Figuras 98A-98B muestran que los residuos catalíticos del dominio C-terminal RuvC de FnCpfl son necesarios para la escisión de ADN. La Figura 98A muestra la estructura de dominios de FnCpfl con los residuos catalíticos de RuvC resaltados. Los residuos catalíticos se identificaron en base a homología de secuencia con RuvC de Thermus thermophilus (PDB ID: 4EP5). La Figura 98B representa un gel de PAGE nativo en TBE que muestra que la mutación de los residuos catalíticos de RuvC de FnCpfl (D917A y E1006A) y la mutación del residuo catalítico de RuvC (D10A)

de SpCas9 previenen la escisión de ADN de doble hebra. El gel de PAGE con TBE-Urea muestra que esa mutación de los residuos catalíticos de RuvC de FnCpfI (D917A y E1006A) previene la actividad de corte de monohebra del ADN, mientras que la mutación del residuo catalítico de RuvC (D10A) de SpCas9 resulta en el corte en monohebra del sitio diana.

Las Figuras 99A-99E muestran los requerimientos de ARNcr para la actividad nucleasa de FnCpfI in vitro. La Figura 99A muestra el efecto de la longitud del espaciador sobre la actividad de escisión de FnCpfI. La Figura 99B muestra el efecto de la falta de coincidencia de ARNcr-ADN diana sobre la actividad de escisión de FnCpfI. La Figura 99C demuestra el efecto de la longitud de las repeticiones directas sobre la actividad de escisión de FnCpfI. La Figura 99D muestra que la actividad de escisión de FnCpfI depende de la estructura secundaria del tallo de la estructura de ARN de repetición directa. La Figura 99E muestra que la actividad de escisión de FnCpfI no se ve afectada por las mutaciones del bucle pero es sensible a la mutación en la base del extremo 3'de la repetición directa. Las SEQ ID NOS 1407-1433, respectivamente, se divulgan en orden de aparición.

Las Figuras 100A-100F proveen un análisis de la diversidad y función de las proteínas de la familia Cpf1. Las Figuras 100A-100B muestran una comparación filogenética de 16 ortólogos de Cpf1 seleccionados mediante análisis funcional. Las secuencias conservadas se muestran en gris oscuro. El domino RuvC, hélice puente, y dedo de cinc están resaltados. La Figura 100C muestra un alineamiento de repeticiones directas de las 16 proteínas de la familia Cpf1. Las secuencias que se eliminan después de la maduración del ARNcr están coloreadas en gris. Las bases no conservadas están coloreadas de rojo. El dúplex tallo se resalta en gris. La Figura 100D representa una predicción de plegamiento de ARN (Lorenz y col., 2011) de la secuencia de repetición directa en el ARNcr maduro. Se muestran las predicciones para FnCpf1 junto con tres ortólogos menos conservados. La Figura 100E muestra que los ARNcrs ortólogos con secuencias repetitivas directas similares son capaces de funcionar con FnCpf1 para mediar la escisión de ADN diana. La Figura 100F muestra las secuencias de PAM para 8 proteínas de la familia Cpf1 que se identificaron usando escisión in vitro de una biblioteca plasmídica que contiene PAM normalizados que flanquean el protoespaciador. Las SEQ ID NOS 1434-1453, respectivamente, se divulgan en orden de aparición.

Las Figuras 101A-101E muestran que Cpf1 media una robusta edición de genoma en líneas celulares humanas. La Figura 101A es un esquema que muestra la expresión de proteínas individuales de la familia Cpf1 en células HEK 293FT usando vectores de expresión dirigidos por CMV. El correspondiente ARNcr se expresa usando un fragmento de PCR que contiene un promotor U6 fusionado a la secuencia de ARNcr. Las células transfectadas se analizaron usando ensayo de nucleasa Surveyor o bien secuenciación profunda direccionada. La Figura 101B (superior) representa la secuencia de ARNcr contra DNMT1, y las lecturas de secuenciación (inferior) muestran los indeles representativos. La Figura 101B divulga las SEQ ID NOS 1454-1465, respectivamente, en orden de aparición. La Figura 101C provee una comparación de actividad de escisión in vitro e in vivo. La región diana de DNMT1 se amplificó por PCR y se usó el fragmento genómico para ensayar la escisión mediada por Cpf1. Las 8 proteínas de la familia Cpf1 mostraron escisión de ADN in vitro (superior). Los candidatos 7 – AsCpf1 y 13 – Lb3Cpf1 facilitaron la formación robusta de indel en células humanas (inferior). La Figura 101D muestra las secuencias diana de Cpf1 y SpCas9 en el locus DNMT1 humano (Las SEQ ID NOS 1466-1473, respectivamente, se divulgan en orden de aparición). La Figura 101E provee una comparación de la eficacia de edición de genoma por Cpf1 y SpCas9. Los sitios diana corresponden a las secuencias que se muestran en la Figura 101D.

Las Figuras 102A-102D muestran un ensayo de agotamiento plasmídico in vivo para identificar PAM de FnCpf1. (Véase también la Figura 95). Figura 102A: Transformación de E. coli que portan pFnCpf1 con una biblioteca de plásmidos que portan secuencias randomizadas de 5' PAM. Se eliminó un subgrupo de plásmidos. El gráfico muestra los niveles de agotamiento en orden jerarquizado. El agotamiento se mide como el log2 negativo de la proporción de veces de abundancia normalizada en comparación con los controles de E. coli con pACYC184. Los PAM con un umbral por arriba de 3,5 se usan para generar logos de secuencia. Figura 102B: Transformación de E. coli que portan pFnCpf1 con una biblioteca de plásmidos que portan secuencias randomizadas de 3' PAM. Se agotó un subgrupo de plásmidos. El gráfico muestra los niveles de agotamiento en orden jerarquizado. El agotamiento se mide como el log2 negativo de la proporción de veces de abundancia normalizada en comparación con los controles de E. coli con pACYC184 y PAM con un umbral por arriba de 3,5 se usan para generar logos de secuencia. Figura 102C: Biblioteca de plásmidos que portan secuencias randomizadas de 5' PAM. El gráfico muestra los niveles de agotamiento en orden jerarquizado. El agotamiento se mide como el log2 negativo de la proporción de veces de abundancia normalizada en comparación con los controles de E. coli con pACYC184. Los PAM con un umbral por arriba de 3,5 se usan para generar logos de secuencia. Figura 102D: Número de PAM únicos que superan el umbral de significancia para combinaciones apareadas de bases en las posiciones 2 y 3 del 5' PAM.

Las Figuras 103A-103D muestran la purificación de proteína FnCpf1. (Véase también la Figura 97). La Figura 103A representa un gel de acrilamida teñido con azul de Coomasie de FnCpf1 que muestra una purificación de a pasos. Una banda justo por arriba de 160 kD eluyó de la columna de Ni-NTA, en consistencia con el tamaño de una fusión MBP-FnCpf1 (189,7 kD). Ante el agregado de TEV proteasa, apareció una banda de menor peso molecular, en consistencia con el tamaño de 147 kD de FnCpf1 libre. Figura 103B: Filtración en gel por exclusión de tamaño de fnCpf1. La FnCpf1 eluyó a un tamaño aproximado de 300 kD (62,65 mL), lo que sugiere que Cpf1 puede existir en solución como dímero. La Figura 103C muestra los estándares de proteína que se usaron para calibrar la columna de Superdex 200. BDex = Dextrano azul (volumen vacío), Ald = Aldolasa (158 kD), Ov = Ovalbúmina (44 kD), RibA = Ribonucleasa A (13,7 kD), Apr = Aprotinina (6,5 kD). Figura 103D: Curva de calibración de la columna Superdex 200. La Ka se calcula como

(volumen de elusión – volumen vacío)/(volumen de columna geométrico– volumen vacío). Los estándares se graficaron y se ajustaron a una curva logarítmica.

Las Figuras 104A-104E muestran patrones de escisión de FnCpfI. (Véase también la Figura 97). Las trazas de secuenciación de Sanger de los ADN blancos digeridos con FnCpfI muestran sobreextensiones escalonadas. El agregado sin molde de una alanina adicional, indicada como N, es un artefacto de la polimerasa que se usó en la secuenciación. Se muestran las trazas de Sanger para diferentes TTN PAM con protoespaciador 1 (Figura 104A), protoespaciador 2 (Figura 104B), y protoespaciador 3 (Figura 104C) y blancos DNMT1 y EMX1 (Figura 104D). La secuencia de hebra (–) se representa como complemento reverso para mostrar la secuencia de la hebra superior. Los sitios de escisión están indicados por triángulos rojos. Los triángulos más pequeños indican sitios de escisión alternativos putativos. La Figura 104E muestra el efecto de la falta de coincidencia de PAM-distal de ARNcr-ADN diana sobre la actividad de escisión de FnCpfI. Las SEQ ID NOS 1474-1494, respectivamente, se divulgan en orden de aparición.

Las Figuras 105A-105B muestran un alineamiento de secuencias de aminoácidos de FnCpfI (SEQ ID NO: 1495), AsCpfI (SEQ ID NO: 1496), y LbCpfI (SEQ ID NO: 1497). (Véase también la Figura 100). Los residuos conservados están resaltados con un fondo rojo y las mutaciones conservadas están resaltadas con un contorno y fuente roja. La predicción de estructura secundaria está resaltada por arriba (FnCpfI) y debajo (LbCpfI) del alineamiento. Las hélices alfa se muestran como símbolo rizado y las hebras beta se muestran como guiones. Los dominios proteicos en la Figura 95A también están resaltados.

Las Figuras 106A-106D proveen mapas de loci genómicos bacterianos que corresponden a las 16 proteínas de la familia CpfI seleccionadas para experimentación en mamíferos. (Véase también la Figura 100). Las Figuras 106A-106D divulgan las SEQ ID NOS 1498-1513, respectivamente, en orden de aparición.

Las Figuras 107A-107E muestran la caracterización in vitro de las proteínas de la familia CpfI. La Figura 107A es un esquema para cribado de PAM in vitro usando proteínas de la familia CpfI. Se escindió una biblioteca de plásmidos que portan secuencias randomizadas de 5' PAM con proteínas individuales de la familia CpfI y sus correspondientes ARNcr. El ADN plasmídico sin escindir se purificó y secuenció para identificar los motivos PAM específicos que se agotaron. La Figura 107B indica el número de secuencias únicas que superan el umbral de significancia para combinaciones apareadas de bases en las posiciones 2 y 3 de 5' PAM para 7 – AsCpfI. La Figura 107C indica el número de PAM únicos que superan el umbral de significancia para combinaciones triples de bases en las posiciones 2, 3, y 4 de 5' PAM para 13 – LbCpfI. Las Figuras 107D-107E E y F muestran trazas de secuenciación de Sanger de 7 – diana digerido con AsCpfI (Figura 107E) y 13 – diana digerido con LbCpf1 (Figura 107F) y muestran sobreextensiones escalonadas. El agregado sin molde de una alanina adicional, indicada como N, es un artefacto de la polimerasa que se usó en la secuenciación. Los sitios de escisión están indicados por triángulos rojos. Los triángulos más pequeños indican sitios de escisión alternativos putativos. La Figura 107D-E divulga las SEQ ID NOS 1514-1519, respectivamente, en orden de aparición.

Las Figuras 108A-108F indican la eficacia de edición de genoma de células humanas en loci adicionales. Los geles de Surveyor muestran la cuantificación de la eficacia de indeles conseguida por cada proteína de la familia CpfI en los sitios blancos DNMT1 1 (Figura 108A), 2 (Figura 108B), y 4 (Figura 108C). Las Figuras 108A-108C indican la eficacia de edición de genoma de células humanas en loci adicionales y la secuenciación de Sanger de los sitios blancos escindidos de DNMT. Los geles de Surveyor muestran la cuantificación de la eficacia de indeles conseguida por cada proteína de la familia CpfI en los sitios diana EMX1 1 (Figura 108D) y 2 (Figura 108E). Distribución de indeles para sitios diana AsCpfI y LbCpfI y DNMT1 2, 3, y 4 (Figura 108F). Las barras en color cian representan la cobertura total de indeles; las barras azules representan la distribución de extremos 3' de los indeles. Para cada diana, la secuencia de PAM está en color rojo y la secuencia de los blancos está en color celeste.

Las Figuras 109A-109C representan un análisis computacional de la estructura primaria de las nucleasas CpfI que revela tres regiones distintas. Primero un dominio C-terminal similar a RuvC, que es el único dominio funcional caracterizado. Segundo una región N-terminal de hélice alfa y tercero una región mixta alfa y beta, localizada entre el dominio similar a RuvC y la región de hélice alfa.

Las Figuras 110A-110E representan un alineamiento de AsCpfI Rad50 (PDB 4W9M). Las SEQ ID NOS 1520 y 1521, respectivamente, se divulgan en orden de aparición. La Figura 110C representa un alineamiento de AsCpfI RuvC (PDB 4LD0). Las SEQ ID NOS 1522 y 1523, respectivamente, se divulgan en orden de aparición. Las Figuras 110D-110E representan un alineamiento de AsCpfI y FnCpfI que identifica al dominio Rad50 en FnCpfI. Las SEQ ID NOS 1524 y 1525, respectivamente, se divulgan en orden de aparición.

La Figura 111 representa una estructura de Rad50 (4W9M) en complejo con ADN. Los residuos que interaccionan con el ADN están resaltados (en color rojo).

La Figura 112 representa una estructura de RuvC (4LD0) en complejo con unión de Holliday. Los residuos que interaccionan con el ADN están resaltados en color rojo.

La Figura 113 representa un blast de AsCpfI alineado a una región de la recombinasa específica de sitio XerD. La región de sitio activo de XerD es LYWTGMR (SEQ ID NO: 1) en donde R es un residuo catalítico. Las SEQ ID NOS

1526-1527, respectivamente, se divulgan en orden de aparición.

La Figura 114 representa una región que está conservada en los ortólogos de Cpf1 (caja amarilla) y a pesar de que R no está conservado, hay un ácido aspártico altamente conservado (caja naranja) en posición justo C-terminal de esta región y una región cercana conservada (caja azul) con una arginina conservada en forma absoluta. El ácido aspártico es D732 en LbCpf1. Las SEQ ID NOS 1204 y 1528-1579, respectivamente, se divulgan en orden de aparición.

La Figura 115A muestran un experimento en donde se plaquearon 150.000 células HEK293T por placa de 24 pocillos, 24 horas antes de la transfección. Se transfectaron las células con 400 ng de plásmido huAsCpf1 y 100 ng de plásmido guía en tándem que comprende una secuencia guía direccionada contra GRIN28 y una direccionada contra EMX1 colocada en tándem detrás del promotor U6, usando Lipofectamin2000. Se cosecharon las células 72 horas después de la transfección y se ensayó la actividad mediada por AsCpf1 por parte de guías en tándem usando el ensayo de nucleasa de SURVEYOR.

La Figura 115B demuestra la formación de INDEL tanto en el gen GRIN28 y como en el gen EMX1.

La Figura 116 muestra la escisión con FnCpf1 de un arreglo con concentraciones crecientes de EDTA (y concentraciones decrecientes de Mg2+). La solución amortiguadora tiene TrisHCl 20 mM a pH 7 (temperatura ambiente), KCl 50 mM, e incluye un inhibidor de RNasa murina para prevenir la degradación de ARN debido a cantidades traza potenciales de ARNasa no específica arrastradas durante la purificación proteica.

Las Figuras de la presente documentación son solamente para propósitos ilustrativos y no necesariamente están dibujadas a escala.

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

En la presente solicitud se describen nuevas endonucleasas dirigidas por ARN (por ejemplo, proteínas efectoras Cpf1) que son funcionalmente diferentes de los sistemas CRISPR-Cas9 descritos previamente y por ende la terminología de los elementos asociados con estas nuevas endonucleasas está modificada de manera concordante en la presente. Los arreglos de CRISPR asociados a Cpf1 que se describen en la presente son procesados en ARNcr maduros sin el requisito de un ARNtracr adicional. Los ARNcr descritos en la presente comprenden una secuencia espaciadora (o secuencia guía) y una secuencia de repetición directa y un complejo Cpf1p-ARNcr es suficiente por sí mismo para clivar eficazmente un ADN diana. La secuencia semilla que se describe en la presente, por ejemplo la secuencia semilla de un ARN guía FnCpf1 se encuentra aproximadamente dentro de los primeros 5 nt en el extremo 5' de la secuencia espaciadora (o secuencia guía) y las mutaciones en la secuencia semilla afectan de manera adversa la actividad de clivaje del complejo de la proteína efectora Cpf1.

En general, un sistema CRISPR está caracterizado por comprender elementos que promueven la formación de un complejo CRISPR en el sitio de una secuencia diana (también denominada protoespaciadora en el contexto de un sistema CRISPR endógeno). En el contexto de la formación de un complejo CRISPR, una "secuencia diana" se refiere a una secuencia para la cual se diseña una secuencia guía que la busca como diana, por ejemplo, de la cual es complementaria, donde la hibridación entre una secuencia diana y una secuencia guía promueve la formación de un complejo CRISPR. La sección de la secuencia guía donde es importante la complementariedad con la secuencia diana para la actividad de clivaje se denomina secuencia semilla en la presente. Una secuencia diana puede comprender cualquier polinucleótido, tales como polinucleótidos de ADN o ARN, y está comprendida dentro de un locus diana de interés. En algunas formas de realización, la secuencia diana está ubicada en el núcleo o en el citoplasma de una célula. La invención descrita en la presente abarca nuevas proteínas efectoras de los sistemas CRISPR-Cas de clase 2, de las cuales Cas9 es un ejemplo de proteína efectora y por ende los términos usados en esta solicitud para describir dichas proteínas efectoras nuevas, se pueden correlacionar con los términos empleados para describir el sistema CRISPR-Cas9.

Los loci de CRISPR-Cas comprenden más de 50 familias de genes y no hay genes estrictamente universales. Por ello, no es factible un árbol evolutivo individual y se necesita un abordaje multidimensional para identificar familias nuevas. Hasta ahora, hay una identificación completa de genes *cas* de 395 perfiles para 93 proteínas Cas. La clasificación incluye perfiles de patrones de genes además de patrones de la arquitectura del locus. En la Figura 1 se propone una clasificación nueva de los sistemas CRISPR-Cas. La clase 1 incluye complejos efectores de ARNcr de múltiples subunidades (cascada) y la clase 2 incluye complejos efectores de ARNcr de subunidades individuales (tipo Cas9). En la Figura 2 se provee una organización molecular de CRISPR-Cas. En la Figura 3 se proveen estructuras de complejos efectores de Tipo I y III: arquitectura común/ancestría común a pesar de una amplia divergencia de secuencias. En la Figura 4 se muestra el CRISPR-Cas como un sistema centrado en el motivo de reconocimiento de ARN (RRM). En la Figura 5 se muestra la filogenia de Cas1, donde la recombinación de adaptación y módulos efectores ARNcr representan un aspecto importante de la evolución de CRISPR-Cas. En la Figura 6 se muestra un censo de CRISPR-Cas, específicamente una distribución de los tipos/subtipos de CRISPR-Cas entre Archaea y bacterias.

La acción del sistema CRISPR-Cas habitualmente se divide en tres etapas: (1) adaptación o integración del espaciador, (2) procesamiento del transcripto primario del locus CRISPR (pre-ARNcr) y maduración del ARNcr que incluye al espaciador y las regiones variables correspondientes a los fragmentos 5' y 3' de repeticiones de CRISPR y

(3) interferencia de ADN (o ARN). Dos proteínas, Cas1 y Cas2, que están presentes en la gran mayoría de los sistemas CRISPR-Cas conocidos, son suficientes para la inserción de espaciadores en los casetes CRISPR. Estas dos proteínas forman un complejo que es necesario para este proceso de adaptación; la actividad endonucleasa de Cas1 es necesaria para la integración del espaciador, en tanto no parece que Cas2 cumpla una función no enzimática. El complejo Cas1-Cas2 representa el módulo de "procesamiento de información" altamente conservado de CRISPR-Cas que parece ser cuasi autónomo del resto del sistema. (Véase *Annotation and Classification of CRISPR-Cas Systems,* Makarova KS, Koonin EV, *Methods Mol Biol.*, 2015;1311: 47-75).

Los sistemas de clase 2 descritos previamente, es decir el Tipo II y el Tipo V putativo, consistían de tan solo tres o cuatro genes en el operón cas, es decir, los 35 genes cas1 y cas2 que comprenden el módulo de adaptación (el par de genes cas1-cas2 no está involucrados en la interferencia), una sola proteína efectora de múltiples dominios que es responsable de la interferencia pero que también contribuye en el procesamiento y la adaptación del pre-ARNcr y a menudo un cuarto gen de funciones no caracterizadas que es dispensable en por lo menos algunos sistemas de Tipo II (y, en algunos casos, el cuarto gen es cas4 (la evidencia bioquímica o *in silico* muestran que Cas4 es una nucleasa de la superfamilia PD-(DE)xK con una agrupación de tres cisteínas C-terminales; tiene actividad exonucleasa ADNhs 5') o csn2, que codifica una ATPasa desactivada). En la mayoría de los casos, el arreglo CRISPR y un gen de una especie de ARN distinto conocido como ARNtracr, un pequeño ARN CRISPR transcodificado, son adyacentes a los operones cas de clase 2. El ARNtracr es parcialmente homólogo de las repeticiones dentro del respectivo arreglo CRISPR y es esencial para el procesamiento del pre-ARNcr que es catalizado por la RNAsa III, una enzima bacteriana ubicua que no está asociada con los loci de CRISPR-Cas.

Cas1 es la proteína más conservada presente en la mayoría de los sistemas CRISPR-Cas y evoluciona más lentamente que otras proteínas Cas. Por lo tanto, se ha utilizado la filogenia de Cas1 como una guía para la clasificación del sistema CRISPR-Cas. La evidencia bioquímica o *in silico* muestra que Cas1 es una desoxirribonucleasa dependiente de metales. La supresión de Cas1 en *E. coli* da como resultado una mayor sensibilidad al daño del ADN y a una segregación cromosómica deteriorada como se describe en "A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair", Babu M y col., *Mol Microbiol* 79: 484-502 (2011). La evidencia bioquímica o in silico muestra que Cas 2 es una ARNasa específica de regiones ricas en U y es una DNasa de hebra doble.

La divulgación describe la identificación y la ingeniería de nuevas proteínas efectoras asociadas con los sistemas CRISPR-Cas de clase 2. En una forma de realización preferida, la proteína efectora comprende un módulo efector de una sola subunidad. En una forma de realización adicional, la proteína efectora es funcional en células procariotas o eucariotas para aplicaciones *in vitro* o *ex vivo*. Un aspecto de la invención abarca métodos y algoritmos computacionales para predecir nuevos sistemas CRISPR-Cas de Clase 2 y para identificar los componentes de los mismos.

En una forma de realización de esta divulgación, se emplea un método en computadora para identificar nuevos loci CRISPR-Cas de Clase 2 que comprende los pasos de: detectar todos los cóntigos que codifican la proteína Cas1; identificar todos los genes codificantes de proteínas predichos dentro de los 20 kB del gen cas1; comparar los genes identificados con los perfiles específicos de la proteína Cas y predecir los arreglos CRISPR; seleccionar loci CRISPR-Cas candidatos no clasificados que contengan proteínas más grandes que 500 aminoácidos (>500 aa); analizar los candidatos seleccionados usando PSI-BLAST y HHPred, para así aislar e identificar nuevos loci CRISPR-Cas de Clase 2. Además de los pasos mencionados precedentemente, se puede realizar un análisis adicional de los candidatos mediante una búsqueda en bases de datos metagenómicos de homólogos adicionales.

En un aspecto de esta divulgación, la detección de todos los contigs que codifican la proteína Cas1 se realiza con GenemarkS, un programa de predicción de genes, que se describe en "GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions", John Besemer, Alexandre Lomsadze y Mark Borodovsky, *Nucleic Acids Research* (2001) 29, páginas 2607-2618.

En un aspecto de esta divulgación, la identificación de todos los genes codificantes de proteínas predichas se lleva a cabo mediante comparación de los genes identificados con perfiles de proteínas específicos de Cas y anotación de los mismos de acuerdo con la base de datos de dominios conservados (CDD) de la NCBI, que es un recurso de anotación de proteínas que consiste en una colección de modelos de alineamientos de múltiples secuencias bien anotados de dominios antiguos y proteínas de longitud completa. Se encuentran disponibles como matrices de calificación específica de la posición (PSSM) para una identificación rápida de los dominios conservados en las secuencias proteicas por medio de RPS-BLAST. El contenido de la CDD incluye dominios curados de NCBI, que incluye información acerca de la estructura tridimensional para definir explícitamente los límites de los dominios y proveer detalles sobre las relaciones de secuencia/estructura/función, así como modelos de dominios importados de numerosas bases de datos de fuentes externas (Pfam, SMART, COG, PRK, TIGRFAM). En un aspecto adicional, se predijeron arreglos CRISPR usando un programa PILER-CR, que es un software de dominio público para hallar las repeticiones CRISPR descritas en "PILER-CR: fast and accurate identification of CRISPR repeats", Edgar, R.C., *BMC Bioinformatics*, 20 de enero; 8: 18(2007).

En un aspecto adicional de esta divulgación, se conduce un análisis por casos usando PSI-BLAST (herramienta básica iterativa de búsqueda por alineamiento local específica de la posición [*Position-Specific Iterative Basic Local Alignment*

Search Tool]). PSI-BLAST deriva una matriz o perfil de calificación específica de la posición (PSSM) a partir del alineamiento de múltiples secuencias de las secuencias detectadas por encima de un puntaje umbral dado usando BLAST de proteína-proteína. Esta PSSM se usa para realizar búsquedas adicionales en la base de datos por coincidencias nuevas, y se actualiza para subsiguientes iteraciones con esas secuencias nuevas recientemente detectadas. Por lo tanto, PSI-BLAST provee un medio para detectar relaciones distantes entre proteínas.

En otro aspecto de esta divulgación, el análisis por casos se efectúa usando HHpred, un método de búsqueda en bases de datos y de predicción de estructuras de secuencias que es fácil de usar, como BLAST o PSI-BLAST, y que al mismo tiempo es mucho más sensible para hallar homólogos remotos. De hecho, la sensibilidad de HHpred compite con los servidores más poderosos para la predicción de estructuras disponibles actualmente. HHpred es el primer servidor que se basa en la comparación apareada de modelos de perfiles ocultos de Harkov (HMM). En tanto la mayoría de los métodos de búsqueda de secuencias convencionales efectúan búsquedas en bases de datos de secuencias, tal como UniProt o NR, HHpred efectúa búsquedas en bases de datos de alineamientos, como Pfam o SMART. Esto simplifica en gran medida la lista de aciertos a una cantidad de familias de secuencias en lugar de un grupo desordenado de secuencias individuales. Todas las principales bases de datos de perfiles y alineamientos disponibles al público, también están disponibles con HHpred. HHpred acepta como entrada una sola secuencia de consulta o un alineamiento múltiple. En tan solo unos pocos minutos devuelve los resultados de la búsqueda en un formato de lectura fácil similar al de PSI-BLAST. Las opciones de búsqueda incluyen un alineamiento local o global y calificación de la similitud de las estructuras secundarias. HHpred puede producir alineamientos apareados de secuencias de consulta-molde, alineamientos múltiples combinados de consulta-molde (por ejemplo, para búsquedas transitivas), así como modelos estructurales tridimensionales calculados con el software MODELLER de los alineamientos HHpred.

El término "sistema de direccionamiento de ácidos nucleicos", en donde el ácido nucleico es ADN o ARN, y en algunos aspectos también puede comprender híbridos de ADN-ARN o derivados de los mismo, se refiere en su conjunto a transcriptos y otros elementos involucrados en la expresión o dirección de la actividad de ADN o el direccionamiento de genes ("Cas") asociados a ARN CRISPR, que pueden incluir secuencias que codifican una proteína Cas de direccionamiento de ADN o ARN y un ARN guía de direccionamiento de ADN o ARN que comprende una secuencia de ARN CRISPR (ARNcr) y (en el sistema CRISPR-Cas9, pero no en todos los sistemas) una secuencia del sistema de ARN CRISPR-Cas transactivante (ARNtracr) u otras secuencias y transcriptos de un locus CRISPR de direccionamiento de ADN o de ARN. En los sistemas de endonucleasas guiadas por ARN de direccionamiento a ADN Cpf1 que se describen en la presente, no se requiere una secuencia ARNtracr. En general, un sistema de direccionamiento de ARN se caracteriza por comprender elementos que promueven la formación de un complejo de direccionamiento de ARN en el sitio de una secuencia de ARN diana. En el contexto de la formación de un complejo de direccionamiento de ADN o de ARN, una "secuencia diana" se refiere a una secuencia de ADN o de ARN para la cual se diseña un ARN guía de direccionamiento de ADN o de ARN complementario, donde la hibridación entre una secuencia diana y el ARN guía de direccionamiento de ARN promueve la formación de un complejo de direccionamiento de ARN. En algunas formas de realización, la secuencia diana está ubicada en el núcleo o en el citoplasma de una célula.

La invención como se detalla en las reivindicaciones se refiere a novedosos sistemas de direccionamiento de ADN, también denominados CRISPR-Cas de direccionamiento de ADN o el sistema CRISPR-Cas de direccionamiento de ADN de la presente solicitud, se basan en proteínas Cas Tipo V(por ejemplo, el subtipo V-A y el subtipo V-B) ya identificadas que no requieren de la generación de proteínas personalizadas para buscar como blancos a secuencias de ADN específicas sino que en su lugar se puede programar una sola proteína efectora o enzima mediante una molécula de ARN para que reconozca un diana de ADN específico; en otras palabras, la enzima se puede reclutar para un diana de ADN específico usando dicha molécula de ARN. Algunos aspectos de la invención se relacionan en particular con los sistemas Cpfl CRISPR guiados por ARN para el direccionamiento de ADN.

En la presente divulgación, los nuevos sistemas de direccionamientos de ARN, también denominados CRISPR-Cas de ARN o de direccionamiento de ARN, o el sistema de direccionamiento de ARN del sistema CRISPR-Cas de la presente solicitud, se basan en proteínas Cas de Tipo VI identificadas, que no requieren la generación de proteínas personalizadas para buscar como blancos secuencias de ARN específicas sino que en su lugar se puede programar una sola enzima mediante una molécula de ARN para que reconozca un diana de ARN específico; en otras palabras la enzima puede ser reclutada para un diana de ARN específico usando dicha molécula de ARN.

Los sistemas de direccionamiento de ácidos nucleicos, los sistemas de vectores, los vectores y las composiciones que se describen en la presente se puede usar en varias aplicaciones de direccionamiento de ácidos nucleicos, para alterar o modificar la síntesis de un producto genético tal como una proteína, el clivaje de ácidos nucleicos, la edición ácidos nucleicos, el corte y empalme de ácidos nucleicos; el tráfico de los ácidos nucleicos diana, seguimiento de los ácidos nucleicos diana, aislamiento de los ácidos nucleicos diana, visualización de los ácidos nucleicos diana, etc.

Según se usa en la presente, una proteína Cas o una enzima CRISPR se refiere a cualquiera de las proteínas presentadas en la clasificación nueva de los sistemas CRISPR-Cas. En una forma de realización ventajosa, la presente invención engloba proteínas efectoras identificadas en loci CRISPR-Cas Tipo V, por ejemplo los loci que codifican Cpfl indicados como subtipo V-A. Actualmente, los loci del subtipo V-A abarcan cas1, cas2, un gen distinto denominado *cpfl* y un arreglo CRISPR. La Cpfl (proteína Cpfl asociada a CRISPR, subtipo PREFRAN) es una proteína grande (de

aproximadamente 1300 aminoácidos) que contiene un dominio nucleasa tipo RuvC que es homólogo del correspondiente dominio de Cas9, junto con una contraparte de la agrupación rica en arginina característica de Cas9. Sin embargo, CpfI no contiene al dominio nucleasa HNH que está presente en todas las proteínas Cas9, y el dominio tipo RuvC es contiguo en la secuencia de CpfI, a diferencia de Cas9 donde contiene insertos largos que incluyen el dominio HNH. Por lo tanto, en formas de realización particulares, la enzima CRISPR-Cas solamente comprende un dominio nucleasa tipo RuvC.

El gen de CpfI está presente en varios genomas bacterianos diversos, típicamente en el mismo locus que los genes cas1, cas2 y cas4 y un casete CRISPR (por ejemplo, 39 FNFX1_1431-FNFX1_1428 de Francisella, véase Fx1 novicida). Por consiguiente, el diseño de este nuevo sistema CRISPR-Cas putativo parece ser similar al del tipo II-B. Además, de manera similar a Cas9, la proteína CpfI contiene una región C-terminal fácilmente identificable que es homóloga del transposón ORF-B e incluye una nucleasa tipo RuvC activa, una región rica en arginina y un dedo de Zn (ausente en Cas9). Sin embargo, a diferencia de Cas9, CpfI también está presente en varios genomas fuera de un contexto CRISPR-Cas y su similitud relativamente alta con ORF-B sugiere que podría ser un componente transposón. Se sugirió que si este era un sistema CRISPR-Cas genuino y CpfI es un análogo funcional de Cas9 se trataría de un nuevo tipo de CRISPR-Cas, es decir, el tipo V (véase *Annotation and Classification of CRISPR-Cas Systems*; Makarova KS, Koonin EV, *Methods Mol Biol.*, 2015; 1311: 47-75). Sin embargo, según se describe en la presente, se ha indicado que CpfI pertenece al subtipo V-A para diferenciarla de C2c1p, que no tiene una estructura de dominio idéntica y por ende se indica que es del subtipo V-B.

En una forma de realización ventajosa, la presente invención abarca composiciones y sistemas que comprenden proteínas efectoras identificadas en los loci CpfI indicados como del subtipo V-A.

Algunos aspectos de la invención también abarcan métodos y usos de las composiciones y los sistemas que se describen en la presente como se detalla en las reivindicaciones en la ingeniería de genomas, por ejemplo, para alterar o manipular la expresión de uno o más genes o dichos uno o más productos genéticos, en células procariotas o eucariotas, *in vitro* o *ex vivo*.

En algunas formas de realización de la invención, los términos ARNcr maduro y ARN guía y ARN guía individual se usan indistintamente, al igual que en los documentos citados con anterioridad, tal como WO 2014/093622 (PCT/US2013/074667). En general, una secuencia guía es cualquier secuencia de polinucleótidos que tiene una complementariedad suficiente con una secuencia polinucleotídica diana como para hibridarse con dicha secuencia diana y dirigir la unión específica de la secuencia de un complejo CRISPR con la secuencia diana. En algunas realizaciones, el grado de complementariedad entre una secuencia guía y su secuencia diana correspondiente, cuando se alinean de manera óptima utilizando un algoritmo de alineación adecuado, es de aproximadamente o más de aproximadamente un 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97,5%, 99% o más. El alineamiento óptimo se podrá determinar utilizando cualquier algoritmo adecuado para alinear secuencias, cuyos ejemplos no taxativos incluyen el algoritmo de Smith-Waterman, el algoritmo de Needleman-Wunsh, los algoritmos basados en la transformada de Burrows-Wheeler (por ejemplo, el alineador de Burrows Wheeler), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; disponible en www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (disponible en soap.genomics.org.en) y Maq (disponible en maq.sourceforge.net). En algunas formas de realización, una secuencia guía tiene una longitud de aproximadamente o más de aproximadamente 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 o más nucleótidos. En algunas formas de realización, una secuencia guía tiene una longitud menor que aproximadamente 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 o menos nucleótidos. Preferentemente, la secuencia guía es de 10-30 nucleótidos de longitud. La capacidad de una secuencia guía para dirigir una unión específica de la secuencia de un complejo CRISPR a una secuencia diana se podrá evaluar mediante cualquier ensayo adecuado. Por ejemplo, se pueden proporcionar los componentes de un sistema CRISPR suficientes para formar un complejo CRISPR, incluida la secuencia guía que se va a evaluar, a una célula huésped que contiene la secuencia diana correspondiente, tal como por transfección con vectores que codifican los componentes de la secuencia CRISPR, y a continuación evaluar el clivaje preferido dentro de la secuencia diana, tal como mediante el ensayo Surveyor como se describe en la presente. De manera similar, se podrá evaluar el clivaje de una secuencia polinucleotídica diana en un tubo de ensayo proporcionando la secuencia diana, los componentes de un complejo CRISPR, incluida la secuencia guía a evaluar y una secuencia guía de control diferente de la secuencia guía de prueba, y comparando la unión o el índice de clivaje en la secuencia diana entre las reacciones de la secuencia guía de control y de prueba. Se pueden emplear otros ensayos, y serán determinados por los especialistas en el arte. Asimismo es posible seleccionar una secuencia guía para buscar como diana a cualquier secuencia diana. En algunas formas de realización, una secuencia diana es una secuencia contenida en el genoma de una célula. Las secuencias diana ilustrativas incluyen aquellas que son únicas en el genoma diana.

En general, y en toda esta Descripción, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido. Los vectores incluyen, de manera no taxativa, moléculas de ácido nucleico que son de una hebra, de hebra doble o parcialmente bicatenarias; moléculas de ácido nucleico que comprenden uno o más extremos libres o sin extremos libres (por ejemplo, circulares); moléculas de ácido nucleico que comprenden ADN, ARN o ambos; y otras variedades de polinucleótidos conocidos en el arte. Otro tipo de vector es un vector viral, donde en dicho vector están presentes las secuencias de ADN o ARN de origen viral para empaquetarlo en un virus (por ejemplo, retrovirus, retrovirus de replicación defectuosa, adenovirus, adenovirus de replicación defectuosa y virus adenoasociados). Los vectores virales también incluyen polinucleótidos llevados por un

virus para la transfección de una célula huésped. Algunos vectores se pueden replicar de manera autónoma en la célula huésped en la que se introducen (por ejemplo, vectores bacterianos que tienen un origen de replicación bacteriano y vectores episómicos de mamífero). Otros vectores (por ejemplo, vectores no episómicos de mamífero) se integran en el genoma de la célula huésped tras la introducción en la célula huésped y, de esta manera, se replican junto con el genoma huésped. Además, algunos vectores pueden dirigir la expresión de genes a los cuales están ligados operativamente. Este tipo de vectores se denominan "vectores de expresión" en la presente. Los vectores que darán como resultado la expresión en una célula eucariota se pueden denominar "vectores de expresión eucariotas" en la presente. Los vectores de expresión comunes que son de utilidad en las técnicas de ADN recombinante a menudo se encuentran en la forma de plásmidos.

Los vectores de expresión recombinantes pueden comprender un ácido nucleico de la invención en una forma adecuada para la expresión del ácido nucleico en una célula huésped, lo que significa que los vectores de expresión recombinantes incluyen uno o más elementos reguladores que se pueden seleccionar en función de las células huésped que se van a utilizar para la expresión, que están ligados operativamente a la secuencia de ácidos nucleicos a expresar. En un vector de expresión recombinante, el término "ligado operativamente" significa que la secuencia de nucleótidos de interés está ligada a dichos uno o más elementos reguladores de una manera que permite la expresión de la secuencia de nucleótidos (por ejemplo, en un sistema de transcripción/traducción in vitro o en una célula huésped cuando el vector se introduce en la célula huésped).

La expresión "elemento regulador" incluye promotores, potenciadores, sitios internos de entrada al ribosoma (IRES, por sus siglas en inglés) y otros elementos de control de la expresión (por ejemplo, señales de terminación de la transcripción tales como señales de poliadenilación y secuencias poli-U). Dichos elementos reguladores se describen, por ejemplo, en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Los elementos reguladores incluyen aquellos que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huésped y aquellos que dirigen la expresión de la secuencia de nucleótidos únicamente en determinadas células huésped (por ejemplo, secuencias reguladoras específicas de tejidos). Un promotor específico de tejidos puede dirigir la expresión principalmente en un tejido deseado de interés, tales como músculos, neuronas, hueso, piel, sangre, órganos específicos (por ejemplo, hígado, páncreas) o tipos celulares particulares (por ejemplo, linfocitos). Los elementos reguladores también pueden dirigir la expresión de una manera dependiente del tiempo, tal como de manera dependiente del ciclo celular o dependiente de la etapa de desarrollo, que también puede ser específica, o no, del tipo celular o del tejido. En algunas formas de realización, un vector comprende uno o más promotores pol III (por ejemplo, 1, 2, 3, 4, 5 o más promotores pol III), uno o más promotores pol II (por ejemplo, 1, 2, 3, 4, 5 o más promotores pol II), uno o más promotores pol I (por ejemplo, 1, 2, 3, 4, 5 o más promotores pol I) o combinaciones de los mismos. Los ejemplos de los promotores pol III incluyen, pero en un sentido no taxativo, los promotores U6 y H1. Los ejemplos de promotores pol II incluyen, pero en un sentido no taxativo, el promotor LTR retroviral del virus del sarcoma de Rous (RSV, por sus siglas en inglés) (opcionalmente con el potenciador del RSV), el promotor del citomegalovirus (CMV) (opcionalmente con el potenciador del CMV) [véase, por ejemplo, Boshart y col., *Cell*, 41: 521-530 (1985)], el promotor del SV40, el promotor de la dihidrofolato-reductasa, el promotor de la β-actina, el promotor de la fosfoglicerol-quinasa (PGK, por sus siglas en inglés) y el promotor EF1α. El término "elemento regulador" también abarca elementos potenciadores, tales como WPRE; potenciadores del CMV; el segmento R-U5' en LTR del HTLV-I (*Mol. Cell. Biol.*, volumen 8(1), páginas 466-472, 1988); el potenciador del SV40; y la secuencia intrónica entre los exones 2 y 3 de la β-globina de conejo (*Proc. Natl. Acad. Sci. USA.*, volumen 78(3), páginas 1527-31, 1981). Los especialistas en el arte podrán apreciar que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped a transformar, el nivel de expresión deseado, etc. Se puede introducir un vector en las células huésped para de esta manera producir transcritos, proteínas o péptidos, inclusive péptidos o proteínas de fusión, codificados por ácidos nucleicos como se describe en la presente (por ejemplo, transcritos de los grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR), proteínas, enzimas, formas mutantes de los mismos, proteínas de fusión de los mismos, etc.).

Los vectores ventajosos incluyen lentivirus y virus adenoasociados, y también se pueden seleccionar tipos de dichos vectores para el direccionamiento a tipos celulares particulares.

Según se usa en la presente, el término "ARNcr" o "ARN guía" o "ARN guía simple" o "ARNgs" o "uno o más componentes de ácidos nucleicos" de una proteína efectora del locus de CRISPR-Cas de Tipo V o de Tipo VI comprende cualquier secuencia de polinucleótidos que tenga complementariedad suficiente con una secuencia de ácido nucleico diana como para hibridarse con la secuencia de ácidos nucleicos diana y dirigir la unión específica de la secuencia de un complejo de direccionamiento de ácidos nucleicos a la secuencia de ácidos nucleicos diana. En algunas formas de realización, el grado de complementariedad, cuando se alinean de manera óptima utilizando un algoritmo de alineamiento adecuado, es de aproximadamente o más de aproximadamente un 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97,5%, 99% o más. El alineamiento óptimo se podrá determinar utilizando cualquier algoritmo adecuado para alinear secuencias, cuyos ejemplos no taxativos incluyen el algoritmo de Smith-Waterman, el algoritmo de Needleman-Wunsh, los algoritmos basados en la transformada de Burrows-Wheeler (por ejemplo, el alineador de Burrows Wheeler), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; disponible en www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (disponible en soap.genomics.org.cn) y Maq (disponible en maq.sourceforge.net). La capacidad de una secuencia guía (en un ARN guía de direccionamiento a ácidos nucleicos) para dirigir la unión específica de secuencias de un complejo de direccionamiento a ácidos nucleicos a una secuencia de ácido nucleico diana se puede evaluar mediante cualquier ensayo adecuado. Por ejemplo, los componentes de un

sistema CRISPR de direccionamiento de ácidos nucleicos suficientes para formar un complejo de direccionamiento de ácidos nucleicos, que incluye la secuencia guía a evaluar, se pueden proporcionar a una célula huésped que contiene la secuencia de ácidos nucleicos diana correspondiente, tal como mediante transfección con vectores que codifican los componentes del complejo de direccionamiento de ácidos nucleicos, seguido por una evaluación del direccionamiento preferencial (por ejemplo, clivaje) dentro de la secuencia de ácidos nucleicos diana, tal como mediante el ensayo Surveyor, que se describe en la presente. De manera similar, se puede evaluar el clivaje de una secuencia de ácidos nucleicos diana en un tubo de ensayo mediante suministro de la secuencia de ácidos nucleicos diana, los componentes de un complejo de direccionamiento de ácidos nucleicos, que incluyen la secuencia guía a evaluar y una secuencia guía de control diferente de la secuencia guía de prueba, y comparación de la unión o el índice de clivaje en la secuencia diana entre las reacciones de la secuencia guía de prueba y de control. Se pueden emplear otros ensayos, y serán determinados por los especialistas en el arte. Se puede seleccionar una secuencia guía, y por ende un ARN guía de direccionamiento de ácidos nucleicos, para buscar como diana a cualquier secuencia de ácidos nucleicos diana. La secuencia diana puede ser ADN. La secuencia diana puede ser cualquier secuencia de ARN. En algunas formas de realización, la secuencia diana puede ser una secuencia dentro de una molécula de ARN que se selecciona del grupo que consiste en ARN mensajero (ARNm), pre-ARNm, ARN ribosómico (ARNr), ARN de transferencia (ARNt), micro-ARN (miARN), ARN pequeño de interferencia (ARNpi), ARN nuclear pequeño (ARNnp), ARN nucleolar pequeño (ARNnop), ARN de hebra doble (ARNhd), ARN no codificante (ARNnc), ARN no codificante largo (ARNncl) y ARN citoplasmático pequeño (ARNcp). En algunas formas de realización preferidas, la secuencia diana puede ser una secuencia dentro de una molécula de ARN que se selecciona del grupo que consiste en ARNm, pre-ARNm y ARNr. En algunas formas de realización preferidas, la secuencia diana puede ser una secuencia dentro de una molécula de ARN seleccionada del grupo que consiste en ARNnc y ARNncl. En algunas formas de realización más preferidas, la secuencia diana puede ser una secuencia dentro de una molécula de ARNm o una molécula de pre-ARNm.

En algunas formas de realización, se selecciona un ARN guía de direccionamiento de ácidos nucleicos para reducir el grado de estructura secundaria dentro del ARN guía de direccionamiento de ARN. En algunas formas de realización, aproximadamente o menos de aproximadamente un 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1% o menos de los nucleótidos del ARN guía de direccionamiento de ácidos nucleicos participan en un apareamiento de bases autocomplementario cuando se pliegan de manera óptima. El plegamiento óptimo se puede determinar mediante cualquier algoritmo de plegamiento de polinucleótidos adecuado. Algunos programas de basan en el cálculo de la energía libre mínima de Gibbs. Un ejemplo de un algoritmo de este tipo es mFold, tal como lo describen Zuker y Stiegler (*Nucleic Acids Res.,* 9 (1981), 133-148). 9 (1981), 133-148). Otro ejemplo de un algoritmo para plegamientos es el servidor de internet en línea RNAfold, desarrollado en el Instituto de Química Teórica de la Universidad de Viena, que utiliza el algoritmo de predicción de estructura centroide (véase, por ejemplo, A.R. Gruber y col., 2008, *Cell* 106(1): 23-24; y PA Carr y GM Church, 2009, *Nature Biotechnology* 27(12): 1151-62).

La secuencia "ARNtracr", o términos análogos, incluye cualquier secuencia de polinucleótidos que tenga suficiente complementariedad con una secuencia ARNcr como para hidridarse. Según se indicó con anterioridad en la presente, en las formas de realización de la presente invención, no se requiere de un ARNtracr para la actividad de clivaje de los complejos de la proteína efectora Cpf1.

Los solicitantes también realizaron un experimento de desafío para verificar el direccionamiento de ADN y la capacidad de clivaje de una proteína de Tipo V/Tipo VI, tal como Cpf1/C2c1 o C2c2. Este experimento es muy similar al trabajo en *E. coli* para la expresión heteróloga de StCas9 (Sapranauskas, R. y col. *Nucleic Acids Res* 39, 9275-9282 (2011)). Los solicitantes introducen un plásmido que contiene tanto un gen PAM como un gen de resistencia en la *E. coli* heteróloga, y luego la plaquean sobre el antibiótico correspondiente. Si hay clivaje del ADN del plásmido, los Solicitantes no observan colonias viables.

Con mayor detalle, el ensayo se conduce de la siguiente manera para un ADN diana. En este ensayo se utilizan dos cepas de E. coli. Uno lleva un plásmido que codifica el locus endógeno de la proteína efectora de la cepa bacteriana. La otra cepa porta un plásmido vacío (por ejemplo, pACYC184, cepa de control). Todas las secuencias de PAM posibles de 7 u 8 pb se presentan en un plásmido de resistencia a un antibiótico (el pUC19 con un gen de resistencia a la ampicilina). El PAM está ubicado cerca de la secuencia del protoespaciador 1 (el ADN diana para el primer espaciador en el locus endógeno de la proteína efectora). Se clonaron dos bibliotecas de PAM. Una tiene 8 pb aleatorios 5' del protoespaciador (por ejemplo, un total de 65536 secuencias de PAM diferentes = complejidad). La otra biblioteca tiene 7 pb aleatorios en posición 3' del protoespaciador (por ejemplo, la complejidad total es de 16384 PAM diferentes). Ambas bibliotecas se clonaron para que tuvieran en promedio 500 plásmidos por PAM posible. La cepa de prueba y la cepa de control se transformaron con bibliotecas 5' PAM y 3'PAM en transformaciones separadas y las células transformadas se plaquearon por separado sobre placas con ampicilina. El reconocimiento y subsiguiente corte/interferencia con el plásmido vuelve a la célula vulnerable a la ampicilina e impide el crecimiento. Aproximadamente 12 h después de la transformación, se cosecharon todas las colonias formadas por las cepas de prueba y de control y se aisló el ADN del plásmido. Se utilizó el ADN del plásmido como molde para la amplificación por PCR y la subsiguiente secuenciación profunda. La representación de todos los PAM de las bibliotecas sin transformar mostró la representación esperada de PAM en las células transformadas. La representación de todos los PAM hallados en las cepas de control mostró la representación real. La representación de todos los PAM de la cepa de prueba mostró los PAM que no son reconocidos por la enzima y una comparación con la cepa de control permite extraer la secuencia del PAM agotado.

En algunas formas de realización de los sistemas CRISPR-Cas9, el grado de complementariedad entre la secuencia ARNtracr y la secuencia ARNcr tiene lugar a lo largo de la extensión del más corto de los dos cuando se alinean de manera óptima. Según se describe en la presente, en las formas de realización de la presente invención, no se requiere un ARNtracr. En algunas formas de realización de los sistemas CRISPR-Cas descritos previamente (por ejemplo, los sistemas CRISPR-Cas9), en los diseños de ARN guía sintéticos quiméricos (ARNgs) se pueden incorporar por lo menos 12 pb de una estructura dúplex entre el ARNcr y el ARNtracr; sin embargo en los sistemas Cpfl CRISPR descritos en la presente, dichos ARN quiméricos (ARNqui) no son posibles ya que el sistema no emplea un ARNtracr.

Para minimizar la toxicidad y el efecto fuera del diana, será importante controlar la concentración suministrada del ARN guía de direccionamiento de ácidos nucleicos. Se pueden determinar las concentraciones óptimas del ARN guía de direccionamiento de ácido nucleico evaluando diferentes concentraciones en un modelo en células o en animales eucariotas no humanos y utilizando una secuenciación profunda para analizar el grado de modificación en los loci genómicos fuera del diana potenciales. Se debería elegir la concentración que proporcione el nivel más elevado de modificación en el diana y a la vez minimice el nivel de modificación fuera del diana para el suministro *in vivo*. El sistema de direccionamiento de ácidos nucleicos deriva ventajosamente de un sistema CRISPR de tipo V/tipo VI. En algunas formas de realización, se obtiene uno o más elementos de un sistema de direccionamiento de ácidos nucleicos a partir de un organismo particular que comprende un sistema de direccionamiento de ARN endógeno. Un sistema de direccionamiento de ARN como se describe en la presente incluye un sistema CRISPR de Tipo V/Tipo VI. Los ejemplos no taxativos de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también conocida como Csn1 y Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homólogos de las mismas o versiones modificadas de las mismas. En algunas formas de realización, la proteína de Tipo V/Tipo VI, tal como Cpf1/C2c1/C2c2 a las cuales se hace referencia en la presente, también abarca un homólogo o un ortólogo de una proteína de Tipo V/Tipo VI, tal como Cpf1/C2c1/C2c2. Los términos "ortólogo" (también indicado como "ortóloga" en la presente) y "homólogo" (también indicado como "homóloga" en la presente) son bien conocidos en el arte. A modo de lineamiento adicional, un "homólogo" de una proteína, según se usa en la presente, es una proteína de la misma especie que cumple la misma función o una función similar a la de la proteína de la cual es homóloga. Las proteínas homólogas pueden estar, pero no necesariamente, relacionadas estructuralmente, o sólo están relacionadas estructuralmente de manera parcial. Un "ortólogo" de una proteína, según se usa en la presente, es una proteína de una especie diferente que cumple la misma función o una función similar que la proteína de la cual es ortóloga. Las proteínas ortólogas pueden, pero no necesariamente, estar relacionadas estructuralmente, o la relación estructural solamente es parcial. Los homólogos y los ortólogos se pueden identificar mediante modelado por homología (véase, por ejemplo, Greer, *Science* volumen 228 (1985) 1055, y Blundell y col., *Eur J Biochem* volumen 172 (1988), 513) o con "BLAST estructural" (Dey F, Cliff Zhang Q, Petrey D, Honig B. *Toward a "structural BLAST": using structural relationships to infer function*; *Protein Sci.*, abril de 2013; 22(4): 359-66. doi: 10.1002/pro.2225.). Véase también Shmakov y col., (2015) por la aplicación en el campo de loci de CRISPR-Cas. Las proteínas homólogas pueden estar, pero no necesariamente, relacionadas estructuralmente, o sólo están relacionadas estructuralmente de manera parcial. En formas de realización particulares, el homólogo u ortólogo de Cpfl al que se hace referencia en la presente tiene una homología o identidad de secuencia de por lo menos un 80%, más preferiblemente por lo menos un 85%, aún más preferiblemente de por lo menos un 90%, tal como, por ejemplo, al menos un 95% con Cpfl. En formas de realización adicionales, el homólogo u ortólogo de Cpfl al que se hace referencia en la presente tiene una identidad de secuencia de por lo menos un 80%, más preferiblemente por lo menos un 85%, aún más preferiblemente de por lo menos un 90%, tal como, por ejemplo, al menos un 95% con Cpfl de tipo salvaje. Cuando la Cpfl contiene una o más mutaciones (está mutada), el homólogo u ortólogo de dicha Cpfl al que se hace referencia en la presente tiene una identidad de secuencia de por lo menos un 80%, más preferiblemente por lo menos un 85%, aún más preferiblemente por lo menos un 90%, tal como, por ejemplo, al menos un 95% con la Cpfl mutada.

En una forma de realización, la proteína Cas tipo V puede ser un ortólogo de un organismo de un género que incluye, pero en un sentido no taxativo, *Acidaminococcus* sp, *Lachnospiraceae bacterium o Moraxella bovoculi*; en formas de realización particulares, la proteína Cas de tipo V puede ser un ortólogo de un organismo de una especie que incluye, pero en un sentido no taxativo *Acidaminococcus sp. BV3L6*; *Lachnospiraceae bacterium ND2006* (LbCpfl) o *Moraxella bovoculi 237.* En formas de realización particulares, es un homólogo u ortólogo de Cpf1 al que se hace referencia en la presente, tiene una homología o identidad de secuencia de por lo menos un 80%, más preferiblemente de por lo menos un 85%, aún más preferiblemente de por lo menos un 90%, tal como por ejemplo de por lo menos un 95%, con una o más de las secuencias divulgadas en la presente. En formas de realización adicionales, el homólogo u ortólogo de una Cpf al que se hace referencia en la presente, tiene una identidad de secuencia de por lo menos un 80%, más preferiblemente de por lo menos un 85%, aún más preferiblemente de por lo menos un 90%, tal como por ejemplo de por lo menos un 95%, con la FnCpfl, AsCpfl o LbCpfl.

En formas de realización particulares, la proteína Cpf1 de la invención tiene una homología o identidad de secuencia de por lo menos un 60%, más particularmente por lo menos 70, tal como por lo menos un 80%, más preferiblemente por lo menos un 85%, aún más preferiblemente por lo menos un 90%, tal como por ejemplo al menos un 95% con FnCpfl, AsCpfl o LbCpfl. En formas de realización adicionales, la proteína Cpfl a la cual se hace referencia en la presente tiene una identidad de secuencia de por lo menos un 60%, tal como por lo menos 70%, más particularmente por lo menos 80%, más preferiblemente por lo menos 85%, aún más preferiblemente por lo menos 90%, tal como por ejemplo por lo menos un 95% con la AsCpfl o LbCpfl de tipo salvaje. En formas de realización particulares, la proteína

Cpfl de la presente invención tiene menos de un 60% de identidad de secuencia con FnCpfl. El especialista comprenderá que esto incluye formas truncadas de la proteína Cpfl, por lo cual la identidad de secuencia se determina sobre la extensión de la forma truncada.

Algunos métodos para identificar ortólogos de las enzimas del sistema CRISPR-Cas pueden incluir la identificación de secuencias tracr en los genomas de interés. La identificación de secuencias tracr se puede relacionar con los siguientes pasos de: Buscar repeticiones directas o secuencias que se corresponden con tracr en una base de datos para identificar una región CRISPR que comprenda una enzima CRISPR. Buscar secuencias homólogas en la región CRISPR que flanquea la enzima CRISPR en ambas direcciones en el sentido de la lectura y antisentido. Buscar terminadores de la transcripción y estructuras secundarias. Identificar cualquier secuencia que no sea una repetición directa o una secuencia que se corresponde con tracr pero que tenga más de un 50% de identidad con la repetición directa o la secuencia que se corresponde con tracr como una secuencia tracr potencial. Tomar la secuencia tracr potencial y analizar según las secuencias terminadoras de la transcripción asociadas a la misma. En este sistema, los datos de secuenciación del ARN revelaron que los potenciales ARNtracr identificados mediante computadora solamente presentaban una expresión baja, lo cual sugería la posibilidad que el ARNtracr no sería necesario para la función del sistema de la presente. Después de una evaluación adicional del locus FnCpfl y de considerar los resultados del clivaje *in vitro*, los Solicitantes concluyeron que el clivaje del ADN diana por un complejo de proteína efectora Cpfl no requiere un ARNtracr. Los Solicitantes determinaron que los complejos de la proteína efectora Cpfl que solamente comprenden una proteína efectora Cpfl y un ARNcr (ARN guía que comprende una secuencia de repetición directa y una secuencia guía) eran suficientes para clivar el ADN diana.

Se podrá apreciar que cualquiera de las funcionalidades que se describen en la presente se puede obtener mediante ingeniería genética en enzimas CRISPR procedentes de otros ortólogos, incluyendo enzimas quiméricas que comprenden fragmentos de múltiples ortólogos. Los ejemplos de dichos ortólogos se describen en otra parte en la presente. Por lo tanto, las enzimas quiméricas pueden comprender fragmentos de enzimas CRISPR ortólogas de un organismo que incluye, pero de manera no taxativa, *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor, Mycoplasma* y *Campylobacter*. Una enzima quimérica puede comprender un primer fragmento y un segundo fragmento, y los fragmentos pueden ser de enzimas CRISPR ortólogas de organismos de los géneros que se han mencionado en la presente o de especies que se han mencionado en la presente, pero ventajosamente los fragmentos son de enzimas CRISPR ortólogas de diferentes especies.

En algunas formas de realización, la proteína efectora de direccionamiento de ARN de Tipo V, en particular la proteína Cpf1/C2c1 a la cual se hace referencia en la presente, también abarca una variante funcional de Cpf1/C2c1 o un homólogo o un ortólogo de la misma. Una "variante funcional" de una proteína según se usa en la presente se refiere a una variante de dicha proteína que retiene por lo menos parcialmente la actividad de dicha proteína. Las variantes funcionales pueden incluir mutantes (que pueden ser mutantes de inserción, supresión, o reemplazo), incluyendo formas polimórficas, etc. Las variantes funcionales también incluyen los productos de fusión de una proteína tal con otro ácido nucleico, proteína, polipéptido o péptido, habitualmente no relacionado. Las variantes funcionales pueden ser de origen natural o pueden ser artificiales. Las formas de realización ventajosas pueden incluir proteínas efectoras de direccionamiento de ARN de Tipo V obtenidas mediante ingeniería genética o no naturales, por ejemplo, Cpf1/C2c1 o un ortólogo u homólogo de las mismas.

En una forma de realización, se pueden optimizar los codones de las moléculas de ácidos nucleicos que codifican la proteína efectora de direccionamiento de ARN de Tipo V, en particular Cpf1/C2c1 o un ortólogo u homólogo de la misma, para su expresión en una célula eucariótica. La célula eucariota puede ser como se describe en la presente. Las moléculas de ácido nucleico pueden ser moléculas manipuladas mediante ingeniería genética o no naturales.

En una forma de realización, la proteína efectora de direccionamiento de ARN de Tipo V, en particular Cpf1/C2c1 o un ortólogo u homólogo de la misma, puede comprender una o más mutaciones (y por lo tanto las moléculas de ácidos nucleicos que codifican la misma pueden contener una o más mutaciones. Las mutaciones pueden ser mutaciones introducidas artificialmente y pueden incluir, pero de manera no taxativa, una o más mutaciones en un dominio catalítico. Algunos ejemplos de dominios catalíticos con referencia a una enzima Cas9 pueden incluir, pero de manera no taxativa, los dominios RuvC I, RuvC II, RuvC III y HNH.

En una forma de realización, la proteína de Tipo V, tal como Cpf1/C2c1 o un ortólogo u homólogo de la misma, se puede utilizar como una proteína genérica de unión a ácidos nucleicos fusionada con un dominio funcional o ligada operativamente a la misma. Los ejemplos de dominios funcionales pueden incluir, pero en un sentido no taxativo, un iniciador de la traducción, un activador de la traducción, un represor de la traducción, nucleasas, en particular ribonucleasas, un *spliceosome*, esferas, un dominio inducible/controlable por luz o un dominio inducible/controlable químicamente.

En algunas formas de realización, la proteína efectora de direccionamiento de ácidos nucleicos no modificada puede presentar actividad de clivaje. En algunas formas de realización, la proteína efectora de direccionamiento de ARN puede dirigir el clivaje de una o ambas hebras de ácido nucleico (ADN o ARN) en la ubicación de una secuencia diana o cerca de la misma, por ejemplo dentro de la secuencia diana y/o dentro del complemento de la secuencia diana o

en las secuencias asociadas a la secuencia diana. En algunas formas de realización, la proteína de direccionamiento de ácidos nucleicos puede dirigir el clivaje de una o ambas hebras de ADN o ARN dentro de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500 o más pares de bases del primer o último nucleótido de una secuencia diana. En algunas formas de realización, el clivaje puede ser escalonado, es decir, que genera extremos adhesivos. En algunas formas de realización, el clivaje es un corte escalonado con una sobreextensión 5'. En algunas formas de realización, el clivaje puede ser un corte escalonado con una sobreextensión en 5' de entre 1 y 5 nucleótidos, preferiblemente de 4 o 5 nucleótidos. En algunas formas de realización, el sitio de clivaje está distante del PAM, por ejemplo, el clivaje tiene lugar después del 18º nucleótido en la hebra no diana y después del 23er nucleótido en la hebra diana (Figura 97A). En algunas formas de realización, el sitio de clivaje tiene lugar después del 18º nucleótido (contado desde el PAM) en la hebra no diana y después del 23er nucleótido (contado desde el PAM) en la hebra diana (Figura 97A). En algunas formas de realización, un vector codifica una proteína efectora de direccionamiento de ácidos nucleicos que puede estar mutada con respecto a una correspondiente enzima de tipo salvaje de manera tal que la proteína de direccionamiento de ácidos nucleicos mutada carece de la capacidad para clivar una o ambas hebras de ADN o ARN de un polinucleótido diana que contiene una secuencia diana. A modo de ejemplo adicional, se pueden mutar dos o más dominios catalíticos de una proteína Cas (por ejemplo, RuvC I, RuvC II y RuvC III o el dominio HNH de una proteína Cas9) para producir una proteína Cas mutada que carece sustancialmente de toda actividad de clivaje del ARN. Según se describe en la presente, los correspondientes dominios catalíticos de una proteína efectora Cpf1 también se pueden mutar para producir una proteína efectora Cpf1 mutada que carece de toda actividad de clivaje de ADN o que tiene una actividad de clivaje de ADN sustancialmente reducida. En algunas formas de realización, se puede considerar que una proteína efectora de direccionamiento de ácidos nucleicos carece sustancialmente de toda actividad de clivaje de ARN cuando la actividad de clivaje del ARN de la enzima mutada no es mayor que aproximadamente un 25%, 10%, 5%, 1%, 0,1%, 0,01% o menos de la actividad de clivaje del ácido nucleico de la forma no mutada de la enzima; un ejemplo puede ser cuando la actividad de clivaje del ácido nucleico de la forma mutada es nula o despreciable en comparación con la forma no mutada. La proteína efectora se podrá identificar con referencia a la clase general de enzimas que comparten homología con la nucleasa más grande con dominios de tipo nucleasa múltiples del sistema CRISPR de Tipo V. Aún más preferiblemente, la proteína efectora es una proteína de Tipo V tal como Cpf1/C2c1. El término derivado significa para los Solicitantes que la enzima se basa en gran medida en una de tipo salvaje enzima, en el sentido que tiene un alto grado de homología de secuencia con la misma, pero que ha sido mutada (modificada) de alguna manera conocida en el arte o como se describe en la presente.

Nuevamente, se podrá apreciar que los términos Cas y enzima CRISPR y proteína CRISPR y proteína Cas generalmente se utilizan como sinónimos y en todos los puntos de referencia de la presente se refieren por analogía a las nuevas proteínas efectoras CRISPR que se describen más ampliamente en esta solicitud, a menos que sea evidente de otra manera, por ejemplo por una referencia específica a Cas9. Como se mencionó previamente, muchas de las numeraciones de residuos que se utilizan en la presente se refieren a la proteína efectora procedente del locus de CRISPR de Tipo V. Sin embargo, se podrá apreciar que esta divulgación incluye muchas más proteínas efectoras de otras especies de microbios. En determinadas formas de realización, las proteínas efectoras pueden estar presentes constitutivamente o indeciblemente o pueden estar presentes o se pueden administrar o suministrar de manera condicional. Se puede usar optimización de la proteína efectora para mejorar la función o bien, para desarrollar nuevas funciones se pueden generar proteínas efectoras quiméricas. Y, según se describe en la presente, se pueden modificar las proteínas efectoras para su uso como proteínas genéricas de unión a ácidos nucleicos.

Típicamente, en el contexto de un sistema de direccionamiento de ácidos nucleicos, la formación de un complejo de direccionamiento de ácidos nucleicos (que comprende un ARN guía que se hibrida con una secuencia diana y forma un complejo con una o más proteínas efectoras de direccionamiento de ácidos nucleicos) da como resultado el clivaje de una o ambas hebras de ADN o de ARN en o cerca de la secuencia diana (por ejemplo, a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 o más pares de bases de la misma). Según se usa en la presente, el término "secuencias asociadas con un locus diana de interés" se refiere a las secuencias que se encuentran próximas a la secuencia diana (por ejemplo, dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 o más pares de bases respecto de la secuencia diana, donde la secuencia diana está comprendida dentro de un locus diana de interés).

Un ejemplo de una secuencia de codones optimizados es, en este caso, una secuencia optimizada para su expresión en eucariotas, por ejemplo, en seres humanos (es decir, que está optimizada para su expresión en seres humanos), o para otro eucariota, animal o mamífero según se expuso en la presente; véase, por ejemplo, la secuencia SaCas9 humana de codones optimizados en WO 2014/093622 (PCT/US2013/074667) como un ejemplo de una secuencia de codones optimizados (según el conocimiento en el arte y esta divulgación, la optimización de codones de moléculas de ácidos nucleicos codificantes, en especial como una proteína efectora (por ejemplo, Cpf1) se encuentra dentro del campo de conocimientos del especialista). Aunque se prefiere el caso anterior, se podrá apreciar que son posibles otros ejemplos y que es conocida la optimización de codones para una especie huésped diferente de la humana, o para la optimización de codones para órganos específicos. En algunas formas de realización, una secuencia codificante de enzimas que codifica una enzima Cas de direccionamiento de ADN/ARN tiene los codones optimizados para su expresión en células particulares, tales como células eucariotas. Las células eucariotas pueden ser o derivar de un organismo particular, o se pueden obtener a partir del mismo, tal como un mamífero, que incluye, pero en un sentido no taxativo, un ser humano, o un eucariota no humano, o un animal o mamífero como se describe en la presente, por ejemplo, ratón, rata, conejo, perro, ganado o un mamífero no humano o un primate. Se excluyen de la invención los procesos para modificar la identidad genética de la línea germinal de seres humanos y los procesos para

modificar la identidad genética de animales que probablemente causen sufrimiento sin ningún beneficio médico sustancial para el hombre o animal, y también los animales que resultan de dichos procesos. En general, la optimización de codones se refiere a un proceso de modificación de una secuencia de ácidos nucleicos para potenciar la expresión en las células huésped de interés reemplazando por lo menos un codón (por ejemplo, aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 o más codones) de la secuencia natural por codones que se utilizan con mayor frecuencia o que son los más utilizados en los genes de dicha célula huésped a la vez que se mantiene la secuencia de aminoácidos nativa. Varias especies presentan un sesgo particular por determinados codones para un aminoácido particular. El sesgo de codones (diferencias en la utilización de codones entre organismos) a menudo se correlaciona con la eficiencia de traducción del ARN mensajero (ARNm), lo cual a su vez se cree que depende, entre otros, de las propiedades de los codones que están siendo traducidos y de la disponibilidad de moléculas de ARN de transferencia particulares (ARNt). La predominancia de los ARNt seleccionados en una célula generalmente es un reflejo de los codones utilizados más frecuentemente en la síntesis de péptidos. Por lo tanto, es posible adaptar los genes para una expresión genética óptima en un organismo dado basado en la optimización de codones. Las tablas de utilización de codones se pueden obtener fácilmente, por ejemplo, en la "Base de datos de utilización de codones" [*Codon Usage Database*] disponible en www.kazusa.orjp/codon/ y estas tablas se pueden adaptar de numerosas maneras. Véase Nakamura, Y., y col., "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" *Nucl. Acids Res.*, 28: 292 (2000). Los algoritmos para computadora para optimizar los codones de una secuencia particular para su expresión en una célula huésped particular también se pueden obtener fácilmente, tal como de Gene Forge (Aptagen; Jacobus, PA). En algunas formas de realización, uno o más de los codones (por ejemplo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50 o más, o todos los codones) de una secuencia que codifica una proteína Cas de direccionamiento de ADN/ARN corresponden a los codones que se utilizan con más frecuencia para un aminoácido en particular. Con respecto a la utilización de codones en levadura, se hace referencia a la base de datos de genomas de levadura en línea, disponible en http://www.yeastgenome.org/community/codon_usage.shtml, o de *Codon selection in yeast*, Bennetzen y Hall, *J Biol Chem.*, 25 de marzo, 1982; 257(6): 3026-31. En cuanto a la utilización de codones en plantas, incluyendo algas, se hace referencia a *Codon usage in higher plants, green algae, and cyanobacteria*, Campbell y Gowri, *Plant Physiol.,* enero de 1990; 92(1): 1-11.; así como *Codon usage in plant genes*, Murray y col., *Nucleic Acids Res.*, 25 de enero, 1989; 17(2): 477-98; o *Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages*, Morton BR, *J Mol Evol.*, abril, 1998; 46(4): 449-59.

Como se ha descrito, pero no reivindicado específicamente en la presente, un vector codifica una proteína efectora de direccionamiento de ácidos nucleicos, tal como la proteína efectora de direccionamiento de ARN de Tipo V, en particular Cpf1/C2c1, o un ortólogo u homólogo de la misma, que comprende una o más secuencias de localización nuclear (NLS), tal como aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más NLS. En algunas formas de realización, la proteína efectora de direccionamiento de ARN comprende aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más NLS en o cerca del extremo amino terminal, aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más NLS en o cerca del extremo carboxilo terminal, o una combinación de ellos (por ejemplo, cero o por lo menos una o más NLS en el extremo amino terminal y una o más NLS en el extremo carboxilo terminal). Cuando hay más de una NLS, cada una se puede seleccionar de manera independiente de las demás, de modo que puede haber una sola NLS en más de una copia y/o en combinación con una o más NLS diferentes presentes en una o más copias. En algunas formas de realización, se considera que una NLS está cerca del extremo N o C cuando el aminoácido de la NLS más cercano está a una distancia del extremo N o C de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50 o más aminoácidos de la cadena de polipéptidos. Los ejemplos no taxativos de NLS incluyen una secuencia de NLS derivada de: la NLS del antígeno T grande del virus SV40, con la secuencia de aminoácidos PKKKRKV (SEQ ID NO: 2); la NLS de la nucleoplasmina (por ejemplo, la NLS bipartita de nucleoplasmina con la secuencia KRPAATKKAGQAKKKK (SEQ ID NO: 3)); la NLS de c-myc con la secuencia de aminoácidos PAAKRVKLD (SEQ ID NO: 4) o RQRRNELKRSP (SEQ ID NO: 5); la NLS de hRNPA1 M9 con la secuencia NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 6); la secuencia RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 7) del dominio IBB de la importina-alfa; las secuencias VSRKRPRP (SEQ ID NO: 8) y PPKKARED (SEQ ID NO: 9) de la proteína mioma T; la secuencia PQPKKKPL (SEQ ID NO: 10) del p53 humano; la secuencia SALIKKKKKMAP (SEQ ID NO: 11) de c-abl IV de ratón; las secuencias DRLRR (SEQ ID NO: 12) y PKQKKRK (SEQ ID NO: 13) de NS1 del virus de la influenza; la secuencia RKLKKKIKKL (SEQ ID NO: 14) del antígeno delta del virus de hepatitis; la secuencia REKKKFLKRR (SEQ ID NO: 15) de la proteína Mx1 de ratón; la secuencia KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 16) de la poli(ADP-ribosa)polimerasa humana; y la secuencia RKCLQAGMNLEARKTKK (SEQ ID NO: 17) de los receptores de la hormona esteroide glucocorticoide (humana). En general, dichas una o más NLS son de una fuerza suficiente como para dirigir la acumulación de la proteína Cas de direccionamiento de ADN/ARN en una cantidad detectable en el núcleo de una célula eucariota. En general, la fuerza de la actividad de la localización nuclear puede derivar del número de NLS en la proteína efectora de direccionamiento de ácidos nucleicos, las NLS particulares utilizadas o una combinación de estos factores. La detección de la acumulación en el núcleo se puede efectuar mediante cualquier técnica adecuada. Por ejemplo, se puede fusionar un marcador detectable con la proteína de direccionamiento de ácidos nucleicos, de manera que se pueda visualizar su ubicación dentro de una célula, tal como en combinación con un medio para detectar la ubicación del núcleo (por ejemplo, una tinción específica para el núcleo tal como DAPI). También se pueden aislar los núcleos celulares de las células, cuyos contenidos se pueden analizar a continuación mediante cualquier proceso adecuado para detectar proteínas, tales como técnicas inmunohistoquímicas, de inmunoelectrotransferencia o ensayos de actividad enzimática. La acumulación en el núcleo también se puede

determinar indirectamente, tal como mediante un ensayo para el efecto de la formación del complejo de direccionamiento de ácidos nucleicos (por ejemplo, un ensayo para el clivaje o la mutación de ADN o ARN en la secuencia diana, o un ensayo para una actividad de expresión genética alterada afectada por la formación del complejo de direccionamiento de ADN o de ARN y/o de la actividad de la proteína Cas de direccionamiento de ADN o de ARN), en comparación con un control que no estuvo expuesto a la proteína Cas de direccionamiento de ácidos nucleicos o al complejo de direccionamiento de ácidos nucleicos, o expuesto a una proteína Cas de direccionamiento de ácidos nucleicos que no incluye dichas una o más NLS. En formas de realización preferidas de los complejos y sistemas de proteínas efectoras Cpf1 y sistemas descritos en la presente, las proteínas efectoras Cpf1 de codones optimizados comprenden una NLS unida al extremo C-terminal de la proteína. En determinadas formas de realización, se pueden fusionar otras marcas de localización a la proteína Cas tal como, sin limitaciones, para localizar la Cas a sitios particulares en una célula, tal como a organelas, como mitocondrias, plástidos, cloroplastos, vesículas, Golgi, membranas (nuclear o celular), ribosomas, nucleolos, ER, citoesqueleto, vacuolas, centrosomas, nucleosomas, gránulos, centríolos, etc.

En algunas formas de realización, se introducen uno o más vectores que dirigen la expresión de uno o más elementos de un sistema de direccionamiento de ácidos nucleicos en una célula huésped de modo que la expresión de los elementos del sistema de direccionamiento de ácidos nucleicos dirige la formación de un complejo de direccionamiento de ácidos nucleicos en uno o más sitios diana. Por ejemplo, una enzima efectora de direccionamiento de ácidos nucleicos y un ARN guía de direccionamiento de ácidos nucleicos se pueden ligar operativamente a elementos reguladores separados en vectores separados. Los ARN del sistema de direccionamiento de ácidos nucleicos se puede suministrar a un animal o mamífero no humano con una proteína efectora de direccionamiento de ácidos nucleicos, por ejemplo, un animal o mamífero no humano que expresa de manera constitutiva o de manera inducible o de manera condicional una proteína efectora de direccionamiento de ácidos nucleicos, o un animal o mamífero no humano que de otra manera expresaría una proteína efectora de direccionamiento de ácidos nucleicos o que tiene células que contienen una proteína efectora de direccionamiento de ácidos nucleicos, por ejemplo por medio de un suministro previo al mismo de uno o más vectores que codifican y expresan *in vivo* una proteína efectora de direccionamiento de ácidos nucleicos. Como alternativa, se pueden combinar dos o más de los elementos expresados a partir de elementos reguladores iguales o diferentes en un solo vector, con uno o más vectores adicionales para proveer todos los componentes del sistema de direccionamiento de ácidos nucleicos que no estén incluidos en el primer vector. Los elementos del sistema de direccionamiento de ácidos nucleicos que se combinan en un único vector se pueden disponer en cualquier orientación apropiada, por ejemplo un elemento situado 5' ("corriente arriba") o 3' ("corriente abajo") con respecto a un segundo elemento. La secuencia codificante de un elemento se puede ubicar en la misma hebra o en la hebra opuesta de la secuencia codificante de un segundo elemento, y orientada en la misma dirección o en una dirección opuesta. En algunas formas de realización, un solo promotor dirige la expresión de un transcripto que codifica una proteína efectora de direccionamiento de ácidos nucleicos y el ARN guía de direccionamiento de ácidos nucleicos, incluido en una o más secuencias de intrón (por ejemplo, cada una en un intrón diferente, dos o más en por lo menos un intrón o todas en un solo intrón). En algunas formas de realización, la proteína efectora de direccionamiento de ácidos nucleicos y el ARN guía de direccionamiento de ácidos nucleicos se pueden ligar operativamente al mismo promotor y expresar a partir del mismo. Los vehículos de suministro, vectores, partículas, nanopartículas, formulaciones y componentes de los mismos para la expresión de uno o más elementos de un sistema de direccionamiento de ácidos nucleicos son los utilizados en documentos anteriores, tal como en WO 2014/093622 (PCT/US2013/074667). En algunas formas de realización, un vector comprende uno o más sitios de inserción, tal como una secuencia de reconocimiento de una endonucleasa de restricción (también denominado "sitio de clonación"). En algunas formas de realización, se ubica uno o más sitios de inserción (por ejemplo, aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más sitios de inserción) corriente arriba y/o corriente abajo con respecto a uno o más elementos de la secuencia de uno o más vectores. Cuando se utilizan múltiples secuencias guía diferentes, se podrá utilizar una sola construcción de expresión para dirigir la actividad de direccionamiento de ácidos nucleicos a múltiples secuencias diana diferentes correspondientes dentro de una célula. Por ejemplo, un vector individual puede comprender aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 o más secuencias guía. En algunas formas de realización, se puede proveer aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más vectores que contienen dichas secuencias guía y, opcionalmente, suministrarlas a una célula. En algunas formas de realización, un vector comprende un elemento regulador ligado operativamente a una secuencia codificante de una enzima que codifica una proteína efectora de direccionamiento de ácidos nucleicos. La proteína efectora de direccionamiento de ácidos nucleicos o el ARN guía de direccionamiento de ácidos nucleicos se pueden suministrar por separado; y, ventajosamente, por lo menos uno de los mismos se suministra mediante un complejo de partículas. El ARNm de la proteína efectora de direccionamiento de ácidos nucleicos se puede suministrar antes del ARN guía de direccionamiento de ácidos nucleicos para dar tiempo a que se exprese la proteína efectora de direccionamiento de ácidos nucleicos. El ARNm de la proteína efectora de direccionamiento de ácidos nucleicos se podría administrar 1-12 horas (preferiblemente, 2-6 horas aproximadamente) antes de la administración del ARN guía de direccionamiento de ácidos nucleicos. Como alternativa, el ARNm de la proteína efectora de direccionamiento de ácidos nucleicos y el ARN guía de direccionamiento de ácidos nucleicos se pueden administrar juntos. Ventajosamente, se puede administrar una segunda dosis de refuerzo de ARN guía 1-12 horas (preferiblemente, 2-6 horas aproximadamente) después de la administración inicial del ARNm de la proteína efectora de direccionamiento de ácidos nucleicos + el ARN guía. Las administraciones adicionales del ARNm de la proteína efectora de direccionamiento de ácidos nucleicos y/o del ARN guía podrían ser útiles para lograr los niveles más eficaces de modificación genómica.

En un aspecto, la invención como se define en las reivindicaciones proporciona métodos para utilizar uno o más elementos de un sistema de direccionamiento de ácidos nucleicos. El complejo de direccionamiento de ácidos nucleicos de la invención provee un medio eficaz para modificar un ADN o un ARN diana (hebra simple o doble, lineal o superenrollado). El complejo de direccionamiento de ácidos nucleicos de la invención tiene una amplia variedad de utilidades que incluyen modificar (por ejemplo, suprimir, insertar, traslocar, inactivar, activar) un ADN o ARN diana en múltiples tipos celulares. En este sentido, el complejo de la invención de direccionamiento de ácidos nucleicos de la invención ofrece un amplio espectro de aplicaciones, por ejemplo, en terapias génicas, en la selección de fármacos, en el diagnóstico y pronóstico de enfermedades. Un ejemplo de un complejo de direccionamiento de ácidos nucleicos comprende una proteína efectora de direccionamiento de ADN o ARN complejada con un ARN guía hibridado con una secuencia diana dentro del locus diana de interés.

En una forma de realización de las reivindicaciones, esta invención implica clivar un ARN diana. El método comprende modificar un ARN diana utilizando un complejo de direccionamiento de ácidos nucleicos que se une al ARN diana y realizar el clivaje de dicho ARN diana. En una forma de realización, cuando el complejo de direccionamiento de ácidos nucleicos de la invención se introduce en una célula puede crear una rotura (por ejemplo, una rotura de hebra single o doble) en la secuencia de ARN. Por ejemplo, el método se puede usar para clivar un ARN de una enfermedad en una célula. Por ejemplo, se puede introducir en una célula un molde de ARN exógeno que comprende una secuencia a integrar flanqueada por una secuencia 5' y una secuencia 3'. Las secuencias 5' y 3' comparten similitud de secuencia con cualquiera de los lados del sitio de integración en el ARN. Si se desea, el ARN donante puede ser un ARNm. El molde de ARN exógeno comprende una secuencia que será integrada (por ejemplo, un ARN mutado). La secuencia para la integración puede ser una secuencia endógena o exógena para la célula. Los ejemplos de una secuencia a integrar incluyen polinucleótidos que codifican una proteína o un ARN no codificante (por ejemplo, un microARN). Por lo tanto, la secuencia para la integración puede estar ligada operativamente a una o más secuencias de control apropiadas. Como alternativa, la secuencia a integrar puede proveer una función reguladora. Las secuencias corriente arriba y corriente abajo en el molde de ARN exógeno se seleccionan para promover la recombinación entre la secuencia de ARN de interés y el ARN donante. La secuencia corriente arriba es una secuencia de ARN que comparte similitud de secuencia con la secuencia del ARN corriente arriba con respecto al sitio buscado para la integración. De manera similar, la secuencia corriente abajo es una secuencia de ARN que comparte similitud de secuencia con la secuencia de ARN corriente abajo de con respecto al sitio de integración. Las secuencias 5' y 3' en el molde de ARN exógeno puede tener aproximadamente un 75%, 80%, 85%, 90%, 95% o un 100% de identidad de secuencia con la secuencia de ARN diana. Preferiblemente, las secuencias corriente arriba y corriente abajo en el molde de polinucleótidos exógeno presenta aproximadamente un 95%, 96%, 97%, 98%, 99% o un 100% de identidad de secuencia con la secuencia de ARN diana. En algunos métodos, las secuencias 5' y 3' en el molde de ARN exógeno tienen aproximadamente un 99% o un 100% de identidad de secuencia con la secuencia de ARN diana. Una secuencia corriente arriba (5') o corriente abajo (3') puede comprender entre aproximadamente 20 pb y aproximadamente 2500 pb, por ejemplo, aproximadamente 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400 o 2500 pb. En algunos métodos, la secuencia 5' o 3' ilustrativa tiene entre aproximadamente 200 pb y aproximadamente 2000 pb, entre aproximadamente 600 pb y aproximadamente 1000 pb o, más particularmente, entre aproximadamente 700 pb y aproximadamente 1000 pb. En algunos métodos, el molde de ARN exógeno puede comprender además un marcador. Un marcador de este tipo puede facilitar la selección de las integraciones diana. Los ejemplos de marcadores adecuados incluyen sitios de restricción, proteínas fluorescentes o marcadores seleccionables. El molde de polinucleótidos exógeno de la invención se puede construir utilizando técnicas recombinantes (véase, por ejemplo, Sambrook y col., 2001 y Ausubel y col., 1996). En un método para modificar un ARN diana mediante la integración de un molde de ARN exógeno, el complejo de direccionamiento de ácidos nucleicos introduce una rotura (por ejemplo, una rotura de hebra doble o de hebra simple en el ADN o el ARN de hebra doble o de hebra simple) en la secuencia de ADN o ARN, la rotura se repara mediante recombinación homóloga con un molde de ARN exógeno de manera tal que el molde se integra en el ARN diana. La presencia de una rotura de hebra doble facilita la integración del molde. En otras formas de realización, esta invención provee un método para modificar la expresión de un polinucleótido en una célula eucariota. El método comprende aumentar o disminuir la expresión de un polinucleótido diana mediante el uso de un complejo de direccionamiento de ácidos nucleicos que se une al ADN o ARN (por ejemplo, ARNm o pre-ARNm). En algunos métodos, se puede inactivar un ARN diana para efectuar la modificación de la expresión en una célula. Por ejemplo, tras la unión de un complejo de direccionamiento de ARN a una secuencia diana en una célula, el ARN diana es activado de modo que no se traduce la secuencia, no se produce la proteína codificada o la secuencia no funciona como lo hace la secuencia no modificada. Por ejemplo, se puede inactivar una proteína o una secuencia codificante de microARN de modo que no se produzca la proteína o el microARN o el transcrito de pre-microARN. El ARN diana de un complejo de direccionamiento de ácidos nucleicos puede ser cualquier ARN endógeno o exógeno para la célula eucariota. Por ejemplo, el ARN diana puede ser un ARN que resida en el núcleo de una célula eucariota. El ARN diana puede ser una secuencia (por ejemplo, ARNm o pre-ARNm) que codifica un producto genético (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, ARNnc, ARNncl, ARNt o ARNr). Los ejemplos de ARN diana incluyen una secuencia asociada con una vía de señalización bioquímica, por ejemplo, un ARN asociado con una vía de señalización bioquímica. Los ejemplos de ARN diana incluyen un ARN asociado con una enfermedad. Un ARN "asociado con una enfermedad" se refiere a cualquier ARN capaz de generar productos de traducción a un nivel anormal o en una forma anormal en células derivadas de un tejido afectado por una enfermedad en comparación con los tejidos o células de un control sin la enfermedad. Puede ser un ARN transcrito a partir de un gen que se expresa a un nivel anormalmente alto; puede ser un ARN que se expresa a un nivel anormalmente bajo, donde la expresión

alterada se correlacione con la presencia y/o evolución de la enfermedad. Un ARN asociado a una enfermedad también se refiere a un ARN transcrito a partir de un gen que comprende una o más mutaciones o una variación genética que es directamente responsable o que está relacionada con un desequilibrio de ligamiento uno o más genes que son responsables de la etiología de una enfermedad. Los productos traducidos pueden ser conocidos o desconocidos y se pueden encontrar a un nivel normal o anormal. El ARN diana de un complejo de direccionamiento de ácidos nucleicos puede ser cualquier ARN endógeno o exógeno para la célula eucariota. Por ejemplo, el ARN diana puede ser un ARN que resida en el núcleo de una célula eucariota. El ARN diana puede ser una secuencia (por ejemplo, ARNm o pre-ARNm) que codifica un producto genético (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, ARNnc, ARNncl, ARNt o ARNr).

En algunas formas de realización, el método puede comprender permitir la unión de un complejo de direccionamiento de ácidos nucleicos al ADN o ARN diana para efectuar el clivaje de dicho ADN o ARN diana, modificando de este modo el ADN o ARN diana, en donde el complejo de direccionamiento de ácidos nucleicos comprende una proteína efectora de direccionamiento de ácidos nucleicos complejada con un ARN guía hibridado con una secuencia diana en dicho ADN o ARN diana. En otro aspecto, la invención provee un método para modificar la expresión de ADN o ARN en una célula eucariota. En algunas formas de realización, el método comprende permitir la unión de un complejo de direccionamiento de ácidos nucleicos al ADN o ARN de manera tal que dicha unión resulte en la expresión aumentada o reducida de dicho ADN o ARN; en donde el complejo de direccionamiento de ácidos nucleicos comprende una proteína efectora de direccionamiento de ácidos nucleicos complejada con un ARN guía. Se aplican las consideraciones y condiciones anteriores para los métodos de modificación de un ADN o ARN diana. De hecho, estas opciones de obtención, cultivo y reintroducción de muestras son válidas en todos los aspectos de la presente invención. En un aspecto, la invención provee métodos para modificar un ADN o ARN diana en una célula eucariota, que es *in vivo* o *in vitro*. En algunas formas de realización, el método comprende obtener muestras de una célula o población de células de un ser humano o un animal no humano y modificar dichas una o más células. El cultivo puede tener lugar en cualquier etapa *ex vivo*. Dichas una o más células incluso se pueden reintroducir en el animal no humano o la planta. En el caso de células reintroducidas, se prefiere particularmente que las células sean células madre.

Aún más, en cualquier aspecto de la invención, el complejo de direccionamiento de ácidos nucleicos puede comprender una proteína efectora de direccionamiento de ácidos nucleicos complejada con un ARN guía hibridado con una secuencia diana.

La divulgación se refiere a la manipulación mediante ingeniería genética y la optimización de los sistemas, métodos y composiciones utilizados para el control de la expresión genética que comprende el direccionamiento de una secuencia de ADN o ARN, que se relacionan con el sistema de direccionamiento de ácidos nucleicos y componentes del mismo. En formas de realización ventajosas de la presente divulgación, la enzima efectora es una proteína de Tipo V/Tipo VI tal como Cpf1/C2c1/C2c2. Una ventaja de los métodos de la presente es que el sistema CRISPR minimiza o evita la unión fuera del diana y sus consiguientes efectos secundarios. Esto se logra utilizando sistemas dispuestos para que tengan un grado elevado de especificidad de secuencia por el ADN o ARN diana.

Con relación a un complejo o sistema de direccionamiento de ácidos nucleicos, la secuencia tracr tiene preferiblemente una o más estructuras de bucle-tallo u horquillas y tiene 30 o más nucleótidos de longitud, 40 o más nucleótidos de longitud o 50 o más nucleótidos de longitud; la secuencia de ARNrc es de entre 10 y 30 nucleótidos de longitud, la proteína efectora de direccionamiento de ácidos nucleicos es una enzima Cas de Tipo V/Tipo VI. En determinadas formas de realización, la secuencia del ARNcr es de entre 42 y 44 nucleótidos de longitud, y las proteína Cas de direccionamiento de ácidos nucleicos es Cpfl de *Francisella tularensis subsp. novocida U112*. En determinadas formas de realización, el ARNcr comprende, consiste esencialmente en o consiste en 19 nucleótidos de una repetición directa y entre 23 y 25 nucleótidos de la secuencia espaciadora, y la proteína Cas de direccionamiento de ácidos nucleicos es Cpfl de *Francisella tularensis subsp. novocida U112*.

El uso de dos aptámeros diferentes (cada uno de los cuales está asociado con ARN guía de direccionamiento de ácidos nucleicos diferentes) permite usar una proteína de fusión adaptadora-activadora y una proteína de fusión represora-adaptadora, con diferentes ARN guía de direccionamiento de ácidos nucleicos, para activar la expresión de un ADN o ARN, a la vez que se reprime otro. Los mismos, junto con sus diferentes ARN guía se pueden administrar juntos, o sustancialmente juntos, en un abordaje multiplexado. Se puede utilizar una gran cantidad de dichos ARN guía de direccionamiento de ácidos nucleicos modificados, todos al mismo tiempo, por ejemplo 10 o 20 o 30 y así sucesivamente, a la vez que solo es necesario suministrar una molécula de proteína efectora (o por lo menos una cantidad mínima de las mismas), ya que se puede utilizar una cantidad comparativamente pequeña de moléculas de proteínas efectoras con una gran cantidad de guías modificadas. La proteína adaptadora puede estar asociada (preferiblemente puede estar ligada o fusionada) a uno o más activadores o uno o más represores. Por ejemplo, la proteína adaptadora puede estar asociada con un primer activador y un segundo activador. El primer y el segundo activadores pueden ser iguales, pero preferiblemente son activadores diferentes. Se pueden usar tres o más o aún cuatro o más activadores (o represores), pero el tamaño del paquete puede limitar un número mayor que 5 dominios funcionales diferentes. Preferiblemente se utilizan conectores, en una fusión directa con la proteína adaptadora, cuando hay dos o más dominios funcionales asociados a la proteína adaptadora. Los conectores apropiados podrían incluir al conector GlySer.

También se contempla que el complejo de ARN guía-proteína efectora de direccionamiento de ácidos nucleicos como

un todo puede estar asociado con dos o más dominios funcionales. Por ejemplo, puede haber dos o más dominios funcionales asociados con la proteína efectora de direccionamiento de ácidos nucleicos o puede haber dos o más dominios funcionales asociados con el ARN guía (a través de una o más proteínas adaptadoras) o puede haber uno o más dominios funcionales asociados con la proteína efectora de direccionamiento de ácidos nucleicos y uno o más dominios funcionales asociados al ARN guía (a través de una o más proteínas adaptadoras).

La fusión entre la proteína adaptadora y la activadora o represora puede incluir un conector. Por ejemplo, se pueden usar conectores GlySer GGGS (SEQ ID NO: 18). Los mismos se pueden utilizar en repeticiones de 3 $((GGGGS)_3)$ o 6, 9 o aún 12 o más, para proveer las longitudes apropiadas, según necesidad. Se pueden usar conectores entre los ARN guía y el dominio funcional (activador o represor), o entre la proteína Cas de direccionamiento de ácidos nucleicos (Cas) y el dominio funcional (activador o represor). Los conectores se pueden usar para modificar cantidades apropiadas de "flexibilidad mecánica".

La invención contempla un complejo de direccionamiento de ácidos nucleicos que comprende una proteína efectora de direccionamiento de ácidos nucleicos y un ARN guía, donde la proteína efectora de direccionamiento de ácidos nucleicos comprende por lo menos una mutación, de manera tal que la proteína Cas de direccionamiento de ácidos nucleicos no tendrá más de un 5% de la actividad de la proteína Cas de direccionamiento de ácidos nucleicos que carece de dicha por lo menos una mutación y, opcionalmente, por lo menos una o más secuencias de localización nuclear; el ARN guía comprende una secuencia guía capaz de hibridarse con una secuencia diana en un ARN de interés en una célula; y donde: la proteína efectora de direccionamiento de ácidos nucleicos se asocia con dos o más dominios funcionales; o se modifica por lo menos un bucle del ARN guía por inserción de una o más secuencias de ARN diferentes que se unen a una o más proteínas adaptadoras, y donde la proteína adaptadora se asocia con dos o más dominios funcionales; o la proteína efectora de direccionamiento de ácidos nucleicos se asocia con uno o más dominios funcionales y se modifica por lo menos un bucle del ARN guía por inserción de secuencias de ARN diferentes que se unen a una o más proteínas adaptadoras, y donde la proteína adaptadora se asocia con uno o más dominios funcionales.

Se describe, pero no se reivindica específicamente en la presente, un método para generar un modelo en células eucariotas que comprenda un gen mutado de una enfermedad. En algunas formas de realización, un gen ligado a una enfermedad es cualquier gen asociado con un aumento del riesgo de padecer o desarrollar una enfermedad. En algunas formas de realización, el método comprende (a) introducir uno o más vectores en una célula eucariota, en donde dichos uno o más vectores dirigen la expresión de uno o más de: una enzima Cpf1 y un ARN guía protegido que comprende una secuencia guía ligada a una secuencia de repetición directa; y (b) permitir la unión de un complejo CRISPR a un polinucleótido diana para efectuar el clivaje del polinucleótido diana en dicho gen de enfermedad, en donde el complejo CRISPR comprende la enzima Cpfl complejada con el ARN guía que comprende la secuencia que se hibrida con la secuencia diana en el polinucleótido diana, generando de esa manera un modelo de una célula eucariota que comprende un gen de enfermedad mutado. En algunas formas de realización, dicho clivaje comprende clivar una o dos hebras en la ubicación de la secuencia diana por dicha enzima Cpfl. En algunas formas de realización, dicho clivaje da como resultado una disminución de la transcripción de un gen diana. En algunas formas de realización, el método comprende además reparar dicho polinucleótido diana clivado mediante mecanismos de inserción de genes basados en la unión de extremos no homólogos (NHEJ) con un polinucleótido molde exógeno, donde dicha reparación da como resultado una mutación que comprende la inserción, supresión o sustitución de uno o más nucleótidos de dicho polinucleótido diana. En algunas formas de realización, dicha mutación da como resultado uno o más cambios de aminoácidos en la expresión de una proteína resultante de un gen que comprende la secuencia diana.

En un aspecto, la invención provee métodos que se describen en la presente, en donde el huésped es una célula eucariota. En un aspecto, la invención provee un método descrito en la presente, en donde el huésped es una célula de mamífero. En un aspecto, la invención provee un método descrito en la presente, en donde el huésped es una célula eucariota no humana. En un aspecto, la invención provee un método descrito en la presente, en donde la célula eucariota no humana es una célula de mamífero no humano. En un aspecto, la invención provee un método descrito en la presente, en donde la célula de mamífero no humano puede ser incluyendo, pero en un sentido no taxativo, una célula de primate, bovina, ovina, porcina, canina, de roedor, de Leporidae, tal como de mono, de vaca, de oveja, de cerdo, de perro, de conejo, de rata o de ratón. En un aspecto, la invención provee un método descrito en la presente, en donde la célula puede ser una célula eucariota no de mamífero, tal como una célula de un ave de corral (por ejemplo, de pollo), de peces vertebrados (por ejemplo, salmón) o de mariscos (por ejemplo, ostras, almejas, langostas o gambas). En un aspecto, la invención provee un método descrito en la presente, en donde la célula eucariota no humana es una célula vegetal. La célula vegetal puede ser de una monocotiledónea o dicotiledónea o de una planta de cultivos o de granos, tal como de mandioca, maíz, sorgo, soja, sorgo, avena o arroz. La célula vegetal también puede ser de algas, árboles o plantas productoras, de frutas o verduras (por ejemplo, árboles tales como árboles de cítricos, por ejemplo, árboles de naranja, pomelo o limón; durazno o nectarina; árboles de manzano o peras; árboles de frutos secos, tales como árboles de almendras o nueces o pistacho; plantas de sombra nocturna; plantas del género *Brassica*; plantas del género *Lactuca*; plantas del género *Spinacia*; plantas del género *Capsicum*; algodón, tabaco, espárrago, zanahoria, repollo, brócoli, coliflor, tomate, berenjena, pimiento, lechuga, espinaca, frutilla, arándano azul, frambuesa, zarzamora, uvas, café, cacao, etc).

Se describe, pero no se reivindica específicamente en la presente, un método para desarrollar un agente biológicamente activo que modula un evento de señalización celular asociado con un gen de una enfermedad. En

algunas formas de realización, un gen ligado a una enfermedad es cualquier gen asociado con un aumento del riesgo de padecer o desarrollar una enfermedad. En algunas formas de realización, el método comprende (a) poner un compuesto de prueba en contacto con un modelo de célula de cualquiera de las formas de realización descritas precedentemente; y (b) detectar un cambio en una lectura que sea indicativo de una reducción o de un aumento de un evento de señalización celular asociado con dicha mutación en dicho gen de la enfermedad, desarrollando de esa manera dicho agente biológicamente activo que modula dicho evento de señalización celular asociado con dicho gen de la enfermedad.

Se describe, pero no se reivindica específicamente en la presente, un método para seleccionar una o más células mediante la introducción de una o más mutaciones en un gen en dichas una o más células, donde dicho método comprende: introducir uno o más vectores en dichas una o más células, en donde dichos uno o más vectores dirigen la expresión de uno o más de: Cpfl, una secuencia guía ligada a una secuencia de repetición directa, y un molde de edición; en donde dicho molde de edición comprende dichas una o más mutaciones que anulan el clivaje de Cpfl; permitir la recombinación homóloga del molde de edición con el polinucleótido diana en dichas una o más células a seleccionar; permitir la unión de un complejo Cpfl CRISPR-Cas a un polinucleótido diana para efectuar el clivaje del polinucleótido diana dentro de dicho gen, en donde el complejo Cpfl CRISPR-Cas comprende la Cpfl complejada con (1) la secuencia guía que se hibrida con la secuencia diana en el polinucleótido diana, y (2) la secuencia de repetición directa, en donde la unión del complejo Cpfl CRISPR-Cas con el polinucleótido diana induce muerte celular, permitiendo de esa manera seleccionar una o más células en las cuales se ha introducido una o más mutaciones; este incluye la Cpfl partida de la presente. En otra forma de realización preferida de la invención, la célula a seleccionar puede ser una célula eucariota. Algunos aspectos de la invención permiten la selección de células específica sin necesidad de un marcador de selección o un proceso de dos pasos que puede incluir un sistema de contraselección. En formas de realización particulares, el modelo de célula eucariota está comprendido en un modelo de organismo eucariota.

En realizaciones particulares, los métodos y composiciones reivindicados pueden implicar un polinucleótido recombinante que comprende una secuencia guía corriente abajo (3') de una secuencia de repetición directa, en donde cuando se expresa la secuencia guía dirige una unión específica de secuencias del complejo Cpfl CRISPR-Cas a una correspondiente secuencia diana presente en una célula eucariota. En algunas formas de realización, la secuencia diana es una secuencia viral presente en una célula eucariota. En algunas formas de realización, la secuencia diana es un proto-oncogén o un oncogén.

También se describe, pero no se reivindica específicamente en la presente, un sistema de vectores o una célula huésped eucariota que comprende (a) un primer elemento regulador ligado operativamente a una secuencia de repetición directa y uno o más sitios de inserción para insertar una o más secuencias guía (incluyendo cualquiera de las secuencias guía modificadas que se describen en la presente) corriente abajo (3') de la secuencia DR, en donde, cuando se expresa, la secuencia guía dirige la unión específica de secuencias de un complejo Cpfl CRISPR-Cas a una secuencia diana en una célula eucariota, en donde dicho complejo Cpfl CRISPR-Cas comprende Cpfl (que incluye cualquiera de las enzimas modificadas descritas en la presente) complejada con la secuencia guía que se hibrida con la secuencia diana (y opcionalmente la secuencia DR); y/o (b) un segundo elemento regulador ligado operativamente a una secuencia codificante de enzimas que codifica dicha enzima Cpfl que comprende una secuencia de localización nuclear y/o NES. En algunas formas de realización, la célula huésped comprende los componentes (a) y (b). En algunas formas de realización, el componente (a), el componente (b) o los componentes (a) y (b) se integran de manera estable en el genoma de la célula huésped eucariota. En algunas formas de realización, el componente (a) comprende además dos o más secuencias guía ligadas operativamente al primer elemento regulador, en donde cuando se expresan, cada una de dichos dos o más secuencias guía dirigen la unión específica de secuencias de un complejo Cpfl CRISPR-Cas con una secuencia diana diferente en una célula eucariota. En algunas formas de realización, la enzima CRISPR comprende una o más secuencias de localización nuclear y/o secuencias de exportación nuclear o NES de fuerza suficiente como para dirigir la acumulación de dicha enzima CRISPR en una cantidad detectable en y/o fuera del núcleo de una célula eucariota. En algunas formas de realización, la enzima Cpfl deriva de una Cpfl de *Francisella tularensis* 1, *Francisella tularensis* subsp. *novicida*, *Prevotella albensis*, *Lachnospiraceae bacterium* MC2017 1, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium* GW2011_GWA2_33_10, *Parcubacteria bacterium* GW2011_GWC2_44_17, *Smithella* sp. SCADC, *Acidaminococcus* sp. BV3L6, *Lachnospiraceae bacterium* MA2020, *Candidatus Metanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi* 237, *Leptospira inadai*, *Lachnospiraceae bacterium* ND2006, *Porphyromonas crevioricanis* 3, *Prevotella disiens* o *Porphyromonas macacae*, incluyendo cualquiera de las enzimas modificadas descritas en la presente, y pueden incluir alteraciones o mutaciones adicionales de la Cpfl y puede ser una Cpfl quimérica. En algunas formas de realización, la enzima CRISPR es de codones optimizados para su expresión en una célula eucariota. En algunas formas de realización, la enzima CRISPR dirige el clivaje de una o dos hebras en la ubicación de la secuencia diana. En una forma de realización preferida, la rotura de hebra es un corte escalonado con una sobreextensión 5'. En algunas formas de realización, la Cpfl carece de actividad de clivaje de hebras de ADN (por ejemplo, no más de un 5% de actividad nucleasa en comparación con una enzima de tipo salvaje o una enzima que no comprende la mutación o alteración que disminuye la actividad nucleasa). En algunas formas de realización, el primer elemento regulador es un promotor de la polimerasa III. En algunas formas de realización, el segundo elemento regulador es un promotor de la polimerasa II. En algunas formas de realización, la repetición directa tiene una longitud mínima de 16 nts y un solo bucle-tallo. En formas de realización adicionales, la repetición directa tiene una longitud mayor que 16 nts,

preferiblemente de más de 17 nts, y tiene más de un bucle-tallo o estructuras secundarias optimizadas. En algunas formas de realización, la secuencia guía es de por lo menos 16, 17, 18, 19, 20, 25 nucleótidos o entre 16-30 o ente 16-25 o entre 16-20 nucleótidos de longitud.

También se describe, pero no se reivindica específicamente en la presente, un conjunto de elementos que comprende uno o más de los componentes que se describen en la presente. En algunas formas de realización, el conjunto de elementos comprende un sistema de vectores o una célula huésped que se describen en la presente y las instrucciones para utilizar dicho conjunto de elementos.

Enzimas Cpf1 modificadas

El análisis mediante computadora de la estructura primaria de las nucleasas Cpfl revela tres regiones distintas (Figura 1). Primero, un dominio tipo RuvC C-terminal, que es el único dominio funcional caracterizado. Segundo, una región alfa-helicoidal N-terminal y tercero, una región alfa y beta mixta, ubicada entre el dominio tipo RuvC y la región alfa-helicoidal.

Se predicen varias extensiones pequeñas de regiones no estructuradas dentro de la estructura primaria de Cpfl. Las regiones no estructuradas, que están expuestas al solvente y que no están conservadas en diferentes ortólogos de Cpf1, son los lados preferidos para cortes e inserciones de secuencias de proteínas pequeñas (Figuras 2 y 3). Además, estos lados se pueden usar para generar proteínas quiméricas entre los ortólogos de Cpfl.

Sobre la base de la información anterior, se pueden generar mutantes que conducirán a la inactivación de la enzima o que modificarán la nucleasa de hebra doble a una actividad de corte monocatenario. En formas de realización alternativas, esta información se usa para desarrollar enzimas con efectos fuera del diana reducidos (descritos en otra parte en la presente)

En el caso algunas de las enzimas Cpfl descritas previamente, la enzima se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones D917, E1006, E1028, D1227, D1255A, N1257, de acuerdo con la proteína FnCpfl o cualquier ortólogo correspondiente. En un aspecto, la invención provee la composición reivindicada, en donde la enzima Cpfl es una enzima inactivada que comprende una o más mutaciones seleccionadas del grupo que consiste en D917A, E1006A, E1028A, D1227A, D1255A, N1257A, D917A, E1006A, E1028A, D1227A, D1255A y N1257A de acuerdo con la proteína FnCpfl o las posiciones correspondientes en un ortólogo de Cpfl. En un aspecto la invención provee la composición reivindicada, en donde la enzima CRISPR comprende D917, o E1006 y D917, o D917 y D1255, de acuerdo con la proteína FnCpfl o la posición correspondiente en un ortólogo de Cpfl.

En algunas de las enzimas Cpfl descritas con anterioridad, la enzima se modifica mediante mutación de uno o más residuos (en el dominio RuvC) que incluyen, pero en un sentido no taxativo, las posiciones R909, R912, R930, R947, K949, R951, R955, K965, K968, K1000, K1002, R1003, K1009, K1017, K1022, K1029, K1035, K1054, K1072, K1086, R1094, K1095, K1109, K1118, K1142, K1150, K1158, K1159, R1220, R1226, R1242 y/o R1252 con referencia a la numeración de posiciones de aminoácidos de AsCpfl (*Acidaminococcus* sp. BV3L6).

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de uno o más residuos (en el dominio RAD50) que incluyen, pero en un sentido no taxativo, las posiciones K324, K335, K337, R331, K369, K370, R386, R392, R393, K400, K404, K406, K408, K414, K429, K436, K438, K459, K460, K464, R670, K675, R681, K686, K689, R699, K705, R725, K729, K739, K748 y/o K752 con referencia a la numeración de posiciones de aminoácidos de AsCpfl (*Acidaminococcus* sp. BV3L6).

En algunas de las enzimas Cpfl, la enzima se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones R912, T923, R947, K949, R951, R955, K965, K968, K1000, R1003, K1009, K1017, K1022, K1029, K1072, K1086, F1103, R1226 y/o R1252 con referencia a la numeración de posiciones de aminoácidos de AsCpfl (*Acidaminococcus* sp. BV3L6).

En determinadas formas de realización, la enzima Cpfl se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, R1138, R1165 y/o R1252 con referencia a la numeración de posiciones de aminoácidos de LbCpfl (*Lachnospiraceae bacterium* ND2006).

En determinadas formas de realización, la enzima Cpfl se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones K15, R18, K26, Q34, R43, K48, K51, R56, R84, K85, K87, N93, R103, N104, T118, K123, K134, R176, K177, R192, K200, K226, K273, K275, T291, R301, K307, K369, S404, V409, K414, K436, K438, K468, D482, K516, R518, K524, K530, K532, K548, K559, K570, R574, K592, D596, K603, K607, K613, C647, R681, K686, H720, K739, K748, K757, T766, K780, R790, P791, K796, K809, K815, T816, K860, R862, R863, K868, K897, R909, R912, T923, R947, K949, R951, R955, K965, K968, K1000, R1003, K1009, K1017, K1022, K1029, A1053, K1072, K1086, F1103, S1209, R1226, R1252, K1273, K1282 y/o K1288 con referencia a la numeración de posiciones de aminoácidos de AsCpfl (*Acidaminococcus* sp. BV3L6).

En determinadas formas de realización, la enzima se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones K15, R18, K26, R34, R43, K48, K51, K56, K87, K88, D90, K96, K106,

K107, K120, Q125, K143, R186, K187, R202, K210, K235, K296, K298, K314, K320, K326, K397, K444, K449, E454, A483, E491, K527, K541, K581, R583, K589, K595, K597, K613, K624, K635, K639, K656, K660, K667, K671, K677, K719, K725, K730, K763, K782, K791, R800, K809, K823, R833, K834, K839, K852, K858, K859, K869, K871, R872, K877, K905, R918, R921, K932, I960, K962, R964, R968, K978, K981, K1013, R1016, K1021, K1029, K1034, K1041, K1065, K1084 y/o K1098 con referencia a la numeración de posiciones de aminoácidos de FnCpf1 (*Francisella novicida* U112).

En determinadas formas de realización, la enzima se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones K15, R18, K26, K34, R43, K48, K51, R56, K83, K84, R86, K92, R102, K103, K116, K121, R158, E159, R174, R182, K206, K251, K253, K269, K271, K278, P342, K380, R385, K390, K415, K421, K457, K471, A506, R508, K514, K520, K522, K538, Y548, K560, K564, K580, K584, K591, K595, K601, K634, K640, R645, K679, K689, K707, T716, K725, R737, R747, R748, K753, K768, K774, K775, K785, K787, R788, Q793, K821, R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, K1121, R1138, R1165, K1190, K1199 y/o K1208 con referencia a la numeración de posiciones de aminoácidos de LbCpf1 (*Lachnospiraceae bacterium* ND2006).

En determinadas formas de realización, la enzima se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones K14, R17, R25, K33, M42, Q47, K50, D55, K85, N86, K88, K94, R104, K105, K118, K123, K131, R174, K175, R190, R198, I221, K267, Q269, K285, K291, K297, K357, K403, K409, K414, K448, K460, K501, K515, K550, R552, K558, K564, K566, K582, K593, K604, K608, K623, K627, K633, K637, E643, K780, Y787, K792, K830, Q846, K858, K867, K876, K890, R900, K901, M906, K921, K927, K928, K937, K939, R940, K945, Q975, R987, R990, K1001, R1034, I1036, R1038, R1042, K1052, K1055, K1087, R1090, K1095, N1103, K1108, K1115, K1139, K1158, R1172, K1188, K1276, R1293, A1319, K1340, K1349 y/o K1356 con referencia a la numeración de posiciones de aminoácidos de MbCpf1 (*Moraxella bovoculi* 237).

Proteína Cpf1 desactivada/inactivada

Cuando la proteína Cpf1 tiene actividad nucleasa, se puede modificar la proteína Cpf1 para que tenga una menor actividad nucleasa, por ejemplo, una inactivación de la nucleasa de por lo menos un 70%, por lo menos un 80%, por lo menos un 90%, por lo menos un 95%, por lo menos un 97% o un 100% en comparación con la enzima de tipo salvaje; o, dicho de otra manera, una enzima Cpf1 tendrá ventajosamente aproximadamente un 0% de la actividad nucleasa de la enzima Cpf1 o la enzima CRISPR no mutada o de tipo salvaje, o no más de aproximadamente un 3% o aproximadamente un 5% o aproximadamente un 10% de la actividad nucleasa de la enzima Cpf1 no mutada o de tipo salvaje, por ejemplo de la forma no mutada o de tipo salvaje de *Francisella novicida U112* (FnCpf1), *Acidaminococcus sp. BV3L6* (AsCpf1), *Lachnospiraceae bacterium ND2006* (LbCpf1) o *Moraxella bovoculi 237* (enzima Cpf1 o enzima CRISPR MbCpf1). Esto es posible mediante introducción de mutaciones en los dominios nucleasa de la Cpf1 y ortólogos de la misma.

Más particularmente, las enzimas Cpf1 inactivadas incluyen enzimas mutadas en las posiciones de aminoácido As908, As993, As1263 de AsCpf1 o las posiciones correspondientes en los ortólogos de Cpf1. Adicionalmente, las enzimas Cpf1 inactivadas incluyen enzimas mutadas en la posición de aminoácido Lb832, 925, 947 o 1180 de LbCpf1 o las posiciones correspondientes en los ortólogos de Cpf1. Más particularmente, las enzimas Cpf1 inactivadas incluyen enzimas que comprenden una o más entre las mutaciones AsD908A, AsE993A, AsD1263A de AsCpf1 o las mutaciones correspondientes en los ortólogos de Cpf1. Adicionalmente, las enzimas Cpf1 inactivadas incluyen enzimas que comprenden una o más entre las mutaciones LbD832A, E925A, D947A o D1180A de LbCpf1 o las mutaciones correspondientes en los ortólogos de Cpf1.

La enzima Cpf1 CRISPR inactivada puede estar asociada (por ejemplo, mediante una proteína de fusión) a uno o más dominios funcionales que incluyen, por ejemplo, uno o más dominios del grupo que comprende, que consiste esencialmente en o que consiste en actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de la transcripción, actividad de modificación de histonas, actividad de clivaje de ARN, actividad de clivaje de ADN, actividad de unión a ácidos nucleicos y de interruptores moleculares (por ejemplo, inducible por luz). Los dominios preferidos son Fok1, VP64, P65, HSF1, MyoD1. En el caso de proveer un Fok1, resulta ventajoso proveer múltiples dominios funcionales Fok1 para permitir un dímero funcional y que se diseñen ARNg que proporcionen un espaciado apropiado para el uso funcional (Fok1) como se describe específicamente en Tsai y col., *Nature Biotechnology*, volumen 32, N° 6, junio de 2014). La proteína adaptadora puede emplear conectores conocidos para unir dichos dominios funcionales. En algunos casos resulta ventajoso proveer adicionalmente por lo menos una NLS. En algunos casos, resulta ventajoso ubicar la NLS en el extremo N-terminal. Cuando se incluye más de un dominio funcional, dichos dominios funcionales pueden ser iguales o diferentes.

En general, la ubicación de dichos uno o más dominios funcionales en la enzima Cpf1 inactivada es aquella que permita una orientación espacial correcta para que el dominio funcional pueda afectar al diana con el efecto funcional atribuido. Por ejemplo, si el dominio funcional es un activador de la transcripción (por ejemplo, VP64 o p65), dicho activador de la transcripción se ubica en una orientación espacial que le permita afectar la transcripción del diana. Asimismo, se ubicará ventajosamente un represor de la transcripción para afectar la transcripción del diana, y se ubicará ventajosamente una nucleasa (por ejemplo, Fok1) para clivar el diana o para clivarlo parcialmente. Pueden incluir

posiciones distintas de los extremos N/C terminales de la enzima CRISPR.

Cpf1 desestabilizada

En determinadas formas de realización, la proteína efectora (enzima CRISPR; Cpfl) usada en los métodos y composiciones de acuerdo con la invención que se describe en la presente está asociada o fusionada con un dominio de desestabilización (DD). En algunas formas de realización, el DD es ER50. El ligando estabilizante correspondiente para este DD es 4HT en algunas formas de realización. Como tal, en algunas formas de realización, dicho por lo menos un DD es ER50 y el ligando estabilizante para el mismo es 4HT o CMP8. En algunas formas de realización, el DD es DHFR50. El ligando estabilizante correspondiente para este DD es TMP en algunas formas de realización. Como tal, en algunas formas de realización, dicho por lo menos un DD es DHFR50 y el ligando estabilizante para el mismo es TMP. En algunas formas de realización, el DD es ER50. El ligando estabilizante correspondiente para este DD es CMP8 en algunas formas de realización. Por lo tanto, CMP8 puede ser un ligando estabilizante alternativo para 4HT en el sistema ER50. Si bien es posible que se puede/se debería utilizar CMP8 y 4HT de una manera competitiva, algunos tipos celulares pueden ser más susceptibles a uno u otro de estos dos ligandos, y a partir de esta divulgación y el conocimiento en el arte, el especialista puede emplear CMP8 y/o 4HT.

En algunas formas de realización, se puede fusionar uno o dos DD al extremo N-terminal de la enzima CRISPR con uno o dos DD fusionados al extremo C-terminal de la enzima CRISPR. En algunas formas de realización, dichos por lo menos dos DD están asociados con la enzima CRISPR y los DD son el mismo DD, es decir, los DD son homólogos. Por consiguiente, ambos (o dos o más) de los DD deberían ser DD ER50. Esto se prefiere en algunas formas de realización. Como alternativa, ambos (o dos o más) de los DD deberían ser DD DHFR50. Esto también se prefiere en algunas formas de realización. En algunas formas de realización, dichos por lo menos dos DD están asociados con la enzima CRISPR y los DD son DD diferentes, es decir, los DD son heterólogos. Por consiguiente, uno de los DD debería ser ER50 en tanto uno o más de los DD o cualquier otro DD debería ser DHFR50. Dos o más DD heterólogos puede resultar ventajoso ya que proporcionaría un mayor nivel de control de la degradación. Una fusión en tándem de más de un DD por el extremo N o C-terminal puede mejorar la degradación; y dicha fusión en tándem puede ser, por ejemplo ER50-ER50- C2c2 o DHFR-DHFR-Cpf1. Se contempla que se observarían niveles altos de degradación en la ausencia de cualquiera de los ligandos estabilizantes, niveles intermedios de degradación en la ausencia de un ligando estabilizante y la presencia del otro ligando estabilizante (o uno diferente), en tanto habría niveles bajos de degradación en la presencia de ambos (o dos o más) ligandos estabilizantes. También se puede tener control con un DD ER50 N-terminal y un DD DHFR50 C-terminal.

En algunas formas de realización, la fusión de la enzima CRISPR con el DD comprende un conector entre el DD y la enzima CRISPR. En algunas formas de realización, el conector es un conector de GlySer. En algunas formas de realización, la fusión de DD-enzima CRISPR comprende además por lo menos una señal de exportación nuclear (NES). En algunas formas de realización, la fusión DD-enzima CRISPR comprende dos o más NES. En algunas formas de realización, la fusión DD-enzima CRISPR comprende por lo menos una señal de localización nuclear (NLS). Esto puede ser además de una NES. En algunas formas de realización, la enzima CRISPR comprende o consiste esencialmente en o consiste en una señal de localización (importación o exportación nuclear) como conector, o como parte de un conector, entre la enzima CRISPR y el DD. Las marcas HA o Flag también se encuentran dentro del ámbito de la invención como conectores. Los Solicitantes emplean NLS y/o NES como conectores y también utilizan conectores de Glicina Serina tan cortos como de GS y hasta (GGGGS)3.

Los dominios desestabilizantes son de utilidad general para conferir inestabilidad a un amplio rango de proteínas; véase, por ejemplo, Miyazaki, *J Am Chem Soc*., 7 de marzo, 2012; 134(9): 3942-3945. CMP8 o 4-hidroxitamoxifeno pueden ser dominios desestabilizantes. En términos más generales, se encontró que un mutante sensible a la temperatura del DHFR de mamífero (DHFRts), un residuo desestabilizante según la regla Nend, era estable a una temperatura permisiva pero inestable a 37 °C. La adición de metotrexato, un ligando de gran afinidad para el DHFR de mamífero, a células que expresan DHFRts inhibió parcialmente la degradación de la proteína. Esta fue una demostración importante que un ligando de molécula pequeña puede estabilizar una proteína por lo demás buscado como diana para su degradación en células. Se usó un derivado de rapamicina para estabilizar un mutante inestable del dominio FRB de mTOR (FRB*) y restablecer la función de la quinasa fusionada, GSK-3β.6.7 Este sistema demostró que la estabilidad dependiente del ligando representaba una estrategia atractiva para regular la función de una proteína específica en un entorno biológico complejo. Un sistema para controlar la actividad proteica puede comprender volver funcional al DD cuando la complementación de ubiquitina tiene lugar por dimerización inducida por rapamicina de la proteína de unión a FK506 y FKBP12. Se pueden manipular mediante ingeniería genética los mutantes de la proteína FKBP12 o ecDHFR humana para que sean metabólicamente inestables en la ausencia de sus ligandos de gran afinidad, Shield-1 o trimetoprim (TMP), respectivamente. Estos mutantes son algunos de los dominios desestabilizantes (DD) posibles de utilidad en la práctica de la invención y la inestabilidad de un DD como una fusión con una enzima CRISPR confiere a la proteína CRISPR la degradación de la proteína de fusión completa por el proteasoma. Ambos Shield-1 y TMP se unen y estabilizan al DD de una manera dependiente de la dosis. El dominio de unión al ligando del receptor de estrógeno (ERLBD, residuos 305-549 de ERS1) también se puede manipular como un dominio desestabilizante. Dado que la vía de señalización del receptor de estrógenos está involucrada en una variedad de enfermedades tal como el cáncer de mama, dicha vía ha sido ampliamente estudiada y se han desarrollado numerosos agonistas y antagonistas del receptor de estrógenos. Por consiguiente, se conocen pares compatibles de ERLBD y fármacos. Hay ligandos que se unen a las formas mutantes pero no a las formas de tipo

salvaje del ERLBD. El uso de uno de estos dominios mutantes que codifican tres mutaciones (L384M, M421G, G521R)12, permite regular la estabilidad de un DD derivado de ERLBD usando un ligando que no afecta las redes sensibles a estrógenos sensibles. Se puede introducir una mutación adicional (Y537S) para desestabilizar adicionalmente al ERLBD y para configurarlo como un potencial candidato de DD. Este tetra-mutante constituye un desarrollo ventajoso de DD. El ERLBD mutante se puede fusionar con una enzima CRISPR y su estabilidad se puede regular o perturbar usando un ligando, con lo cual la enzima CRISPR tiene un DD. Otro DD puede ser una marca de 12 kDa (107 aminoácidos) basada en una proteína FKBP mutada, estabilizada por el ligando Shield1; véase, por ejemplo, *Nature Methods* 5, (2008). Por ejemplo, el DD puede ser una proteína de unión a FK506 12 (FKBP12) modificada que se une y es estabilizada de manera reversible por una molécula pequeña sintética, biológicamente inerte, Shield-1; véase, por ejemplo, Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ, *A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules*; *Cell*, 2006; 126: 995-1004; Banaszynski LA, Sellmyer MA, Contag CH, Wandless TJ, Thorne SH, *Chemical control of protein stability and function in living mice*; *Nat Med.* 2008; 14: 1123-1127; Maynard-Smith LA, Chen LC, Banaszynski LA, Ooi AG, Wandless TJ, *A directed approach for engineering conditional protein stability using biologically silent small molecules*, *The Journal of Biological Chemistry*; 2007; 282: 24866-24872; y Rodriguez, *Chem Biol.*, 23 de marzo, 2012; 19(3): 391-398; todas las cuales se pueden emplear en la práctica de la invención en la selección de un DD para asociarlo con una enzima CRISPR en la práctica de esta invención. Como se puede observar, el conocimiento en el arte incluye numerosos DD, y el DD se puede asociar, por ejemplo, fusionar, ventajosamente con un conector, a una enzima CRISPR, con lo cual el DD se puede estabilizar en la presencia de un ligando y cuando no está presente, el DD se puede desestabilizar, con lo cual toda la enzima CRISPR es desestabilizada, o el DD se puede estabilizar en la ausencia de un ligando y cuando el ligando está presente el DD se puede desestabilizar; el DD permite que la enzima CRISPR, y por ende el complejo o sistema CRISPR-Cas, sea regulada o controlada –encendida o apagada por decirlo así, para así proveer un medio para la regulación o el control del sistema, por ejemplo, en un entorno *in vivo* o *in vitro*. Por ejemplo, cuando una proteína de interés se expresa como una fusión con la marca DD, es desestabilizada y degradada rápidamente en la célula, por ejemplo, por proteasomas. Por consiguiente, la ausencia del ligando estabilizante conduce a la degradación de una D asociada a Cas. Cuando se fusiona un DD nuevo a una proteína de interés, su inestabilidad es conferida a la proteína de interés, dando como resultado la degradación rápida de toda la proteína de fusión. La actividad pico de Cas a veces es beneficiosa para reducir los efectos fuera del diana. Por consiguiente, se prefieren los estallidos cortos de gran actividad. La presente invención puede proveer dichos picos. En algunos contextos, el sistema es inducible. En algunos otros contextos, el sistema es reprimido en la ausencia del ligando estabilizante y es desreprimido en la presencia del ligando estabilizante.

Mutaciones de la enzima que reducen los efectos fuera del diana

En un aspecto, los métodos y composiciones de la invención reivindicada proveen el uso de una enzima CRISPR no natural o modificada, que es una enzima CRISPR de clase 2 de tipo V descrita en la presente tal como preferiblemente, pero en un sentido no taxativo, Cpf1 descrita en otra parte en la presente, que tiene una o más mutaciones que dan como resultado efectos fuera del diana reducidos, es decir enzimas CRISPR mejoradas para emplearlas en la modificación de los loci diana pero que reducen o eliminan la actividad fuera del diana, tal como cuando forma complejos con ARN guía, así como enzimas CRISPR mejoradas para aumentar la actividad de las enzimas CRISPR, tal como cuando forma complejos con ARN guía. Se comprenderá que las enzimas mutadas que se describen más adelante en la presente se pueden usar en cualquiera de los métodos de acuerdo con la invención reivindicada como se describe en otra parte en la presente. Cualquiera de los métodos, productos, composiciones y usos descritos en otra parte en la presente son igualmente aplicables utilizando las enzimas CRISPR mutadas como se describirá con mayor detalle más adelante. Se comprenderá que en los aspectos y las formas de realización descritos en la presente, cuando se hace referencia o se indica a Cpf1 como la enzima CRISPR, la reconstitución de un sistema CRISPR-Cas funcional preferiblemente no requiere o no depende de una secuencia tracr y/o la repetición directa está 5' (corriente arriba) con respecto a la secuencia guía (diana o espaciadora).

A modo de lineamiento, se proveen los siguientes aspectos y formas de realización particulares.

Los inventores han determinado sorprendentemente que se pueden efectuar modificaciones en las enzimas CRISPR que confieren una actividad fuera de diana reducida en comparación con las enzimas CRISPR no modificadas y/o una actividad aumentada en el diana en comparación con las enzimas CRISPR no modificadas. Por consiguiente, en determinados aspectos de la invención reivindicada en la presente se incluye el uso de enzimas CRISPR mejoradas que pueden ser de utilidad en una gran variedad de aplicaciones modificadoras de genes. En la presente también se proveen complejos, composiciones y sistemas CRISPR, así como métodos y usos, donde todos comprenden las enzimas CRISPR modificadas divulgadas en la presente.

En esta divulgación, el término "Cas" puede significar "Cpf1" o una enzima CRISPR. En el contexto de este aspecto de la invención, se muta o se modifica una enzima CRISPR o Cpf1, "con lo cual la enzima en el complejo CRISPR tendrá una capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada" (o expresiones parecidas); y, al leer esta descripción, los términos "Cpf1" o "Cas" o "enzima CRISPR y semejantes pretenden incluir las formas mutadas o modificadas de Cpf1 o Cas o la enzima CRISPR de acuerdo con la invención, es decir, "con lo cual la enzima en el complejo CRISPR tiene una capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada" (o expresiones parecidas).

En un aspecto, se provee una proteína Cpfl modificada mediante ingeniería definida en la presente, tal como Cpfl, en donde la proteína forma complejos con una molécula de ácido nucleico que comprende ARN para formar un complejo CRISPR, en donde cuando está en el complejo CRISPR, la molécula de ácido nucleico busca como diana uno o más loci de polinucleótidos diana, la proteína comprende por lo menos una modificación en comparación con una proteína Cpfl no modificada, y en donde el complejo CRISPR que comprende la proteína modificada tiene una actividad alterada en comparación con el complejo que comprende la proteína Cpfl no modificada. Se comprenderá que cuando en la presente se hace referencia a una "proteína" CRISPR, la proteína Cpfl preferiblemente es una enzima CRISPR modificada (por ejemplo, con una mayor o menor (o ninguna) actividad enzimática tal como, en un sentido no taxativo, inclusive Cpfl. El término "proteína CRISPR" se puede usar indistintamente con el término "enzima CRISPR", independientemente de si la proteína CRISPR tiene una actividad enzimática alterada, tal como aumentada o disminuida (o ninguna), en comparación con la proteína CRISPR de tipo salvaje.

En un aspecto, la actividad alterada de la proteína CRISPR modificada comprende una propiedad de unión alterada respecto de la molécula de ácido nucleico que comprende ARN o los locus de polinucleótidos diana, una cinética de unión alterada respecto de la molécula de ácido nucleico que comprende ARN o los locus de polinucleótidos diana o una especificidad de unión alterada respecto de la molécula de ácido nucleico que comprende ARN o los locus de polinucleótidos diana en comparación con los locus de polinucleótidos fuera del diana.

En algunas formas de realización, la Cas no modificada presenta actividad de clivaje de ADN, tal como Cpfl. En algunas formas de realización, la Cas dirige el clivaje de una o ambas hebras en la ubicación de una secuencia diana, tal como dentro de la secuencia diana y/o dentro del complemento de la secuencia diana. En algunas formas de realización, la Cas dirige el clivaje de una o ambas hebras dentro de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500 o más pares de bases desde el primer o último nucleótido de una secuencia diana. En algunas formas de realización, un vector codifica una Cas que está mutada con respecto a una correspondiente enzima de tipo salvaje de manera tal que la Cas mutada no tiene la capacidad para clivar una o ambas hebras de un polinucleótido diana que contiene una secuencia diana. En algunas formas de realización, se considera que Cas carece sustancialmente de toda la actividad de clivaje de ADN cuando dicha actividad de clivaje de ADN de la enzima mutada no es superior a aproximadamente un 25%, 10%, 5%, 1%, 0,1%, 0,01%, o menos de la actividad de clivaje del ADN de la forma no mutada de la enzima; un ejemplo de ello puede ser cuando la actividad de clivaje de ADN de la forma mutada es nula o despreciable con la forma no mutada. Por lo tanto, la Cas9 puede comprender una o más mutaciones y se puede usar como una proteína genérica de unión a ADN con o sin fusión a un dominio funcional. Las mutaciones pueden ser mutaciones introducidas artificialmente o mutaciones de ganancia o pérdida de función. En un aspecto de los métodos y composiciones de la invención, la enzima Cas se puede fusionar a una proteína, por ejemplo, una TAG, y/o a un dominio inducible/controlable, tal como un dominio inducible/controlable químicamente. La proteína Cas en los métodos y composiciones de la invención puede ser una proteína Cas quimérica; por ejemplo, una Cas con una función mejorada por el hecho de ser una quimera. Las proteínas Cas quiméricas pueden ser Cas nuevas que contienen fragmentos de más de una Cas natural. Pueden comprender fusiones de uno o más fragmentos N-terminales de un homólogo de Cas9 con uno o más fragmentos C-terminales de otro homólogo de Cas9. La Cas se puede suministrar en la célula en la forma de ARNm. La expresión de Cas puede estar bajo el control de un promotor inducible. Aún más, la frase "con lo cual la enzima en el complejo CRISPR tiene una capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada y/o con lo cual la enzima en el complejo CRISPR tiene una mayor capacidad para modificar dichos uno o más loci diana en comparación con una enzima no modificada" (o expresiones parecidas) no pretende indicar mutaciones que solamente dan como resultado una nickasa o una Cas muerta o mutaciones conocidas de Cas9. Sin embargo, esto no significa que los presentes métodos o composiciones de las modificaciones o mutaciones de la presente invención "con la cuales la enzima en el complejo CRISPR tiene una capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada y/o con las cuales la enzima en el complejo CRISPR tiene una mayor capacidad para modificar dichos uno o más loci diana en comparación con una enzima no modificada" (o expresiones parecidas) no se puedan combinar con mutaciones que dan como resultado que la enzima sea una nickasa o una enzima muerta. Dicha enzima muerta puede ser un agente de unión de moléculas de ácido nucleico mejorado. Y dicha nickasa puede ser una nickasa muerta. Por ejemplo, el cambio de uno o más aminoácidos neutros en y/o cerca del surco y/u otros residuos con carga en otras ubicaciones en Cas que están muy próximos a un ácido nucleico (por ejemplo, ADN, ADNc, ARN, ARNg) por uno o más aminoácidos de carga positiva pueden dar como resultado "que la enzima en el complejo CRISPR tenga una capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada y/o que la enzima en el complejo CRISPR tenga una mayor capacidad para modificar dichos uno o más loci diana en comparación con una enzima no modificada", por ejemplo, más cortes. Dado que pueden ser cortes mejorados tanto en y como fuera del diana (una Cpfl supercortadora), se puede utilizar con lo que se conoce en el arte como una guía-tru o un ARNgs-tru (véase, por ejemplo, Fu y col., "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs", *Nature Biotechnology* 32, 279-284 (2014) doi:10.1038/nbt.2808, recibido el 17 de noviembre de 2013, aceptado el 6 de enero de 2014, publicado en línea el 26 de enero de 2014, corregido en línea el 29 de enero de 2014) para que tenga una mayor actividad en el diana sin un aumento de cortes fuera del diana o para elaborar nickasas supercortadoras, o para su combinación con una mutación que vuelve a la Cas muerta para un super agente de unión. En determinadas formas de realización, la actividad alterada de la proteína Cpfl modificada comprende una eficacia de direccionamiento aumentada o una unión fuera del diana reducida. En determinadas formas de realización, la actividad alterada de la proteína Cpfl modificada comprende una actividad de clivaje modificada.

En determinadas formas de realización, la actividad alterada comprende una propiedad de unión alterada respecto de la molécula de ácido nucleico que comprende ARN o los locus de polinucleótidos diana, una cinética de unión alterada respecto de la molécula de ácido nucleico que comprende ARN o los locus de polinucleótidos diana o una especificidad de unión alterada respecto de la molécula de ácido nucleico que comprende ARN o los locus de polinucleótidos diana en comparación con los locus de polinucleótidos fuera del diana.

En determinadas formas de realización, la actividad alterada comprende una eficacia de direccionamiento aumentada o una unión fuera del diana reducida. En determinadas formas de realización, la actividad alterada comprende una actividad de clivaje modificada. En determinadas formas de realización, la actividad alterada comprende una mayor actividad de clivaje respecto de los loci de polinucleótidos diana. En determinadas formas de realización, la actividad alterada comprende una actividad de clivaje reducida respecto de los loci de polinucleótidos diana. En determinadas formas de realización, la actividad alterada comprende una actividad de clivaje reducida respecto de los loci de polinucleótidos fuera del diana. En determinadas formas de realización, la actividad alterada comprende una actividad de clivaje aumentada respecto de los loci de polinucleótidos fuera del diana.

Por lo tanto, en determinadas formas de realización de los métodos y composiciones reivindicados de la invención, hay una mayor especificidad por los loci de polinucleótidos diana en comparación con los loci de polinucleótidos fuera del diana. En otras formas de realización, hay una especificidad reducida por los loci de polinucleótidos diana en comparación con loci de polinucleótidos fuera del diana.

En un aspecto de los métodos y composiciones reivindicados de la invención, la actividad alterada de la proteína CRISPR modificada comprende una cinética de helicasa alterada.

En un aspecto de los métodos y composiciones reivindicados de la invención, la proteína Cpfl modificada comprende una modificación que altera la asociación de la proteína con la molécula de ácido nucleico que comprende ARN, o una hebra de los loci de polinucleótidos diana, o una hebra de los loci de polinucleótidos fuera del diana. En un aspecto de la invención, la proteína Cpfl modificada comprende una modificación que altera la formación del complejo CRISPR.

En determinadas formas de realización de los métodos y composiciones reivindicados, la proteína Cpfl modificada comprende una modificación que altera el direccionamiento de la molécula de ácido nucleico a los loci de polinucleótidos. En determinadas formas de realización, la modificación comprende una mutación en una región de la proteína que se asocia con la molécula de ácido nucleico. En determinadas formas de realización, la modificación comprende una mutación en una región de la proteína que se asocia con una hebra de los loci de polinucleótidos diana. En determinadas formas de realización, la modificación comprende una mutación en una región de la proteína que se asocia con una hebra de los loci de polinucleótidos fuera del diana. En determinadas formas de realización, la modificación o mutación comprende menos cargas positivas en una región de la proteína que se asocia con la molécula de ácido nucleico que comprende ARN, o una hebra de los loci de polinucleótidos diana, o una hebra de los loci de polinucleótidos fuera del diana. En determinadas formas de realización, la modificación o mutación comprende menos cargas negativas en una región de la proteína que se asocia con la molécula de ácido nucleico que comprende ARN, o una hebra de los loci de polinucleótidos diana, o una hebra de los loci de polinucleótidos fuera del diana. En determinadas formas de realización, la modificación o mutación comprende más cargas positivas en una región de la proteína que se asocia con la molécula de ácido nucleico que comprende ARN, o una hebra de los loci de polinucleótidos diana, o una hebra de los loci de polinucleótidos fuera del diana. En determinadas formas de realización, la modificación o mutación comprende más cargas negativas en una región de la proteína que se asocia con la molécula de ácido nucleico que comprende ARN, o una hebra de los loci de polinucleótidos diana, o una hebra de los loci de polinucleótidos fuera del diana. En determinadas formas de realización, la modificación o mutación aumenta el impedimento estérico entre la proteína y la molécula de ácido nucleico que comprende ARN, o una hebra de los loci de polinucleótidos diana, o una hebra de loci de polinucleótidos fuera del diana. En determinadas formas de realización, la modificación o mutación comprende una sustitución de Lys, His, Arg, Glu, Asp, Ser, Gly o Thr. En determinadas formas de realización, la modificación o mutación comprende una sustitución con Gly, Ala, Ile, Glu o Asp. En determinadas formas de realización, la modificación o mutación comprende una sustitución de aminoácidos en un surco de unión.

En un aspecto, la solicitud provee:

una enzima CRISPR no natural definida en la presente, tal como Cpf1, en donde:

la enzima forma complejos con el ARN guía para formar un complejo CRISPR,

cuando está en el complejo CRISPR, el ARN guía busca como diana uno o más loci de polinucleótidos diana y la enzima altera los loci de polinucleótidos, y

la enzima comprende por lo menos una modificación,

con lo cual la enzima en el complejo CRISPR tiene una capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada y/o con lo cual la enzima en el complejo CRISPR tiene una mayor capacidad para modificar dichos uno o más loci diana en comparación con una enzima no modificada.

En cualquiera de dichas enzimas CRISPR no naturales, la modificación puede comprender la modificación de uno o más residuos de aminoácidos de la enzima.

En cualquiera de dichas enzimas CRISPR no naturales, la modificación puede comprender la modificación de uno o más residuos de aminoácidos localizados en una región que comprende residuos de carga positiva en la enzima no modificada.

En cualquier enzima CRISPR no natural, la modificación puede comprender la modificación de uno o más residuos de aminoácidos de carga positiva en la enzima no modificada.

En cualquier enzima CRISPR no natural, la modificación puede comprender la modificación de uno o más residuos de aminoácidos de carga no positiva en la enzima no modificada.

La modificación puede comprender la modificación de uno o más residuos de aminoácidos sin carga en la enzima no modificada.

La modificación puede comprender la modificación de uno o más residuos de aminoácidos de carga negativa en la enzima no modificada.

La modificación puede comprender la modificación de uno o más residuos de aminoácidos hidrofóbicos en la enzima no modificada.

La modificación puede comprender la modificación de uno o más residuos de aminoácidos polares en la enzima no modificada.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la modificación puede comprender la modificación de uno o más residuos ubicados en un surco.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la modificación puede comprender la modificación de uno o más residuos ubicados fuera de un surco.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la modificación comprende una modificación de uno o más residuos, en donde dichos uno o más residuos comprenden arginina, histidina o lisina.

En cualquiera de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se puede modificar mediante mutación de dichos uno o más residuos.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dicho uno o más residuos, y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada con un residuo de alanina.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dicho uno o más residuos, y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada con ácido aspártico o ácido glutámico.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dicho uno o más residuos, y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada con serina, treonina, asparagina o glutamina.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dicho uno o más residuos, y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada con alanina, glicina, isoleucina, leucina, metionina, fenilalanina, triptófano, tirosina o valina.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dicho uno o más residuos, y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada con un residuo de aminoácido polar.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dichos uno o más residuos y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada por un residuo de aminoácido que no es un residuo de aminoácido polar.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dichos uno o más residuos y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada por un residuo de aminoácido cargado negativamente.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dichos uno o más residuos y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada por un residuo de aminoácido que no es un residuo de aminoácido cargado negativamente.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación

de dichos uno o más residuos y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada por un residuo de aminoácido sin carga.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dichos uno o más residuos y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada por un residuo de aminoácido que no es un residuo de aminoácido sin carga.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dichos uno o más residuos y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada por un residuo de aminoácido hidrofóbico.

En algunas de las enzimas CRISPR no naturales descritas con anterioridad, la enzima se modifica mediante mutación de dichos uno o más residuos y en donde la mutación comprende la sustitución de un residuo en la enzima no modificada por un residuo de aminoácido que no es un residuo de aminoácido hidrofóbico.

En algunas formas de realización, la enzima CRISPR, tal como preferiblemente la enzima Cpfl, deriva de una Cpfl de *Francisella tularensis* 1, *Francisella tularensis* subsp. *novicida*, *Prevotella albensis*, *Lachnospiraceae bacterium* MC2017 1, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium* GW2011_GWA2_33_10, *Parcubacteria bacterium* GW2011_GWC2_44_17, *Smithella* sp. SCADC, *Acidaminococcus* sp. BV3L6, *Lachnospiraceae bacterium* MA2020, *Candidatus Metanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi* 237, *Leptospira inadai*, *Lachnospiraceae bacterium* ND2006, *Porphyromonas crevioricanis* 3, *Prevotella disiens* o *Porphyromonas macacae* (por ejemplo, una Cpfl de uno de estos organismos modificados como se describe en la presente), y puede incluir otras mutaciones o alteraciones o puede ser una Cpf1 quimérica.

En determinadas formas de realización, la proteína Cpf1 comprende uno o más dominios de señales de localización nuclear (NLS). En determinadas formas de realización, la proteína Cpf1 comprende por lo menos dos o más NLS.

En determinadas formas de realización, la proteína Cpf1 comprende una proteína CRISPR quimérica, que comprende un primer fragmento de un primer ortólogo de CRISPR y un segundo fragmento de un segundo ortólogo de CRISPR, y el primer y segundo ortólogos de CRISPR son diferentes.

En determinadas formas de realización, la enzima se modifica o comprende una modificación, por ejemplo, comprende, consiste esencialmente en o consiste en una modificación mediante mutación de cualquiera de los residuos enumerados en la presente o un residuo correspondiente en el ortólogo respectivo; o la enzima comprende, consiste esencialmente en o consiste en la modificación en cualquiera entre una (simple), dos (doble), tres (triple), cuatro (cuádruple) o más posiciones de acuerdo con la divulgación de toda esta solicitud, o un residuo correspondiente o una posición en la enzima ortóloga de CRISPR, por ejemplo, una enzima que comprende, que consiste esencialmente en o que consiste en la modificación en cualquiera de los residuos de Cpf1 indicados en la presente, o un residuo o una posición correspondiente en la enzima ortóloga de CRISPR. En dicha enzima, cada residuo se puede modificar mediante sustitución por un residuo de alanina.

Los solicitantes describieron recientemente un método para la generación de ortólogos de Cas9 con una mayor especificidad (Slaymaker y col., 2015 “Rationally engineered Cas9 nucleases with improved specificity”). Esta estrategia se puede usar para mejorar la especificidad de los ortólogos de Cpf1. Los residuos primarios para la mutagénesis son preferiblemente todos residuos de carga positiva dentro del dominio RuvC. Otros residuos adicionales son residuos de carga positiva que están conservados entre diferentes ortólogos.

En determinadas formas de realización, se puede mejorar la especificidad de Cpf1 mediante mutación de residuos que estabilizan la hebra de ADN no buscada como diana.

En algunas de las enzimas Cpf1 no naturales descritas con anterioridad, la enzima se modifica mediante mutación de uno o más residuos (en el dominio RuvC) que incluyen, pero en un sentido no taxativo, las posiciones R909, R912, R930, R947, K949, R951, R955, K965, K968, K1000, K1002, R1003, K1009, K1017, K1022, K1029, K1035, K1054, K1072, K1086, R1094, K1095, K1109, K1118, K1142, K1150, K1158, K1159, R1220, R1226, R1242 y/o R1252 con referencia a la numeración de posiciones de aminoácidos de AsCpfl (*Acidaminococcus* sp. BV3L6).

En algunas de las enzimas Cpfl no naturales descritas con anterioridad, la enzima se modifica mediante mutación de uno o más residuos (en el dominio RAD50) que incluyen, pero en un sentido no taxativo, las posiciones K324, K335, K337, R331, K369, K370, R386, R392, R393, K400, K404, K406, K408, K414, K429, K436, K438, K459, K460, K464, R670, K675, R681, K686, K689, R699, K705, R725, K729, K739, K748 y/o K752 con referencia a la numeración de posiciones de aminoácidos de AsCpf1 (*Acidaminococcus* sp. BV3L6).

En algunas de las enzimas Cpfl no naturales descritas previamente, la enzima se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones R912, T923, R947, K949, R951, R955, K965, K968, K1000, R1003, K1009, K1017, K1022, K1029, K1072, K1086, F1103, R1226 y/o R1252 con referencia a la numeración de posiciones de aminoácidos de AsCpf1 (*Acidaminococcus* sp. BV3L6).

En determinadas formas de realización, la enzima se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935,

K940, K948, K953, K960, K984, K1003, K1017, R1033, R1138, R1165 y/o R1252 con referencia a la numeración de posiciones de aminoácidos de LbCpf1 (*Lachnospiraceae bacterium* ND2006).

En determinadas formas de realización, la enzima Cpfl se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones K15, R18, K26, Q34, R43, K48, K51, R56, R84, K85, K87, N93, R103, N104, T118, K123, K134, R176, K177, R192, K200, K226, K273, K275, T291, R301, K307, K369, S404, V409, K414, K436, K438, K468, D482, K516, R518, K524, K530, K532, K548, K559, K570, R574, K592, D596, K603, K607, K613, C647, R681, K686, H720, K739, K748, K757, T766, K780, R790, P791, K796, K809, K815, T816, K860, R862, R863, K868, K897, R909, R912, T923, R947, K949, R951, R955, K965, K968, K1000, R1003, K1009, K1017, K1022, K1029, A1053, K1072, K1086, F1103, S1209, R1226, R1252, K1273, K1282 y/o K1288 con referencia a la numeración de posiciones de aminoácidos de AsCpf1 (*Acidaminococcus* sp. BV3L6).

En determinadas formas de realización, la enzima Cpfl se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones K15, R18, K26, R34, R43, K48, K51, K56, K87, K88, D90, K96, K106, K107, K120, Q125, K143, R186, K187, R202, K210, K235, K296, K298, K314, K320, K326, K397, K444, K449, E454, A483, E491, K527, K541, K581, R583, K589, K595, K597, K613, K624, K635, K639, K656, K660, K667, K671, K677, K719, K725, K730, K763, K782, K791, R800, K809, K823, R833, K834, K839, K852, K858, K859, K869, K871, R872, K877, K905, R918, R921, K932, I960, K962, R964, R968, K978, K981, K1013, R1016, K1021, K1029, K1034, K1041, K1065, K1084 y/o K1098 con referencia a la numeración de posiciones de aminoácidos de FnCpfl (*Francisella novicida* U112).

En determinadas formas de realización, la enzima Cpfl se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones K15, R18, K26, K34, R43, K48, K51, R56, K83, K84, R86, K92, R102, K103, K116, K121, R158, E159, R174, R182, K206, K251, K253, K269, K271, K278, P342, K380, R385, K390, K415, K421, K457, K471, A506, R508, K514, K520, K522, K538, Y548, K560, K564, K580, K584, K591, K595, K601, K634, K640, R645, K679, K689, K707, T716, K725, R737, R747, R748, K753, K768, K774, K775, K785, K787, R788, Q793, K821, R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, K1121, R1138, R1165, K1190, K1199, y/o K1208 con referencia a la numeración de posiciones de aminoácidos de LbCpf1 (*Lachnospiraceae bacterium* ND2006).

En determinadas formas de realización, la enzima se modifica mediante mutación de uno o más residuos que incluyen, pero en un sentido no taxativo, las posiciones K14, R17, R25, K33, M42, Q47, K50, D55, K85, N86, K88, K94, R104, K105, K118, K123, K131, R174, K175, R190, R198, I221, K267, Q269, K285, K291, K297, K357, K403, K409, K414, K448, K460, K501, K515, K550, R552, K558, K564, K566, K582, K593, K604, K608, K623, K627, K633, K637, E643, K780, Y787, K792, K830, Q846, K858, K867, K876, K890, R900, K901, M906, K921, K927, K928, K937, K939, R940, K945, Q975, R987, R990, K1001, R1034, I1036, R1038, R1042, K1052, K1055, K1087, R1090, K1095, N1103, K1108, K1115, K1139, K1158, R1172, K1188, K1276, R1293, A1319, K1340, K1349, y/o K1356 con referencia a la numeración de posiciones de aminoácidos de MbCpf1 (*Moraxella bovoculi* 237).

En cualquiera de las enzimas CRISPR no naturales:

puede existir una única falta de coincidencia entre el diana y una secuencia correspondiente de dichos uno o más loci fuera del diana; y/o

pueden existir dos, tres o cuatro o más faltas de coincidencia entre el diana y una secuencia correspondiente de dichos uno o más loci fuera del diana y/o

en donde en (ii) dichos dos, tres o cuatro o más faltas de coincidencia son contiguas.

En cualquiera de las enzimas CRISPR no naturales, la enzima en el complejo CRISPR puede tener capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada y en donde la enzima en el complejo CRISPR tiene una capacidad aumentada para modificar dichos loci diana en comparación con una enzima no modificada.

En cualquiera de las enzimas CRISPR no naturales, cuando está en el complejo de CRISPR, la diferencia relativa de la capacidad modificadora de la enzima entre el diana y por lo menos un locus fuera del diana puede estar aumentada en comparación con la deferencia relativa de una enzima no modificada.

En cualquiera de las enzimas CRISPR no naturales, la enzima CRISPR puede comprender una o más mutaciones adicionales, en donde dichas una o más mutaciones están en uno o más dominios catalíticamente activos.

En dichas enzimas CRISPR no naturales, la enzima CRISPR puede tener una actividad nucleasa reducida o suprimida en comparación con una enzima que carece de dichas una o más mutaciones adicionales.

En dichas enzimas CRISPR no naturales, la enzima CRISPR no dirige el clivaje de una u otra hebra de ADN en la ubicación de la secuencia diana.

En los casos donde la enzima CRISPR comprende una o más mutaciones adicionales en uno o más dominios catalíticamente activos, dichas una o más mutaciones adicionales pueden estar en un dominio catalíticamente activo

de la enzima CRISPR que comprende RuvCI, RuvCII o RuvCIII.

Sin limitaciones por la teoría, en un aspecto de la invención, los métodos y las mutaciones descritas proveen un reordenamiento de conformación mejorado de los dominios de la enzima CRISPR (por ejemplo, los dominios Cpf1) a posiciones que dan como resultado el clivaje en sitios en el diana y evitan dichos estados de conformación en sitios fuera del diana. Las enzimas CRISPR clivan el ADN diana en una serie de pasos coordinados. En primer lugar, el dominio que interactúa con PAM reconoce la secuencia PAM 5' del ADN diana. Después de la unión a PAM, se muestrean los primeros 10 a 12 nucleótidos de la secuencia diana (secuencia semilla) para la complementariedad de ARNsg:ADN, un proceso dependiente de la separación del dúplex de ADN. Si los nucleótidos de la secuencia semilla son complementarios del ARNg, se desenrolla el resto del ADN y toda la extensión del ARNg se hibrida con la hebra de ADN diana. Los surcos que no son del diana pueden estabilizar la hebra de ADN no buscada como diana y facilitar el desenrollado por medio de interacciones no específicas con cargas positivas del esqueleto fosfato de ADN. Las interacciones de ARN:ADNc y enzima CRISPR:ADNnc dirigen el desenrollamiento del ADN en franca competencia con la rehibridación de ADNc:ADNnc. Hay otros dominios de la enzima CRISPR que también pueden afectar la conformación de los dominios nucleasa, por ejemplo, conectores que unen diferentes dominios. Por lo tanto, los métodos y las mutaciones provistos abarcan, en un sentido no taxativo, RuvCI, RuvCIII, RuvCIII y conectores. Los cambios de conformación, por ejemplo, en Cpf1 producidos por la unión al ADN diana, incluyendo la interacción con la secuencia semilla, y las interacciones con la hebra de ADN diana y no diana determinarán si los dominios están ubicados como para desencadenar una actividad nucleasa. Por lo tanto, las mutaciones y los métodos provistos en la presente demuestran y permiten modificaciones que van más allá del reconocimiento de PAM y el apareamiento de bases de ARN-ADN.

En un aspecto, los métodos y composiciones reivindicados de la invención proveen las nucleasas CRISPR definidas en la presente, tal como Cpfl, que comprenden un equilibrio mejorado hacia las conformaciones asociadas con la actividad de clivaje cuando están involucradas en las interacciones en el diana y/o un equilibrio mejorado fuera de las conformaciones asociadas con la actividad de clivaje cuando están involucradas en las interacciones fuera del diana. En un aspecto, los métodos y composiciones reivindicados de la invención proveen nucleasas Cas (por ejemplo, Cpf1) con una función de corrección de lectura mejorada, es decir una nucleasa Cas (por ejemplo, Cpf1) que adopta una conformación que comprende actividad nucleasa en un sitio en el diana, y donde dicha conformación es más desfavorable en un sitio fuera del diana. Sternberg y col., *Nature* 527(7576): 110-3, doi:10.1038/nature15544, publicada en línea el 28 de octubre 2015; publicación electrónica el 28 de octubre de 2015, emplearon experimentos con transferencia de energía por resonancia de Förster (FRET) para detectar las orientaciones relativas de los dominios catalíticos de Cas (por ejemplo, Cpf1) cuando se asocian con ADN en y fuera del diana, y que se pueden extrapolar a las enzimas CRISPR de los métodos y composiciones reivindicados de la presente invención (por ejemplo, Cpf1).

Los métodos y composiciones reivindicados de la invención proveen además el uso de métodos y mutaciones para modular la actividad y/o especificidad de una nucleasa usando ARN guía modificados. Según se expuso antes, se puede aumentar o reducir la actividad nucleasa en el diana. Asimismo, se puede aumentar o reducir la actividad nucleasa fuera del diana. Además, puede haber una especificidad aumentada o reducida en cuanto a la actividad en el diana frente a la actividad fuera del diana. Los ARN guía modificados incluyen, en un sentido no taxativo, ARN guía truncados, ARN guía desactivados, ARN guía modificados químicamente, ARN guía asociados con dominios funcionales, ARN guía modificados que comprenden dominios funcionales, ARN guía modificados que comprenden aptámeros, ARN guía modificados que comprenden proteínas adaptadoras y ARN guía que comprenden bucles agregados o modificados. En algunas formas de realización, se asocia uno o más dominios funcionales con un ARNg muerto (ARNd). En algunas formas de realización, un complejo de ARNd con la enzima CRISPR dirige la regulación de genes mediante un dominio funcional en un locus genético, en tanto un ARNg dirige el clivaje del ADN mediante la enzima CRISPR en otro locus. En algunas formas de realización, los ARNd se seleccionan para maximizar la selectividad de la regulación para un locus genético de interés en comparación con una regulación fuera del diana. En algunas formas de realización, los ARNd se seleccionan para maximizar la regulación del gen diana y minimizar el clivaje diana.

A los efectos de la siguiente descripción, la referencia a un dominio funcional podría ser un dominio funcional asociado con la enzima CRISPR o un dominio funcional asociado con la proteína adaptadora.

En la práctica de la invención, se pueden extender bucles del ARNg, sin colisionar con la proteína Cas (por ejemplo, Cpf1) mediante la inserción de uno o más bucles de ARN distintos o de secuencias distintas que pueden reclutar proteínas adaptadoras que se podrán unir a los distintos bucles de ARN o a secuencias distintas. Las proteínas adaptadoras pueden incluir, pero en un sentido no taxativo, combinaciones ortogonales de la proteína de unión a ARN/aptámero que existen como parte de la diversidad de proteínas de la envoltura de bacteriófagos. Un listado de dichas proteínas de la envoltura incluye, pero en un sentido no taxativo: Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, φCb5, φCb8r, φCb12r, φCb23r, 7s y PRR1. Estas proteínas adaptadoras o proteínas de unión a ARN ortogonales también pueden reclutar proteínas o fusiones efectoras que comprenden uno o más dominios funcionales. En algunas formas de realización, el dominio funcional se puede seleccionar del grupo que consiste en: dominio transposasa, dominio integrasa, dominio recombinasa, dominio resolvasa, dominio invertasa, dominio proteasa, dominio ADN metiltransferasa, dominio ADN hidroxilmetilasa, dominio ADN desmetilasa, dominio histona acetilasa, dominio histona desacetilasa, dominio nucleasa, dominio

represor, dominio activador, dominios de señales de localización nuclear, dominio de la proteína reguladora de la transcripción (o reclutamiento del complejo de transcripción), dominio asociado a la actividad de captación celular, dominio de unión a ácidos nucleicos, dominio de presentación de anticuerpos, enzimas modificadoras de histonas, reclutadores de enzimas modificadoras de histonas; inhibidores de enzimas modificadoras de histonas, histona metiltransferasa, histona desmetilasa, histona quinasa, histona fosfatasa, histona ribosilasa, histona desribosilasa, histona ubiquitinasa, histona desubiquitinasa, histona biotinasa e histona cola proteasa. En algunas formas de realización preferidas, el dominio funcional es un dominio de activación de la transcripción tal como, en un sentido no taxativo, VP64, p65, MyoD1, HSF1, RTA, SET7/9 o una histona acetiltransferasa. En algunas formas de realización, el dominio funcional es un dominio de represión de la transcripción, preferiblemente KRAB. En algunas formas de realización, el dominio de represión de la transcripción es SID o concatámeros de SID (por ejemplo, SID4X). En algunas formas de realización, el dominio funcional es un dominio de modificación epigenética, de modo que se provee una enzima de modificación epigenética. En algunas formas de realización, el dominio funcional es un dominio de activación, que puede ser el dominio de activación P65. En algunas formas de realización, el dominio funcional es una desaminasa, tal como una citidina desaminasa. La citidina desaminasa se puede dirigir a un ácido nucleico diana donde dirige la conversión de citidina en uridina, dando como resultado sustituciones de C a T (G a A en la hebra complementaria). En una forma de realización tal, las sustituciones de nucleótidos pueden efectuarse sin clivaje de ADN.

En un aspecto, los métodos y composiciones reivindicados de la invención también proveen el uso de métodos y mutaciones para modular la actividad de unión y/o especificidad de unión de Cas (por ejemplo, Cpf1). En determinadas formas de realización, se utilizan proteínas Cas (por ejemplo, Cpf1) que carecen de actividad nucleasa. En determinadas formas de realización, se emplean ARN guía modificados que promueven la unión pero no la actividad nucleasa de una nucleasa Cas (por ejemplo, Cpf1). En dichas formas de realización, se puede aumentar o reducir la unión en el diana. Asimismo, en dichas formas de realización se puede aumentar o reducir la unión fuera del diana. Además, puede haber una especificidad aumentada o reducida en cuanto a la unión en el diana frente a la unión fuera del diana.

En formas de realización particulares, la reducción del clivaje fuera del diana se puede asegurar mediante separación de la hebra desestabilizante, más particularmente mediante introducción de mutaciones en la enzima Cpf1 que disminuyan la carga positiva en las regiones de interacción con el ADN (como se describe en la presente y es ejemplificado también para Cas9 por Slaymaker y col., 2016 (*Science*, 1; 351(6268): 84-8). En formas de realización adicionales, la reducción del clivaje fuera del diana se asegura mediante la introducción de mutaciones en la enzima Cpf1 que afecta la interacción entre la hebra diana y la secuencia de ARN guía, más particularmente, mediante alteración de las interacciones entre Cpf1 y el esqueleto de fosfato de la hebra de ADN diana, de una manera tal que se retiene la actividad específica del diana pero se reduce la actividad fuera del diana (como se describe para Cas9 en Kleinstiver y col., 2016, *Nature*, 28; 529(7587): 490-5). En formas de realización particulares, la actividad fuera del diana se reduce mediante una Cpf1 modificada, en donde se modifica tanto la interacción con la hebra diana como con la hebra no diana, en comparación con la Cpf1 de tipo salvaje.

Los métodos y las mutaciones que se pueden emplear en varias combinaciones para aumentar o reducir la actividad y/o la especificidad de la actividad en el diana frente a fuera del diana, o para aumentar o reducir la unión y/o la especificidad de la unión en el diana frente a fuera del diana, se pueden usar para compensar o potenciar las mutaciones o modificaciones producidas para promover otros efectos. Dichas mutaciones o modificaciones realizadas para promover otros efectos incluyen mutaciones o modificación en la proteína Cas (por ejemplo, Cpf1) y/o una mutación o modificación efectuada en un ARN guía. En determinadas formas de realización, los métodos y las mutaciones se usan con ARN guía modificados químicamente. Los ejemplos de modificaciones químicas del ARN guía incluyen, en un sentido no taxativo, la incorporación de 2'-O-metilo (M), 3'fosforotioato de 2'-O-metilo (MS) o 3'tioPACE de 2'-O-metilo (MSP) en uno o más nucleótidos terminales. Dichos ARN guía modificados pueden comprender una estabilidad aumentada y una actividad aumentada en comparación con los ARN guía no modificados, aunque la especificidad en el diana versus fuera del diana no es predecible. (Véase Hendel, 2015, *Nat. Biotechnol.,* 33(9): 985-9, doi: 10.1038/nbt.3290, publicado en línea el 29 de junio de 2015). Los ARN guía químicamente modificados incluyen además, en un sentido no taxativo, ARN con enlaces fosforotioato y nucleótidos de ácidos nucleicos bloqueados (LNA) que comprenden un puente metileno entre los carbonos 2' y 4' del anillo de ribosa. Los métodos y las mutaciones de la invención se usan para modular la actividad nucleasa de Cas (por ejemplo, Cpf1) y/o la unión con ARN guía modificados químicamente.

En un aspecto, los métodos y composiciones reivindicados de la invención proveen el uso de métodos y mutaciones para modular la unión y/o la especificidad de unión de las proteínas Cas (por ejemplo, Cpf1) de acuerdo con la invención definidas en la presente, que comprenden dominios funcionales tales como nucleasas, activadores de la transcripción, represores de la transcripción y semejantes. Por ejemplo, una proteína Cas (por ejemplo, Cpf1) se puede hacer nula para nucleasa o puede tener una actividad nucleasa alterada o reducida mediante la introducción de mutaciones tales como, por ejemplo, las mutaciones de Cpf1 que se describen en otra parte en la presente, y que incluyen, por ejemplo, D917A, E1006A, E1028A, D1227A, D1255A, N1257A, D917A, E1006A, E1028A, D1227A, D1255A y N1257A con referencia a las posiciones de aminoácidos en el dominio FnCpf1p RuvC; o, por ejemplo, N580A, N584A, T587A, W609A, D610A, K613A, E614A, D616A, K624A, D625A, K627A y Y629A con referencia al segundo dominio nucleasa putativo que se describe en otra parte en la presente. Las proteínas Cas (por ejemplo, Cpf1) deficientes en nucleasa son de utilidad para un suministro dependiente de la secuencia diana guiado por ARN

de los dominios funcionales. Los métodos y composiciones reivindicados de la invención proveen el uso de métodos y mutaciones para modular la unión de las proteínas Cas (por ejemplo, Cpf1). En una forma de realización, el dominio funcional comprende VP64, que proporciona un factor de transcripción guiado por ARN. En otra forma de realización, el dominio funcional comprende Fok I, que provee actividad nucleasa guiada por ARN. Se menciona la publicación de Patente de los EE.UU. N°: 2014/0356959, la publicación de Patente de los EE.UU. N°: 2014/0342456, la publicación de Patente de los EE.UU. N°: 2015/0031132, y Mali, P. y col., 2013, *Science* 339(6121): 823-6, doi:10.1126/*Science*.1232033, publicada en línea el 3 de enero de 2013 y las divulgaciones de presente la invención comprende métodos y materiales de estos documentos aplicados junto con las descripciones de la presente. En determinadas formas de realización, se aumentó la unión en el diana. En determinadas formas de realización, se disminuye la unión fuera del diana. En determinadas formas de realización, se disminuye la unión en el diana. En determinadas formas de realización, se aumenta la unión fuera del diana. Por lo tanto, los métodos y composiciones reivindicados de la invención también proveen el aumento o la disminución de la especificidad de la unión en el diana versus la unión fuera del diana de las proteínas de unión Cas (por ejemplo, Cpf1) funcionalizadas.

El uso de Cas (por ejemplo, Cpf1) como una proteína de unión guiada por ARN no se limita a la Cas (por ejemplo, Cpf1) nula para nucleasas. Las enzimas Cas (por ejemplo, Cpf1) que comprenden actividad nucleasa también pueden funcionar como proteínas de unión guiadas por ARN cuando se usan con determinados ARN guía. Por ejemplo, los ARN guía cortos y los ARN guía que comprenden nucleótidos no coincidentes con el diana pueden promover la unión de Cas (por ejemplo, Cpf1) dirigida por ARN a una secuencia diana con poco o ningún clivaje del diana. (Véase, por ejemplo, Dahlman, 2015, *Nat. Biotechnol.*, 33(11): 1159-1161, doi: 10.1038/nbt.3390, publicada en línea el 5 de octubre, 2015). En un aspecto, los métodos y composiciones reivindicados de la invención proveen el uso de métodos y mutaciones para modular la unión de proteínas Cas (por ejemplo, Cpf1) que comprenden actividad nucleasa. En determinadas formas de realización, se aumentó la unión en el diana. En determinadas formas de realización, se disminuye la unión fuera del diana. En determinadas formas de realización, se disminuye la unión en el diana. En determinadas formas de realización, se aumenta la unión fuera del diana. En determinadas formas de realización, hay una especificidad de unión aumentada o reducida en el diana con respecto a la unión fuera del diana. En determinadas formas de realización, también se modula la actividad nucleasa de la enzima Cas (por ejemplo, Cpf1)-ARN guía.

La formación del heterodúplex de ARN-ADN es importante para la actividad y especificidad de clivaje por toda la región diana, no solo la secuencia de la región semilla más próxima al PAM. Por lo tanto, los ARN guía truncados muestran una actividad y especificidad de clivaje reducidas. En un aspecto, los métodos y composiciones reivindicados de la invención proveen el uso de métodos y mutaciones para aumentar la actividad y especificidad de clivaje usando ARN guía alterados.

La divulgación también demuestra que las modificaciones de la especificidad de nucleasa Cas (por ejemplo, Cpf1) se pueden efectuar de acuerdo con las modificaciones en el rango de direccionamiento. Se pueden diseñar mutantes (por ejemplo, Cpf1) que tengan una especificidad aumentada por el diana así como modificaciones adaptadas para el reconocimiento de PAM, por ejemplo, mediante la selección de mutaciones que alteran la especificidad de PAM y que combinan esas mutaciones con mutaciones en el surco que no es del diana que aumentan (o si se desea disminuyen) la especificidad por las secuencias en el diana en comparación con las secuencias fuera del diana. En una forma de realización similar, se muta un residuo de dominio PI para adaptar el reconocimiento de una secuencia PAM deseada en tanto se mutan uno o más aminoácidos del surco que no es del diana para alterar la especificidad por el diana. Los métodos y las modificaciones de Ca (por ejemplo, Cpf1) descritos en la presente se pueden usar para contrarrestar la pérdida de especificidad debida a la alteración del reconocimiento de PAM, mejorar la ganancia de especificidad debida a la alteración del reconocimiento de PAM, contrarrestar la ganancia de especificidad debida a la alteración del reconocimiento de PAM o mejorar la pérdida de especificidad debida a la alteración del reconocimiento de PAM.

Los métodos y las mutaciones se pueden usar con cualquier enzima Cas (por ejemplo, Cpf1) con un reconocimiento de PAM alterado. Los ejemplos no taxativos de los PAM incluidos se describen en otra parte en la presente.

En formas de realización adicionales, en los métodos y las mutaciones se emplean proteínas modificadas.

En cualquiera de las enzimas CRISPR no naturales, la enzima CRISPR puede comprender uno o más dominios funcionales heterólogos.

Dichos uno o más dominios funcionales heterólogos pueden comprender uno o más dominios de señales de localización nuclear (NLS). Dichos uno o más dominios funcionales heterólogos pueden comprender al menos dos o más NLS.

Dichos uno o más dominios funcionales heterólogos pueden comprender uno o más dominios de activación de la transcripción. Un dominio de activación de la transcripción puede comprender VP64.

Dichos uno o más dominios funcionales heterólogos pueden comprender uno o más dominios de represión de la transcripción. Un dominio de represión de la transcripción puede comprender un dominio KRAB o un dominio SID.

Dichos uno o más dominios funcionales heterólogos pueden comprender uno o más dominios de nucleasa. Dichos uno o más dominios de nucleasa pueden comprender Fok1.

Dichos uno o más dominios funcionales heterólogos pueden presentar una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad de factor de liberación de la transcripción, actividad de modificación de histonas, actividad nucleasa, actividad de clivaje de un ARN de hebra doble, actividad de clivaje de ARN de hebra doble, actividad de clivaje de ADN de hebra simple, actividad de clivaje de ADN de hebra doble y actividad de unión a ácidos nucleicos.

Dichos por lo menos uno o más dominios funcionales heterólogos pueden estar en o cerca del extremo amino terminal de la enzima y/o en o cerca del extremo carboxilo terminal de la enzima.

Dichos uno o más dominios funcionales heterólogos pueden fusionarse a la enzima CRISPR o anclarse a la enzima CRISPR o ligarse a la enzima CRISPR mediante un residuo conector.

En cualquiera de las enzimas CRISPR no naturales, la enzima CRISPR puede comprender una enzima CRISPR de un organismo perteneciente a un género que comprende *Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Metanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, o Porphyromonas macacae* (por ejemplo, una Cpf1 de uno de estos organismos modificada como se describe en la presente), y pueden incluir otras mutaciones o alteraciones o pueden ser una Cas quimérica (por ejemplo, Cpf1).

En cualquiera de las enzimas CRISPR no naturales, la enzima CRISPR puede comprender una enzima Cas (por ejemplo, Cpf1) quimérica que comprende un primer fragmento de un primer ortólogo de Cas (por ejemplo, Cpf1) y un segundo fragmento de un segundo ortólogo de Cas (por ejemplo, Cpf1), y el primer y segundo ortólogos de Cas (por ejemplo, Cpf1) son diferentes. Por lo menos uno entre el primer y el segundo ortólogos de Cas (por ejemplo, Cpf1) puede comprender una Cas (por ejemplo, Cpf1) de un organismo que comprende *Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Metanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens o Porphyromonas macacae*.

En cualquiera de las enzimas CRISPR no naturales, se pueden optimizar los codones de la secuencia de nucleótidos que codifica la enzima CRISPR para su expresión en un eucariota.

En cualquiera de las enzimas CRISPR no naturales, la célula puede ser una célula eucariota o una célula procariota; en donde el complejo CRISPR es funcional en la célula y con lo cual la enzima del complejo CRISPR tiene una capacidad reducida para modificar uno o más loci fuera del diana de la célula en comparación con una enzima no modificada y/o con lo cual la enzima en el complejo CRISPR tiene una capacidad aumentada para modificar dichos uno o más loci diana en comparación con una enzima no modificada.

Por lo tanto, se desvela pero no se reivindica específicamente en la presente una célula eucariota que comprende la proteína CRISPR modificada o el sistema definidos en la presente.

En determinadas formas de realización, los métodos que se describen en la presente pueden comprender proveer una célula transgénica Cas (por ejemplo, Cpf1) en la cual se provee uno o más ácidos nucleicos que codifica uno o más ARN guía o se introducen ligados operativamente en la célula con un elemento regulador que comprende un promotor de uno o más genes de interés. Según se usa en la presente, el término "célula transgénica Cas" se refiere a una célula, tal como una célula eucariota, en la cual se ha integrado genómicamente un gen Cas. La naturaleza, el tipo o el origen de la célula no son particularmente limitantes de acuerdo con la presente invención. Además, la manera en que se introduce el transgén Cas en la célula puede variar y puede ser mediante cualquier método conocido en el arte. En determinadas formas de realización, la célula transgénica Cas se obtiene mediante introducción del transgén Cas en una célula aislada. En algunas otras formas de realización, la célula transgénica Cas se obtiene mediante aislamiento de células a partir de un organismo transgénico Cas. A modo de ejemplo, y sin limitaciones, la célula transgénica Cas a la cual se hace referencia en la presente puede derivar de un eucariota transgénico Cas, tal como un eucariota Cas *knock-in*. Se hace referencia a WO 2014/093622 (PCT/US13/74667). Los métodos de las publicaciones de Patentes de los EE.UU. n.º: 20120017290 y 20110265198 asignadas a Sangamo BioSciences, Inc. están dirigidos al direccionamiento al locus Rosa, se pueden modificar para utilizar el sistema CRISPR Cas de la presente invención. Los métodos de la publicación de Patente de los EE.UU. n.º: 20130236946 asignado a Cellectis están dirigidos al direccionamiento al locus Rosa, también se pueden modificar para utilizar el sistema CRISPR Cas de la presente invención. A modo de un ejemplo adicional, se hace referencia a Platt y col., (Cell; 159(2): 440-455 (2014)), que describen un *knock-in* de Cas9 de ratón, y que se puede extrapolar a las enzimas CRISPR de la presente invención como se define en la presente. El transgén Cas puede comprender además un casete Lox-Stop-poliA-Lox(LSL) que vuelve inducible a la expresión de Cas mediante la Cre recombinasa. Como alternativa, la célula transgénica Cas se puede obtener mediante introducción del transgén Cas en una célula aislada. Los sistemas de suministro de transgenes son bien conocidos en el arte. A modo de ejemplo, el transgen Cas se puede suministrar, por ejemplo, en una célula eucariota mediante suministro de un vector (por ejemplo, AAV, adenovirus, lentivirus) y/o

una partícula y/o una nanopartícula, como también se describe en otra parte en la presente.

El especialista comprenderá que la célula, tal como la célula transgénica Cas, a la cual se hace referencia en la presente, puede comprender además alteraciones genómicas además de haber integrado un gen Cas o las mutaciones debidas a la acción de Cas específica de la secuencia cuando forma complejos con un ARN capaz de guiar a Cas hacia un locus diana tal como, por ejemplo, una o más mutaciones oncogénicas como se describe, por ejemplo y en un sentido no limitante, en Platt y col. (2014), Chen y col. (2014) o Kumar y col. (2009).

La invención también reivindica una composición que comprende las proteína CRISPR manipulada descrita en la presente, tal como se describe en esta sección.

La invención también describe, pero no proporciona específicamente, una composición no natural modificada que comprende un complejo CRISPR-Cas que comprende cualquiera de las enzimas CRISPR no naturales descritas anteriormente.

La solicitud describe un sistema de vectores que comprende uno o más vectores, en donde dichos uno o más vectores comprende:

a) un primer elemento regulador ligado operativamente a una secuencia de nucleótidos que codifica la proteína CRISPR modificada definida en la presente; y opcionalmente

b) un segundo elemento regulador ligado operativamente a una o más secuencias de nucleótidos que codifican una o más moléculas de ácidos nucleicos que comprenden un ARN guía que comprende una secuencia guía, una secuencia de repetición directa, en donde opcionalmente los componentes (a) y (b) están ubicados en el mismo vector o en vectores diferentes.

La solicitud también describe una composición modificada no natural que comprende:

un sistema de suministro configurado operativamente para suministrar componentes del complejo CRISPR-Cas o una o más secuencias de polinucleótidos que comprenden o codifican dichos componentes en una célula y en donde dicho complejo CRISPR-Cas es funcional en la célula,

componentes del complejo CRISPR-Cas o una o más secuencias de polinucleótidos que codifican la transcripción y/o la traducción en la célula de los componentes del complejo CRISPR-Cas, que comprende:

(I) la enzima CRISPR no natural (por ejemplo, una Cpf1 modificada) descrita en la presente;

(II) un ARN guía CRISPR-Cas que comprende:

la secuencia guía, y

una secuencia de repetición directa,

en donde la enzima en el complejo CRISPR tiene una capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada y/o con lo cual la enzima en el complejo CRISPR tiene una mayor capacidad para modificar dichos uno o más loci diana en comparación con una enzima no modificada.

En un aspecto, los métodos y composiciones reivindicados de la invención también proveen el uso de un sistema que comprende la proteína CRISPR modificada descrita en la presente, tal como se describe en esta sección.

En cualquiera de dichas composiciones, el sistema de suministro puede comprender un sistema de levadura, un sistema de lipofección, un sistema de microinyección, un sistema de biolístico, virosomas, liposomas, inmunoliposomas, policationes, conjugados de lípidos:ácidos nucleicos o viriones artificiales, definidos en otra parte en la presente.

En cualquiera de dichas composiciones, el sistema de suministro puede comprender un sistema de vectores que comprende uno o más vectores, y en donde el componente (II) comprende un primer elemento regulador ligado operativamente a una secuencia de polinucleótidos que comprende la secuencia guía, la secuencia de repetición directa y opcionalmente, y en donde el componente (I) comprende un segundo elemento regulador ligado operativamente a una secuencia de polinucleótidos que codifica la enzima CRISPR.

En cualquiera de dichas composiciones, el sistema de suministro puede comprender un sistema de vectores que comprende uno o más vectores, y en donde el componente (II) comprende un primer elemento regulador ligado operativamente a la secuencia guía y la secuencia de repetición directa, y en donde el componente (I) comprende un segundo elemento regulador ligado operativamente a una secuencia de polinucleótidos que codifica la enzima CRISPR.

En cualquiera de dichas composiciones, la composición puede comprender más de un ARN guía y cada ARN guía tiene un diana diferente, con lo cual hay multiplexación.

En cualquiera de dichas composiciones, dichas una o más secuencias de polinucleótidos pueden estar en un vector.

Los métodos y composiciones reivindicados de la invención también engloban el uso de sistema de vectores (CRISPR-Cas) de (Cas) asociada a CRISP de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR) modificados, no naturales, que comprenden uno o más vectores que comprenden:

a) un primer elemento regulador ligado operativamente a una secuencia de nucleótidos que codifica una enzima CRISPR no natural de una cualquiera de las construcciones de la invención en la presente; y

b) un segundo elemento regulador ligado operativamente a una o más secuencias de nucleótidos que codifican uno o más de los ARN guía, el ARN guía que comprende una secuencia guía, una secuencia de repetición directa,

en donde:

los componentes (a) y (b) están ubicados en vectores iguales o diferentes,

se forma el complejo CRISPR;

el ARN guía busca al locus de polinucleótidos diana y la enzima altera los loci de polinucleótidos y

la enzima en el complejo CRISPR tiene una capacidad reducida para modificar uno o más loci fuera del diana en comparación con una enzima no modificada y/o con lo cual la enzima en el complejo CRISPR tiene una mayor capacidad para modificar dichos uno o más loci diana en comparación con una enzima no modificada.

En dicho sistema, el componente (II) puede comprender un primer elemento regulador ligado operativamente a una secuencia de polinucleótidos que comprende la secuencia guía, la secuencia de repetición directa, y en donde el componente (II) puede comprender un segundo elemento regulador ligado operativamente a una secuencia de polinucleótidos que codifica la enzima CRISPR. En un sistema tal, cuando fuera aplicable el ARN guía puede comprender un ARN quimérico.

En un sistema tal, el componente (I) puede comprender un primer elemento regulador ligado operativamente a la secuencia guía y la secuencia de repetición directa, y en donde el componente (II) puede comprender un segundo elemento regulador ligado operativamente a una secuencia de polinucleótidos que codifica la enzima CRISPR. Dicho sistema puede comprender más de un ARN guía y cada ARN guía tiene un diana diferente, con lo cual hay multiplexación. Los componentes (a) y (b) pueden estar en el mismo vector.

En cualquiera de dichos sistemas que comprenden vectores, dichos uno o más vectores pueden comprender uno o más vectores virales, tales como uno o más retrovirus, lentivirus, adenovirus, virus adenoasociados o virus del herpes simple.

En cualquiera de dichos sistemas que comprenden elementos reguladores, por lo menos uno de dichos elementos reguladores puede comprender un promotor específico de tejidos. El promotor específico de tejidos puede dirigir la expresión en una célula sanguínea de mamífero, en una célula de hígado de mamífero o en un ojo de mamífero.

En cualquiera de las composiciones o los sistemas descritos previamente, la secuencia de repetición directa puede comprender uno o más aptámeros de ARN que interactúa con proteínas. Dichos uno o más aptámeros se pueden localizar en el tetrabucle. Dichos uno o más aptámeros pueden tener la capacidad para unirse a la proteína de envoltura del bacteriófago MS2.

En cualquiera de las composiciones o los sistemas descritos con anterioridad, la célula puede ser una célula eucariota o una célula procariota; en donde el complejo CRISPR es funcional en la célula y con lo cual la enzima del complejo CRISPR tiene una capacidad reducida para modificar uno o más loci fuera del diana de la célula en comparación con una enzima no modificada y/o con lo cual la enzima en el complejo CRISPR tiene una capacidad aumentada para modificar dichos uno o más loci diana en comparación con una enzima no modificada.

Los métodos y composiciones reivindicados de la invención también proporcionan el uso de un complejo CRISPR de una cualquiera de las composiciones descritas con anterioridad o de cualquiera de los sistemas descritos previamente.

La solicitud también describe un método para modificar un locus de interés en una célula que comprende poner la célula en contacto con cualquiera de las enzimas CRISPR modificadas que se describen en la presente (por ejemplo, una Cpfl modificada), composiciones o cualquier de los sistemas o sistemas de vectores que se describen en la presente, o en donde la célula comprende cualquiera de los complejos CRISPR que se describen en la presente que están presentes en la célula. En dichos métodos la célula puede ser una célula procariota o eucariota, preferiblemente una célula eucariota. no es un ser humano.

Cualquiera de dichos métodos es *ex vivo* o *in vitro*.

En determinadas formas de realización, una secuencia de nucleótidos que codifica por lo menos uno entre dicho ARN guía o proteína Cas está unido operativamente en la célula a un elemento regulador que comprende un promotor de un gen de interés, con lo cual la expresión de por lo menos un componente del sistema CRISPR-Cas será dirigida por

el promotor del gen de interés. "Unido operativamente" significa que la secuencia de nucleótidos que codifica el ARN guía y/o la Cas está ligada a uno o más elementos reguladores de una manera que permitirá la expresión de la secuencia de nucleótidos, como también se menciona en otra parte en la presente. El término "elemento regulador" también se describe en otra parte en la presente. El elemento regulador comprende un promotor de un gen de interés tal como, preferiblemente, un promotor de un gen endógeno de interés. En determinadas formas de realización, el promotor se encuentra en su ubicación genómica endógena. En dichas formas de realización, el ácido nucleico que codifica CRISPR y/o Cas se encuentra bajo el control de transcripción del promotor del gen de interés en su ubicación genómica nativa. En algunas otras formas de realización, el promotor se provee en una molécula de ácido nucleico (separada), tal como un vector o un plásmido, u otro ácido nucleico extracromosómico, es decir, el promotor no se provee en su ubicación genómica nativa. En determinadas formas de realización, el promotor está integrado genómicamente en una ubicación genómica no nativa.

En cualquiera de dichos métodos, dicha modificación puede comprender modular la expresión genética. Dicha modulación de la expresión genética puede comprender activar la expresión genética y/o reprimir la expresión genética. Por lo tanto, en un aspecto, los métodos y composiciones de la invención pueden comprender un método para modular la expresión genética, en donde el método comprende introducir la proteína CRISPR o el sistema modificados como se describe en la presente en una célula.

Los métodos y composiciones reivindicados por la invención también proveen el uso de cualquiera de las enzimas CRISPR modificadas (por ejemplo, una Cpf1 modificada), composiciones, sistemas o complejo CRISPR descritos en anteriormente para la edición génica o del genoma *in vitro* o *ex vivo* en una muestra.

Los métodos y composiciones reivindicados de la invención también comprenden un método *in vitro* o *ex vivo* para alterar la expresión de un locus genómico de interés en una célula de mamífero en una muestra que comprende poner la célula en contacto con las enzimas CRISPR modificadas (por ejemplo, una Cpf1 modificada), composiciones, sistemas o complejos CRISPR descritos en la presente y de esa manera suministrar el CRISPR-Cas (vector) y permitir la formación del complejo CRISPR-Cas y su unión al diana, y determinar si la expresión del locus genómico ha sido alterado, tal como si la expresión ha aumentado o disminuido, o si hubo una modificación del producto genético.

En determinadas formas de realización, la actividad de las enzimas CRISPR modificadas (por ejemplo, una Cpf1 modificada) descritas en la presente comprende el clivaje de ADN genómico, lo que opcionalmente da como resultado una transcripción disminuida de un gen.

También se describe pero no se reivindica específicamente en la presente una expresión alterada de un locus genómico con el método descrito en la presente, en donde la expresión alterada se compara con una célula que no fue sometida al método para alterar la expresión del locus genómico. En un aspecto relacionado, la invención provee en una línea celular establecida a partir de una célula tal.

También se describe pero no se reivindica en la presente un método para modificar un organismo o un organismo no humano mediante manipulación de una secuencia diana en un locus genómico de interés, tal como de una HSC (célula madre hematopoyética), por ejemplo, en donde el locus genómico de interés está asociado con una mutación asociada con la expresión de una proteína aberrante o con una condición o estado de enfermedad que comprende:

suministrar a una HSC, por ejemplo, mediante el contacto de una HSC con una partícula que contiene una composición no natural o modificada que comprende:

I. una secuencia de polinucleótidos de ARN guía (ARNg) del sistema CRISPR-Cas, que comprende:

(a) una secuencia guía capaz de hibridarse con una secuencia diana en una HSC,

(b) una secuencia de repetición directa y

II. una enzima CRISPR que opcionalmente comprende por lo menos una o más secuencias de localización nuclear,

en donde, la secuencia guía dirige una unión específica de la secuencia de un complejo CRISPR a la secuencia diana, y

en donde el complejo CRISPR comprende la enzima CRISPR complejada con (1) la secuencia guía que se hibrida con la secuencia diana; y

opcionalmente, el método también puede incluir suministrar un molde de HDR, por ejemplo, mediante el contacto de la partícula con la HSC o el contacto de la HSC con otra partícula que contiene el molde de HDR, en donde dicho molde de HDR provee expresión de una forma normal o menos aberrante de la proteína; en donde "normal" se refiere al tipo salvaje y "aberrante" puede ser una expresión de una proteína que origina una condición o estado de enfermedad; y

el método puede incluir opcionalmente aislar u obtener HSC del organismo o de un organismo no humano, opcionalmente expandir la población de HSC, llevar a cabo el contacto de una o más partículas con la HSC para

obtener una población de HSC modificadas, opcionalmente expandir la población de HSC modificadas y opcionalmente administrar las HSC modificadas al organismo o a un organismo no humano.

También se describe pero no se reivindica en la presente un método para modificar un organismo o un organismo no humano mediante manipulación de una secuencia diana en un locus genómico de interés, tal como de una HSC, por ejemplo, en donde el locus genómico de interés está asociado con una mutación asociada con la expresión de una proteína aberrante o con una condición o estado de enfermedad que comprende: suministrar a una HSC, por ejemplo, mediante el contacto de una HSC con una partícula que contiene una composición no natural o modificada que comprende: I. (a) una secuencia guía capaz de hibridarse con una secuencia diana en una HSC, y (b) por lo menos una o más secuencias de repetición directa, y II. una enzima CRISPR que opcionalmente contiene una o más NLS, y la secuencia guía dirige la unión específica de la secuencia de un complejo CRISPR a la secuencia diana, y en donde el complejo CRISPR comprende la enzima CRISPR complejada con la secuencia guía que se hibrida con la secuencia diana; y opcionalmente,

el método también puede incluir suministrar un molde de HDR, por ejemplo, mediante el contacto de la partícula con la HSC o el contacto de la HSC con otra partícula que contiene el molde de HDR, en donde dicho molde de HDR provee expresión de una forma normal o menos aberrante de la proteína; en donde "normal" se refiere al tipo salvaje y "aberrante" puede ser una expresión de una proteína que origina una condición o estado de enfermedad; y

el método puede incluir opcionalmente aislar u obtener HSC del organismo o de un organismo no humano, opcionalmente expandir la población de HSC, llevar a cabo el contacto de una o más partículas con la HSC para obtener una población de HSC modificadas, opcionalmente expandir la población de HSC modificadas y opcionalmente administrar las HSC modificadas al organismo o a un organismo no humano.

El suministro puede comprender uno o más polinucleótidos que codifican cualquiera entre uno o más o todos los complejos CRISPR, ventajosamente unidos a uno o más elementos reguladores para una expresión in vivo, por ejemplo, mediante partículas que contienen un vector que contiene los polinucleótidos ligados operativamente a dichos elementos reguladores. Cualquiera o todas las secuencias de polinucleótidos que codifican una enzima CRISPR, una secuencia guía, una secuencia de repetición directa, pueden ser ARN. Se podrá apreciar que cuando se hace referencia a un polinucleótido, que es ARN y del que se dice que 'comprende' una característica, tal como una secuencia de repetición directa, la secuencia de ARN incluye dicha característica. Cuando el polinucleótido es ADN y se dice que comprende una característica, tal como una secuencia de repetición directa, la secuencia de ADN se transcribe o se puede transcribir en el ARN que incluye la característica en cuestión. Cuando la característica es una proteína, tal como la enzima CRISPR, la secuencia de ADN o ARN a la que se hace referencia se traduce o se puede traducir (y en el caso del ADN, primero se transcribe).

También se describe pero no se reivindica en la presente un método para modificar un organismo, por ejemplo, un mamífero incluyendo un mamífero humano o no humano mediante manipulación de una secuencia diana en un locus genómico de interés de una HSC, por ejemplo, en donde el locus genómico de interés está asociado con una mutación asociada a la expresión de una proteína aberrante o con una condición o estado de enfermedad, que comprende el suministro, por ejemplo mediante el contacto de una composición no natural o modificada con la HSC, en donde la composición comprende una o más partículas que comprenden vectores virales, plasmídicos o de moléculas de ácido nucleico (por ejemplo, ARN) que codifican operativamente una composición para la expresión de la misma, en donde dicha composición comprende: (A) I. un primer elemento regulador ligado operativamente a secuencia de polinucleótidos de ARN de un sistema CRISPR-Cas, en donde la secuencia de polinucleótidos comprende (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota, (b) una secuencia de repetición directa y II. un segundo elemento regulador ligado operativamente a una secuencia codificante de una enzima que codifica una enzima CRISPR que comprende por lo menos una o más secuencias de localización nuclear (u opcionalmente por lo menos una o más secuencias de localización nuclear ya que algunas formas de realización no comprenden una NLS), en donde (a), (b) y (c) se disponen en una orientación 5' a 3', en donde los componentes I y II están ubicados en el mismo vector o en vectores diferentes del sistema, en donde cuando se transcribe y la secuencia guía dirige la unión específica de secuencias de un complejo CRISPR a la secuencia diana, y en donde el complejo CRISPR comprende la enzima CRISPR complejada con la secuencia guía que se hibrida con la secuencia diana, o (B) una composición no natural o manipulada que comprende un sistema de vectores que comprende uno o más vectores que comprenden I. un primer elemento regulador ligado operativamente a (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota, y (b) por lo menos una o más secuencias de repetición directa, II. un segundo elemento regulador ligado operativamente a una secuencia codificante de una enzima que codifica una enzima CRISPR y, opcionalmente, cuando fuera aplicable, en donde los componentes I y II están ubicados en los mismos vectores o en vectores diferentes del sistema, en donde cuando se transcribe, la secuencia guía dirige una unión específica de secuencias de un complejo CRISPR a la secuencia diana, y en donde el complejo CRISPR comprende la enzima CRISPR complejada con la secuencia guía que se hibrida con la secuencia diana; el método también puede incluir opcionalmente suministrar un molde de HDR, por ejemplo, mediante el contacto de partículas con la HSC que lo contiene o mediante el contacto de la HSC con otra partícula que contiene al molde de HDR, en donde el molde de HDR provee la expresión de una forma normal o menos aberrante de la proteína; en donde "normal" se refiere al tipo salvaje, y "aberrante" puede ser la expresión de una proteína que origina una condición o estados de enfermedad; y opcionalmente el método pueden incluir aislar u obtener HSC del organismo o de un organismo no

humano, opcionalmente expandir la población de HSC, efectuar el contacto de dichas una o más partículas con las HSC para obtener una población de HSC modificada, opcionalmente expandir la población de HSC modificadas, y opcionalmente administrar las HSC modificadas al organismo o a un organismo no humano. En algunas formas de realización, los componentes I, II y III se encuentran en el mismo vector. En otras formas de realización, los componentes I y II están en el mismo vector, en tanto el componente III se encuentra en otro vector. En otras formas de realización, los componentes I y III se encuentran en el mismo vector, en tanto el componente II está en otro vector. En otras formas de realización, los componentes II y III se encuentran en el mismo vector, en tanto el componente I está en otro vector. En otras formas de realización, cada uno de los componentes I, II y III se encuentra en vectores diferentes. También se describe en la presente un sistema de vectores virales o plasmídicos que se describe en la presente.

Los solicitantes también se refieren a la manipulación epigenética de una secuencia diana cuando mencionan la expresión manipulación de una secuencia diana. Esto puede ser para el estado de la cromatina de una secuencia diana, tal como mediante una modificación del estado de metilación de la secuencia diana (es decir, adición o eliminación de la metilación o de patrones de metilación o islas de CpG), modificación de histonas, un aumento o disminución de la accesibilidad a la secuencia diana o favoreciendo el plegamiento tridimensional. Se podrá apreciar que cuando se hace referencia a un método para modificar un organismo o mamífero, incluyendo un ser humano o un mamífero u organismo no humano, mediante manipulación de una secuencia diana en un locus genómico de interés, esto se puede aplicar al organismo (o mamífero) como un todo o solo a una sola célula o población de células de ese organismo (si el organismo es multicelular). En el caso de seres humanos, por ejemplo, los Solicitantes prevén, entre otros, que se pueda modificar una sola célula o una población de células y que preferiblemente se pueden modificar *ex vivo* y luego se pueden volver a introducir. En este caso, puede ser necesaria una biopsia u otra muestra de un fluido o tejido biológico. En este sentido, también se prefieren particularmente las células madre. Y la invención es especialmente ventajosa para las HSC.

También se describe pero no se reivindica en la presente un método para modificar un organismo o un organismo no humano mediante manipulación de una primera y una segunda secuencia diana en hebras opuestas de un dúplex de ADN en un locus genómico de interés en una HSC, por ejemplo, donde el locus genómico de interés está asociado con una mutación asociada con la expresión de una proteína aberrante o con una condición o estado de enfermedad, que comprende administrar, por ejemplo, mediante el contacto de HSC con una o más partículas que comprenden una composición no natural o modificada que comprende:

I. una primera secuencia de polinucleótidos de ARN del sistema CRISPR-Cas (por ejemplo, Cpfl), en donde dicha primera secuencia de polinucleótidos comprende:

(a) una primera secuencia guía capaz de hibridarse con la primera secuencia diana,

(b) una primera secuencia de repetición directa, y

II. una segunda secuencia de polinucleótidos de ARN guía del sistema CRISPR-Cas (por ejemplo, Cpf1), en donde dicha segunda secuencia de polinucleótidos comprende:

(a)una segunda secuencia guía capaz de hibridarse con la segunda secuencia diana,

(b)una segunda secuencia de repetición directa, y

III. una secuencia de polinucleótidos que codifica una enzima CRISPR que comprende por lo menos una o más secuencias de localización nuclear y que comprende una o más mutaciones, en donde (a), (b) y (c) están dispuestas en una orientación 5' a 3'; o

IV. productos de expresión de uno o más de I. a III., por ejemplo, la primera y segunda secuencia de repetición directa, la enzima CRISPR;

en donde cuando se transcribe, la primera y la segunda secuencia guía dirige la unión específica de secuencias de un primer y un segundo complejo CRISPR a la primera y segunda secuencias diana respectivamente, en donde el primer complejo CRISPR comprende la enzima CRISPR complejada con (1) la primera secuencia guía que se hibrida con la primera secuencia diana, en donde el segundo complejo CRISPR comprende la enzima CRISPR complejada con (1) la segunda secuencia guía que se hibrida con la segunda secuencia diana, en donde la secuencia de polinucleótidos que codifica una enzima CRISPR es de ADN o de ARN, y en donde la primera secuencia guía dirige el clivaje de una hebra del dúplex de ADN cerca de la primera secuencia diana y la segunda secuencia guía dirige el clivaje de la otra hebra cerca de la segunda secuencia diana induciendo así una rotura de hebra doble, de esa manera se modifica el organismo o un organismo no humano; y el método también puede incluir opcionalmente suministrar un molde de HDR, por ejemplo, por medio del contacto de partículas con las HSC que los contienen o del contacto de las HSC con otras partículas que contienen al molde de HDR en donde el molde de HDR provee la expresión de una forma normal o menos aberrante de la proteína; en donde "normal" es el de tipo salvaje, y "aberrante" puede ser una expresión de la proteína que origina una condición o estado de enfermedad; y opcionalmente el método puede incluir aislar u obtener HSC del organismo o de un organismo no humano, opcionalmente expandir la población de HSC, realizar el contacto de una o más partículas con las HSC para obtener una población de HSC modificadas, opcionalmente expandir la

población de HSC modificadas, y opcionalmente administrar las HSC modificadas al organismo o un organismo no humano. En algunos métodos descritos en la presente, cualquiera o todas las secuencias de polinucleótidos que codifican la enzima CRISPR, la primera y la segunda secuencia guía, la primera y la segunda secuencia de repetición directa. En formas de realización adicionales descritas en la presente, los polinucleótidos que codifican la secuencia que codifica la enzima CRISPR, la primera y la segunda secuencia guía, la primera y la segunda secuencia de repetición directa, es/son ARN y se suministran por medio de liposomas, nanopartículas, exosomas, microvesículas o una pistola de genes; pero, es ventajoso que el suministro sea por medio de una partícula. En determinadas formas de realización del método descrito en la presente, la primera y la segunda secuencia de repetición directa comparten un 100% de identidad. En algunas formas de realización, los polinucleótidos pueden estar comprendidos dentro de un sistema de vectores que comprende uno o más vectores. En formas de realización preferidas, la primera enzima CRISPR tiene una o más mutaciones de modo que la enzima es una enzima de corte monocatenario de hebra complementaria y una segunda enzima CRISPR tiene una o más mutaciones de modo que la enzima es una enzima de corte monocatenario de hebra no complementaria. Como alternativa, la primera enzima puede ser una enzima de corte monocatenario de hebra no complementaria y la segunda enzima puede ser una enzima de corte monocatenario de hebra complementaria. En métodos preferidos descritos en la presente, la primera secuencia guía dirige el clivaje de una hebra del dúplex de ADN cerca de la primera secuencia diana y la segunda secuencia guía dirige el clivaje de la otra hebra cerca de la segunda secuencia diana lo que da como resultado una sobreextensión 5'. En formas de realización descritas en la presente, la sobreextensión 5' tiene como mucho 200 pares de bases, preferiblemente, como mucho 100 pares de bases o más preferiblemente como mucho 50 pares de bases. En formas de realización descritas en la presente, la sobreextensión 5' tiene como mínimo 26 pares de bases, preferiblemente, como mínimo 30 pares de bases o más preferiblemente como mínimo 34-50 pares de bases.

También se describe, pero no se reivindica en la presente, un método para modificar un organismo o un organismo no humano mediante manipulación de una primera y una segunda secuencia diana en hebras opuestas de un dúplex de ADN en un locus genómico de interés tal como en una HSC, por ejemplo, donde el locus genómico de interés está asociado con una mutación asociada con la expresión de una proteína aberrante o con una condición o estado de enfermedad, que comprende administrar, por ejemplo, mediante el contacto de HSC con una o más partículas que comprenden una composición no natural o modificada que comprende:

I. un primer elemento regulador ligado operativamente a

(a) una primera secuencia guía capaz de hibridarse con la primera secuencia diana, y

(b) por lo menos una o más secuencias de repetición directa,

II. un segundo elemento regulador ligado operativamente a

(a) una segunda secuencia guía capaz de hibridarse con la segunda secuencia diana, y

(b) por lo menos una o más secuencias de repetición directa,

III. un tercer elemento regulador ligado operativamente a una secuencia codificante de una enzima que codifica una enzima CRISPR (por ejemplo, Cpf1), y

V. productos de expresión de uno o más de I. a IV., por ejemplo, la primera y segunda secuencia de repetición directa, la enzima CRISPR;

en donde los componentes I, II, III y IV están ubicados en los mismos vectores o en vectores diferentes del sistema, cuando se transcriben, y la primera y la segunda secuencia guía dirigen una unión específica de secuencias de un primer y un segundo complejo CRISPR a la primera y la segunda secuencias diana respectivamente, en donde el primer complejo CRISPR comprende la enzima CRISPR complejada con (1) la primera secuencia guía que se hibrida con la primera secuencia diana, en donde el segundo complejo CRISPR comprende la enzima CRISPR complejada con la segunda secuencia guía que se hibrida con la segunda secuencia diana, en donde la secuencia de polinucleótidos que codifica una enzima CRISPR es de ADN o de ARN, y en donde la primera secuencia guía dirige el clivaje de una hebra del dúplex de ADN cerca de la primera secuencia diana y la segunda secuencia guía dirige el clivaje de la otra hebra cerca la segunda secuencia diana induciendo así una rotura de hebra doble, de esa manera se modifica el organismo o un organismo no humano; y el método también puede incluir opcionalmente suministrar un molde de HDR, por ejemplo, por medio del contacto de una partícula con las HSC que lo contiene o el contacto de HSC con otra partícula que contiene el molde de HDR, en donde el molde de HDR provee la expresión de una forma normal o menos aberrante de la proteína; en donde "normal" es el tipo salvaje, y "aberrante" puede ser la expresión de una proteína que origina una condición o estado de enfermedad; y opcionalmente el método pueden incluir aislar u obtener HSC del organismo o de un organismo no humano, opcionalmente expandir la población de HSC, realizar el contacto de una o más partículas con las HSC para obtener una población de HSC modificadas, opcionalmente expandir la población de HSC modificadas, y opcionalmente administrar las HSC modificadas al organismo u organismo no humano.

También se describe, pero no se reivindica, un sistema de vectores descrito en la presente. El sistema puede comprender uno, dos, tres o cuatro vectores diferentes. Los componentes I, II, III y IV se pueden ubicar entonces en

uno, dos, tres o cuatro vectores diferentes, y en la presente se prevén todas las combinaciones de ubicaciones posibles de los componentes, por ejemplo: los componentes I, II, III y IV en el mismo vector; los componentes I, II, III y IV pueden ubicarse, cada uno, en vectores diferentes; los componentes I, II, II I y IV se pueden ubicar en un total de dos o tres vectores diferentes, y se prevén todas las combinaciones de ubicaciones, etc. En algunos métodos del sistema de vectores descrito en la presente, cualquiera o todas las secuencias de polinucleótidos que codifican la enzima CRISPR, la primera y la segunda secuencia guía, la primera y la segunda secuencia de repetición directa es/son ARN. En formas de realización adicionales del sistema de vectores descrito en la presente, la primera y la segunda secuencia de repetición directa comparten un 100% de identidad. En formas de realización preferidas, la primera enzima CRISPR tiene una o más mutaciones de modo que la enzima es una enzima de corte monocatenario de hebra complementaria y una segunda enzima CRISPR tiene una o más mutaciones de modo que la enzima es una enzima de corte monocatenario de hebra no complementaria. Como alternativa, la primera enzima puede ser una enzima de corte monocatenario de hebra no complementaria y la segunda enzima puede ser una enzima de corte monocatenario de hebra complementaria. En una forma de realización adicional del sistema de vectores descrito en la presente, uno o más de los vectores virales se suministra mediante liposomas, nanopartículas, exosomas, microvesículas o una pistola de genes, pero es ventajoso el suministro de partículas.

En métodos preferidos del sistema de vectores descrito en la presente, la primera secuencia guía dirige el clivaje de una hebra del dúplex de ADN cerca de la primera secuencia diana y la segunda secuencia guía dirige el clivaje de la otra hebra cerca de la segunda secuencia diana lo que da como resultado una sobreextensión 5'. En formas de realización de la invención, la sobreextensión 5' tiene como mucho 200 pares de bases, preferiblemente, como mucho 100 pares de bases o más preferiblemente como mucho 50 pares de bases. En formas de realización del sistema de vectores, la sobreextensión 5' tiene como mínimo 26 pares de bases, preferiblemente, como mínimo 30 pares de bases o más preferiblemente como mínimo 34-50 pares de bases.

La invención como se reivindica en la presente comprende en lagunas formas de realización un locus genómico de interés, tal como en HSC por ejemplo, donde el locus genómico de interés está asociado con una mutación asociada con una expresión aberrante de la proteína o con una condición o estado de enfermedad, mediante la introducción en HSC, por ejemplo, mediante el contacto de las HSC con una o más partículas que comprenden una proteína Cas que tiene una o más mutaciones y dos ARN guía que dirigen una primera hebra y una segunda hebra de la molécula de ADN, respectivamente, en las HSC, con lo cual el ARN guía dirige la molécula de ADN y la proteína Cas corta monocatenariamente cada una de las primeras hebras y las segundas hebras de la molécula de ADN, con lo cual se altera el diana en las HSC; y, en donde la proteína Cas y los dos ARN guía no aparecen juntos naturalmente y el método también puede incluir opcionalmente suministrar un molde de HDR, por ejemplo, por medio del contacto de las partículas que las HSC que las contienen o el contacto de las HSC con otra partícula que contiene el molde de HDR, en donde el molde de HDR provee la expresión de una forma normal o menos aberrante de la proteína; en donde “normal” es el tipo salvaje, y “aberrante” puede ser la expresión de una proteína que origina una condición o estado de enfermedad; y opcionalmente el método pueden incluir aislar u obtener HSC del organismo o de un organismo no humano, opcionalmente expandir la población de HSC, realizar el contacto de una o más partículas con las HSC para obtener una población de HSC modificadas, opcionalmente expandir la población de HSC modificadas, y opcionalmente administrar las HSC modificadas al organismo o al organismo no humano. En métodos preferidos de la invención, la proteína Cas corta cada una entre la primera hebra y la segunda hebra de la molécula de ADN da como resultado una sobreextensión 5'. En formas de realización de la invención, la sobreextensión 5' tiene como mucho 200 pares de bases, preferiblemente, como mucho 100 pares de bases o más preferiblemente como mucho 50 pares de bases. En formas de realización de la invención, la sobreextensión 5' tiene como mínimo 26 pares de bases, preferiblemente, como mínimo 30 pares de bases o más preferiblemente como mínimo 34-50 pares de bases. En un aspecto de la invención, se optimizan los codones de la proteína Cas para la expresión en una célula eucariota, preferiblemente una célula de mamífero o una célula humana. Algunos aspectos de la invención se refieren a la disminución de la expresión de un producto génico o a la introducción adicional de un polinucleótido molde en la molécula de ADN que codifica el producto génico o al recorte de una secuencia interviniente de manera precisa al permitir el alineamiento o ligamiento de las dos sobreextensiones 5' o la alteración de la actividad o función del producto génico o el aumento de la expresión del producto génico. En una forma de realización de la invención, el producto génico es una proteína.

La invención reivindicada en la presente en algunas realizaciones comprende modificar un locus genómico de interés tal como en HSC, por ejemplo, en donde el locus genómico de interés está asociado con una mutación asociada con una expresión de proteína aberrante o con una condición o estado de enfermedad, mediante la introducción en la HSC, por ejemplo, con el contacto de la HSC con partículas que comprenden,

a) un primer elemento regulador ligado operativamente a cada uno de los dos ARN guía del sistema CRISPR-Cas que tienen como diana una primera hebra y una segunda hebra, respectivamente, de una molécula de ADN de hebra doble de la HSC y

b) un segundo elemento regulador ligado operativamente a una proteína Cas (por ejemplo, Cpf1), o

c) uno o más productos de expresión de a) o b),

en donde los componentes (a) y (b) están localizados en el mismo vector o en vectores distintos del sistema, con lo

cual los ARN guía buscan como diana a la molécula de ADN de la HSC y la proteína Cas efectúa cortes monocatenarios de la primera hebra y la segunda hebra de la molécula de ADN de la HSC; y en donde la proteína Cas y los dos ARN guía no aparecen juntos de manera natural; y el método también puede incluir opcionalmente suministrar un molde de HDR, por ejemplo, mediante contacto de la partícula con la HSC que contiene a la misma o contacto de la HSC con otra partícula que contiene el molde de HDR, en donde el molde de HDR provee la expresión de una forma normal o menos aberrante de la proteína; en donde "normal" se refiere al tipo salvaje y "aberrante" puede ser la expresión de una proteína que da lugar a una condición o estado de enfermedad; y, opcionalmente, el método puede incluir aislar u obtener HSC del organismo o de un organismo no humano, opcionalmente expandir la población de HSC, poner una o más partículas en contacto con la HSC para obtener una población de HSC modificada, opcionalmente expandir la población de HSC modificada y opcionalmente administrar las HSC al organismo o a un organismo no humano. En aspectos de la invención, el ARN guía puede comprender una secuencia guía fusionada a una secuencia de repetición directa. Algunos aspectos de la invención se refieren a la disminución de la expresión de un producto génico o a la introducción adicional de un polinucleótido molde en la molécula de ADN que codifica el producto génico o al recorte de una secuencia interviniente de manera precisa al permitir el alineamiento o ligamiento de las dos sobreextensiones 5' o la alteración de la actividad o función del producto génico o el aumento de la expresión del producto génico. En una forma de realización de la invención, el producto génico es una proteína. En las formas de realización preferidas de la invención, los vectores del sistema son vectores virales. En una forma de realización adicional, los vectores del sistema se suministran mediante liposomas, nanopartículas, exosomas, microvesículas o una pistola de genes; y se prefieren las partículas. En un aspecto, los métodos reivindicados de la invención comprenden modificar un polinucleótido diana en una HSC. En algunas formas de realización, el método comprende permitir la unión de un complejo CRISPR al polinucleótido diana para efectuar el clivaje de dicho polinucleótido diana, modificando de esa manera el polinucleótido diana; en donde el complejo CRISPR comprende una enzima CRISPR complejada con una secuencia guía hibrida con una secuencia diana contenida en dicho polinucleótido diana, en donde dicha secuencia guía está ligada a una secuencia de repetición directa. En algunas formas de realización, dicho clivaje comprende clivar una o dos hebras en la ubicación de la secuencia diana por dicha enzima CRISPR. En algunas formas de realización, dicho clivaje da como resultado una disminución de la transcripción de un gen diana. En algunas formas de realización, el método comprende además reparar dicho polinucleótido diana clivado mediante recombinación homóloga con un polinucleótido molde exógeno, donde dicha reparación da como resultado una mutación que comprende una inserción, supresión o sustitución de uno o más nucleótidos de dicho polinucleótido diana. En algunas formas de realización, dicha mutación da como resultado uno o más cambios de aminoácidos en una proteína expresada a partir de un gen que comprende la secuencia diana. En algunas formas de realización, el método comprende además suministrar uno o más vectores o productos de expresión de los mismos, por ejemplo, por medio de una o más partículas, a por ejemplo dichas HSC, en donde dichos uno o más vectores dirigen la expresión de una o más entre: la enzima CRISPR, la secuencia guía ligada a la secuencia de repetición directa. En algunas formas de realización, dichos vectores se suministran, por ejemplo, a las HSC en un sujeto. En algunas formas de realización, dicha modificación tiene lugar en dichas HSC en un cultivo celular. En algunas formas de realización, el método comprende además aislar dichas HSC de un sujeto antes de dicha modificación. En algunas formas de realización, el método comprende además volver a introducir dichas HSC y/o células derivadas de las mismas a dicho sujeto.

También se describe, pero no se reivindica específicamente en la presente, un método para generar, por ejemplo, una HSC que comprende un gen mutado de una enfermedad. En algunas formas de realización, un gen ligado a una enfermedad es cualquier gen asociado con un aumento del riesgo de padecer o desarrollar una enfermedad. En algunas formas de realización, el método comprende (a) introducir uno o más vectores o productos de expresión de los mismos, por ejemplo, por medio de una o más partículas, en una HSC, en donde dichos uno o más vectores dirigen la expresión de una o más entre: una enzima CRISPR, una secuencia guía ligada a una secuencia de repetición directa; y (b) permitir la unión de un complejo CRISPR a un polinucleótido diana para efectuar el clivaje del polinucleótido diana en dicho gen de una enfermedad, en donde el complejo CRISPR comprende la enzima CRISPR complejada con la secuencia guía que se hibrida con la secuencia diana en el polinucleótido diana y, opcionalmente, cuando fuera aplicable, para así generar una HSC que comprende un gen mutado de una enfermedad. En algunas formas de realización, dicho clivaje comprende clivar una o dos hebras en la ubicación de la secuencia diana por dicha enzima CRISPR. En algunas formas de realización, dicho clivaje da como resultado una disminución de la transcripción de un gen diana. En algunas formas de realización, el método comprende además reparar dicho polinucleótido diana clivado mediante recombinación homóloga con un polinucleótido molde exógeno, donde dicha reparación da como resultado una mutación que comprende una inserción, supresión o sustitución de uno o más nucleótidos de dicho polinucleótido diana. En algunas formas de realización, dicha mutación da como resultado uno o más cambios de aminoácidos en la expresión de una proteína resultante de un gen que comprende la secuencia diana. En algunas formas de realización, la HSC modificada se administra a un animal para generar de este modo un modelo animal.

En un aspecto, los métodos reivindicados de la invención proveen modificar un polinucleótido diana, por ejemplo, en una HSC. En algunas formas de realización, el método comprende permitir la unión de un complejo CRISPR al polinucleótido diana para efectuar el clivaje de dicho polinucleótido diana, modificando de esa manera el polinucleótido diana; en donde el complejo CRISPR comprende una enzima CRISPR complejada con una secuencia guía hibridada con una secuencia diana contenida en dicho polinucleótido diana, en donde dicha secuencia guía está ligada a una secuencia de repetición directa. En otras formas de realización, la invención reivindicada implica modificar la expresión de un polinucleótido en una célula eucariota que proviene, por ejemplo, de una HSC. El método comprende aumentar

o reducir la expresión de un polinucleótido diana usando un complejo CRISPR que se une al polinucleótido en la HSC; ventajosamente, el complejo CRISPR se suministra mediante una o más partículas.

En algunos métodos, se puede inactivar un polinucleótido diana para efectuar la modificación de la expresión, por ejemplo, en una HSC. Por ejemplo, tras la unión de un complejo CRISPR a la secuencia diana en una célula, el polinucleótido diana se inactiva de modo que no se transcribirá la secuencia, no se producirá la proteína codificada o la secuencia no funcionará como lo hace la secuencia no modificada.

En algunas formas de realización, se puede modificar el ARN del sistema CRISPR-Cas, por ejemplo, el ARN guía o ARNgs; por ejemplo, para que incluya un aptámero o un dominio funcional. Un aptámero es un oligonucleótido sintético que se une a una molécula diana específica; por ejemplo, una molécula de ácido nucleico que se ha modificado mediante rondas repetidas de selección in vitro o SELEX (evolución sistemática de ligandos mediante enriquecimiento exponencial) para unirse a varios blancos moleculares, tales como moléculas pequeñas, proteínas, ácidos nucleicos e incluso células, tejidos y organismos. Los aptámeros son útiles por cuanto ofrecen propiedades de reconocimiento molecular que rivalizan con las de los anticuerpos. Además de su reconocimiento discriminado, los aptámeros ofrecen ventajas frente a los anticuerpos, que incluyen el hecho que generan poca o ninguna inmunogenicidad en las aplicaciones terapéuticas. Por consiguiente, en la práctica de la invención, cualquiera o ambas enzima y ARN pueden incluir un dominio funcional.

En algunas formas de realización, el dominio funcional es un dominio de activación de la transcripción, preferiblemente VP64. En algunas formas de realización, el dominio funcional es un dominio de represión de la transcripción, preferiblemente KRAB. En algunas formas de realización, el dominio de represión de la transcripción es SID o concatámeros de SID (por ejemplo, SID4X). En algunas formas de realización, el dominio funcional es un dominio de modificación epigenética, de modo que se provee una enzima de modificación epigenética. En algunas formas de realización, el dominio funcional es un dominio de activación, que puede ser el dominio de activación P65. En algunas formas de realización, el dominio funcional comprende actividad nucleasa. En dicha forma de realización, el dominio funcional comprende Fok1.

Los métodos de la invención dan lugar a una célula *in vitro* o *ex vivo* que comprende cualquiera de las enzimas CRISPR, composiciones, sistemas o complejos modificados que se describieron previamente, o de cualquiera de los métodos descritos con anterioridad. La célula puede ser una célula eucariota o una célula procariota. La descripción también prevé la progenie de dichas células. La descripción también prevé un producto de cualquiera de dichas células o de cualquiera de dicha progenie, en donde el producto es un producto de dichos uno o más loci diana modificados por la enzima CRISPR modificada del complejo CRISPR. El producto puede ser un péptido, polipéptido o proteína. Algunos de dichos productos pueden ser modificados por la enzima CRISPR modificada del complejo CRISPR. En algunos de dichos productos modificados, el producto del locus diana es físicamente distinto del producto de dicho locus diana que no fue modificado por dicha enzima CRISPR modificada.

Los métodos y composiciones reivindicados de la invención también proveen el uso de una molécula de polinucleótidos que codifica cualquiera de las enzimas CRISPR no naturales descritas anteriormente.

Cualquiera de dichos polinucleótidos puede comprender además uno o más elementos reguladores que están ligados operativamente a la secuencia de polinucleótidos que codifica la enzima CRISPR no natural.

En cualquiera de dichos polinucleótidos que comprenden uno o más elementos reguladores, dichos uno o más elementos reguladores se pueden configurar operativamente para la expresión de la enzima CRISPR no natural en una célula eucariota. La célula eucariota puede ser una célula humana. La célula eucariota puede ser una célula de roedor, opcionalmente una célula de ratón. La célula eucariota puede ser una célula de levadura. La célula eucariota puede ser una célula de ovario de hámster chino (CHO). La célula eucariota puede ser una célula de insecto.

En cualquiera de dichos polinucleótidos que comprenden uno o más elementos reguladores, dichos uno o más elementos reguladores se pueden configurar operativamente para la expresión de la enzima CRISPR no natural en una célula procariota.

En cualquiera de dichos polinucleótidos que comprenden uno o más elementos reguladores, dichos uno o más elementos reguladores se pueden configurar operativamente para la expresión de la enzima CRISPR no natural en un sistema *in vitro*.

Los métodos y composiciones reivindicados de la invención también proveen el uso de un vector de expresión que comprende cualquiera de las moléculas de polinucleótidos descritas con anterioridad. La invención también provee dichas moléculas de polinucleótidos, por ejemplo, moléculas de polinucleótidos configuradas operativamente para expresar la proteína y/o los componentes de ácidos nucleicos, así como dichos vectores.

Los métodos y composiciones reivindicados de la invención también pueden implicar efectuar mutaciones en una Cas (por ejemplo, Cpf1) o una Cas mutada o modificada (por ejemplo, Cpf1) que es un ortólogo de las enzimas CRISPR de acuerdo con la invención descritas en la presente, que comprende determinar uno o más aminoácidos en dicho ortólogo que pueden estar muy próximos o pueden tocar una molécula de ácido nucleico, por ejemplo, ADN, ARN, ARNg, etc., y/o uno o más aminoácidos análogos o correspondientes a uno o más aminoácidos identificados en la

presente en las enzimas CRISPR de acuerdo con la invención como se describe en la presente para la modificación y/o mutación, y sintetizar o preparar o expresar el ortólogo que comprende, que consiste en o que consiste esencialmente en una o más modificaciones y/o mutaciones o mutar como se describe en la presente, por ejemplo, modificar, por ejemplo, cambiar o mutar, un aminoácido neutro por un aminoácido con carga, por ejemplo, con carga positiva, por ejemplo, de alanina, por ejemplo, por lisina. El ortólogo modificado de este modo se puede usar en los sistemas CRISPR-Cas; y las moléculas de ácido nucleico que lo expresan se pueden usar en vectores o en otros sistemas de suministro que suministren moléculas o que codifiquen los componentes del sistema CRISPR-Cas que se describe en la presente.

En un aspecto, los métodos y composiciones reivindicados de la invención implican proveer una actividad eficaz en el diana y minimiza la actividad fuera del diana. En un aspecto, los métodos y composiciones reivindicados de la invención proveen una clivaje eficaz en el diana por una proteína CRISPR y minimiza el clivaje fuera del diana por la proteína CRISPR. En un aspecto, los métodos y composiciones reivindicados de la invención proveen el uso de una unión específica de la guía de una proteína CRISPR en un locus génico sin clivaje de ADN. En un aspecto, los métodos y composiciones reivindicados de la invención una implican una guía eficaz dirigida por la guía en el diana de una proteína CRISPR en un locus génico y minimiza la unión fuera del diana de la proteína CRISPR. Por consiguiente, los métodos y composiciones reivindicados de la invención pueden implicar una regulación génica específica de diana. En un aspecto, los métodos y composiciones reivindicados de la invención proveen unión específica de la guía de una enzima CRISPR en un locus génico sin clivaje de ADN. Por consiguiente, en un aspecto, la invención provee el clivaje de un locus génico y la regulación génica en un locus génico diferente usando una sola enzima CRISPR. En un aspecto, los métodos y composiciones reivindicados de la invención proveen una activación y/o inhibición y/o clivaje ortogonal de múltiples blancos usando una o más proteínas y/o enzimas CRISPR.

También se describe, pero no se reivindica en la presente, un método de selección funcional de genes en un genoma en un grupo de células ex vivo o in vivo que comprende la administración o expresión de una biblioteca que comprende múltiples ARN guía (ARNsg) del sistema CRISPR-Cas y en donde la selección comprende además el uso de una enzima CRISPR, en donde el complejo CRISPR se modifica para que comprenda un dominio heterólogo funcional. En un aspecto que no es parte de la invención se provee un método para seleccionar un genoma que comprende la administración a un huésped o la expresión en un huésped in vivo de una biblioteca. En un aspecto, la solicitud desvela un método como se describe en la presente que además comprende un activador administrado al huésped o expresado en el huésped. En un aspecto, la solicitud desvela un método descrito en la presente en donde el activador se une a una proteína CRISPR. En un aspecto, la solicitud desvela un método descrito en la presente, en donde el activador se une al extremo N-terminal o al extremo C-terminal de la proteína CRISPR. En un aspecto, la solicitud desvela un método descrito en la presente, en donde el activador está unido a un bucle de ARNsg. En un aspecto, la solicitud desvela un método como se describe en la presente que además comprende un represor administrado al huésped o expresado en el huésped. En un aspecto, la solicitud desvela un método descrito en la presente, en donde la selección comprende afectar y detectar la activación génica, la inhibición génica o el clivaje en el locus.

En un aspecto, la invención provee un método descrito en la presente, en donde el huésped es una célula eucariota. En un aspecto, la invención provee un método descrito en la presente, en donde el huésped es una célula de mamífero. En un aspecto, la invención provee un método descrito en la presente, en donde el huésped es una célula eucariota no humana. En un aspecto, la invención provee un método descrito en la presente, en donde la célula eucariota no humana es una célula de mamífero no humano. En un aspecto, la invención provee un método descrito en la presente, en donde la célula de mamífero no humano puede ser incluyendo, pero en un sentido no taxativo, una célula de primate, bovina, ovina, porcina, canina, de roedor, de Leporidae, tal como de mono, de vaca, de oveja, de cerdo, de perro, de conejo, de rata o de ratón. En un aspecto, la invención provee un método descrito en la presente, en donde la célula puede ser una célula eucariota no de mamífero, tal como una célula de un ave de corral (por ejemplo, de pollo), de peces vertebrados (por ejemplo, salmón) o de mariscos (por ejemplo, ostras, almejas, langostas o gambas). En un aspecto, la invención provee un método descrito en la presente, en donde la célula eucariota no humana es una célula vegetal. La célula vegetal puede ser de una monocotiledónea o dicotiledónea o de una planta de cultivos o de granos, tal como de mandioca, maíz, sorgo, soja, sorgo, avena o arroz. La célula vegetal también puede ser de algas, árboles o plantas productoras, de frutas o verduras (por ejemplo, árboles tales como árboles de cítricos, por ejemplo, árboles de naranja, pomelo o limón; durazno o nectarina; árboles de manzano o peras; árboles de frutos secos, tales como árboles de almendras o nueces o pistacho; plantas de sombra nocturna; plantas del género *Brassica*; plantas del género *Lactuca*; plantas del género *Spinacia*; plantas del género *Capsicum*; algodón, tabaco, espárrago, zanahoria, repollo, brócoli, coliflor, tomate, berenjena, pimiento, lechuga, espinaca, frutilla, arándano azul, frambuesa, zarzamora, uvas, café, cacao, etc).

Un aspecto que no es parte de la invención provee un método descrito en la presente que comprende el suministro de los complejos CRISPR-cas, o componentes del mismo, o moléculas de ácido nucleico que codifican a los mismos, en donde dichas moléculas de ácido nucleico están ligadas operativamente a secuencias reguladores y se expresan in vivo. Un aspecto que no es parte de la invención provee un método descrito en la presente, en donde la expresión in vivo es mediante un lentivirus, un adenovirus o un AAV. En un aspecto, la invención provee un método descrito en la presente, en donde el suministro es mediante una partícula, una nanopartícula, un lípido o un péptido penetrador en células (CPP).

En formas de realización particulares puede ser de interés dirigir el complejo CRISPR-Cas al cloroplasto. En muchos

casos, este direccionamiento se puede lograr mediante la presencia de una extensión N-terminal, denominado péptido de tránsito a cloroplastos (CTP) o péptido de tránsito a plástidos. Los transgenes cromosómicos de fuentes bacterianas deben tener una secuencia que codifica una secuencia CTP fusionada a una secuencia que codifica un polipéptido expresado si dicho polipéptido expresado será compartimentalizado en el plástido vegetal (por ejemplo, un cloroplasto). Por lo tanto, la localización de un polipéptido exógeno en un cloroplasto a menudo se lleva a cabo por medio de la unión operativa de una secuencia de polinucleótidos que codifica una secuencia CTP a la región 5' de un polinucleótido que codifica el polipéptido exógeno. El CTP es eliminado en un paso de proceso durante la traslocación al plástido. Sin embargo, la eficacia de procesamiento puede ser afectada por la secuencia de aminoácidos del CTP y secuencias próximas en el extremo NH2 terminal del péptido. Otras opciones para el direccionamiento al cloroplasto que ya fueron descritas comprenden la secuencia señal cab-m7 de maíz (Patente de los EE.UU. n.°: 7.022.896, WO 97/41228), una secuencia señal de la glutatión reductasa de guisante (WO 97/41228) y el CTP descrito en US2009029861.

En un aspecto, los métodos y composiciones reivindicados de la invención pueden comprender proveer un par de complejos CRISPR-Cas, cada uno de los cuales comprende un ARN guía (ARNsg) que comprende una secuencia guía capaz de hibridarse con una secuencia diana en un locus genómico de interés en una célula, en donde por lo menos un bucle de cada ARNsg está modificado mediante la inserción de distintas secuencias de ARN que se unen a una o más proteínas adaptadoras y en donde la proteína adaptadora se asocia con uno o más dominios funcionales, en donde cada ARNsg de cada CRISPR-Cas comprende un dominio funcional que tiene actividad de clivaje de ADN. En un aspecto, los métodos y composiciones reivindicados de la invención pueden comprender proveer complejos CRISPR-Cas apareados como se describe en la presente, en donde la actividad de clivaje de ADN se debe a una nucleasa Fok1.

En un aspecto, los métodos y composiciones reivindicados que no forman parte de la invención pueden implicar cortar una secuencia diana en un locus genómico de interés que comprende el suministro a una célula de los complejos CRISPR-Cas, o componentes de los mismos, o moléculas de ácido nucleico que codifican los mismos, en donde dichas moléculas de ácido nucleico están ligadas operativamente a secuencias reguladoras y se expresan in vivo. En un aspecto, la invención provee un método descrito en la presente, en donde el suministro es mediante un lentivirus, un adenovirus o un AAV. En un aspecto, la invención provee un método descrito en la presente o complejos CRISPR-Cas apareados descritos en la presente, en donde la secuencia diana de un primer complejo del par se encuentra en una primera hebra de ADN bicatenario y la secuencia diana para el segundo complejo del par se encuentra en una segunda hebra de ADN bicatenario. En un aspecto, la invención provee un método descrito en la presente o complejos CRISPR-Cas apareados descritos en la presente, en donde las secuencias diana del primer y segundo complejo se encuentran próximas entre sí de tal forma que el ADN se corta de un modo tal que facilita una reparación dirigida por homología. En un aspecto, un método de la presente puede incluir además introducir un ADN molde en una célula. En un aspecto, un método de la presente puede abarcar los complejos CRISPR-Cas apareados de la presente, en donde cada complejo CRISPR-Cas tiene una enzima CRISPR que está mutada de tal forma que no tiene más que aproximadamente un 5% de la actividad nucleasa de la enzima CRISPR que no está mutada.

También se describe, pero no se reivindica específicamente en la presente, una biblioteca, un método o un complejo, como se describe en la presente, en donde el ARNsg se modifica para que contenga por lo menos un bucle funcional no codificante, por ejemplo, en donde dicho por lo menos un bucle funcional no codificante es represor; por ejemplo, en donde dicho por lo menos un bucle funcional no codificante comprende Alu.

En un aspecto, los métodos y composiciones reivindicados de la invención implican alterar o modificar la expresión de un producto génico. Dicho método puede comprender introducir en una célula que contiene y expresa una molécula de ADN que codifica un producto génico un sistema CRISPR-Cas modificado no natural que comprende una proteína Cas y un ARN guía que dirige la molécula de ADN, con lo cual el ARN guía dirige a la molécula de ADN que codifica el producto génico y la proteína Cas cliva la molécula de ADN que codifica el producto génico, con lo cual se altera la expresión del producto génico; y donde la proteína Cas y el ARN guía no aparecen juntos de manera natural. La invención comprende además la proteína Cas de codones optimizados para su expresión en una célula eucariota. En una forma de realización preferida, la célula eucariota es una célula de mamífero y en una forma de realización más preferida, la célula de mamífero es una célula humana. En una forma de realización adicional de la invención, la expresión del producto génico está reducida.

Los métodos reivindicados pueden dar lugar a células alteradas y la progenie de dichas células, así como los productos producidos por las células. Los sistemas y las proteínas CRISPR-Cas (por ejemplo, Cpf1) de la invención se usan para producir células que comprenden un locus diana modificado. En algunas formas de realización, el método puede comprender permitir la unión de un complejo de direccionamiento de ácidos nucleicos al ADN o ARN diana para efectuar el clivaje de dicho ADN o ARN diana, modificando de este modo el ADN o ARN diana, en donde el complejo de direccionamiento de ácidos nucleicos comprende una proteína efectora de direccionamiento de ácidos nucleicos complejada con un ARN guía hibridado con una secuencia diana en dicho ADN o ARN diana. En un aspecto, los métodos y composiciones reivindicados de la invención implican reparar un locus genético en una célula. En otro aspecto, los métodos y composiciones reivindicados de la invención implican modificar la expresión de ADN o ARN en una célula eucariota. En algunas formas de realización, el método comprende permitir la unión de un complejo de direccionamiento de ácidos nucleicos al ADN o ARN de manera tal que dicha unión resulte en la expresión aumentada o reducida de dicho ADN o ARN; en donde el complejo de direccionamiento de ácidos nucleicos comprende una

proteína efectora de direccionamiento de ácidos nucleicos complejada con un ARN guía. Se aplican las consideraciones y condiciones anteriores para los métodos de modificación de un ADN o ARN diana. De hecho, estas opciones de obtención, cultivo y reintroducción de muestras son válidas en todos los aspectos de la presente invención. En un aspecto, los métodos reivindicados de la invención implican modificar un ADN o ARN diana en una célula eucariota, *ex vivo* o *in vitro*. En algunas formas de realización, el método comprende obtener muestras de una célula o población de células de un ser humano o un animal no humano y modificar dichas una o más células. El cultivo puede tener lugar en cualquier etapa *ex vivo*. Dichas células pueden ser, en un sentido no taxativo, células vegetales, células animales, tipos celulares particulares de cualquier organismo, incluyendo células madre, células inmunológicas, linfocitos T, linfocitos B, células dendríticas, células cardiovasculares, células epiteliales, células madre y semejantes. Las células se pueden modificar de acuerdo con la invención para producir productos génicos, por ejemplo, en cantidades controladas que se pueden aumentar o disminuir, dependiendo del uso, y/o mutar. En determinadas formas de realización, se repara un locus genético de la célula. Dichas una o más células incluso se pueden reintroducir en el animal no humano o la planta. En el caso de células reintroducidas, se prefiere que las células sean células madre. Los métodos reivindicados de la invención no son métodos de modificación de la identidad genética de la línea germinal de los seres humanos.

En un aspecto, los métodos reivindicados de la invención proveen células que comprenden de manera transitoria sistemas, o componentes, CRISPR. Por ejemplo, las proteínas o enzimas CRISPR y los ácidos nucleicos se proveen transitoriamente en una célula y se altera un locus genético, seguido por una reducción de la cantidad de uno o más componentes del sistema CRISPR. A continuación, las células, la progenie de las células y los organismos que comprenden las células, que han adquirido una alteración genética mediada por CRISPR, comprenden una cantidad reducida de uno o más componentes del sistema CRISPR o ya no contienen dichos uno o más componentes del sistema CRISPR. Un ejemplo no limitante es un sistema CRISPR-Cas autoinactivante, que también se describe en la presente. Por lo tanto, los métodos de la invención pueden dar lugar a células que comprenden uno o más loci genéticos alterados por el sistema CRISPR-Cas, pero que carecen esencialmente de uno o más componentes del sistema CRISPR. En determinadas formas de realización, los componentes del sistema CRISPR están sustancialmente ausentes. Dichas células comprenden ventajosamente una alteración genética deseada o seleccionada pero que han perdido componentes de CRISPR-Cas o remanentes de los mismos que potencialmente pueden actuar de manera no específica, conducir a problemas de seguridad o impedir la aprobación regulatoria.

Sistemas inducibles Cpf1 CRISPR-Cas ("Cpf1 doble")

En un aspecto los métodos reivindicados de la invención proporcionan el uso de un sistema inducible no natural o modificado Cpf1 CRISPR-Cas, que comprende:

una primera construcción de fusión de Cpf1 unida a una primera mitad de un dímero inducible y

una segunda construcción de fusión de Cpf1 unida a una segunda mitad del dímero inducible,

en donde la primera construcción de fusión de Cpf1 está operativamente conectada a una o más señales de localización nuclear,

en donde la segunda construcción de fusión de Cpf1 está operativamente conectada a una o más señales de exportación nuclear,

en donde el contacto con una fuente de energía inductora pone juntas a la primera y segunda mitad del dímero inducible,

en donde poner juntas a la primera y segunda mitad del dímero inducible les permite a la primera y segunda construcción de fusión de Cpf1 constituir un sistema funcional Cpf1 CRISPR-Cas,

en donde el sistema Cpf1 CRISPR-Cas comprende un ARN guía (ARNg) que comprende una secuencia guía capaz de hibridar con una secuencia diana en un locus genómico de interés en una célula, y

en donde el sistema funcional Cpf1 CRISPR-Cas se une a la secuencia diana y, opcionalmente, edita el locus genómico para alterar la expresión génica.

En un aspecto de la invención en el sistema inducible Cpf1 CRISPR-Cas, el dímero inducible es o comprende o consiste esencialmente en o consiste en un heterodímero inducible. En un aspecto, en el sistema inducible Cpf1 CRISPR-Cas, la primera mitad o una primera porción o un primer fragmento del heterodímero inducible es o comprende o consiste en o consiste esencialmente en una FKBP, opcionalmente FKBP12. En un aspecto de la invención, en el sistema inducible Cpf1 CRISPR-Cas, la segunda mitad o una segunda porción o un segundo fragmento del heterodímero inducible es o comprende o consiste en o consiste esencialmente en FRB. En un aspecto de la invención, en el sistema inducible Cpf1 CRISPR-Cas, el arreglo de la primera construcción de fusión de Cpf1 es o comprende o consiste en o consiste esencialmente en La parte N' terminal de Cpf1-FRB-NES. En un aspecto de la invención, en el sistema inducible Cpf1 CRISPR-Cas, el arreglo de la primera construcción de fusión de Cpf1 es o comprende o consiste en o consiste esencialmente en NES-la parte N' terminal de Cpf1-FRB-NES. En un aspecto de la invención, en el sistema inducible Cpf1 CRISPR-Cas, el arreglo de la segunda construcción de fusión de Cpf1 es o comprende o consiste esencialmente en o consiste en la parte C' terminal de Cpf1-FKBP-NLS. En un aspecto la

invención provee el sistema inducible Cpf1 CRISPR-Cas, en donde el arreglo de la segunda construcción de fusión de Cpf1 es o comprende o consiste en o consiste esencialmente en NLS-la parte C' terminal de Cpf1-FKBP-NLS. En un aspecto, en el sistema inducible Cpf1 CRISPR-Cas puede haber un conector que separa la parte de Cpf1 de la mitad o porción o fragmento del dímero inducible. En un aspecto, en el sistema inducible Cpf1 CRISPR-Cas, la fuente de energía inductora es o comprende o consiste esencialmente en o consiste en rapamicina. En un aspecto, en el sistema inducible Cpf1 CRISPR-Cas, el dímero inducible es un homodímero inducible. En un aspecto, en el sistema inducible Cpf1 CRISPR-Cas, el Cpf1 es FnCpf1. En un aspecto, en el sistema inducible Cpf1 CRISPR-Cas, uno o más dominios funcionales se asocian con una o ambas partes de Cpf1, por ejemplo, los dominios funcionales opcionalmente incluyendo un activador transcripcional, un transcripcional o una nucleasa tal como una nucleasa Fok1. En un aspecto, en el sistema inducible Cpf1 CRISPR-Cas, el sistema funcional Cpf1 CRISPR-Cas se une a la secuencia diana y la enzima es una dead-Cpf1, opcionalmente teniendo una menor actividad nucleasa de por lo menos 97%, o 100% (o no más de 3% y ventajosamente 0% de actividad nucleasa) en comparación con el Cpf1 que no tiene la por lo menos una mutación. La invención además comprende y un aspecto de la invención provee, un polinucleótido que codifica para el sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación.

En un aspecto, los métodos y composiciones reivindicados de la invención proveen el uso de un vector para el suministro de la primera construcción de fusión de Cpf1, unida a una primera mitad o porción o fragmento de un dímero inducible y operativamente conectada a una o más señales de localización nuclear, como se divulga en la presente documentación. En un aspecto, la invención provee un vector para el suministro de la segunda construcción de fusión de Cpf1, unida a una segunda mitad o porción o fragmento de un dímero inducible y operativamente conectada a una o más señales de exportación nuclear.

En un aspecto, los métodos y composiciones reivindicados de la invención proveen el uso de un vector para el suministro tanto de la primera construcción de fusión de Cpf1, unida a una primera mitad o porción o fragmento de un dímero inducible y operativamente conectada a una o más señales de localización nuclear, como se divulga en la presente documentación; como de la segunda construcción de fusión de Cpf1, unida a una segunda mitad o porción o fragmento de un dímero inducible y operativamente conectada a una o más señales de exportación nuclear, como se divulga en la presente documentación.

En un aspecto, el vector puede ser un plásmido individual o un casete de expresión.

Los métodos reivindicados de la invención pueden, en un aspecto, proveer una célula o línea celular huésped eucariota transformada con cualquiera de los vectores que se divulgan en la presente o que expresa el sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación.

La solicitud describe además, pero no reivindica, un organismo transgénico transformado con cualquiera de los vectores que se divulgan en la presente o que expresa el sistema inducible Cpf1 CRISPR-Cas que se divulga en la presente documentación, o la progenie del mismo. También se describe en la presente un organismo modelo que expresa constitutivamente el sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación.

En un aspecto, los métodos y composiciones reivindicados de la invención proveen el uso de un sistema inducible Cpf1 CRISPR-Cas no natural o modificado, que comprende:

una primera construcción de fusión de Cpf1 unida a una primera mitad de un heterodímero inducible y

una segunda construcción de fusión de Cpf1 unida a una segunda mitad del heterodímero inducible,

en donde la primera construcción de fusión de Cpf1 está operativamente conectada a una o más señales de localización nuclear,

en donde la segunda construcción de fusión de Cpf1 está operativamente conectada a una señal de exportación nuclear,

en donde el contacto con una fuente de energía inductora pone juntas a la primera y a la segunda mitad del heterodímero inducible,

en donde poner juntas a la primera y a la segunda mitad del heterodímero inducible les permite a la primera y segunda construcción de fusión de Cpf1 constituir un sistema funcional Cpf1 CRISPR-Cas,

en donde el sistema Cpf1 CRISPR-Cas comprende un ARN guía (ARNg) que comprende una secuencia guía capaz de hibridar con una secuencia diana en un locus genómico de interés en una célula, y

en donde el sistema funcional Cpf1 CRISPR-Cas edita el locus genómico para alterar la expresión génica.

También se describe, pero no se reivindica en la presente, un método para tratar a un sujeto que lo necesite, que comprende inducir edición génica mediante la transformación del sujeto con el polinucleótido como se divulga en la presente documentación o cualquiera de los vectores que se divulgan en la presente y administrar una fuente de energía inductora al sujeto. También se describe, pero no se reivindica en la presente, un polinucleótido o vector en la fabricación de un medicamento, por ejemplo, tal como un medicamento para tratar a un sujeto o para dicho método

para tratar a un sujeto. También se describe, pero no se reivindica en la presente, el polinucleótido como se divulga en la presente documentación o cualquiera de los vectores que se divulgan en la presente para su uso en un método para tratar a un sujeto que lo necesite que comprende inducir edición génica, en donde el método además comprende administrar una fuente de energía inductora al sujeto. En un aspecto de los métodos descritos en la presente, también se provee un molde de reparación, por ejemplo administrado mediante un vector que comprende dicho molde de reparación.

También se describe, pero no se reivindica en la presente, método para tratar a un sujeto que lo necesite, que comprende inducir la activación o represión de la transcripción mediante la transformación del sujeto con el polinucleótido que se discute en la presente o cualquiera de los vectores que se divulgan en la presente, en donde dicho polinucleótido o vector codifica o comprende la Cpf1 catalíticamente inactiva y uno o más dominios funcionales asociadas como se divulga en la presente documentación; el método que además comprende administrar una fuente de energía inductora al sujeto. La invención también provee el polinucleótido que se discute en la presente o cualquiera de los vectores que se divulgan en la presente para su uso en un método para tratar a un sujeto que lo necesite que comprende inducir la activación de la transcripción o represión, en donde el método además comprende administrar una fuente de energía inductora al sujeto.

Por lo tanto, la invención comprende *inter alia* el uso de homodímeros, así como heterodímeros, Cpf1 muerta o Cpf1 que esencialmente no tiene actividad nucleasa, por ejemplo, a través de mutación, sistemas o complejos en donde hay uno o más NLS y/o uno o más NES; dominio funcional(s) conectados a Cpf1 doble; métodos, incluyendo métodos de tratamiento, y usos.

Se podrá apreciar que cuando se hace referencia en la presente a Cpf1, proteína Cpf1 o enzima Cpf1, esto incluye la presente Cpf1 doble. En un aspecto, los métodos y composiciones reivindicados de la invención pueden comprender un método para alterar o modificar la expresión de un producto genético. Dicho método puede comprender introducir en una célula que contiene y que expresa una molécula de ADN que codifica para el producto genético un sistema Cpf1 CRISPR-Cas modificado no natural que comprende una proteína Cpf1 y ARN guía que se direcciona contra la molécula de ADN, mediante lo cual el ARN guía se direcciona contra la molécula de ADN que codifica para el producto genético y la proteína Cpf1 escinde la molécula de ADN que codifica para el producto genético, mediante lo cual se altera la expresión del producto genético; y, en donde la proteína Cpf1 y el ARN guía no existen juntos en forma natural. La invención comprende el ARN guía que comprende una secuencia guía conectada a una secuencia de repetición directa (DR). La invención puede además comprender la proteína Cpf1 que se optimiza por codones para su expresión en una célula eucariota. En una forma de realización preferida la célula eucariota es una célula de mamífero y en una forma de realización más preferida la célula de mamífero es una célula humana. En una forma de realización adicional de la invención, la expresión del producto genético disminuye.

La solicitud proporciona además, pero no reivindica explícitamente, un sistema Cpf1 CRISPR-Cas manipulado que no existen de forma natural que comprende una proteína Cpf1 y un ARN guía que dirige una molécula de ADN que codifica un producto génico en una célula, por lo que el ARN guía se dirige a la molécula de ADN que codifica el producto génico y la proteína Cpf1 escinde la molécula de ADN que codifica el producto génico, por lo cual se altera la expresión del producto génico; y, en donde la proteína Cpf1 y el ARN guía no ocurren juntos naturalmente; incluyendo esto la presente división Cpf1. La solicitud también describe el ARN guía que comprende una secuencia guía unida a una secuencia de DR. La solicitud describe además la proteína Cpf1 que está optimizada en codones para la expresión en una célula eucariota. La célula eucariota puede ser una célula de mamífero y en una forma de realización más preferida la célula de mamífero es una célula humana. En una forma de realización adicional de la invención, la expresión del producto génico es reducida.

También se describe, pero no se reivindica específicamente en la presente, un sistema de vectores modificado no natural que comprende uno o más vectores que comprenden un primer elemento regulador operativamente conectado a un sistema Cpf1 CRISPR-Cas ARN guía que se direcciona contra una molécula de ADN que codifica para un producto genético y un segundo elemento regulador operativamente conectado a una proteína Cpf1; esto incluye la presente Cpf1 doble. Los componentes (a) y (b) se pueden localizar en el mismo o en diferentes vectores del sistema. El ARN guía se direcciona contra la molécula de ADN que codifica para el producto genético en una célula y la proteína Cpf1 escinde la molécula de ADN que codifica para el producto genético, mediante lo cual se altera la expresión del producto genético; y, en donde la proteína Cpf1 y el ARN guía no existen juntos en forma natural. El sistema de vectores comprende el ARN guía que comprende una secuencia guía conectada a una secuencia de DR. El sistema de vectores además comprende la proteína Cpf1 que se optimiza por codones para su expresión en una célula eucariota. En una forma de realización preferida la célula eucariota es una célula de mamífero y en una forma de realización más preferida la célula de mamífero es una célula humana. En una forma de realización adicional de la invención, la expresión del producto genético disminuye.

También se describe, pero no se reivindica específicamente en la presente, un sistema de vectores que comprende uno o más vectores. En algunas formas de realización, el sistema comprende:(a) un primer elemento regulador operativamente conectado a una secuencia de DR y uno o más sitios de inserción para insertar una o más secuencias guía corriente abajo de la secuencia de DR, que cuando se expresa, la secuencia guía dirige la unión específica de secuencia de un complejo Cpf1 CRISPR-Cas a una secuencia diana en una célula eucariota, en donde el complejo Cpf1 CRISPR-Cas comprende Cpf1 complejada con (1) la secuencia guía que se hibrida a la secuencia diana, y (2)

la secuencia de DR; y (b) un segundo elemento regulador operativamente conectado a una secuencia que codifica para una enzima que codifica para dicha enzima Cpf1 que comprende una secuencia de localización nuclear; en donde los componentes (a) y (b) se localizan en el mismo o en diferentes vectores del sistema; esto incluye la presente Cpf1 doble. En algunas formas de realización, el componente (a) además comprende dos o más secuencias guía operativamente conectadas al primer elemento regulador, que cuando se expresan, cada una de las dos o más secuencias guía dirige la unión específica de secuencia de un complejo Cpf1 CRISPR-Cas a una secuencia diana diferente en una célula eucariota.

En algunas formas de realización, el complejo Cpf1 CRISPR-Cas comprende una o más secuencias de localización nuclear de suficiente fuerza para conducir la acumulación de dicho complejo Cpf1 CRISPR-Cas en una cantidad detectable en el núcleo de una célula eucariota. Sin pretender estar limitados por la teoría, se cree que una secuencia de localización nuclear no es necesaria para la actividad del complejo Cpf1 CRISPR-Cas en eucariotas, pero que la inclusión de dichas secuencias aumenta la actividad del sistema, en especial para direccionarse contra moléculas de ácido nucleico en el núcleo.

En algunas formas de realización, la enzima Cpf1 es Cpf1 de una especie bacteriana seleccionada del grupo que consiste en Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Metanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, y Porphyromonas macacae, y puede incluir a CPf1 mutadas derivadas de estos organismos. La enzima puede ser un homólogo u ortólogo de Cpf1. En algunas formas de realización, la Cpf1 se optimiza por codones para su expresión en una célula eucariota. En algunas formas de realización, la Cpf1 dirige la escisión de una o dos hebras en la ubicación de la secuencia diana. En una forma de realización preferida, la rotura de hebra es un corte escalonado con una sobreextensión 5'. En algunas formas de realización, el primer elemento regulador es un promotor de polimerasa III. En algunas formas de realización, el segundo elemento regulador es un promotor de polimerasa II. En algunas formas de realización, la repetición directa tiene una longitud mínima de 16 nts y un bucle-tallo individual. En formas de realización adicionales la repetición directa tiene una longitud mayor a 16 nts, preferiblemente más de 17 nts, y tiene más de un bucle-tallo o estructuras secundarias optimizadas.

También se describe, pero no se reivindica específicamente, una célula huésped eucariota que comprende (a) un primer elemento regulador operativamente conectado a una secuencia de repetición directa y uno o más sitios de inserción para insertar una o más secuencias guía corriente abajo de la secuencia de DR, que cuando se expresa, la secuencia guía dirige la unión específica de secuencia de un complejo Cpf1 CRISPR-Cas a una secuencia diana en una célula eucariota, en donde el complejo Cpf1 CRISPR-Cas comprende Cpf1 complejada con (1) la secuencia guía que se hibrida a la secuencia diana, y (2) la secuencia de DR; y/o (b) un segundo elemento regulador operativamente conectado a una secuencia que codifica para una enzima que codifica para dicha enzima Cpf1 que comprende una secuencia de localización nuclear. En algunas formas de realización, la célula huésped comprende componentes (a) y (b); esto incluye la presente Cpf1 doble. En algunas formas de realización, el componente (a), el componente (b), o los componentes (a) y (b) se integran establemente al genoma de la célula huésped eucariota. En algunas formas de realización, el componente (a) además comprende dos o más secuencias guía operativamente conectadas al primer elemento regulador, que cuando se expresan, cada una de las dos o más secuencias guía dirige la unión específica de secuencia de un complejo Cpf1 CRISPR-Cas a una secuencia diana diferente en una célula eucariota. En algunas formas de realización, la CPf1 se optimiza por codones para su expresión en una célula eucariota. En algunas formas de realización, la Cpf1 dirige la escisión de una o dos hebras en la ubicación de la secuencia diana. En una forma de realización preferida, la rotura de hebra es un corte escalonado con una sobreextensión 5'. En algunas formas de realización, la Cpf1 carece de actividad de escisión de hebra de ADN. En algunas formas de realización, el primer elemento regulador es un promotor de polimerasa III. En algunas formas de realización, la repetición directa tiene una longitud mínima de 16 nts y un bucle-tallo individual. En formas de realización adicionales la repetición directa tiene una longitud mayor a 16 nts, preferiblemente más de 17 nts, y tiene más de un bucle-tallo o estructuras secundarias optimizadas. También se describe en la presente, pero no se reivindica, un organismo eucariota no humano; tal como un organismo eucariota multicelular, que comprende una célula huésped eucariota según cualquiera de las realizaciones descritas. También se describe, pero no se reivindica en la presente, un organismo eucariota; tal como un organismo eucariota multicelular, que comprende una célula huésped eucariota según cualquiera de las realizaciones descritas. El organismo en algunas formas de realización de estos aspectos puede ser un animal; por ejemplo, un mamífero. Por tanto, el organismo puede ser un artrópodo, tal como un insecto. El organismo también puede ser una planta. Además, el organismo puede ser un hongo.

También se describe, pero no se reivindica específicamente en la presente, un conjunto de elementos que comprende uno o más de los componentes descritos en la presente. En algunas formas de realización, el conjunto de elementos comprende un sistema de vectores e instrucciones para usar el conjunto de elementos. En algunas formas de realización, el sistema de vectores comprende (a) un primer elemento regulador operativamente conectado a una secuencia de repetición directa y uno o más sitios de inserción para insertar una o más secuencias guía corriente abajo de la secuencia de DR, que cuando se expresa, la secuencia guía dirige la unión específica de secuencia de un complejo Cpf1 CRISPR-Cas a una secuencia diana en una célula eucariota, en donde el complejo Cpf1 CRISPR-Cas comprende Cpf1 complejada con (1) la secuencia guía que se hibrida a la secuencia diana, y (2) la secuencia de DR;

y/o (b) un segundo elemento regulador operativamente conectado a una secuencia que codifica para una enzima que codifica para dicha enzima Cpf1 que comprende una secuencia de localización nuclear y ventajosamente esto incluye la presente Cpf1 doble. En algunas formas de realización, el conjunto de elementos comprende los componentes (a) y (b) localizados en el mismo o en diferentes vectores del sistema. En algunas formas de realización, el componente (a) además comprende dos o más secuencias guía operativamente conectadas al primer elemento regulador, que cuando se expresan, cada una de las dos o más secuencias guía dirige la unión específica de secuencia de un complejo Cpf1 CRISPR-Cas a una secuencia diana diferente en una célula eucariota. En algunas formas de realización, la Cpf1 comprende una o más secuencias de localización nuclear de suficiente fuerza para conducir la acumulación de dicho Cpf1 en una cantidad detectable en el núcleo de una célula eucariota. En algunas formas de realización, la enzima Cpf1 es Cpf1 de una especie bacteriana seleccionada del grupo que consiste en Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Metanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, y Porphyromonas macacae, y puede incluir a CPf1 mutadas derivadas de estos organismos. La enzima puede ser un homólogo u ortólogo de Cpf1. En algunas formas de realización, la Cpf1 se optimiza por codones para su expresión en una célula eucariota. En algunas formas de realización, la Cpf1 dirige la escisión de una o dos hebras en la ubicación de la secuencia diana. En una forma de realización preferida, la rotura de hebra es un corte escalonado con una sobreextensión 5'. En algunas formas de realización, la enzima CRISPR carece de actividad de escisión de hebra de ADN. En algunas formas de realización, la repetición directa tiene una longitud mínima de 16 nts y un bucle-tallo individual. En formas de realización adicionales la repetición directa tiene una longitud mayor a 16 nts, preferiblemente más de 17 nts, y tiene más de un bucle-tallo o estructuras secundarias optimizadas.

En un aspecto, la invención provee un método para modificar un polinucleótido diana en una célula eucariota. En algunas formas de realización, el método comprende permitir que un complejo Cpf1 CRISPR-Cas se una al polinucleótido diana para efectuar la escisión de dicho polinucleótido diana modificando de esa manera al polinucleótido diana, en donde el complejo Cpf1 CRISPR-Cas comprende Cpf1 complejada con una secuencia guía hibridada a una secuencia diana dentro de dicho polinucleótido diana, en donde dicha secuencia guía está conectada a una secuencia de repetición directa. En algunas formas de realización, dicha escisión comprende escindir una o dos hebras en la ubicación de la secuencia diana mediante dicha Cpf1; esto incluye la presente Cpf1 doble. En algunas formas de realización, dicha escisión da como resultado una menor transcripción de un gen diana. En algunas formas de realización, el método además comprende reparar dicho polinucleótido diana escindido mediante la recombinación homóloga con un polinucleótido molde exógeno, en donde dicha reparación da como resultado dicha mutación que comprende una inserción, supresión o sustitución de uno o más nucleótidos de dicho polinucleótido diana. En algunas formas de realización, dicha mutación da como resultado uno o más cambios de aminoácido en una proteína expresada a partir de un gen que comprende la secuencia diana. En algunas formas de realización, el método además comprende suministrar uno o más vectores a dicha célula eucariota, en donde dichos uno o más vectores dirigen la expresión de uno o más de: la Cpf1, y la secuencia guía conectada a la secuencia de DR. En algunas formas de realización, dichos vectores se administran a la célula eucariota en un sujeto. En algunas formas de realización, dicha modificación tiene lugar en dicha célula eucariota en una célula cultivo. En algunas formas de realización, el método además comprende aislar dicha célula eucariota a partir de un sujeto antes de dicha modificación. En algunas formas de realización, el método además comprende retornar dicha célula eucariota y/o células derivadas de dicho sujeto.

En un aspecto, los métodos de la invención implican modificar la expresión de un polinucleótido en una célula eucariota. En algunas formas de realización, el método comprende permitir que un complejo Cpf1 CRISPR-Cas se una al polinucleótido de forma que dicha unión da como resultado una mayor o una menor expresión de dicho polinucleótido; en donde el complejo Cpf1 CRISPR-Cas comprende Cpf1 complejada con una secuencia guía hibridada a una secuencia diana dentro de dicho polinucleótido, en donde dicha secuencia guía está conectada a una secuencia de repetición directa; esto incluye la presente Cpf1 doble. En algunas formas de realización, el método además comprende suministrar uno o más vectores a dichas células eucariotas, en donde dichos uno o más vectores dirigen la expresión de uno o más de: la Cpf1, y la secuencia guía conectada a la secuencia de DR.

También se describe, pero no se reivindica específicamente en la presente, un método para generar una célula eucariota modelo que comprende un gen de enfermedad mutado. En algunas formas de realización, un gen de enfermedad es cualquier gen asociado con un incremento en el riesgo de tener o desarrollar una enfermedad. En algunas formas de realización, el método comprende (a) introducir uno o más vectores dentro de una célula eucariota, en donde dichos uno o más vectores dirigen la expresión de uno o más de: Cpf1, y una secuencia guía conectada a una secuencia de repetición directa; y (b) permitir que un complejo Cpf1 CRISPR-Cas se una a un polinucleótido diana para efectuar la escisión del polinucleótido diana dentro de dicho gen de enfermedad, en donde el complejo Cpf1 CRISPR-Cas comprende la Cpf1 complejada con (1) la secuencia guía que se hibrida a la secuencia diana dentro del polinucleótido diana, y (2) la secuencia de DR, generando de esa forma una célula eucariota modelo que comprende un gen de enfermedad mutado; esto incluye la presente Cpf1 doble. En algunas formas de realización, dicha escisión comprende escindir una o dos hebras en la ubicación de la secuencia diana mediante dicha Cpf1. En una forma de realización preferida, la rotura de hebra es un corte escalonado con una sobreextensión 5'. En algunas formas de realización, dicha escisión da como resultado una menor transcripción de un gen diana. En algunas formas de

realización, el método además comprende reparar dicho polinucleótido diana escindido mediante la recombinación homóloga con un polinucleótido molde exógeno, en donde dicha reparación da como resultado dicha mutación que comprende una inserción, supresión o sustitución de uno o más nucleótidos de dicho polinucleótido diana. En algunas formas de realización, dicha mutación da como resultado uno o más cambios de aminoácido en una proteína expresión a partir de un gen que comprende la secuencia diana.

También se describe, pero no se reivindica específicamente en la presente, un método para desarrollar un agente biológicamente activo que modula un evento de señalización celular asociado con un gen de enfermedad. En algunas formas de realización, un gen de enfermedad es cualquier gen asociado con un incremento en el riesgo de tener o desarrollar una enfermedad. En algunas formas de realización, el método comprende (a) poner en contacto un compuesto de prueba con una célula modelo de una cualquiera de las formas de realización descritas; y (b) detectar un cambio en un resultado que sea indicativo de una reducción o un aumento de un evento de señalización celular asociado con dicha mutación en dicho gen de enfermedad, desarrollando de esa manera dicho agente biológicamente activo que modula dicho evento de señalización celular asociado con dicho gen de enfermedad.

En un aspecto, los métodos y composiciones reivindicados de la invención hacen uso de un polinucleótido recombinante que comprende una secuencia guía corriente abajo de una secuencia de repetición directa, en donde cuando se expresa la secuencia guía dirige la unión específica de secuencia de un complejo Cpf1 CRISPR-Cas a una correspondiente secuencia diana presente en una célula eucariota. En algunas formas de realización, la secuencia diana es una secuencia viral presente en una célula eucariota. En algunas formas de realización, la secuencia diana es un protooncogén o un oncogén.

También se describe, pero no se reivindica específicamente en la presente, un método para seleccionar una o más célula(s) mediante la introducción de una o más mutaciones en un gen en la una o más célula(s), donde dicho método comprende: introducir uno o más vectores en la o las células, en donde dichos uno o más vectores dirigen la expresión de uno o más de: Cpf1, una secuencia guía conectada a una secuencia de repetición directa, y un molde de edición; en donde el molde de edición comprende dichas una o más mutaciones que impiden la escisión de Cpf1; permitir la recombinación homóloga del molde de edición con el polinucleótido diana en la o las célula(s) a seleccionar; permitir que un complejo Cpf1 CRISPR-Cas se una a un polinucleótido diana para efectuar la escisión del polinucleótido diana dentro de dicho gen, en donde el complejo Cpf1 CRISPR-Cas comprende la Cpf1 complejada con (1) la secuencia guía que se hibrida a la secuencia diana dentro del polinucleótido diana, y (2) la secuencia de repetición directa, en donde la unión del complejo Cpf1 CRISPR-Cas al polinucleótido diana induce muerte celular, permitiendo de esa forma seleccionar una o más célula(s) en las que se han introducido una o más mutaciones; esto incluye la presente Cpf1 doble. La célula a seleccionar puede ser una célula eucariota. La selección de células específicas puede ser sin requerir un marcador de selección o un proceso de dos pasos que puede incluir un sistema de contra selección.

En la presente existe la frase “esto incluye la presente Cpf1 doble” o texto similar; y, esto indica que la Cpf1 en formas de realización en la presente puede ser una Cpf1 doble como se divulga en la presente documentación.

También se describe, pero no se reivindica específicamente en la presente, un sistema inducible Cpf1 CRISPR-Cas no natural o modificado, que comprende una primera construcción de fusión de Cpf1 unida a una primera mitad de un heterodímero inducible y una segunda construcción de fusión de Cpf1 unida a una segunda mitad del heterodímero inducible, en donde la primera construcción de fusión de Cpf1 está operativamente conectada a una o más señales de localización nuclear, en donde la segunda construcción de fusión de Cpf1 está operativamente conectada a una señal de exportación nuclear, en donde el contacto con una fuente de energía inductora pone juntas a la primera y a la segunda mitad del heterodímero inducible, en donde poner juntas a la primera y a la segunda mitad del heterodímero inducible les permite a la primera y segunda construcción de fusión de Cpf1 constituir un sistema funcional Cpf1 CRISPR-Cas, en donde el sistema Cpf1 CRISPR-Cas comprende un ARN guía (ARNg) que comprende una secuencia guía capaz de hibridar con una secuencia diana en un locus genómico de interés en una célula, y en donde el sistema funcional Cpf1 CRISPR-Cas edita el locus genómico para alterar la expresión génica. La primera mitad del heterodímero inducible puede ser FKBP12 y la segunda mitad del heterodímero inducible puede ser FRB. La fuente de energía inductora puede ser rapamicina.

Una fuente de energía inductora puede considerarse como simplemente un inductor o un agente dimerizante. El término ‘fuente de energía inductora’ se usa a lo largo de la presente para consistencia. La fuente de energía inductora (o inductor) actúa para reconstituir la Cpf1. En algunas formas de realización, la fuente de energía inductora pone juntas las dos partes de la Cpf1 a través de la acción de las dos mitades del dímero inducible. Las dos mitades del dímero inducible por ello que juntan en presencia de la fuente de energía inductora. Las dos mitades del dímero no formarán un dímero (dimerizar) sin la fuente de energía inductora.

Por consiguiente, las dos mitades del dímero inducible cooperan con la fuente de energía inductora para dimerizar el dímero. Esto a su vez reconstituye la Cpf1 poniendo juntas la primera y la segunda parte de la Cpf1.

Cada una de las construcciones de fusión de enzima CRISPR comprende una parte de la Cpf1 doble. Estas se fusionan, preferiblemente a través de un conector tal como un conector GlySer descrito en la presente, a una de las dos mitades del dímero. Las dos mitades del dímero pueden ser sustancialmente los dos monómeros que formen juntos el heterodímero, o pueden ser monómeros diferentes que forman juntos el heterodímero. Como tal, los dos monómeros pueden interpretarse como una mitad del dímero completo.

La Cpf1 es doble en el sentido de que las dos partes de la enzima Cpf1 sustancialmente comprenden una Cpf1 funcional. La Cpf1 puede funcionar como una enzima de edición de genomas (cuando forma un complejo con el ADN diana y la guía), tal como una *nickasa* o una nucleasa (que escinde ambas hebras del ADN), o puede ser una Cpf1 muerta que es esencialmente una proteína de unión a ADN con muy poca o nada de actividad catalítica, debido típicamente a una o más mutaciones en sus dominios catalíticos.

Las dos partes de la Cpf1 doble pueden interpretarse como la parte N' terminal y la parte C' terminal de la Cpf1 doble. La fusión típicamente es en el punto de división de la Cpf1. En otras palabras, el extremo C' terminal de la parte N' terminal de la Cpf1 doble se fusiona a una de las mitades del dímero, mientras que el extremo N' terminal de la parte C' terminal se fusiona a la otra mitad del dímero.

La Cpf1 no tiene que ser dividida en el sentido en que se crea una nueva rotura. EL punto de división típicamente se diseña *in silico* y se clona en las construcciones. Juntas, las dos partes de la Cpf1 doble, las partes N' terminal y C' terminal, forman una Cpf1 completa, que comprende preferiblemente por lo menos 70% o más de los aminoácidos de tipo salvaje (o nucleótidos que los codifican), preferiblemente por lo menos 80% o más, preferiblemente por lo menos 90% o más, preferiblemente por lo menos 95% o más, y más preferiblemente por lo menos 99% o más de los aminoácidos de tipo salvaje (o nucleótidos que los codifican). Puede ser posible algo de recorte, y se prevén mutantes. Los dominios no funcionales se pueden eliminar completamente. Lo que es importante es que las dos partes pueden ponerse juntas y que la función deseada de Cpf1 se restituya o reconstituya.

El dímero puede ser un homodímero o un heterodímero.

Se pueden usar uno o más, preferiblemente dos, NLS en conexión operativa con la primera construcción de Cpf1. Se pueden usar uno o más, preferiblemente dos, NES en conexión operativa con la primera construcción de Cpf1. Las NLS y/o NES preferiblemente flanquean la doble fusión de dímero de Cpf1 (es decir, medio dímero), es decir, una NLS puede posicionarse en el extremo N' terminal de la primera construcción de Cpf1 y una NLS puede estar en el extremo C' terminal de la primera construcción de Cpf1. De manera similar, una NES puede posicionarse en el extremo N' terminal de la segunda construcción de Cpf1 y una NES puede estar en el extremo C' terminal de la segunda construcción de Cpf1. Cuando se hace referencia a extremos N' o C' terminales, se podrá apreciar que estos corresponden a los extremos 5' y 3' de la correspondiente secuencia de nucleótidos.

Un arreglo preferido es que la primera construcción de Cpf1 se arregle como 5'-NLS-(parte N' terminal de Cpf1)-conector-(primera mitad del dímero)-NLS-3'. Un arreglo preferido es que la segunda construcción de Cpf1 se arregle 5'-NES--(segunda mitad del dímero)-conector-(parte C' terminal de Cpf1)-NES-3'. Un promotor adecuado está preferiblemente corriente arriba de cada una de estas construcciones. Las dos construcciones se pueden administrar separadas o juntas.

En algunas formas de realización, una o todas las NES en conexión operativa con la segunda construcción de CPf1 se puede intercambiar con una NLS. Sin embargo, esto típicamente puede ser no preferido y, en otras formas de realización, la señal de localización en conexión operativa con la segunda construcción de Cpf1 es una o más NES.

También ha de apreciarse que la NES puede estar operativamente conectada al fragmento N' terminal de la Cpf1 doble y que la NLS puede estar operativamente conectada al fragmento C' terminal de la Cpf1 doble. Sin embargo, el arreglo donde la NLS está operativamente conectada al fragmento N' terminal de la Cpf1 doble y en donde la NES está operativamente conectada al fragmento C' terminal de la Cpf1 doble, puede ser preferido.

La NES funciona para localizar la segunda construcción de fusión de Cpf1 por fuera del núcleo, por lo menos hasta que se provea la fuente de energía inductora (por ejemplo, por lo menos hasta que la fuente de energía se provea al inductor para llevar a cabo su función). La presencia del inductor estimula la dimerización de las dos fusiones de Cpf1 dentro del citoplasma y hace termodinámicamente posible que las fusiones de Cpf1 dimerizadas, primera y segunda, se localicen en el núcleo. Sin estar limitados por la teoría, los Solicitantes creen que la NES secuestra la segunda fusión de Cpf1 en el citoplasma (es decir, por fuera del núcleo). La NLS de la primera fusión de Cpf1 la localiza en el núcleo. En ambos casos, los Solicitantes usan la NES o la NLS para desplazar el equilibrio (el equilibrio del transporte nuclear) en una dirección deseada. La dimerización típicamente ocurre por fuera del núcleo (una muy pequeña fracción podría ocurrir en el núcleo) y las NLS en el complejo dimerizado desplazan el equilibrio de transporte nuclear a localización nuclear, de forma que la Cpf1 dimerizada y por ende reconstituida entra al núcleo.

Beneficiosamente, los Solicitantes son capaces de reconstituir la función en la Cpf1 doble. Se usa transfección transitoria para probar el concepto y ocurre dimerización de fondo en presencia de la fuente de energía inductora. No se ve actividad con los fragmentos separados de la Cpf1. Luego se usa expresión estable a través de suministro lentiviral para desarrollar esto y mostrar que se puede usar un abordaje de Cpf1 doble.

Este presente abordaje de Cpf1 doble es beneficioso ya que permite que la actividad de Cpf1 sea inducible, por consiguiente, permitiendo un control temporal. Aún más, se pueden usar diferentes secuencias de localización (es decir, la NES y la NLS como preferidas) para reducir la actividad de fondo de los complejos autoensamblados. También se pueden usar promotores específicos de tejido, por ejemplo uno para cada una de la primera y segunda construcción de fusión de Cpf1, para un direccionamiento específico de tejido, proporcionando por consiguiente un control espacial. Se pueden usar dos promotores específicos de tejido diferentes para ejercer un grado más fino de control, si así se lo

requiere. El mismo abordaje se puede usar con respecto a promotores específicos de etapa o puede haber una mezcla de promotores específicos de etapa y de tejido, en donde una de la primera y la segunda construcción de fusión de Cpf1 está bajo el control (es decir operativamente conectada, o comprendiendo) un promotor específico de tejido, mientras que la otra de la primera y segunda construcción de fusión de Cpf1 está bajo el control (es decir operativamente conectada, o comprendiendo) un promotor específico de etapa.

El sistema inducible Cpf1 CRISPR-Cas comprende una o más secuencias de localización nuclear (NLS), como se describe en la presente, por ejemplo como operativamente conectada a la primera construcción de fusión de Cpf1. Estas secuencias de localización nuclear idealmente tienen la fuerza suficiente como para conducir la acumulación de dicho primera construcción de Cpf1 de fusión en una cantidad detectable en el núcleo de una célula eucariota. Sin pretender estar limitados por la teoría, se cree que una secuencia de localización nuclear no es necesaria para la actividad del complejo Cpf1 CRISPR-Cas en eucariotas, pero que la inclusión de dichas secuencias aumenta la actividad del sistema, en especial para direccionarse contra moléculas de ácido nucleico en el núcleo, y asiste en la operación del presente sistema de 2 partes.

Equivalentemente, la segunda construcción de fusión de Cpf1 está operativamente conectada a una secuencia de exportación nuclear (NES). De hecho, la misma se puede conectar a una o más secuencias de exportación nuclear. En otras palabras, el número de secuencias de exportación que se usan con la segunda construcción de fusión de Cpf1 es preferiblemente 1 o 2 o 3. Típicamente se prefiere 2, pero 1 es suficiente y es lo preferido en algunas formas de realización. En el arte se conocen ejemplos adecuados de NLS y NES. Por ejemplo, una señal de exportación nuclear (NES) preferida es la proteína tirosina quinasa 2 humana. Las señales preferidas serán específicas de especie.

Cuando se usa el sistema de FRB y FKBP, la FKBP preferiblemente está flanqueada por secuencias de localización nuclear (NLS). Cuando se usa el sistema de FRB y FKBP, el arreglo preferido es N' terminal Cpf1 – FRB – NES :C' terminal Cpf1-FKBP-NLS. Por consiguiente, la primera construcción de fusión de Cpf1 comprendería la parte C' terminal de Cpf1 y la segunda construcción de fusión de Cpf1 comprendería la parte N' terminal de Cpf1.

Otro aspecto beneficioso para la presente invención es que la misma se puede activar rápidamente, es decir que la misma tiene una respuesta rápida. Se cree, sin estar limitados por la teoría, que la actividad de Cpf1 puede inducirse a través de la dimerización de construcciones de fusión existentes (ya presentes) (a través del contacto con la fuente de energía inductora) más rápidamente que a través de la expresión (especialmente traducción) de nuevas construcciones de fusión. Como tal, la primera y la segunda construcciones de fusión de Cpf1 pueden expresarse en la célula diana antes de tiempo, es decir antes de que se requiera la actividad de Cpf1. La actividad de Cpf1 entonces se puede controlar temporalmente y luego constituirse rápidamente a través de la adición de la fuente de energía inductora, que idealmente actúa más rápidamente (para dimerizar el heterodímero y por lo tanto proveer la actividad de Cpf1) que a través de la expresión (incluyendo inducción de transcripción) de la Cpf1 administrada mediante un vector, por ejemplo.

Los términos Cpf1 o enzima Cpf1 y enzima CRISPR se usan indistintamente en la presente a menos que sea evidente de otra manera.

Los Solicitantes demuestran que CPf1 puede dividirse en dos componentes, que reconstituyen una nucleasa funcional cuando se ponen juntos. El empleo de dominios de dimerización sensibles a rapamicina, los Solicitantes generan una Cpf1 químicamente inducible para control temporal de edición de genomas y transcripción modulación mediados por Cpf1. En otras palabras, los Solicitantes demuestran que la Cpf1 se puede hacer químicamente inducible mediante su partición en dos fragmentos y que los dominios de dimerización sensibles a rapamicina se pueden usar para un reensamblaje controlado de la Cpf1. Los Solicitantes muestran que la Cpf1 reensamblada se puede usar para mediar la edición de genomas (a través de actividad nucleasa/nickasa) así como para modulación de la transcripción (como dominio de unión a ADN, la llamada "Cpf1 muerta").

Como tal, el uso de dominios de dimerización sensibles a rapamicina es lo preferido. El reensamblaje de la Cpf1 es lo preferido. El reensamblaje se puede determinar mediante la restitución de la actividad de unión. Cuando la Cpf1 es un nickasa o induce una rotura de doble hebra, se describen porcentajes de comparación adecuados en comparación con una versión de tipo salvaje como se describe en la presente.

Los tratamientos con rapamicina pueden durar 12 días. La dosis puede ser de 200 nM. Esta dosificación temporal y/o molar es un ejemplo de una dosis apropiada para la línea celular de riñón de embrión humano 293FT (HEK293FT) y esto también se puede usar en otras líneas celulares. Esta figura puede extrapolarse a un uso terapéutico *in vivo*, por ejemplo, en mg/kg. Sin embargo, también se prevé usar aquí la dosificación estándar para administrar rapamicina a un sujeto. Como "dosificación estándar", se designa a la dosificación bajo un uso o indicación primaria normal de rapamicina (es decir la dosis que se usa cuando se administra rapamicina para usar en la prevención de rechazo de órganos).

Es de destacar que el arreglo preferido de las piezas de Cpf1-FRB/FKBP es por separado e inactivo hasta que la dimerización de FRB y FKBP inducida por la rapamicina da como resultado el reensamblaje de una nucleasa de Cpf1 funcional de longitud completa. Por consiguiente, se prefiere que la primera construcción de Cpf1 de fusión unida a una primera mitad de un heterodímero inducible se suministre separadamente y/o que se localice separadamente de la segunda construcción de fusión de Cpf1 unida a una primera mitad de un heterodímero inducible.

Para secuestrar el fragmento Cpf1(N)-FRB en el citoplasma, donde es menos probable que se dimerice con el fragmento Cpf1(C)-FKBP localizado en el núcleo, es preferible usar en Cpf1(N)-FRB una secuencia de exportación nuclear (NES) individual de la proteína tirosina quinasa 2 humana (Cpf1(N)-FRB-NES). En presencia de rapamicina, la Cpf1(N)-FRB-NES se dimeriza con Cpf1(C)-FKBP-2xNLS para reconstituir una proteína Cpf1 completa, que altera el balance de tráfico nuclear hacia la importación nuclear y permite el direccionamiento contra el ADN.

Una alta dosificación de Cpf1 puede exacerbar la frecuencia de indeles y de secuencias fuera de diana (OT) que exhiben pocas faltas de coincidencia con la hebra guía. Dichas secuencias son especialmente susceptibles, si las faltas de coincidencia son no consecutivas y/o están por fuera de la región semilla de la guía. Por lo tanto, se puede usar un control temporal de la actividad de Cpf1 para reducir la dosificación de los experimentos de expresión a largo plazo y ello da como resultado menor cantidad de indeles fuera de diana en comparación con la Cpf1 constitutivamente activa.

La administración por virus es lo preferido. En particular, se prevé el uso de un vector de administración lentiviral o AAV. Los Solicitantes generan una construcción lentiviral con Cpf1 doble, similar al plásmido con lentiCRISPR. Las piezas divididas deben ser lo suficientemente pequeñas para caber dentro de límite de tamaño de ~4,7kb del AAV.

Los Solicitantes demuestran que se puede usar una expresión estable, de bajo número de copias de Cpf1 doble para inducir indeles sustanciales en un locus diana sin mutación significativa de sitios fuera de diana. Los Solicitantes clonan fragmentos de Cpf1 (2 partes en base a la división 5, descrita en la presente).

También se puede usar una Cpf1, que comprende un dominio de transactivación VP64, por ejemplo agregado a Cpf1(C)-FKBP-2xNLS (Cpf1muerta (C)-FKBP-2xNLS-VP64). Estos fragmentos reconstituyen una fusión Cpf1-VP64 catalíticamente inactiva (Cpf1 muerta-VP64). La activación de la transcripción se induce por VP64 en presencia de rapamicina para inducir la dimerización de la fusión de Cpf1(C)-FKBP y la fusión de Cpf1(N)-FRB. En otras palabras, los Solicitantes prueban la capacidad de inducción de Cpf1 muerta-VP64 doble y muestran que se induce la activación de la transcripción de la Cpf1 muerta-VP64 doble en presencia de rapamicina. Como tal, la presente Cpf1 inducible se puede asociar con uno o más dominios funcionales, tales como un activador o represor transcripcional o una nucleasa (tal como Fok1). Se puede unir o fusionar un dominio funcional a una parte de la Cpf1 doble.

Un arreglo preferido es que la primera construcción de Cpf1 se arregle 5'-Primera Señal de Localización-(parte N' terminal de Cpf1)-conector-(primera mitad del dímero)-Primera Señal de Localización-3' y que la segunda construcción de Cpf1 se arregle 5'- Segunda Señal de Localización--(segunda mitad del dímero)-conector-(parte C' terminal de Cpf1)-Segunda Señal de Localización-Dominio funcional-3'. Aquí, se coloca un dominio funcional en el extremo 3' de la segunda construcción de Cpf1. Como alternativa, se puede colocar un dominio funcional en el extremo 5' de la primera construcción de Cpf1. Uno o más dominios funcionales se puede usar en el extremo 3' o en el extremo 5' o en ambos extremos. Preferiblemente hay un promotor adecuado corriente arriba de cada una de estas construcciones. Las dos construcciones se pueden administrar separadas o juntas. Las Señales de Localización pueden ser una NLS o una NES, con la condición de que las mismas no se mezclen entre las distintas construcciones.

En un aspecto, los métodos y composiciones reivindicados de la invención pueden implicar un sistema inducible Cpf1 CRISPR-Cas en donde la Cpf1 tiene una actividad nucleasa disminuida en por lo menos 97%, o 100% en comparación con la enzima Cpf1 que no tiene la por lo menos una mutación.

Por lo tanto, también se prefiere que la Cpf1 sea una Cpf1 muerta. Idealmente, siempre se debe hacer la división de modo que el (los) dominio(s) catalíticos no se vean afectados. Para la Cpf1 muerta, la intención es que ocurra la unión al ADN, pero que la misma no muestre actividad de escisión o nickasa.

En un aspecto, los métodos y composiciones de la invención proveen un sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación en donde uno o más dominios funcionales se asocian con la Cpf1. Este dominio funcional se puede asociar (es decir unir o fusionar) con una parte de la Cpf1 doble, o con ambas. Puede haber uno asociado con cada una de las dos partes de la Cpf1 doble. Por ello típicamente estos se pueden proveer como parte de la primera y/o segunda construcción de fusión de Cpf1, como fusiones dentro de la construcción. Los dominios funcionales típicamente se fusionan a través de un conector, tal como conector GlySer, como se divulga en la presente. El uno o más dominios funcionales pueden ser dominios de activación o de represión de la transcripción. A pesar de que pueden ser dominios diferentes, se prefiere que todos los dominios funcionales sean o bien activadores o bien represores, y que no se use una mezcla de los dos.

El dominio de activación de la transcripción puede comprender a VP64, p65, MyoD1, HSF1, RTA o SET7/9.

En un aspecto, los métodos y composiciones de la invención proveen un sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación en donde el uno o más dominios funcionales asociadas con la Cpf1 son un dominio represor transcripcional.

En un aspecto, los métodos y composiciones proveen un sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación en donde el dominio represor transcripcional es un dominio KRAB.

En un aspecto, los métodos y composiciones reivindicados proveen un sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación en donde el dominio represor transcripcional es un dominio NuE, dominio NcoR,

dominio SID o un dominio SID4X.

En un aspecto, los métodos y composiciones proveen un sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación en donde el uno o más dominios funcionales asociadas con la proteína adaptadora tienen una o más actividades que comprenden actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad de factor de liberación de la transcripción, actividad de modificación de histonas, actividad de escisión de ARN, actividad de escisión de ADN, actividad de integración de ADN o actividad de unión a ácidos nucleicos.

Los dominios modificadores de histonas también se prefieren en algunas formas de realización. Los ejemplos de dominios modificadores de histonas se divulgan más adelante. Los dominios transposasa, dominios de maquinaria de HR (recombinación homóloga), dominios recombinasa, y/o dominios integrasa también se prefieren como los dominios funcionales presentes. En algunas formas de realización, la actividad de integración de ADN incluye dominios de maquinaria de HR, dominios integrasa, dominios recombinasa y/o dominios transposasa.

En un aspecto, los métodos y composiciones reivindicados proveen un sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación en donde la actividad de escisión de ADN es debida a una nucleasa.

En un aspecto, los métodos y composiciones reivindicados proveen un sistema inducible Cpf1 CRISPR-Cas como se divulga en la presente documentación en donde la nucleasa comprende una nucleasa Fok1.

El uso de dichos dominios funcionales, que se prefieren con el presente sistema de Cpf1 doble, también se expone en detalle en Konermann y col., ("Genome-scale transcriptional activation with an engineered CRISPR-Cas9 complex" *Nature* publicado el 11 de diciembre de 2014).

El presente sistema se puede usar con cualquier guía.

Se pueden usar guías modificadas en determinadas formas de realización. Particularmente preferidas son las guías que representan los contenidos del artículo de Konermann *Nature* del 11 de diciembre de 2014 que se mencionó previamente. Estas guías se modifican de forma que se agregan porciones de ARN de unión a proteína (tales como aptámeros). Dicha o dichas porciones pueden reemplazar a una porción de la guía. Se pueden usar proteínas de unión a dominios de ARN para luego reconocer el ARN y reclutar los dominios funcionales, tales como los que se describen en la presente, hacia la guía. Principalmente esto es para usar con la Cpf1 muerta que conduce a la activación o represión de la transcripción o escisión de ADN a través de nucleasas tales como Fok1. El uso de dichas guías en combinación con Cpf1 muerta es poderoso, y es especialmente poderoso si la Cpf1 misma también se asocia con su propio dominio funcional, como se divulga en la presente. Cuando se induce una Cpf1 muerta (con o sin su propio dominio funcional asociado) para reconstituirse de acuerdo con la presente invención, es decir es una Cpf1 doble, entonces la herramienta es especialmente poderosa.

Un ARN guía (ARNg), también preferido para su uso en la presente invención, puede comprender una secuencia guía capaz de hibridar con una secuencia diana en un locus genómico de interés en una célula, en donde el ARNg se modifica mediante la inserción de distintas secuencia(s) de ARN que se unen a una o más proteínas adaptadoras, y en donde la proteína adaptadora se asocia con uno o más dominios funcionales. La Cpf1 puede comprender por lo menos una mutación, de forma que la enzima Cpf1 no tenga más de un 5% de la actividad nucleasa de la enzima Cpf1 que no tiene la por lo menos una mutación; y/o por lo menos una o más secuencias de localización nuclear. También se provee una composición no natural o manipulada que comprende: uno o más ARN guías (ARNg) que comprenden una secuencia guía capaz de hibridar con una secuencia diana en un locus genómico de interés en una célula, una enzima Cpf1 que comprende por lo menos una o más secuencias de localización nuclear, en donde la enzima Cpf1 comprende por lo menos una mutación, de forma que la enzima Cpf1 no tiene más de un 5% de la actividad nucleasa de la enzima Cpf1 que no tiene la por lo menos una mutación, en donde el por lo menos un ARNg se modifica mediante la inserción de distintas secuencia(s) de ARN que se unen a una o más proteínas adaptadoras, y en donde la proteína adaptadora se asocia con uno o más dominios funcionales.

El ARNg preferiblemente se modifica mediante la inserción de distintas secuencia(s) de ARN que se unen a una o más proteínas adaptadoras. La inserción de distintas secuencia(s) de ARN que se unen a una o más proteínas adaptadoras preferiblemente es de una secuencia de aptámero o dos o más secuencias de aptámero específicas con la misma o con diferentes proteínas adaptadora(s). La proteína adaptadora preferiblemente comprende MS2, PP7, Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, φCb5, φCb8r, φCb12r, φCb23r, 7s, PRR1. Las líneas celulares que expresan establemente *inter alia* Cpf1 muerta doble pueden ser útiles.

Los Solicitantes demuestran que la Cpf1 se puede dividir en dos fragmentos distintos, que reconstituyen una nucleasa Cpf1 funcional de longitud completa cuando se juntan usando inducción química. La arquitectura de Cpf1 doble será útil para una variedad de aplicaciones. Por ejemplo, la CPf1 doble puede permitir estrategias genéticas para restringir la actividad de Cpf1 a poblaciones de célula interseccionales poniendo cada fragmento bajo un promotor específico de tejido diferente. Adicionalmente, también se pueden emplear dominios de dimerización inducibles químicamente diferentes tales como APA y giberelina.

La fuente de energía inductora es preferiblemente inducción química.

La posición o ubicación de la división es el punto en donde la primera parte de la enzima Cpf1 se separa de la segunda parte. En algunas formas de realización, la primera parte comprenderá o codificará para los aminoácidos 1 a X, mientras que la segunda parte comprenderá o codificará para los aminoácidos X+1 hasta el final. En este ejemplo, la numeración es continua, pero esto puede no ser siempre necesario ya que pueden recortarse aminoácidos (o los nucleótidos que los codifican) desde el extremo de alguna de las partes, con la condición de que se retenga una actividad de unión a ADN suficiente y, si se requiere, actividad nickasa o de escisión de ADN, por ejemplo por lo menos 40%, 50%, 60%, 70%, 80%, 90% o 95% actividad en comparación con la Cpf1 de tipo salvaje.

La numeración ejemplificativa que se provee en la presente puede ser en referencia a la proteína de tipo salvaje, preferiblemente la FnCpf1 de tipo salvaje. Sin embargo, se contempla el posible uso de mutantes de la Cpf1 de tipo salvaje tal como la proteína FnCpf1. La numeración también puede no seguir exactamente la numeración de FnCpf1 ya que, por ejemplo, se pueden usar algunas truncaciones o supresiones en N' o C' terminal, pero esto se puede solucionar usando herramientas estándares de alineamiento de secuencia. Los ortólogos también se prefieren como herramienta de alineamiento de secuencia.

Por consiguiente, la posición de división se puede seleccionar usando la experiencia ordinaria en el arte, por ejemplo en base a datos cristalinos y/o predicciones computacionales de estructura.

Por ejemplo, el análisis computacional de la estructura primaria de las nucleasas Cpf1 revela tres regiones distintas (Figura 1). Primero, un dominio similar a C-terminal RuvC, que es el único dominio funcional caracterizado. Segundo, una región de hélice alfa N-terminal y tercero una región mixta alfa y beta, localizada entre el dominio similar a RuvC y la región de hélice alfa. Se predicen varios tramos pequeños de regiones no estructuradas dentro de la estructura primaria de Cpf1. Las regiones no estructuradas, que se exponen al solvente y que no están conservadas dentro de los diferentes ortólogos de Cpf1, pueden representar lugares preferidos para las divisiones (Figura 2 y Figura 3).

La siguiente tabla presenta regiones de división potencial no limitantes dentro de As y LbCpf1. Un sitio de división dentro de una región tal puede ser oportuno.

Región de división	**AsCpf1**	**LbCpf1**
1	575-588	566-571
2	631-645	754-757
3	653-664	-
4	818-844	-

Para los mutantes Fn, As y Lb Cpf1, es altamente evidente que la correspondiente posición para un sitio potencial de división es, por ejemplo, en base a un alineamiento de secuencia. Para las enzimas distintas a Fn, As y Lb, uno puede usar la estructura cristalina de un ortólogo si es que existe un grado de homología relativamente alto entre el ortólogo y la Cpf1 de interés, o sino uno pude usar predicción computacional.

Idealmente, la posición de división se debe localizar dentro de una región o bucle. Preferiblemente, la posición de división está en un lugar en donde una interrupción de la secuencia de aminoácidos no da como resultado la destrucción parcial o completa de una característica estructural (por ejemplo hélices alfa u hojas beta). Las regiones no estructuradas (regiones que no aparecen en la estructura cristalina debido a que estas regiones no tienen la suficiente estructura para ser "congeladas" en un cristal) a menudo son las opciones preferidas. Los Solicitantes por ejemplo pueden hacer divisiones en regiones no estructuradas que se exponen a la superficie de Cpf1.

Los Solicitantes pueden seguir los siguientes procedimientos que se proveen como ejemplo preferido y como guía. Debido a que las regiones no estructuradas no aparecen en la estructura cristalina, los Solicitantes hacen una referencia cruzada de la secuencia de aminoácidos que rodea al cristal con la secuencia de aminoácidos primaria de la Cpf1. Cada región no estructurada puede hacerse de por ejemplo entre aproximadamente 3 y 10 aminoácidos, la cual no aparece en el cristal. Por ello los Solicitantes hacen la división entre estos aminoácidos. Para incluir más lugares de división potenciales, los Solicitantes incluyen divisiones localizadas en bucles en el exterior de Cpf1 usando los mismos criterios que con las regiones no estructuradas.

En algunas formas de realización, la posición de división está en un bucle externo de la Cpf1. En otras formas de realización preferidas, la posición de división está en una región no estructurada de la Cpf1. Una región no estructurada típicamente es un bucle externo altamente flexible cuya estructura no se puede determinar fácilmente a partir del patrón cristalino.

Una vez que la posición de división se ha identificado, se pueden diseñar construcciones adecuadas.

Típicamente, se posiciona una NES en el extremo N' terminal del aminoácido de división de la primera parte (o el extremo 5' del nucleótido que lo codifica). En este caso, se posiciona una NLS en el extremo C' terminal del aminoácido de división de la segunda parte (o el extremo 3' del nucleótido que lo codifica). De esta forma, la primera construcción de fusión de Cpf1 puede estar operativamente conectada a una o más señales de exportación nuclear y la segunda construcción de fusión de Cpf1 puede estar operativamente conectada a una señal de localización nuclear.

Por supuesto, se puede proveer el arreglo inverso, en donde se posiciona una NLS en el extremo N' terminal del aminoácido de división de la primera parte (o el extremo 5' del nucleótido que lo codifica). En este caso, se posiciona una NES en el extremo C' terminal del aminoácido de división de la segunda parte (o el extremo 3' del nucleótido que lo codifica). Por consiguiente, la primera construcción de fusión de Cpf1 puede estar operativamente conectada a una o más señales de localización nuclear y la segunda construcción de fusión de Cpf1 puede estar operativamente conectada a una señal de exportación nuclear.

Las proteínas dobles que mantienen las dos partes (a cada lado de la división) con prácticamente la misma longitud pueden ser ventajosas para propósitos de empaquetamiento. Por ejemplo, se cree que es m\as fácil mantener la estequiometría entre ambas piezas cuando los transcriptos tienen aproximadamente el mismo tamaño.

En ciertos ejemplos, las piezas N- y C-terminales de la Cpf1 humana optimizada por codones tal como FnCpf1 se fusionan a los dominios de dimerización FRB y FKBP, respectivamente. Este arreglo puede ser el preferido. Los mismos se pueden intercambiar (es decir N' terminal con FKBP y C' terminal con FRB).

Preferentemente se usan conectores tales como $(GGGGS)_3$ en la presente para separar el fragmento de la Cpf1 del dominio de dimerización. Se prefiere a $(GGGGS)_3$ debido a que el mismo es un conector relativamente largo (15 aminoácidos). Los residuos de glicina son los más flexibles y los residuos de serina aumentan la posibilidad de que el conector esté en el exterior de la proteína. $(GGGGS)_6$ $(GGGGS)_9$ o $(GGGGS)_{12}$ pueden usarse preferiblemente como alternativas. Otras alternativas preferidas son $(GGGGS)_1$, $(GGGGS)_2$, $(GGGGS)_4$, $(GGGGS)_5$, $(GGGGS)_7$, $(GGGGS)_8$, $(GGGGS)_{10}$, o $(GGGGS)_{11}$.

Por ejemplo, se puede incluir $(GGGGS)_3$ entre el fragmento de Cpf1 N' terminal y FRB. Por ejemplo, se puede incluir $(GGGGS)_3$ entre FKB y el fragmento de Cpf1 C' terminal.

Hay disponibles conectores alternativos, pero se cree que los conectores altamente flexibles funcionan mejor permitiendo un máximo de oportunidades para que las 2 partes de la Cpf1 se junten y por consiguiente reconstituyan la actividad de Cpf1. Una alternativa es que la NLS de nucleoplasmina se pueda usar como conector.

También se puede usar un conector entre la Cpf1 y cualquier dominio funcional. Otra vez, se puede usar el conector $(GGGGS)_3$ aquí (o las versiones de 6, 9, o 12 repeticiones) o se puede usar la NLS de nucleoplasmina como conector entre la CPf1 y el dominio funcional.

Se prevén alternativas para el sistema FRB/FKBP. Por ejemplo el sistema de ABA y giberelina.

Por lo tanto, los ejemplos preferidos de la familia FKBP son cualquiera de los siguientes sistemas inducibles. FKBP que se dimeriza con CalcineurinA (CNA), en presencia de FK506; FKBP que se dimeriza con CyP-Fas, en presencia de FKCsA; FKBP que se dimeriza con FRB, en presencia de Rapamicina; GyrB que se dimeriza con GryB, en presencia de Coumermicina; GAI que se dimeriza con GID1, en presencia de Giberelina; o Snap-tag que se dimeriza con HaloTag, en presencia de HaXS.

Las alternativas dentro de la familia de FKBP misma son también preferidas. Por ejemplo, FKBP, que se homodimeriza (es decir una FKBP se dimeriza con otra FKBP) en presencia de FK1012. Por consiguiente, también se provee un sistema inducible Cpf1 CRISPR-Cas no natural o modificado, que comprende:

una primera construcción de fusión de Cpf1 unida a una primera mitad de un homodímero inducible y

una segunda construcción de fusión de Cpf1 unida a una segunda mitad del homodímero inducible,

en donde la primera construcción de fusión de Cpf1 está operativamente conectada a una o más señales de localización nuclear,

en donde la segunda construcción de fusión de Cpf1 está operativamente conectada a una (opcionalmente una o más) señal(es) de exportación nuclear,

en donde el contacto con una fuente de energía inductora junta la primera mitad y la segunda mitad del homodímero inducible,

en donde juntar la primera mitad y la segunda mitad del homodímero inducible les permite a la primera y segunda construcción de fusión de Cpf1 constituir un sistema funcional Cpf1 CRISPR-Cas,

en donde el sistema Cpf1 CRISPR-Cas comprende un ARN guía (ARNg) que comprende una secuencia guía capaz de hibridar con una secuencia diana en un locus genómico de interés en una célula, y

en donde el sistema funcional Cpf1 CRISPR-Cas se une a la secuencia diana y, opcionalmente, edita el locus genómico para alterar la expresión génica.

En una forma de realización, el homodímero preferiblemente es FKBP y la fuente de energía inductora es preferiblemente FK1012. En otra forma de realización, el homodímero preferiblemente es GryB y la fuente de energía inductora es preferiblemente Coumermicina. En otra forma de realización, el homodímero preferiblemente es ABA y la fuente de energía inductora es preferiblemente Giberelina.

En otras formas de realización, el dímero es un heterodímero. Los ejemplos preferidos de heterodímeros son cualquiera de los siguientes sistemas inducibles: FKBP que se dimeriza con CalcineurinA (CNA), en presencia de FK506; FKBP que se dimeriza con CyP-Fas, en presencia de FKCsA; FKBP que se dimeriza con FRB, en presencia de Rapamicina, en presencia de Coumermicina; GAI que se dimeriza con GID1, en presencia de Giberelina; o Snap-tag que se dimeriza con HaloTag, en presencia de HaXS.

Los solicitantes usaron FKBP/FRB debido a que está bien caracterizado y ambos dominios son lo suficientemente pequeños (<100 aminoácidos) para ayudar al empaquetamiento. Aún más, se ha usado rapamicina durante un largo tiempo y sus efectos secundarios se comprenden bien. Los dominios de dimerización grandes (>300 aa) también deberían funcionar, pero pueden requerir conectores más largos para hacer posible la reconstitución de Cpf1.

Paulmurugan y Gambhir (Cancer Res, agosto 15, 2005 65; 7413) exponen las características de fondo del sistema FRB/FKBP/Rapamicina. Otro artículo útil es el artículo de Crabtree y col., (Chemistry & Biology 13, 99-107, enero 2006).

En un ejemplo, se construye un vector individual, un casete de expresión (plásmido). El ARNg está bajo el control de un promotor U6. Se usan dos Cpf1 dobles diferentes. La construcción de Cpf1 doble se basa en la primera construcción de fusión de Cpf1, flanqueada por NLS, con FKBP fusionada a la parte C-terminal de la CPf1 doble a través de un conector GlySer; y una segunda construcción de fusión de CPf1, flanqueada por NES, con FRB fusionada con la parte N-terminal de la CPf1 doble a través de un conector GlySer. Para separar la primera y la segunda construcción de fusión de Cpf1, se usa P2A que se divide durante la transcripción. La Cpf1 doble muestra una formación de indeles similar a la de tipo salvaje en presencia de rapamicina, pero una formación de indeles marcadamente inferior que la de tipo salvaje en ausencia de rapamicina.

Por lo tanto, se provee un vector individual. El vector comprende:

una primera construcción de fusión de Cpf1 unida a una primera mitad de un dímero inducible y

una segunda construcción de fusión de Cpf1 unida a una segunda mitad del dímero inducible,

en donde la primera construcción de fusión de Cpf1 está operativamente conectada a una o más señales de localización nuclear,

en donde la segunda construcción de fusión de Cpf1 está operativamente conectada a una o más señales de exportación nuclear,

en donde el contacto con una fuente de energía inductora pone juntas a la primera y a la segunda mitad del heterodímero inducible,

en donde poner juntas a la primera y a la segunda mitad del heterodímero inducible les permite a la primera y segunda construcción de fusión de Cpf1 constituir un sistema funcional Cpf1 CRISPR-Cas,

en donde el sistema Cpf1 CRISPR-Cas comprende un ARN guía (ARNg) que comprende una secuencia guía capaz de hibridar con una secuencia diana en un locus genómico de interés en una célula, y

en donde el sistema funcional Cpf1 CRISPR-Cas se une a la secuencia diana y, opcionalmente, edita el locus genómico para alterar la expresión génica. Estos elementos se proveen preferiblemente en una construcción individual, por ejemplo, un casete de expresión.

La primera construcción de fusión de Cpf1 preferiblemente está flanqueada por al menos una señal de localización nuclear en cada extremo. La segunda construcción de fusión de Cpf1 preferiblemente está flanqueada por al menos una señal de exportación nuclear en cada extremo.

También se provee un método para tratar a un sujeto que lo necesite, que comprende inducir edición génica mediante la transformación del sujeto con el polinucleótido que codifica para el sistema o cualquiera de los presentes vectores y administrar una fuente de energía inductora al sujeto. También se puede proveer un molde de reparación adecuado, por ejemplo, administrado mediante un vector que comprende dicho molde de reparación.

También se provee un método para tratar a un sujeto que lo necesite, que comprende inducir la activación o represión de la transcripción mediante la transformación del sujeto con el polinucleótido que codifica para el presente sistema o cualquiera de los presentes vectores, en donde dicho polinucleótido o vector codifica o comprende la Cpf1 catalíticamente inactiva y uno o más asociadas dominios funcionales; el método que además comprende administrar

una fuente de energía inductora al sujeto.

También se proveen composiciones que comprenden el presente sistema para su uso en dicho método de tratamiento. También se provee el uso del presente sistema en la fabricación de un medicamento para dichos métodos de tratamiento.

Los ejemplos de condiciones tratables mediante el presente sistema se describen en la presente o en documentos citados en la presente.

El vector individual puede comprender un agente de división de transcriptos, por ejemplo, P2A. La P2A divide al transcripto en dos, para separar la primera y la segunda construcción de fusión de CPf1. La división es debía a un "salteo ribosomal". En esencia, el ribosoma se saltea un aminoácido durante la traducción, lo que rompe la cadena proteica y da como resultado dos polipéptidos/proteínas separadas. El vector individual también es útil para aplicaciones en donde una baja actividad de fondo no es un problema, pero donde se desea una alta actividad inducible.

Un ejemplo sería la generación de líneas de células madres embrionarias no humanas clonales. El procedimiento normal es una transfección transitoria con plásmidos que codifican para las CPf1 wt o Cpf1. Estos plásmidos producen moléculas de Cpf1, que se mantienen activas durante varios días y tienen muchas posibilidades de actividad fuera de diana. Usando el vector de expresión individual para Cpf1 doble, se permite restringir la "alta" actividad de Cpf1 durante una ventana de tiempo más corta (por ejemplo, una dosis de un inductor, tal como rapamicina). Sin tratamientos con inductor (por ejemplo, rapamicina) continuo (diario) la actividad de los vectores de expresión individuales de Cpf1 doble es baja y presenta una menor posibilidad de causar efectos no deseados fuera de diana.

Un pico de actividad de Cpf1 inducida es beneficioso en algunas formas de realización y se puede producir muy fácilmente usando un vector de suministro individual, pero también es posible a través de un sistema de vectores duales (cada vector entrega una mitad de la CPf1 doble). El pico puede tener una alta actividad y durante un periodo corto de tiempo, típicamente durante la vida útil del inductor.

Por lo tanto, se provee un método para la generación de líneas de células madre embrionarias no humanas clonales, que comprende transfectar una o más célula madre embrionarias no humanas con un polinucleótido que codifica para el presente sistema o uno de los presentes vectores para expresar la presente Cpf1 doble y administrar o poner en contacto la una o más células madres con la presente fuente de energía inductora para inducir la reconstitución de la Cpf1. Se puede proveer un molde de reparación.

Como con todos los métodos que se describen en la presente, se podrá apreciar que se requerirán ARNg o guides adecuados.

Cuando los dominios funcionales y semejantes se "asocian" con una u otra parte de la enzima, estas son típicamente fusiones. El término "asociadas con" se usa aquí en relación a que una molécula 'se asocia' con respecto a otra, por ejemplo, entre partes de la Cpf1 y un dominio funcional. En el caso de dichas interacciones proteína-proteína, esta asociación se puede ver en términos de reconocimiento en la vía de que un anticuerpo reconoce a un epítopo. Como alternativa, una proteína se puede asociar con otra proteína a través de una fusión de las dos, por ejemplo, una subunidad fusionada a otra subunidad. La fusión típicamente ocurre por adición de la secuencia de aminoácidos de una a la de la otra, por ejemplo. a través de corte y empalme de las secuencias de nucleótidos que codifican para cada proteína o subunidad. Como alternativa, esto se puede ver esencialmente como la unión o conexión directa entre dos moléculas, tal como una proteína de fusión. En cualquier caso, la proteína de fusión puede incluir un conector entre las dos subunidades de interés (es decir entre la enzima y el dominio funcional o entre la proteína adaptadora y el dominio funcional). Por consiguiente, en algunas formas de realización, la parte de la CPf1 se asocia con un dominio funcional mediante unión al mismo. En otras formas de realización, la CPf1 se asocia con un dominio funcional debido a que los dos se fusionan juntos, opcionalmente a través de un conector intermediario. Los ejemplos de conectores incluyen a los conectores GlySer que se exponen en la presente.

Otros ejemplos de inductores incluyen a luz y hormonas. Para la luz, los dímeros inducibles pueden ser heterodímeros e incluyen una primera mitad inducible por luz de un dímero y una segunda (y complementaria) mitad inducible por luz de un dímero. Un ejemplo preferido de primera y segunda mitad inducibles por luz es el sistema de CIB1 y CRY2. El dominio CIB1 es un compañero de unión heterodimérica de la Criptocromo 2 sensible a la luz (CRY2).

En otro ejemplo, el sistema de dimerización Magnet de respuesta a la luz azul (pMag y nMag) se puede fusionar a las dos partes de una proteína Cpf1 doble. En respuesta a la estimulación por luz, pMag y nMag se dimerizan y se reensambla la Cpf1. Por ejemplo, dicho sistema se describe en conexión con Cas9 en Nihongaki y col., (Nat. Biotechnol. 33, 755–790, 2015).

La invención comprende que la fuente de energía inductora puede ser calor, ultrasonido, energía electromagnética o química. En una forma de realización preferida de la invención, la fuente de energía inductora puede ser un antibiótico, una molécula pequeña, una hormona, un derivado de hormona, un esteroide o un derivado de esteroide. En una forma de realización más preferida, la fuente de energía inductora puede ser ácido abscísico (ABA), doxiciclina (DOX), cumato, rapamicina, 4-hidroxitamoxifeno (4OHT), estrógeno o ecdisona. La invención provee que el por lo menos un

interruptor se puede seleccionar del grupo que consiste en antibióticos basados en sistemas inducibles, sistemas inducibles basados en energía electromagnética, sistemas inducibles basados en moléculas pequeñas, sistemas inducibles basados en receptores nucleares y sistemas inducibles basados en hormona. En una forma de realización más preferida el por lo menos un interruptor se puede seleccionar del grupo que consiste en sistemas inducibles por tetraciclina (Tet)/DOX, sistemas inducibles por luz, sistemas inducibles por ABA, sistemas de represor/operador de cumato, sistemas inducibles por 4OHT/estrógeno, sistemas inducibles basados en ecdisona y sistemas inducibles por FKBP12/FRAP (complejo de FKBP12-rapamicina). Dichos inductores también se exponen en la presente y en PCT/US2013/051418.

En general, cualquier uso que se pueda hacer de una Cpf1, ya sea wt, nickasa o una Cpf1 muerta (con o sin dominios funcionales asociadas) se puede realizar usando el presente abordaje de Cpf1 doble. El beneficio mantiene la naturaleza inducible de la actividad de la Cpf1.

Como otro ejemplo, se pueden hacer fusiones de CPf1 doble con proteínas fluorescentes como GFP. Esto permitiría la adquisición de imágenes de loci genómicos (véase "Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System" Chen B y col., Cell 2013), pero de una forma inducible. Como tal, en algunas formas de realización, uno o más partes de la Cpf1 se pueden asociar (y en particular fusionar) una proteína fluorescente, por ejemplo, GFP.

Otros experimentos investigan si existe una diferencia en el corte fuera de diana entre de la Cpf1 tipo salvaje (wt) y la doble, cuando el corte fuera de diana es al mismo nivel. Para hacer esto, los Solicitantes usan transfección transitoria de plásmidos con Cpf1 wt y doble y cosechan a diferentes puntos de tiempo. Los Solicitantes buscan una activación fuera de diana después de hallar un grupo de muestras en donde el corte fuera de diana está dentro de +/- 5%. Los Solicitantes elaboran líneas celulares con expresión estable de Cpf1 wt o doble sin guías (usando lentivirus). Después de la selección con antibióticos, se administran las guías con un lentivirus separado y se cosecha a diferentes puntos de tiempo para medir el corte en el diana o fuera de diana.

Los Solicitantes introducen una secuencia desestabilizante (PLAGA, véase "Use of mRNA- y protein-destabilizing elements to develop a highly responsive reporter system" Voon DC y col. Nucleic Acids Research 2005) en el fragmento FRB(N)Cpf1-NES para facilitar una degradación más rápida y por ello una estabilidad reducida del complejo Cpf1 doble muerta-VP64.

Dichas secuencias desestabilizantes como se describen en otra parte en esta descripción (incluyendo PLAGA) pueden ser ventajosas para usar con los sistemas de Cpf1 doble.

Se generan líneas celulares que expresan establemente Cpf1 doble muerta-VP64 y MS2-p65-HSF1 + guía. Un cribado por resistencia a PLX puede demostrar que una activación de la transcripción controlada por tiempo, no reversible, puede ser útil para cribado de fármacos. Este abordaje puede ser ventajoso cuando la Cpf1 doble muerta-VP64 no es reversible.

En un aspecto los métodos y composiciones reivindicados de la invención pueden implicar el uso de un sistema no natural o modificado Cpf1 CRISPR-Cas que puede comprender por lo menos un interruptor en donde la actividad de dicho Sistema Cpf1 CRISPR-Cas se controla mediante el contacto con por lo menos una fuente de energía inductora actuando como interruptor. En una forma de realización de la invención el control con por lo menos un interruptor o la actividad de dicho Sistema Cpf1 CRISPR-Cas se puede activar, aumentar, terminar o reprimir. El contacto con la por lo menos una fuente de energía inductora puede dar como resultado un primer efecto y un segundo efecto. El primer efecto puede ser uno o más de importación nuclear, exportación nuclear, reclutamiento de un componente secundario (tal como una molécula efectora), cambio conformacional (de proteína, ADN o ARN), escisión, liberación de carga (tal como una molécula o un cofactor encerrado), asociación o disociación. El segundo efecto puede ser uno o más de activación, potenciación, terminación o represión del control con el por lo menos un interruptor o de la actividad de dicho Sistema Cpf1 CRISPR-Cas. En una forma de realización el primer efecto y el segundo efecto pueden ocurrir en cascada.

En otro aspecto de los métodos y composiciones reivindicados de la invención, el sistema Cpf1 CRISPR-Cas puede comprender además por lo menos uno o más de señal de localización nuclear (NLS), señal de exportación nuclear (NES), dominio funcional, flexible conector, mutación, supresión, alteración o truncación. El uno o más de la NLS, la NES o el dominio funcional se pueden activar o inactivar en forma condicional. En otra forma de realización, la mutación puede ser una o más de una mutación en una región de homología de factor de transcripción, a una mutación en un dominio de unión a ADN (tal como la mutación de residuos básicos de una hélice bucle hélice básico), una mutación en una NLS endógena o una mutación en una NES endógena. La invención comprende que la fuente de energía inductora puede ser calor, ultrasonido, energía electromagnética o química. En una forma de realización preferida de la invención, la fuente de energía inductora puede ser un antibiótico, una molécula pequeña, una hormona, un derivado de hormona, un esteroide o un derivado de esteroide. En una forma de realización más preferida, la fuente de energía inductora puede ser ácido abscísico (ABA), doxiciclina (DOX), cumato, rapamicina, 4-hidroxitamoxifeno (4OHT), estrógeno o ecdisona. La invención provee que el por lo menos un interruptor se puede seleccionar del grupo que consiste en sistemas inducibles basados en antibióticos, sistemas inducibles basados en energía electromagnética, sistemas inducibles basados en moléculas pequeñas, sistemas inducibles basados en receptores nucleares y sistemas inducibles basados en hormonas. En una forma de realización más preferida el por lo menos un interruptor se puede

seleccionar del grupo que consiste en sistemas inducibles por tetraciclina (Tet)/DOX, sistemas inducibles por luz, sistemas inducibles por ABA, sistemas de represor/operador de cumato, sistemas inducibles por 4OHT/estrógeno, sistemas inducibles basados en ecdisoma y sistemas inducibles por FKBP12/FRAP (complejo FKBP12-rapamicina).

Los aspectos del control tal como se detallan en esta solicitud se relacionan con por lo menos uno o más interruptor(es). El término "interruptor" según se usa en la presente se refiere a un sistema o un grupo de componentes que actúan en forma coordinada para efectuar un cambio, abarcando todos los aspectos de función biológica tales como activación, represión, potenciación o terminación de esa función. En un aspecto el término interruptor abarca interruptores genéticos que comprenden los componentes básicos de proteínas de regulación génica y las secuencias de ADN específicas que estas proteínas reconocen. En un aspecto, los interruptores se relacionan con sistemas inducibles y reprimibles como usados en la regulación de genes. En general, un sistema inducible puede estar apagado a menos que existe la presencia de alguna molécula (llamada inductor) que permite la expresión génica. Se dice que la molécula "induces la expresión". La forma en que esto sucede depende de los mecanismos de control, así como de diferencias de tipo celular. Un sistema reprimible está encendido excepto en presencia de alguna molécula (llamada corepresor) que suprime la expresión génica. Se dice que la molécula "reprime la expresión". La forma en que esto sucede depende de los mecanismos de control, así como de diferencias de tipo celular. El término "inducible" según se usa en la presente puede abarcar todos los aspectos de un interruptor, independientemente del mecanismo molecular implicado. Por lo tanto, un interruptor tal como se interpreta en la invención pueden incluir, pero en un sentido no taxativo, a sistemas inducibles basados en antibióticos, sistemas inducibles basados en energía electromagnética, sistemas inducibles basados en moléculas pequeñas, sistemas inducibles basados en receptores nucleares y sistemas inducibles basados en hormonas. En formas de realización preferidas el interruptor puede ser un sistema inducible por tetraciclina (Tet)/DOX, un sistema inducible por luz, un sistema inducible por ácido abscísico (ABA), un sistema de represor/operador de cumato, un sistema inducible por 4OHT/estrógeno, un sistema inducible basado en ecdisona o un inducible sistema por FKBP12/FRAP (complejo FKBP12-rapamicina).

El presente sistema Cpf1 CRISPR-Cas se puede diseñar para modular o alterar la expresión de genes endógenos individuales en una forma espacial y temporal precisa. El sistema Cpf1 CRISPR-Cas se puede diseñar para unirse a la secuencia promotora del gen de interés para alterar la expresión génica. La Cpf1 se puede dividir en dos, en donde una mitad se fusiona a una mitad del heterodímero de criptocromo (criptocromo-2 o CIB1), mientras que el restante socio de criptocromo se fusiona con la otra mitad de la Cpf1. En algunos aspectos, también se puede incluir un dominio efector transcripcional en el sistema Cpf1 CRISPR-Cas. Los dominios efectores pueden ser activadores, tales como VP16, VP64, o p65, o represores, tales como KRAB, EnR, o SID. En estado no estimulado, una mitad de la proteína Cpf1-criptocromo2 se localiza en el promotor del gen de interés, pero no se une a la CIB1-proteína efectora. Ante la estimulación con luz de espectro azul, el criptocromo-2 se activa, sufre un cambio a conformacional, y expone su dominio de unión. La CIB1, a su vez, se une al criptocromo-2 resultando en la localización de la segunda mitad de la Cpf1 en la región promotora del gen de interés e iniciando la edición de genomas que puede dar como resultado una sobreexpresión o un silenciamiento génico. Otros aspectos de LITE se describen adicionalmente en Liu, H y col., *Science*, 2008 y Kennedy M y col., *Nature Methods* 2010.

Los dominios activador y represor que además pueden modular la función se pueden seleccionar en base a especie, fuerza, mecanismo, duración, tamaño, o cualquier número de otros parámetros. Los dominios efectores preferidos incluyen, pero en un sentido no taxativo, a dominio transposasa, dominio integrase, dominio recombinasa, dominio resolvasa, dominio invertasa, dominio proteasa, dominio ADN metiltransferasa, dominio ADN desmetilasa, dominio histona acetilasa, dominio histona desacetilasa, dominio nucleasa, dominio represor, dominio activador, dominios de localización de señal nuclear, dominio de reclutamiento de proteínas de transcripción, dominio asociados con la activación de captación celular, dominio de unión a ácidos nucleicos o dominio de presentación de anticuerpos.

Existen varios modos diferentes para generar sistemas inducibles químicos también: 1. Sistema basado en ABI-PILO inducible por Ácido Abscísico (ABA) (véase, por ejemplo, sitio de internet en stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2), 2. Sistema basado en FKBP-FRB inducible por rapamicina (o químicos relacionados basados en rapamicina) (véase, por ejemplo, sitio de internet en nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. Sistema basado en GID1-GAI inducible por Giberelina (GA) (véase, por ejemplo, sitio de internet en nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

Otro sistema contemplado por la presente invención es un sistema inducible químico basado en un cambio de localización subcelular. Los Solicitantes también divulgan un sistema inducible Cpf1 CRISPR-Cas modificado para direccionarse contra un locus genómico de interés en donde la enzima Cpf1 se divide en dos construcciones de fusión que se conectan además a diferentes partes de una proteína sensible a químico o energía. Esta proteína sensible a químico o energía conducirá a un cambio en la localización subcelular de cada mitad de la enzima Cpf1 (es decir transporte de alguna de las mitades de la enzima Cpf1 del citoplasma al el núcleo de las células) ante la unión de un químico o la transferencia de energía a la proteína sensible al químico o energía. Este transporte de construcciones de fusión de un compartimento subcelular u organela, en donde se secuestra su actividad debido a la falta de sustrato para el sistema Cpf1 CRISPR-Cas reconstituido, a otro en donde el sustrato está presente, le permitiría a los componentes estar juntos y reconstituir la actividad funcional, y luego entrar en contacto con su sustrato deseado (es decir ADN genómico en el núcleo del mamífero) y dar como resultado la activación o la represión de la expresión del gen diana.

Se contemplan otros sistemas inducibles tales como, pero en un sentido no taxativo, regulación por metales pesados [Mayo KE y col., Cell 1982, 29:99-108; Searle PF y col., Mol Cell Biol 1985, 5:1480-1489 y Brinster RL y col., *Nature* (Londres) 1982, 296:39-42], hormonas esteroides [Hynes NE y col., *Proc Natl Acad Sci USA* 1981, 78:2038-2042; Klock G y col., *Nature* (Londres) 1987, 329:734-736 y Lee F y col., *Nature* (Londres) 1981, 294:228-232.], choque térmico [Nouer L:Heat Shock Response. Boca Raton, FL:CRC; 1991] y otros reactivos que se han desarrollado [Mullick A, Massie B:Transcription, translation and the control of gen expression. En Encyclopedia de Cell Tecnology Editado por: Speir RE. Wiley; 2000:1140-1164 y Fussenegger M., Biotechnol Prog 2001, 17:1-51]. Sin embargo, existen limitaciones con estos promotores inducibles de mamífero tales como "pérdidas" de estado "apagado" y efectos pleiotrópicos de los inductores (choque térmico, metales pesados, glucocorticoides etc.). Se ha propuesto el uso de hormonas de insecto (ecdisona) en un intento por reducir la interferencia con procesos celulares en células de mamífero [No D y col., *Proc Natl Acad Sci USA* 1996, 93:3346-3351]. Otro sistema elegante usa rapamicina como inductor [Rivera VM y col., Nat Med 1996, 2:1028-1032] pero el role de la rapamicina como inmunosupresor fue una limitación importante para su uso in vivo y por ello fue necesario hallar un compuesto biológicamente inerte [Saez E y col., *Proc Natl Acad Sci USA* 2000, 97:14512-14517] para el control de la expresión génica.

En formas de realización particulares, el sistema de edición génica que se describe en la presente se coloca bajo el control de un interruptor de muerte con contraseña, que es un mecanismo que mata eficazmente la célula huésped cuando se alteran las condiciones de la célula. Esto se asegura mediante la introducción de factores de transcripción híbridos de la familia LacI-GalR, que requieren la presencia de IPTG para activarse (Chan y col., 2015 *Nature Nature Chemistry Biology* doi:10.1038/nchembio.1979 que se puede usar para dirigir a un gen que codifica para una enzima crítica para la supervivencia celular. Mediante la combinación de diferentes factores de transcripción sensibles a diferentes químicos, se puede generar un "código". Este sistema se puede usar para control espacial y temporalmente la extensión de las modificaciones genéticas inducidas por CRISPR, lo que puede ser de interés en diferentes campos incluyendo aplicaciones terapéuticas y también puede ser de interés para evitar el "escape" de las GMO de su entorno pretendido.

Sistemas autoinactivantes

Una vez que se han editado todas las copias de un gen en el genoma de una célula, la expresión continuada de CRISPRP/Cpf1p es esta célula no es más necesaria. De hecho, la expresión sostenida sería indeseable en el caso de los efectos fuera de diana en sitios genómicos no pretendidos, *etc.* Por lo tanto sería útil una expresión limitada en el tiempo. La expresión inducible ofrece un abordaje, pero además los Solicitantes prevén un sistema de CRISPR-Cpf1 autoinactivante que se basa en el uso de una secuencia diana guía no codificante dentro del vector mismo de CRISPR. De esta manera, tras comenzar la expresión, el sistema CRISPR conducirá a su propia destrucción, pero antes de que se complete la destrucción tendrá tiempo de editar las copias genómicas del gen diana (que, con una mutación puntual normal en una célula diploide, requiere cuando menos dos ediciones). Sencillamente, el sistema CRISPR-Cas autoinactivante incluye ARN adicional (es decir, ARN guía) que dirige la secuencia de codificación para la propia enzima CRISP o que dirige una o más secuencias guía no codificantes complementarias a las secuencias individuales presentes en una o más de los siguientes:

(a) dentro del promotor, lo que impulsa la expresión de los elementos de ARN no codificantes,

(b) dentro del promotor conduciendo la expresión del gen de Cpf1,

(c) dentro de las 100 pb desde el ATG del codón de comienzo de traducción en la secuencia codificante de Cpf1,

(d) dentro de la repetición terminal invertida (iTR) de un vector de suministro viral, por ejemplo, en el genoma de AAV.

Además, este ARN puede suministrarse mediante un vector, por ejemplo, un vector separado o el mismo vector que codifica para el complejo CRISPR. Cuando se provee mediante un vector separado, el ARN de CRISPR que dirige la expresión de Cpf1 puede administrase secuencial o simultáneamente. Cuando se administra secuencialmente, el ARN de CRISPR que dirige la expresión de Cpf1 ha de suministrarse después del ARN de CRISPR que se pretende para, por ejemplo, editar o modificar el gen. Este periodo puede ser un periodo de minutos (por ejemplo, 5 minutos, 10 minutos, 20 minutos, 30 minutos, 45 minutos, 60 minutos). Este periodo puede ser un periodo de horas (por ejemplo, 2 horas, 4 horas, 6 horas, 8 horas, 12 horas, 24 horas). Este periodo puede ser un periodo de días (por ejemplo, 2 días, 3 días, 4 días, 7 días). Este periodo puede ser un periodo de semanas (por ejemplo, 2 semanas, 3 semanas, 4 semanas). Este periodo puede ser un periodo de meses (por ejemplo, 2 meses, 4 meses, 8 meses, 12 meses). Este periodo puede ser un periodo de años (2 años, 3 años, 4 años). De esta forma, la enzima Cas se asocia con un primer ARNg capaz de hibridar con un primer diana, tal como un locus o loci genómicos de interés y se encarga de la o las funciones deseadas del sistema CRISPR-Cas (por ejemplo, la modificación de genes); y subsiguientemente la enzima Cpf1 se puede asociar entonces con el segundo ARNg capaz de hibridar con la secuencia que comprende por lo menos parte del casete Cpf1 o CRISPR. Cuando el ARNg se direcciona contra las secuencias que codifican para la expresión de la proteína Cpf1, la enzima se vuelve inactiva y el sistema se autoinactiva. De la misma manera, el ARN de CRISPR que dirige la expresión de Cpf1 se aplica mediante, por ejemplo, liposomas, lipofección, nanopartículas, microvesículas como se explica en la presente, se puede administrar secuencial o simultáneamente. De forma similar, se puede usar la autoinactivación para la inactivación de uno o más ARN guía usados para dirigirse a uno o más blancos.

En algunos aspectos, se proporciona un ARNg individual que es capaz de hibridarse a una secuencia en la dirección 3' de un codón de inicio de la enzima CRISPR, por lo cual, tras un periodo de tiempo existe una pérdida de expresión de la enzima CRISPR. En algunos aspectos, se proporcionan uno o más ARNg que son capaces de hibridación a una o más regiones codificantes o no codificantes del polinucleótido que codifica el sistema CRISPR-Cas, por lo cual, tras un periodo de tiempo existe la inactivación de uno o más, o en algunos casos todos, los sistemas CRISPR-Cas. En algunos aspectos del sistema, y no para quedar limitado por teoría alguna, la célula puede comprender una pluralidad de complejos CRISPR-Cas, en donde un primer subconjunto de complejos CRISPR comprende un primer ARNg capaz de dirigirse a un locus genómico o a loci genómicos que se van a editar, y un segundo subconjunto de complejos CRISPR comprende al menos un segundo ARNg capaz de dirigirse al polinucleótido que codifica el sistema CRISPR-Cas, en donde el primer subconjunto de complejos CRISPR-Cas media en la edición del locus genómico o de los loci genómicos blancos y el segundo subconjunto de complejos CRISPR inactiva eventualmente el sistema CRISPR-Cas, inactivando por tanto la expresión de CRISPR-Cas adicional en la célula.

De esta manera, los métodos y composiciones reivindicados de la invención pueden englobar proveer un sistema CRISPR-Cas que comprende uno o más vectores para el suministro a una célula eucariota, donde el vector o los vectores codifican: (i) una enzima CRISPR, más particularmente Cpf1; (ii) un primer ARN guía capaz de hibridar con una secuencia diana en la célula; y (iii) un segundo ARN guía capaz de hibridarse con uno o más secuencia dianas en el vector que codifica para la enzima CRISPR, Cuando se expresa en la célula, el primer ARN guía dirige la unión específica de secuencia de un primer complejo CRISPR a la secuencia diana en la célula; el segundo ARN guía dirige la unión específica de secuencia de un segundo complejo CRISPR a la secuencia diana en el vector que codifica para la enzima CRISPR; los complejos CRISPR comprenden una enzima CRISPR unida a un ARN guía, mediante lo cual un ARN guía se puede hibridar con su secuencia diana; y el segundo complejo CRISPR inactiva al sistema CRISPR-Cas para prevenir una expresión continua de la enzima CRISPR por parte de la célula.

Otras características de los vectores, la enzima codificada, las secuencias guía, *etc.* se divulgan en otra parte en la presente. El sistema puede codificar para (i) una enzima CRISPR, más particularmente Cpf1; (ii) un primer ARNg que comprende una secuencia capaz de hibridarse con una primera secuencia diana en la célula, (iii) un segundo ARN guía capaz de hibridarse con el vector que codifica para la enzima CRISPR. De manera similar, la enzima puede incluir uno o más NLS, *etc.*

Las diversas secuencias codificantes (enzima CRISPR, ARN guías) se pueden incluir en un vector individual o en múltiples vectores. Por ejemplo, es posible codificar la enzima en un vector y las diversas secuencias de ARN en otro vector, o codificar la enzima y un ARNg en un vector, y el ARNg restante en otro vector, o cualquier otra permutación. En general, se prefiere un sistema que utiliza un total de uno o dos vectores diferentes.

Cuando se usan múltiples vectores, es posible suministrarlos en número desigual, e idealmente con un exceso de un vector que codifica el primer ARN guía con respecto al segundo ARN guía, ayudando por tanto al retraso final de la inactivación del sistema CRISPR hasta que la edición del genoma ha tenido la oportunidad de producirse.

El primer ARN guía puede dirigir cualquier secuencia de interés en el genoma como se describe en otra parte en la presente. El segundo ARN guía se direcciona contra una secuencia en el vector que codifica la enzima Cas9 de CRISPR, e inactiva por tanto la expresión de la enzima procedente de este vector. De esta manera, la secuencia diana en el vector debe ser capaz de inactivar la expresión. Las secuencias diana adecuadas pueden estar, por ejemplo, próximas o comprendidas en el codón de inicio de la traducción para la secuencia que codifica Cpf1, en una secuencia no codificante en el promotor que impulsa la expresión de elementos de ARN no codificantes, en el promotor que impulsa la expresión del gen Cpf1, dentro de los 100 pb del codón ATG de inicio de la traducción en la secuencia que codifica Cpf1, y/o en la repetición terminal invertida (iTR) de un vector de suministro viral, por ejemplo, en el genoma de AAV. Una rotura bicatenaria próxima a esta región puede inducir un desplazamiento del marco en la secuencia que codifica Cpf1, produciendo una pérdida de expresión de la proteína. Una secuencia diana alternativa para el ARN guía "autoinactivante" tendría como objetivo editar/inactivar las regiones/secuencias reguladoras necesarias para la expresión del sistema CRISPR-Cpf1 o para la estabilidad del vector. Por ejemplo, si el promotor de la secuencia de codificación de Cpf1 está perturbado, entonces se puede inhibir o evitar la transcripción. De forma similar, si un vector incluye secuencias para la replicación, el mantenimiento o la estabilidad, entonces es posible dirigir estas. Por ejemplo, en un vector de AAV la secuencia diana útil está en la iTR. Otras secuencias útiles para el direccionamiento pueden las secuencias promotoras, los sitios de poliadenilación, etc.

Además, si los ARN guía se expresan en formato de matriz, los ARN guía "autoinactivantes" que se dirigen a ambos promotores simultáneamente darán como resultado la escisión de los nucleótidos intermedios de la construcción de expresión CRISPR-Cas, conduciendo eficazmente a su completa inactivación. De forma similar, la escisión de los nucleótidos intervinientes dará resultado cuando los ARN guía se dirigen a ambas iTR, o se dirigen a dos o más componentes de CRISPR-Cas diferentes simultáneamente. La autoinactivación como se explica en la presente es aplicable, en general, con sistemas CRISPR-Cpf1 con el objetivo de proveer una regulación de CRISPR-Cpf1. Por ejemplo, la autoinactivación como se explica en la presente puede aplicarse a la reparación CRISPR de mutaciones, por ejemplo, trastornos de extensión, como se explica en la presente. Como resultado de esta autoinactivación, la reparación de CRISPR está solo transitoriamente activa.

La adición de nucleótidos de no direccionamiento al extremo 5' (por ejemplo, entre 1 y 10 nucleótidos, preferiblemente

entre 1 y 5 nucleótidos) del ARN guía "autoinactivante" se puede usar para retrasar su procesamiento y/o modificar su eficacia como un medio para asegurar la edición en el locus genómico diana antes de la inactivación de CRISPR-Cpf1.

En un aspecto de la autoinactivación del sistema AAV-CRISPR-Capf1, los plásmidos que expresan simultáneamente uno o más ARNg contra secuencias genómicas de interés (por ejemplo, 1-2, 1-5, 1-10, 1 -15, 1-20, 1-30) pueden establecerse con ARNg "autoinactivantes" que se dirigen a una secuencia SpCas9 en, o cerca del sitio de inicio ATG manipulado (por ejemplo, dentro de los 5 nucleótidos, dentro de los 15 nucleótidos, dentro de los 30 nucleótidos, dentro de los 50 nucleótidos, dentro de los 100 nucleótidos). Una secuencia reguladora en la región del promotor U6 puede ser diana también de un ARNsg. Los ARNg impulsados por U6 pueden diseñarse en un formato de matriz de tal manera que se puedan liberar simultáneamente múltiples secuencias de ARNg. Cuando se administran primero a los tejidos/células diana (célula izquierda) los ARNg comienzan a acumularse mientras que los niveles de Cpf1 se elevan en el núcleo. La Cpf1 se compleja con todos los ARNg para mediar la edición de genoma y la autoinactivación de los plásmidos con CRISPR-Cpf1.

Un aspecto de un sistema CRISPR-Cpf1 de autoinactivación es la expresión de un formato de matriz individual o en tándem a partir de entre 1 y 4 o más secuencias guía diferentes, por ejemplo, hasta aproximadamente 20 o aproximadamente 30 secuencias guía. Cada secuencia guía autoinactivante individual puede dirigirse a una diana diferente. La mencionada puede procesarse a partir de, por ejemplo, un transcrito pol3 quimérico. Se puede usar promotores Pol3 tales como promotores U6 o H1. Promotores Pol2 tales como los mencionado a través de la presente. Las secuencias repetidas terminales invertidas (iTR) pueden flanquear al promotor de Pol3 - gRNA(s)-promotor de Pol2 - Cpf1.

Un aspecto de un transcrito quimérico de matriz en tándem es que una o más guías editan el uno o más blancos mientras que una o más guías autoinactivantes inactivan el sistema CRISPR/Cpf1. Por lo tanto, por ejemplo, el sistema CRISPR-Cpf1 descrito para la reparación de trastornos de expansión, se puede combinar convenientemente con el sistema CRISPR-Cpf1 autoinactivante que se describe en la presente. Dicho sistema puede, por ejemplo, tener dos guías dirigidas contra la región diana para la reparación así como también al menos una tercera guía dirigida a la autoinactivación del CRISPR-Cpf1. Se hace referencia a la Solicitud de No de Acta PCT/US2014/069897, titulada "Compositions And Methods Of Use Of Crispr-Cas Systems In Nucleotide Repeat Disorders," publicada el 12 de diciembre de 2014 como WO/2015/089351.

Edición Génica o Alteración de Loci Blancos con Cpf1

La rotura de hebra doble o la rotura de hebra individual en una de las hebras ventajosamente debería ser suficientemente cercana la posición diana de forma que ocurra la corrección. En una forma de realización, la distancia no es más de 50, 100, 200, 300, 350 o 400 nucleótidos. Sin pretender estar limitados por la teoría, se cree que la rotura debería ser suficientemente cercana a la posición diana de forma que la rotura esté dentro de la región que se somete a eliminación mediada por exonucleasa durante la resección final. Si la distancia entre la posición diana y una rotura es muy grande, la mutación puede no estar incluida en la resección final y, por ello, puede no ser corregida, ya que la secuencia de templado de ácidos nucleicos solo se puede usar para corregir la secuencia dentro de la región de resección final.

En una forma de realización, en la que un ARN guía y una molécula de Tipo V/Tipo VI, en particular Cpf1/C2c1/C2c2 o un ortólogo u homólogo de la misma, preferiblemente una nucleasa Cpf1, induce una rotura de hebra doble con el propósito de inducir una corrección mediada por HDR, el sitio de escisión está entre 0 y 200 pb (por ejemplo, entre 0 y 175, entre 0 y 150, entre 0 y 125, entre 0 y 100, entre 0 y 75, entre 0 y 50, entre 0 y 25, entre 25 y 200, entre 25 y 175, entre 25 y 150, entre 25 y 125, entre 25 y 100, entre 25 y 75, entre 25 y 50, entre 50 y 200, entre 50 y 175, entre 50 y 150, entre 50 y 125, entre 50 y 100, entre 50 y 75, entre 75 y 200, entre 75 y 175, entre 75 y 150, entre 75 y 1 25, entre 75 y 100 pb) por fuera de la posición diana. En una forma de realización, el sitio de escisión está entre 0 y 100 pb (por ejemplo, entre 0 y 75, entre 0 y 50, entre 0 y 25, entre 25 y 100, entre 25 y 75, entre 25 y 50, entre 50 y 100, entre 50 y 75 o 75 y 100 pb) por fuera de la posición diana. En una forma de realización adicional, dos o más ARN guías complejadas con la Cpf1 o un ortólogo u homólogo de la misma, se pueden usar para inducir múltiples roturas con el propósito de inducir una corrección mediada por HDR.

El brazo de homología se debe extender por lo menos tanto como la región en la que va a ocurrir la resección final, por ejemplo, con el objetivo de permitir que la sobreextensión de la hebra individual reseccionada halle una región complementaria dentro del molde donante. La longitud global puede estar limitada por parámetros tales como el tamaño del plásmido o los límites de empaquetamiento viral. En una forma de realización, un brazo de homología puede no extenderse en elementos repetitivos. Los ejemplos de longitudes de brazos de homología incluyen a por lo menos 50, 100, 250, 500, 750 o 1000 nucleótidos.

La posición diana, según se usa en la presente, se refiere a un sitio sobre un ácido nucleico diana o gen diana (por ejemplo, el cromosoma) que se modifica mediante un proceso dependiente de molécula de Tipo V/Tipo VI, en particular Cpf1/C2c1/C2c2 o un ortólogo u homólogo de las mismas, preferiblemente una molécula Cpf1. Por ejemplo, la posición diana puede ser una escisión con molécula Cpf1 modificada del ácido nucleico diana y una modificación dirigida por ácido nucleico diana, por ejemplo, corrección, de la posición diana. En una forma de realización, una posición diana puede ser un sitio entre dos nucleótidos, por ejemplo, nucleótidos adyacentes, sobre el ácido nucleico diana en el que se agregan uno o más nucleótidos. La posición diana puede comprender uno o más nucleótidos que se alteran, por

ejemplo, corrigen, mediante un molde de ácido nucleico. En una forma de realización, la posición diana está dentro de una secuencia diana (por ejemplo, la secuencia a la que se une el ARN guía). En una forma de realización, una posición diana está corriente arriba o corriente abajo de una secuencia diana (por ejemplo, la secuencia a la que se une el ARN guía).

Un ácido nucleico molde, tal como se usa el término en la presente, se refiere a una secuencia de ácidos nucleicos que se puede usar en conjunción con una molécula de Tipo V/Tipo VI, en particular Cpf1/C2c1/C2c2 o un ortólogo u homólogo de la misma, preferiblemente una molécula de Cpf1 y una molécula de ARN guía para alterar la estructura de una posición diana. En una forma de realización, el ácido nucleico diana se modifica para tener alguna parte o toda la secuencia del ácido nucleico molde, típicamente en o cerca del sitio(s) de escisión. En una forma de realización, el ácido nucleico molde es de hebra individual. En una forma de realización alternativa, el ácido nucleico molde es de hebra doble. En una forma de realización, el ácido nucleico molde es ADN, por ejemplo, ADN de hebra doble. En una forma de realización alternativa, el ácido nucleico molde es ADN de hebra individual.

En una forma de realización, el ácido nucleico molde altera la estructura de la posición diana mediante la participación en la recombinación homóloga. En una forma de realización, el ácido nucleico molde altera la secuencia de la posición diana. En una forma de realización, el ácido nucleico molde da como resultado la incorporación de una base modificada, o no natural en el ácido nucleico diana.

La secuencia de molde puede sufrir una recombinación mediada o catalizada por rotura con la secuencia diana. En una forma de realización, el ácido nucleico molde puede incluir una secuencia que corresponde a un sitio de la secuencia diana que se escinde mediante un evento de escisión mediado por Cpf1. En una forma de realización, el ácido nucleico molde puede incluir una secuencia que corresponde tanto a un primer sitio de la secuencia diana que se escinde en un primer evento mediado por Cpf1, como a un segundo sitio de la secuencia diana que se escinde en un segundo evento mediado por Cpf1.

En determinadas formas de realización, el ácido nucleico molde puede incluir una secuencia que da como resultado una alteración en la secuencia codificante de una secuencia traducida, por ejemplo, una que da como resultado la sustitución de un aminoácido por otro en una proteína producto, por ejemplo, transformando un alelo mutante en uno de tipo salvaje, transformando un alelo de tipo salvaje en un alelo mutante, y/o introduciendo un codón de terminación, la inserción de un residuo de aminoácido, la supresión de un residuo de aminoácido, o una mutación sin sentido. En determinadas formas de realización, el ácido nucleico molde puede incluir una secuencia que da como resultado una alteración en una secuencia no codificante, por ejemplo, una alteración en un exón o en una región no traducida en 5' o 3' o no transcripta. Dichas alteraciones incluyen una alteración en un elemento de control, por ejemplo, un promotor, potenciador, y una alteración en un elemento de control con acción en cis o trans.

Un ácido nucleico molde que tiene homología con una posición diana en un gen diana se puede usar para alterar la estructura de una secuencia diana. La secuencia molde se puede usar para alterar una estructura no deseada, por ejemplo, un nucleótido no deseado o mutante. El ácido nucleico molde puede incluir una secuencia que, cuando se integra, da como resultado: disminución de la actividad de un elemento de control positivo; incremento de la actividad de un elemento de control positivo; disminución de la actividad de un elemento de control negativo; aumento de la actividad de un elemento de control negativo; disminución de la expresión de un gen; incremento de la expresión de un gen; incremento de la resistencia a un trastorno o enfermedad; incremento de la resistencia a entrada de virus; corrección de una mutación o alteración de un residuo de aminoácido indeseado que confiere, incrementa, suprime o disminuye una propiedad biológica de un producto genético, por ejemplo, incrementando la actividad enzimática de una enzima, o incrementado la capacidad de un producto genético para interaccionar con otra molécula.

El ácido nucleico molde pueden incluir una secuencia que da como resultado: un cambio en secuencia de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o más nucleótidos de la secuencia diana. En una forma de realización, el ácido nucleico molde puede ser de 20+/- 10, 30+/- 10, 40+/- 10, 50+/- 10, 60+/- 10, 70+/- 10, 80+/- 10, 90+/- 10, 100+/- 10, 1 10+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 1 80+/- 10, 190+/- 10, 200+/- 10, 210+/-10, de 220+/- 10 nucleótidos de longitud. En una forma de realización, el ácido nucleico molde puede ser de 30+/-20, 40+/-20, 50+/-20, 60+/-20, 70+/- 20, 80+/-20, 90+/-20, 100+/-20, 1 10+/-20, 120+/-20, 130+/-20, 140+/-20, l 50+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/-20, 200+/-20, 210+/-20, de 220+/-20 nucleótidos de longitud. En una forma de realización, el ácido nucleico molde es de entre 10 y 1,000, entre 20 y 900, entre 30 y 800, entre 40 y 700, entre 50 y 600, entre 50 y 500, entre 50 y 400, entre 50 y 300, entre 50 y 200, o entre 50 y 100 nucleótidos de longitud.

Un ácido nucleico molde comprende los siguientes componentes: [5' brazo de homología]-[secuencia de reemplazo]-[3' brazo de homología]. Los brazos de homología proveen la recombinación en el cromosoma, por consiguiente reemplazando el elemento no deseado, por ejemplo, una mutación o signatura, con la secuencia de reemplazo. En una forma de realización, los brazos de homología flanquean los sitios de escisión más distales. En una forma de realización, el extremo 3' del 5' brazo de homología es la posición siguiente al extremo 5' de la secuencia de reemplazo. En una forma de realización, el 5' brazo de homología se puede extender por lo menos a 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, o 2000 nucleótidos del lado 5' del extremo 5' de la secuencia de reemplazo. En una forma de realización, el extremo 5' del 3' brazo de homología es la posición siguiente al extremo 3' de la secuencia de reemplazo. En una forma de realización, el 3' brazo de homología se puede extender a por lo menos 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, o 2000 nucleótidos 3' desde el

extremo 3' de la secuencia de reemplazo.

En determinadas formas de realización, uno o ambos brazos de homología pueden acortarse para evitar incluir ciertos elementos de secuencia repetitiva. Por ejemplo, un 5' brazo de homología puede acortarse para evitar un elemento de secuencia repetitiva. En otras formas de realización, un 3' brazo de homología puede acortarse para evitar un elemento de secuencia repetitiva. En algunas formas de realización, ambos brazos de homología a 5' y 3' pueden acortarse para evitar incluir ciertos elementos de secuencia repetitiva.

En determinadas formas de realización, se puede diseñar un molde de ácido nucleico para corregir una mutación para usar como oligonucleótido de hebra individual. Cuando se usa un oligonucleótido de hebra individual, los brazos de homología a 5' y 3' pueden estar en el rango de hasta aproximadamente 200 pares de bases (pb) de longitud, por ejemplo, por lo menos 25, 50, 75, 100, 125, 150, 175, o 200 pb de longitud.

Unión de extremos no homólogos promovida por el sistema de complejo con proteína efectora Cpf1

En determinadas formas de realización, se puede usar la unión de extremos no homólogos inducida por nucleasa (NHEJ) para noqueos específicos de gen diana. La NHEJ inducida por nucleasa también se puede usar para eliminar (por ejemplo, delecionar) la secuencia en un gen de interés. En general, la NHEJ repara una rotura de doble hebra en el ADN por la unión de dos extremos; sin embargo, en general, la secuencia original se restituye solamente si dos extremos compatibles, exactamente como se formaron por la rotura de doble hebra, se ligan perfectamente. Los extremos de ADN de la rotura de doble hebra con frecuencia se someten a procesamiento enzimático, resultando en la adición o eliminación de nucleótidos, en una o ambas hebras, antes de unir nuevamente los extremos. Esto da como resultado la presencia de mutaciones de inserción y/o supresión (indel) en la secuencia de ADN en el sitio de la reparación por NHEJ. Dos tercios de estas mutaciones típicamente alteran el marco de lectura y, por ello, producen una proteína no funcional. Adicionalmente, las mutaciones que mantienen el marco de lectura, pero que insertan o eliminan una cantidad significativa de secuencia, pueden destruir la funcionalidad de la proteína. Esto es dependiente de locus ya que las mutaciones en dominios funcionales críticos probablemente son menos tolerables que las mutaciones en las regiones no críticas de la proteína. Las mutaciones por indel generadas por la NHEJ son no predecibles en la naturaleza; sin embargo, en un sitio dado de rotura, ciertas secuencias de indel están favorecidas y están sobre representadas en la población, probablemente debido a pequeñas regiones de microhomología. Las longitudes de las supresiones pueden variar ampliamente; más comúnmente en el rango entre 1 y 50 pb, pero fácilmente pueden ser mayores de 50 pb, por ejemplo, fácilmente pueden alcanzar tamaños mayores de aproximadamente entre 100 y 200 pb. Las inserciones tienden a ser más cortas y a menudo incluyen duplicaciones cortas de secuencia inmediatamente alrededor del sitio de rotura. Sin embargo, es posible obtener inserciones grandes, y en estos casos, la secuencia insertada a menudo se ha trazado en otras regiones del genoma o en el ADN plasmídico presente en las células.

Debido a que la NHEJ es un proceso mutagénico, también se puede usar para eliminar pequeños motivos de secuencia con la condición de que la generación de una secuencia final específica no sea algo requerido. Si se ataca una rotura de doble hebra cerca de una secuencia diana corta, las mutaciones de supresión causadas por la reparación por NHEJ a menudo abarcan, y por ello eliminan, los nucleótidos no deseados. Para la supresión de segmentos más grandes de ADN, la introducción de dos roturas de hebra doble, una en cada lado de la secuencia, puede dar como resultado una NHEJ entre los extremos con eliminación de toda la secuencia intermedia. Estos dos abordajes se pueden usar para eliminar secuencias de ADN específicas; sin embargo, la naturaleza propensa a error de la NHEJ aún puede producir mutaciones por indel en el sitio de reparación.

Tanto las moléculas de escisión de doble hebra de Tipo V/Tipo VI, en particular las Cpf1/C2c1/C2c2 o un ortólogo u homólogo de las mismas, preferiblemente moléculas de Cpf1 como las moléculas de escisión de hebra individual, o nickasa, de Tipo V/Tipo VI, en particular Cpf1/C2c1/C2c2 o un ortólogo u homólogo de las mismas, preferiblemente moléculas de Cpf1 se pueden usar en los métodos y composiciones que se describen en la presente para generar indeles mediados por NHEJ. Los indeles mediados por NHEJ contra el gen, por ejemplo, una región codificante, por ejemplo, una región codificante temprana de un gen de interés, se pueden usar para noquear (es decir, eliminar la expresión) de un gen de interés. Por ejemplo, la región codificante temprana de un gen de interés incluye la secuencia inmediatamente contigua al sitio de inicio de la transcripción, dentro de un primer exón de la secuencia codificante, o dentro de las 500 pb del sitio de inicio de la transcripción (por ejemplo, menos de 500, 450, 400, 350, 300, 250, 200, 150, 100 o 50 pb).

En una forma de realización, en la que un ARN guía y una molécula de Tipo V, en particular Cpf1/C2c1 o un ortólogo u homólogo de la misma, preferiblemente una nucleasa Cpf1 generan una rotura de doble hebra con el propósito de inducir indeles mediados por NHEJ, se puede configurar un ARN para posicionar una rotura de doble hebra en la proximidad de un nucleótido de la posición diana. En una forma de realización, el sitio de escisión puede estar entre 0 y 500 pb por fuera de la posición diana (por ejemplo, menos de 500, 400, 300, 200, 100, 50, 40, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 pb desde la posición diana).

En una forma de realización, en la que dos ARN guías complejados con moléculas de Tipo V, en particular Cpf1/C2c o un ortólogo u homólogo de la misma, preferiblemente nickasas Cpf1, inducen dos roturas de hebra individual con el propósito de inducir indeles mediados por NHEJ, se puede configurar dos ARN guías para posicionar dos roturas de hebra individual para proveer la reparación por NHEJ de un nucleótido de la posición diana.

Los complejos de proteína efectora Cpf1 pueden suministrar efectores funcionales

Al contrario del noqueo del gen mediado por CRISPR-Cas, que elimina permanentemente la expresión por mutación del gen a nivel de ADN, el noqueo de CRISPR-Cas permite una reducción temporal de la expresión de un gen usando factores de transcripción artificiales. La mutación de residuos clave en ambos dominios de clivaje del ADN de la proteína Cpf1, tal como la proteína FnCpf1 (por ejemplo, las mutaciones D917A y H1006A) dan como resultado la generación de una Cpf1 catalíticamente inactiva. Una Cpf1 catalíticamente inactiva se compleja con un ARN guía y se localiza en la secuencia de ADN especificada por dicho dominio de direccionamiento del ARN guía, sin embargo, este no escinde el ADN diana. La fusión de la proteína Cpf1 inactiva, tal como la proteína FnCpf1 (por ejemplo, las mutaciones D917A y H1006A de la proteína FnCpf1 o D908A, E993A, D1263A de acuerdo a la proteína AsCpf1 o D832A, E925A, D947A o D1180A de acuerdo a la proteína LbCpf1) con un dominio efector, por ejemplo, un dominio de represión de la transcripción, permite el reclutamiento del efector a cualquier sitio del ADN especificado por el ARN guía. En ciertas formas de realización, la Cpf1 se puede fusionar con un dominio de represión de la transcripción y se puede reclutar hacia la región promotora de un gen. Especialmente para la represión del gen, aquí se contempla que el bloqueo del sitio de unión de un factor de transcripción endógeno puede contribuir a disminuir la expresión del gen. En otra forma de realización, se puede fusionar una Cpf1 inactiva a una proteína modificadora de la cromatina. El estado de alteración de la cromatina puede dar como resultado una reducción de la expresión del gen diana.

En una forma de realización, una molécula de ARN guía se puede dirigir hacia elementos de respuesta a la transcripción conocidos (por ejemplo, promotores, intensificadores, etc.), una secuencia activadora en dirección 5' conocida, y/o secuencias con una función desconocida o conocida que se sospecha que son capaces de controlar la expresión del ADN diana.

En algunos métodos, se puede desactivar un polinucleótido diana para efectuar la modificación de la expresión en una célula. Por ejemplo, tras la unión de un complejo CRISPR a la secuencia diana en una célula, el polinucleótido diana se desactiva de modo que la secuencia no se transcriba, la proteína codificada no se produce o la secuencia no funciona como lo hace la secuencia no modificada. Por ejemplo, una proteína o una secuencia codificante de microARN se podrán desactivar de modo que no se produzca la proteína.

En ciertas formas de realización, la enzima CRISPR comprende una o más mutaciones seleccionadas entre el grupo constituido por D917A, E1006A y D1225A y/o la una o más mutaciones es en un dominio RuvC de la enzima CRISPR o de otra manera es una mutación como las que se describen en la presente. En algunas formas de realización, la enzima CRISPR tiene una o más mutaciones en un dominio catalítico, donde, cuando se transcribe, la repetición directa de la secuencia forma un único tallo-bucle y la secuencia guía dirige la unión específica de una secuencia de un complejo CRISPR a la secuencia diana, y donde la enzima comprende además un dominio funcional. En algunas formas de realización, el dominio funcional es un dominio de activación transcripcional, preferiblemente VP64. En algunas formas de realización, el dominio funcional es un dominio de represión transcripcional, preferentemente KRAB. En algunas formas de realización, el dominio de represión transcripcional es SID, o concatémeros de SID (por ejemplo, SID4X). En algunas formas de realización, el dominio funcional es un dominio de modificación epigenética, de modo que se proporciona una enzima de modificación epigenética. En algunas formas de realización, el dominio funcional es un dominio de activación, que podrá ser el dominio de activación P65.

Suministro del complejo de proteína Cpf1 efectora o componentes del mismo.

Mediante la presente divulgación y el conocimiento de la técnica, los sistemas CRISPR-Cas, específicamente los nuevos sistemas CRIPSR que se describen en la presente, o los componentes del mismo (incluyendo, por ejemplo, el molde de HDR) o las moléculas de ácido nucleico del mismo o las moléculas de ácido nucleico que codifican o proporcionan componentes del mismo pueden suministrarse mediante un sistema de suministro descrito en el presente documento, tanto de manera general como detallada.

Suministro del vector, por ejemplo, viral o plásmido: La enzima CRISPR, por ejemplo, una Cpf1 y/o cualquiera de los ARN de la presente, por ejemplo, un ARN guía, se pueden suministrar utilizando cualquier vector adecuado, por ejemplo vectores plasmídicos o virales, tales como un virus adenoasociado (AAV), lentivirus, adenovirus u otros tipos de vectores virales o combinaciones de estos. Se puede empaqueta a Cpf1 y uno o más ARN guía en uno o más vectores, por ejemplo, vectores plasmídicos o virales. En algunas formas de realización el vector, por ejemplo, el plásmido o el vector viral, se suministra al tejido de interés mediante, por ejemplo, inyección intramuscular, mientras que otras veces el suministro es por medio de métodos de suministro intravenoso, transdérmico, intranasal, oral, mucosal o de otro tipo. Este tipo de suministro podrá ser mediante una única dosis o múltiples dosis. El experto en la técnica entenderá que la dosificación real que se ha de suministrar en la presente podrá variar enormemente dependiendo de varios factores tales como la elección del vector, la célula, organismo o tejido diana, el estado general del sujeto que se va a tratar, el grado de transformación/modificación que se busca, la vía de administración, el modo de administración, el tipo de transformación/modificación buscada, etc.

Una dosificación de este tipo podrá contener además, por ejemplo, un portador (agua, solución salina, etanol, glicerol, lactosa, sacarosa, fosfato de calcio, gelatina, dextrano, agar, pectina, aceite de cacahuete, aceite de sésamo, etc.), un diluyente, un portador farmacéuticamente aceptable (por ejemplo, solución salina de pH regulado con fosfato), un excipiente farmacéuticamente aceptable, y/u otros compuestos conocidos en la técnica. La dosificación podrá contener además una o más sales farmacéuticamente aceptables tales como, por ejemplo, una sal de un ácido mineral tal como

un clorhidrato, un bromhidrato, un fosfato, un sulfato, etc.; y las sales de ácidos orgánicos tales como acetatos, propionatos, malonatos, benzoatos, etc. Adicionalmente, también podrán estar presentes en este documento sustancias auxiliares, tales como agentes humectantes o emulsionantes, sustancias reguladoras del pH, geles o materiales gelatinizantes, saborizantes, colorantes, microesferas, polímeros, agentes de suspensión, etc. Además, también podrán estar presentes uno o más ingredientes farmacéuticos convencionales tales como conservantes, humectantes, agentes de suspensión, tensioactivos, antioxidantes, agentes antiapelmazantes, cargas, agentes quelantes, agentes de recubrimiento, estabilizantes químicos, etc., especialmente si la forma farmacéutica es una forma reconstituible. Los ingredientes ilustrativos adecuados incluyen celulosa microcristalina, carboximetilcelulosa de sodio, polisorbato 80, alcohol feniletílico, clorobutanol, sorbato de potasio, ácido sórbico, dióxido de azufre, galato de propilo, los parabenos, etilvanillina, glicerina, fenol, paraclorofenol, gelatina, albúmina y combinaciones de estos. Se puede consultar una discusión exhaustiva de excipientes farmacéuticamente aceptables en REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991).

En una forma de realización de la presente, el suministro tiene lugar mediante un adenovirus, que podrá ser a una dosis de refuerzo única que contenga al menos 1×10^5 partículas (también denominadas unidades de partículas, up) de un vector adenoviral. En una forma de realización de la presente, preferentemente la dosis es de al menos aproximadamente 1×10^6 partículas (por ejemplo, aproximadamente 1×10^6-1×10^{12} partículas), más preferentemente al menos aproximadamente 1×10^7 partículas, más preferentemente al menos aproximadamente 1×10^8 partículas (por ejemplo, aproximadamente 1×10^8-1×10^{11} partículas o aproximadamente 1×10^8-1×10^{12} partículas) y de la manera más preferida al menos aproximadamente 1×10^0 partículas (por ejemplo, aproximadamente 1×10^9 - 1×10^{10} partículas o aproximadamente 1×10^9 - 1×10^{12} partículas) o incluso al menos aproximadamente 1×10^{10} partículas (por ejemplo, aproximadamente 1×10^{10} - 1×10^{12} partículas) del vector adenoviral. Como alternativa, la dosis comprende no más de aproximadamente 1×10^{14} partículas, preferentemente no más de aproximadamente 1×10^{13} partículas, aún más preferentemente no más de aproximadamente 1×10^{12} partículas, aún más preferentemente no más de aproximadamente 1×10^{11} partículas y de la manera más preferida no más de aproximadamente 1×10^{10} partículas (por ejemplo, no más de aproximadamente 1×10^9 partículas). Por lo tanto, la dosis podrá contener una dosis única del vector adenoviral con, por ejemplo, aproximadamente 1×10^6 unidades de partículas (up), aproximadamente 2×10^6 up, aproximadamente 4×10^6 up, aproximadamente 1×10^7 up, aproximadamente 2×10^7 up, aproximadamente 4×10^7 up, aproximadamente 1×10^8 up, aproximadamente 2×10^8 up, aproximadamente 4×10^8 up, aproximadamente 1×10^9 up, aproximadamente 2×10^9 up, aproximadamente 4×10^9 up, aproximadamente 1×10^{10} up, aproximadamente 2×10^{10} up, aproximadamente 4×10^{10} up, aproximadamente 1×10^{11} up, aproximadamente 2×10^{11} up, aproximadamente 4×10^{11} up, aproximadamente 1×10^{12} up, aproximadamente 2×10^{12} up o aproximadamente 4×10^{12} up del vector adenoviral. Véase, por ejemplo, los vectores adenovirales de la patente de los EE.UU. n.º 8.454.972 B2 de Nabel *y col.*, otorgada el 4 de junio de 2013; y las dosificaciones de la col 29, líneas 36-58 de esta. En una forma de realización de la presente, se suministra el adenovirus mediante múltiples dosis.

En una forma de realización de la presente, el suministro se realiza mediante un AAV. Se cree que una dosificación terapéuticamente eficaz para el suministro *in vivo* del AAV a un ser humano está comprendida en el intervalo de aproximadamente 20 a aproximadamente 50 mL de solución salina que contiene de aproximadamente 1×10^{10} a aproximadamente 1×10^{10} AAV funcionales/mL de solución. La dosificación se podrá ajustar para encontrar el equilibrio entre el beneficio terapéutico y cualesquiera efectos secundarios. En una forma de realización de la presente, la dosis de AAV está comprendida generalmente en el intervalo de concentraciones de aproximadamente 1×10^5 a 1×10^{50} genomas de AAV, entre aproximadamente 1×10^8 y 1×10^{20} genomas de AAV, entre aproximadamente 1×10^{10} y aproximadamente 1×10^{16} genomas o entre aproximadamente 1×10^{11} y aproximadamente 1×10^{16} genomas de AAV. Una dosificación para seres humanos podrá ser de aproximadamente 1×10^{13} genomas de AAV. Se podrán suministrar este tipo de concentraciones en una cantidad de entre aproximadamente 0,001 mL y aproximadamente 100 mL, entre aproximadamente 0,05 y aproximadamente 50 mL o entre aproximadamente 10 y aproximadamente 25 mL de una solución portadora. El experto en la técnica podrá determinar fácilmente otras dosificaciones eficaces mediante ensayos habituales determinando las curvas de dosis de respuesta. Véase, por ejemplo, la patente de los EE.UU. con n.º 8.404.658 B2 de Hajjar *y col.*, otorgada el 26 de marzo de 2013, en la col. 27, líneas 45-60.

En una forma de realización de la presente, el suministro se realiza mediante un plásmido. En tales composiciones plasmídicas, la dosificación debería ser una cantidad de plásmido suficiente para suscitar una respuesta. Por ejemplo, las cantidades adecuadas de ADN plasmídico en las composiciones plasmídicas podrán estar comprendidas entre aproximadamente 0,1 y aproximadamente 2 mg o entre aproximadamente 1 µg y aproximadamente 10 µg por individuo de 70 kg. Los plásmidos adecuados para llevar a cabo la invención comprenderán generalmente (i) un promotor; (ii) una secuencia que codifica una enzima CRISPR, ligada operativamente a dicho promotor; (iii) un marcador seleccionable; (iv) un origen de replicación; y (v) un terminador de la transcripción en la dirección 3' y ligado operativamente a (ii). El plásmido puede codificar también los componentes de ARN de un complejo CRISPR, pero uno o más de estos pueden a su vez estar codificados en un vector diferente.

En la presente, las dosis se basan en un individuo de 70 kg de promedio. La frecuencia de administración está comprendida en las competencias del facultativo médico o veterinario (por ejemplo, médico, veterinario) o científico experto en la técnica. Se notifica también que los ratones usados en los experimentos suelen tener aproximadamente 20 g y los experimentos con ratones se pueden escalar hasta un individuo de 70 kg.

La dosificación que se usa para las composiciones que se proveen en la presente incluyen dosificaciones para administración repetida o dosificación repetida. En formas de realización particulares, la administración se repite dentro de un periodo de varias semanas, meses, o años. Se pueden llevar a cabo ensayos adecuados para obtener un régimen de dosificación óptimo. La administración repetida puede permitir el uso de una dosificación inferior, lo que puede afectar de forma positiva a las modificaciones fuera de diana.

En algunas formas de realización, las moléculas de ARN usadas en los métodos y composiciones reivindicados de la invención se suministran en formulaciones de liposomas o lipofectina y similares y se pueden preparar mediante métodos muy conocidos por los expertos en la técnica. Tales métodos se describen, por ejemplo, en las patentes de los EE.UU. con n.º 5.593.972, 5.589.466 y 5.580.859. Se han desarrollado sistemas de suministro orientados específicamente al suministro potenciado y mejorado de ARNip en células de mamífero (véase a, por ejemplo, Shen y col. FEBS Let. 2003, 539:111-114; Xia y col., Nat. Biotech. 2002, 20:1006-1010; Reich y col., Mol. Vision. 2003, 9: 210-216; Sorensen y col., J. Mol. Biol. 2003, 327: 761-766; Lewis y col., Nat. Gen. 2002, 32: 107-108 y Simeoni *y col.*, *NAR* 2003, 31, 11: 2717-2724) y se podrán aplicar a la presente invención. Recientemente se ha utilizado con éxito ARNip para la inhibición de la expresión génica en primates (véase a, por ejemplo, Tolentino *y col.*, *Retina* 24(4):660 que también se podrá aplicar a la presente invención.

A su vez, el suministro de ARN también es un método útil para el suministro *in vivo*. Es posible administrar Cpf1 y ARN guía ARNg (y, por ejemplo, molde de reparación HR) al interior de las células usando liposomas o nanopartículas. Por lo tanto, el suministro de la enzima CRISPR, por ejemplo, una Cpf1 y/o la administración de los ARN de la invención puede ser en forma de ARN y mediante microvesículas, liposomas o partículas. Por ejemplo, el ARNg y el ARNm de Cpf1 se pueden empaquetar en partículas liposomales para el suministro *in vivo*. Los reactivos de transfección liposomal tales como lipofectamina de Life Technologies y otros reactivos comercializados pueden suministrar de manera eficaz moléculas de ARN en el hígado.

También se prefieren otros medios de suministro de ARN que incluyen el suministro de ARN mediante partículas (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. y Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, Advanced Functional Materials, 19: 3112-3118, 2010) o exosomas (Schroeder, A., Levins, C., Cortez, C., Langer, R., y Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641). Ciertamente, se ha demostrado que los exosomas son particularmente útiles en el suministro de ARNip, un sistema con ciertos paralelismos con el sistema CRISPR. Por ejemplo, El-Andaloussi S, y col. ("Exosome-mediated delivery of siRNA *in vitro* and *in vivo*." *Nat Protoc.* diciembre de 2012;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 15 de noviembre de 2012) describen cómo los exosomas son herramientas prometedoras para el suministro de fármacos a través de diferentes barreras biológicas y se pueden aprovechar para el suministro de ARNip *in vitro* e *in vivo*. Su estrategia consiste en generar exosomas dirigidos mediante la transfección con un vector de expresión que comprende una proteína exosómica fusionada con un ligando peptídico. A continuación, se purifican los exosomas y se caracterizan a partir del sobrenadante de las células transfectadas, a continuación se carga el ARN en los exosomas. El suministro o administración de acuerdo con la invención se puede realizar con exosomas en particular, de manera no taxativa, al cerebro. La vitamina E (α-tocoferol) se puede conjugar con la CRISPR Cas y se puede suministrar al cerebro junto con una lipoproteína de alta densidad (HDL), por ejemplo, de manera similar a como hicieron Uno y col. (HUMAN GENE THERAPY 22:711-719 (junio de 2011) para el suministro del ARN de interferencia pequeño (ARNip) al cerebro. Se administró una infusión a ratones mediante minibombas osmóticas (modelo 1007D; Alzet, Cupertino, CA) rellenada con solución salina de pH regulado con fosfato (PBS) o TocBACEip libre o Toc-BACEip/HDL y se conectó con el conjunto de elementos 3 de infusión cerebral (Alzet). Se colocó una cánula de infusión al cerebro aproximadamente 0,5 mm detrás del bregma en la línea central para la infusión al tercer ventrículo dorsal. Uno *y col.* observaron que tan solo 3 nmol de Toc-ARNip con HDL pueden inducir una reducción objetivo de un grado comparable mediante el mismo método de infusión ICV. En la presente invención se puede contemplar una dosificación similar de CRISPR Cas conjugada con α-tocoferol y se puede administrarse conjuntamente con HDL dirigido al cerebro para los seres humanos, por ejemplo, se podrán contemplar entre aproximadamente 3 nmol y aproximadamente 3 µmol de CRISPR Cas dirigida hacia el cerebro. Zou y col. (HUMAN GENE THERAPY 22:465-475 (abril de 2011)) describen un método de suministro mediado por lentivirus de ARN de horquilla corta dirigido hacia PKCγ para el silenciamiento génico *in vivo* en la médula espinal de ratas. Zou *y col.* administraron aproximadamente 10 µL de un lentivirus recombinante que tenía un título de 1×10^9 unidades de transducción (UT)/mL mediante un catéter intratecal. En la presente invención se podrá contemplar una dosificación similar de CRISPR Cas expresado en un vector lentiviral dirigido al cerebro para los seres humanos, por ejemplo, se podrán contemplar aproximadamente 10-50 ml de CRISPR Cas dirigida al cerebro en un lentivirus que tiene un título de 1×10^9 unidades de transducción (UT)/ml

Los complejos de CRISPR-Cpf1 recombinante preensamblados que comprenden Cpf1 y ARNcr se pueden transfectar, por ejemplo, por electroporación, lo que resulta en altas tasas de mutación y ausencia de mutaciones fuera de diana detectables. Hur, J.K. y col., Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596. [publicación electrónica en proceso de impresión]

En términos de suministro local al cerebro, esto se puede conseguir de diversos modos. Por ejemplo, se puede suministrar material por vía intraestrial, por ejemplo, mediante inyección. Se puede llevar a cabo la inyección estereotácticamente mediante una craneotomía.

También es útil la mejora de la eficacia de la RH o NHEJ para el suministro. Se prefiere mejorar la eficacia de la NHEJ mediante la coexpresión de enzimas que procesan el extremo tal como Trex2 (Dumitrache y col., Genetics. agosto de 2011; 188(4): 787-797). Se prefiere incrementar la eficacia de la RH inhibiendo de manera transitoria la maquinaria de la NHEJ tal como Ku70 y Ku86. También se puede incrementar la eficacia de la RH coexpresando enzimas de recombinación homóloga eucariotas o procariotas tales como RecBCD, RecA

Empaquetamiento y promotores

Las maneras de empaquetar las moléculas de ácido nucleico que codifican para Cpf1 de la invención, por ejemplo, ADN, en vectores, por ejemplo, vectores virales, para mediar la modificación del genoma *in vivo* incluyen:

- Para lograr la inactivación génica mediada por NHEJ:
- Vector viral único:
- Vector que contiene dos o más casetes de expresión:
- Promotor- molécula de ácido nucleico que codifica Cpf1s-terminador
- Promotor- ARN guía 1-terminador
- Promotor-ARN guía 2 – terminador
- Promotor- ARNg (N)-terminador (hasta el tamaño límite del vector)
- Vector viral doble:
- Vector 1 que contiene un casete de expresión para impulsar la expresión de Cpf1
- Promotor- molécula de ácido nucleico que codifica una Cpf1-terminador
- Vector 2 que contiene uno o más casetes de expresión para impulsar la expresión de uno o más ARN guía
- Promotor- ARN1 guía-terminador
- Promotor - ARNg (N)-terminador (hasta el tamaño límite del vector)
- Para mediar la reparación dirigida por homología.
- Además de las estrategias del vector viral único y doble descritas anteriormente, se puede usar un vector adicional para suministrar un molde de reparación dirigida por homología.

El promotor que se utiliza para impulsar la expresión de la molécula de ácido nucleico que codifican para Cpf1) puede incluir:

- La ITR de AAV puede actuar como promotor: esto es conveniente para eliminar la necesidad de un elemento promotor adicional (que puede ocupar espacio en el vector). El espacio adicional liberado se puede utilizar para impulsar la expresión de elementos adicionales (ARNg, etc.). También, la actividad de las ITR es relativamente más débil, de manera tal que se pueden utilizar para reducir la potencial toxicidad debido a la sobreexpresión de Cpf1.
- Para la expresión ubicua, los promotores que se pueden utilizar incluyen: CMV, CAG, CBh, PGK, SV40, cadena ligera o pesada de la ferritina, etc.

Para la expresión en el cerebro u otra expresión diferente en el SNC, se pueden usar los promotores: Sinapsinal para todas las neuronas, CaMKIIalfa para las neuronas excitatorias, GAD67 o GAD65 o VGAT para las neuronas GABAérgicas, etc.

Para la expresión en el hígado, se puede utilizar el promotor de la albúmina.

Para la expresión en el pulmón, se puede utilizar SP-B.

Para las células endoteliales, se puede utilizar ICAM.

Para las células hematopoyéticas, se puede utilizar IFNbeta o CD45.

Para los osteoblastos, se puede utilizar OG-2.

El promotor utilizado para impulsar el ARN guía puede incluir:

- Promotores de Pol III tal como U6 o H1

- Uso del promotor de Pol II y casetes intrónicos para expresar el ARNg

Virus adenoasociado (AAV)

Se pueden suministrar la Cpf1 y uno o más ARN guías utilizando un virus adenoasociado (AAV), lentivirus, adenovirus u otro plásmido o tipos de vectores virales, en particular, utilizando las formulaciones y dosis, por ejemplo, de las patentes de los EE.UU. n.º 8.454.972 (formulaciones, dosis para adenovirus), 8.404.658 (formulaciones, dosis para AAV) y 5.846.946 (formulaciones, dosis para plásmidos de ADN) y de ensayos clínicos y publicaciones acerca de ensayos clínicos en los que se utilizan lentivirus, AAV y adenovirus. Por ejemplo, para AAV, la vía de administración, formulación y dosis puede ser como en la patente de los EE.UU. con n.º 8.454.972 y como en los ensayos clínicos en los que se utilizan AAV. Para adenovirus, la vía de administración, formulación y dosis puede ser como en la patente de los EE.UU. con n.º 8.404.658 y como en los ensayos clínicos en los que se utilizan adenovirus. Para el suministro con plásmidos, la vía de administración, formulación y dosis puede ser como en la patente de los EE.UU. con n.º 5.846.946 y como en los ensayos clínicos en los que se utilizan plásmidos. Las dosis se pueden basar o extrapolar a un individuo promedio de 70 kg, (por ejemplo, un ser humano adulto varón), y se pueden ajustar para pacientes, sujetos, mamíferos con pesos diferentes y de especies diferentes. La frecuencia de administración está comprendida en las competencias del facultativo médico o veterinario (por ejemplo, médico, veterinario), dependiendo de factores normales que incluyen la edad, sexo, salud general, otras afecciones del sujeto o paciente y la afección o síntomas particulares que se están abordando. Los vectores virales se pueden inyectar en el tejido de interés. Para la modificación específica del tipo de célula genoma, la expresión de Cpf1 puede ser impulsada por un promotor específico para el tipo de célula. Por ejemplo, la expresión hepatoespecífica podrá utilizar el promotor de la albúmina y la expresión específica de las neuronas (por ejemplo, para dirigirse a los trastornos del SNC) podría utilizar el promotor de la sinapsina I.

En términos de suministro *in vivo*, AAV es más conveniente respecto de otros vectores virales por un par de razones:

Baja toxicidad (esta se puede deber a que el método de purificación no requiere ultracentrifugación de partículas celulares que puedan activar la respuesta inmunitaria) y Baja probabilidad de provocar mutagénesis por inserción debido a que no se integra en el genoma del huésped.

AAV tiene un límite de empaquetamiento de 4,5 o 4,75 Kb. Esto significa que la Cpf1 así como un promotor y un terminador de transcripción deben caber todos dentro del mismo vector viral. Las construcciones mayores de 4,5 o 4,75 Kb conducirán a una producción de virus significativamente reducida. SpCas9 es bastante grande, el gen mismo tiene más de 4,1 Kb, lo que hace difícil su empaquetamiento en AAV. Por lo tanto, algunas formas de realización de la invención incluyen el uso de homólogos de Cpf1 que son más cortos.

En lo que se refiere a AAV, el AAV puede ser AAV1, AAV2, AAV5 o cualquier combinación de estos. Se puede seleccionar el AAV de los AAV teniendo en cuenta las células que se quieren modificar de manera dirigida; por ejemplo, se pueden seleccionar los serotipos 1, 2, 5 de AAV o una cápside híbrida de AAV1, AAV2, AAV5 o cualquier combinación de estos para que tengan como diana células neuronales o del cerebro; y se puede seleccionar AAV4 para que tenga como diana el tejido cardiaco. AAV8 es útil para el suministro al hígado. Los promotores y vectores de la presente se prefieren individualmente. Una tabulación de algunos serotipos de AAV en lo que se refiere a estas células (véase a Grimm, D. y col., J. Virol. 82: 5887-5911 (2008)) es como sigue:

Línea celular	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2,5	0,0	0,1	10	0,7	0,0
HEK293	25	100	2,5	0,1	0,1	5	0,7	0,1
HeLa	3	100	2,0	0,1	6,7	1	0,2	0,1
HepG2	3	100	16,7	0,3	1,7	5	0,3	ND
Hep1A	20	100	0,2	1,0	0,1	1	0,2	0,0
911	17	100	11	0,2	0,1	17	0,1	ND
CHO	100	100	14	1,4	333	50	10	1,0
COS	33	100	33	3,3	5,0	14	2,0	0,5
MeWo	10	100	20	0,3	6,7	10	1,0	0,2
NIH3T3	10	100	2,9	2,9	0,3	10	0,3	ND

A549	14	100	20	ND	0,5	10	0,5	0,1
HT1180	20	100	10	0,1	0,3	33	0,5	0,1
Monocitos	1111	100	ND	ND	125	1429	ND	ND
DC inmaduras	2500	100	ND	ND	222	2857	ND	ND
DC maduras	2222	100	ND	ND	333	3333	ND	ND

Lentivirus

Los lentivirus son retrovirus complejos que tienen la capacidad de infectar y expresar sus genes tanto en células mitóticas y posmitóticas. El lentivirus conocido más comúnmente es el virus de la inmunodeficiencia humana (VIH), que utiliza glucoproteínas de la envoltura de otros virus para actuar sobre una gama amplia de tipos celulares.

Se podrán preparar lentivirus de la siguiente manera. Después de clonar pCasES10 (que contiene un esqueleto plasmídico de transferencia lentiviral), se sembraron HEK293FT que habían sido sometidas a pocos pases (p=5) en un matraz T-75 hasta alcanzar una confluencia de un 50% el día antes de la transfección en DMEM con un 10% de suero bovino fetal y sin antibióticos. Se cambió el medio después de 20 horas por medio OptiMEM (exento de suero) y se realizó la transfección 4 horas más tarde. Se transfectaron las células con 10 µg de plásmido de transferencia lentiviral (pCasES10) y los siguientes plásmidos de empaquetamiento: 5 µg de pMD2.G (pseudotipo VSV-g) y 7,5 µg de psPAX2 (gag/pol/rev/tat). Se realizó la transfección en 4 mL de OptiMEM con un agente de suministro lipídico catiónico (50 µL de Lipofectamine 2000 y 100 µL de reactivo Plus). Después de 6 horas, se cambió el medio por DMEM exento de antibióticos con un 10% de suero bovino fetal. Estos métodos usan suero durante el cultivo celular, pero se prefieren métodos exentos de suero.

Se podrán purificar lentivirus de la siguiente manera. Se recolectaron los sobrenadantes virales después de 48 horas. Se limpiaron los sobrenadantes de desechos en primer lugar y se filtraron a través de un filtro con una unión a proteínas baja (PVDF, por sus siglas en inglés) de 0,45 µm. A continuación, se centrifugaron en una ultracentrífuga durante 2 horas a 24,000 rpm. Se resuspendieron los sedimentos virales en 50 µL de DMEM durante toda la noche a 4 ºC. Seguidamente se distribuyeron en alícuotas y se congelaron inmediatamente a -80 ºC.

En otra forma de realización, también se contemplan vectores lentivirales mínimos que no son de primates basados en el virus de la anemia infecciosa equina (EIAV, por sus siglas en inglés), especialmente para la terapia génica ocular (véase, por ejemplo, Balagaan, J Gene Med 2006; 8: 275 – 285). En otra forma de realización, también se contempla al RetinoStat®, un vector para la terapia génica lentiviral derivado del virus de la anemia infecciosa equina que expresa las proteínas angiostáticas endostatina y angiostatina que se suministra mediante inyección subretinal para el tratamiento de la forma húmeda de degeneración macular senil (véase, por ejemplo, Binley y col., HUMAN GENE THERAPY 23:980–991 (septiembre de 2012)) y este vector se puede modificar para el sistema CRISPR-Cas de la presente invención.

En otra forma de realización, se pueden usar vectores lentivirales autoinactivantes con un ARNip que se dirigen a un exón común compartido por tat/rev de VIH, un señuelo TAR de ubicación nucleolar y una ribozima de tipo cabeza de martillo específica dirigida contra CCR5 (véase, por ejemplo, DiGiusto y col. (2010) Sci Transl Med 2:36ra43) y/o adaptarse al sistema CRISPR Cas de la presente invención. Se podrán recoger un mínimo de $2,5 \times 10^6$ células CD34+ por kilogramo de peso del paciente y preestimular durante 16 a 20 horas en medio X-VIVO 15 (Lonza) que contiene L-glutamina 2 µmol, factor de células madre (100 ng/ml), ligando Flt-3 (Flt-3L) (100 ng/ml) y trombopoyetina (10 ng/ml) (CellGenix) con una densidad de 2×10^6 células/ml. Las células preestimuladas se podrán transducir con lentivirus con una multiplicidad de infección de 5 durante 16 a 24 horas en matraces de cultivo tisular de 75 cm^2 recubiertos con fibronectina (25 mg/cm^2) (RetroNectin, Takara Bio Inc.).

Se han divulgado vectores lentivirales en el tratamiento de la enfermedad de Parkinson, véase, por ejemplo, la patente de los EE.UU. con n.º de publicación 20120295960 y las patentes de los EE.UU. con nos 7303910 y 7351585. También se han divulgado vectores lentivirales para el tratamiento de enfermedades oculares, véase, por ejemplo, las patentes de los EE.UU. con n.º de publicación 20060281180, 20090007284, US20110117189; US20090017543; US20070054961, US20100317109. También se han descrito vectores lentivirales para el suministro al cerebro, véase, por ejemplo, las patentes de los EE.UU. n.º de publicación US20110293571; US20110293571, US20040013648, US20070025970, US20090111106 y la patente de los EE.UU. con n.º US7259015

Suministro de ARN

Suministro de ARN: La enzima CRISPR, por ejemplo, una Cpf1, y/o cualquiera de los presentes ARN, por ejemplo un ARN guía, también se puede suministrar en forma de ARN. El ARNm de Cpf1 se puede generar usando una transcripción *in vitro*. Por ejemplo, el ARNm de Cpf1 se puede sintetizar usando un casete de PCR que contiene los siguientes elementos: promotor_T7-secuencia kozak (GCCACC)-Cpf1-3'UTR de la cola de poliA de globina beta (una cadena de 120 o más adeninas). El casete puede ser utilizado para la transcripción por la polimerasa T7. También se pueden transcribir el ARN guía utilizando la transcripción *in vitro* de un casete que contiene la secuencia promotor_T7-

GG-ARN guía.

Para potenciar la expresión y reducir la posible toxicidad, la secuencia que codifica para la enzima CRISPR y/o el ARN guía se pueden modificar para incluir uno o más nucleósidos manipulados, por ejemplo, utilizando pseudo-U o 5-metil-C.

Los métodos de suministro de ARNm son especialmente prometedores en la actualidad para el suministro al hígado.

Muchos trabajos clínicos sobre el suministro de ARN se han centrado en la iARN o de sentido contrario, pero estos sistemas se pueden adaptar para el suministro del ARN para implementar la presente invención. Por consiguiente, deben leerse las siguientes referencias a la iARN, etc.

Sistemas y/o formulaciones de suministro de partículas:

Se conocen varios tipos de sistemas y/o formulaciones de suministro de partículas que son útiles en un espectro diverso de aplicaciones biomédicas. En general, una partícula se define como un objeto pequeño que se comporta como una unidad completa con respecto de su transporte y propiedades. Las partículas se clasifican además de acuerdo con su diámetro. Las partículas gruesas abarcan un intervalo entre 2.500 y 10.000 nanómetros. Las partículas finas tienen un tamaño entre 100 y 2.500 nanómetros. Las partículas ultrafinas, o nanopartículas, tienen generalmente un tamaño entre 1 y 100 nanómetros. La base del límite de 100 nm es el hecho de que novedosas propiedades que diferencian las partículas del material a granel aparezcan normalmente en una escala de longitud crítica por debajo de 100 nm.

Como se usa en la presente, un sistema/formulación de suministro de partículas se define como cualquier sistema/formulación de suministro biológico que incluye una partícula adecuada para su uso de acuerdo con los métodos y composiciones reivindicados de la presente invención. Una partícula de acuerdo con la presente invención es cualquier entidad que tiene una dimensión mayor (por ejemplo, diámetro) de menos de 100 micrómetros (µm). En algunas formas de realización, las partículas de la invención tienen una dimensión mayor de menos de 10 µm. En algunas formas de realización, las partículas de la invención tienen una dimensión mayor de menos de 2000 nanómetros (nm). En algunas formas de realización, las partículas de la invención tienen una dimensión mayor de menos de 1000 nanómetros (nm). En algunas formas de realización, las partículas de la invención tienen una dimensión mayor de menos de 900 nm, 800 nm, 700 nm, 600 nm, 500 nm, 400 nm, 300 nm, 200 nm, o 100 nm. Normalmente, las partículas de la invención tienen una dimensión mayor (por ejemplo, el diámetro) de 500 nm o menos. En algunas formas de realización, las partículas de la invención tienen una dimensión mayor (por ejemplo, el diámetro) de 250 nm o menos. En algunas formas de realización, las partículas de la invención tienen una dimensión mayor (por ejemplo, el diámetro) de 200 nm o menos. En algunas formas de realización, las partículas de la invención tienen una dimensión mayor (por ejemplo, el diámetro) de 150 nm o menos. En algunas formas de realización, las partículas de la invención tienen una dimensión mayor (por ejemplo, el diámetro) de 100 nm o menos. Las partículas más pequeñas, por ejemplo, que tienen una dimensión mayor de 50 nm o menos se usan en algunas formas de realización de la invención. En algunas formas de realización, las partículas de la invención tienen una dimensión mayor que varía entre 25 nm y 200 nm.

La caracterización de partículas (incluyendo, por ejemplo, la caracterización de la morfología, dimensión, etc.) se lleva a cabo usando una variedad de técnicas diferentes. Las técnicas comunes son la microscopía electrónica (TEM, SEM), microscopía de fuerza atómica (AFM), dispersión dinámica de luz (DLS), espectroscopía fotoelectrónica de rayos X (XPS), difracción de rayos × en polvo (XRD), espectroscopía infrarroja por transformada de Fourier (FTIR), espectrometría de masas mediante desorción/ionización láser asistida por matriz con tiempo de vuelo (MALDI-TOF), espectroscopía en el ultravioleta visible, interferometría de doble polarización y resonancia magnética nuclear (RMN). Se puede hacer la caracterización (medidas de dimensión) en lo que se refiere a las partículas nativas (es decir, la precarga) o tras cargar el material de carga (el material de carga en la presente se refiere a, por ejemplo, uno o más componentes del sistema CRISPR-Cas, por ejemplo, la enzima CRISPR o el ARNm o el ARN guía, o cualquier combinación de estos, y puede incluir portadores y/o excipientes adicionales) para proporcionar partículas de un tamaño óptimo para el suministro para cualquier aplicación *in vitro* y/o *ex vivo* de la presente invención. En ciertas formas de realización preferidas, la caracterización de la dimensión de las partículas (por ejemplo, el diámetro) se basa en medidas que utilizan la dispersión de láser dinámica (DLS). Se menciona la patente de Estados Unidos n.º 8.709.843; patente de Estados Unidos n.º 6.007.845; patente de Estados Unidos n.º 5.855.913; patente de Estados Unidos n.º 5.985.309; patente de Estados Unidos 5.543.158; y la publicación de James E. Dahlman y Carmen Barnes y col. Nature Nanotechnology (2014) publicada en línea el 11 de mayo de 2014, doi:10.1038/nnano.2014.84, que se refiere a partículas, métodos de preparación y de utilización y medidas de los mismos.

Los sistemas de suministro de partículas comprendidos en el alcance de la presente invención pueden proporcionarse en cualquier forma, incluido, sin limitación, partículas sólidas, semisólidas, en emulsión o partículas coloidales. Como cualquiera de los sistemas de suministro descritos en la presente, incluyendo pero sin limitación, se pueden proporcionar por ejemplo, sistemas basados en lípidos, liposomas, micelas, microvesículas, exosomas, o un cañón de genes como sistemas de suministro de partículas comprendidos en el alcance de la presente invención.

Partículas

Ha de apreciarse que la referencia que se hace en la presente a las partículas o a las nanopartículas, puede ser intercambiable, cuando sea apropiado. El ARNm de la enzima CRISPR y el ARN guía se pueden suministrar simultáneamente usando partículas o envolturas lipídicas; por ejemplo, se pueden suministrar la enzima CRISPR y el ARN usado en los métodos y composiciones reivindicados de la invención, por ejemplo, como un complejo, mediante una partícula como en Dahlman y col., WO2015089419 A2 y documentos que se citan allí, por ejemplo 7C1 (véase, por ejemplo, James E. Dahlman y Carmen Barnes y col. Nature Nanotechnology (2014) publicado en línea el 11 de mayo de 2014, doi:10.1038/nnano.2014.84), por ejemplo, suministro de partículas que comprenden un lípido o lipidoide y polímero hidrofílico, por ejemplo, un lípido catiónico y polímero hidrofílico, por ejemplo donde el lipido catiónico comprende 1,2-dioleoil-3-trimetilamonio-propano (DOTAP) o 1,2-ditetradecanoil-*sn*-glicero-3-fosfocolina (DMPC) y/o donde el polímero hidrofílico comprende etilenglicol o polietilenglicol (PEG); y/o donde la partícula comprende además colesterol (por ejemplo, la partícula de la formulación 1 = DOTAP 100, DMPC 0, PEG 0, Colesterol 0; formulación número 2 = DOTAP 90, DMPC 0, PEG 10, Colesterol 0; formulación número 3 = DOTAP 90, DMPC 0, PEG 5, Colesterol 5), donde las partículas se forman usando un eficiente, proceso multipasos donde primero se mezclan entre sí la proteína efectora y el ARN, por ejemplo, en una proporción molar 1:1, por ejemplo, a la temperatura ambiente, por ejemplo, durante 30 minutos, por ejemplo, en 1X PBS estéril y libre de nucleasa, y por separado, se disuelven en alcohol, por ejemplo, etanol al 100%, DOTAP, DMPC, PEG, y colesterol según corresponda para la formulación; y, las dos soluciones se mezclan entre sí para formar partículas que contienen los complejos).

El ARNm de proteínas efectoras del direccionamiento hacia ácidos nucleicos (tales como una proteína de tipo V, por ejemplo, Cpf1) y el ARN guía se pueden suministrar simultáneamente usando partículas o envolturas lipídicas. Los ejemplos de partículas adecuadas incluyen, a título enunciativo no taxativo, los que se describen en US 9.301.923.

Por ejemplo, Su X, Fricke J, Kavanagh DG, Irvine DJ ("*In vitro* and *in vivo* mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. junio de 2011 6;8(3):774-87. doi: 10.1021/mp100390w. La publicación electrónica del 1 de abril de 2011) describe nanopartículas biodegradables con una estructura núcleo-coraza con un núcleo de poli(β-aminoéster)(PBAE) envuelto por una coraza de tipo bicapa fosfolipídica. Estas se desarrollaron para el suministro de ARNm *in vivo*. Se escogió el componente de tipo PBAE sensible al pH para promover la desorganización de endosomas, mientras que la capa de la superficie lipídica se seleccionó para minimizar la toxicidad del núcleo policatiónico. Estos son, por lo tanto, los sistemas preferidos para el suministro de ARN de la presente invención.

En una forma de realización, se contemplan partículas/nanopartículas basadas en polímeros bioadhesivos autoensamblables, que se podrán aplicar para el suministro oral de péptidos, suministro intravenoso de péptidos y suministro nasal de péptidos, en todos los casos al cerebro. También se contemplan otras formas de realización, tales como la absorción oral y el suministro ocular de fármacos hidrofóbicos. En la tecnología de la envoltura molecular participa una envoltura polimérica manipulada que se protege y se suministra en el sitio de la enfermedad (véase, por ejemplo, Mazza, M. y col., ACSNano, 2013. 7(2): 1016-1026; Siew, A., y col. Mol Pharm, 2012. 9(1):14-28; Lalatsa, A., y col. J Contr Rel, 2012. 161(2):523-36; Lalatsa, A., *y col.*, *Mol Pharm*, 2012. 9(6):1665-80; Lalatsa, A., y col. Mol Pharm, 2012. 9(6):1764-74; Garrett, N.L., y col. J Biophotonics, 2012. 5(5-6):458-68; Garrett, N.L., y col. J Raman Spect, 2012. 43(5):681-688; Ahmad, S., y col. J Royal Soc Interface 2010. 7:S423-33; Uchegbu, I.F. *Expert Opin Drug Deliv*, 2006. 3(5):629-40; Qu, X. *y col., Biomacromolecules*, 2006. 7(12):3452-9 y Uchegbu, I.F., y col. Int J Pharm, 2001. 224:185-199). Se contemplan dosis de aproximadamente 5 mg/kg, con dosis únicas o múltiples, dependiendo del tejido diana.

En una forma de realización, se podrán utilizar y/o adaptar al sistema CRISPR-Cas usado en los métodos y composiciones reivindicados de la presente invención, partículas/nanopartículas que pueden suministrar ARN a una célula cancerosa para detener el crecimiento tumoral, desarrolladas por el laboratorio de Dan Anderson en el MIT. En particular, el laboratorio de Anderson ha desarrollado sistemas combinatorios, totalmente automáticos para la síntesis, purificación, caracterización y formulación de nuevos biomateriales y nanoformulaciones. Véase, por ejemplo, Alabi *y col.*, *Proc Natl Acad Sci U S A.* 6 de agosto de 2013;110(32):12881-6; Zhang *y col.*, *Adv Mater.* 6 de septiembre de 2013;25(33):4641-5; Jiang y col., Nano Lett. 13 de marzo de 2013;13 (3):1059-64; Karagiannis y col., ACS Nano. 23 de octubre de 2012;6(10):8484-7; Whitehead *y col.*, ACS Nano. 28 de agosto de 2012;6(8):6922-9 y Lee *y col.*, *Nat Nanotechnol.* 3 de junio de 2012;7(6):389-93.

La solicitud de patente de los EE.UU. 20110293703 se refiere a compuestos lipídicos que también son particularmente útiles en la administración de polinucleótidos, que se podrán aplicar al suministro del sistema CRISPR-Cas usado en los métodos y composiciones reivindicados de la presente invención. En un aspecto, los compuestos lipídicos aminoalcohólicos se combinan con un agente que va a ser suministrado a una célula o un sujeto para formar micropartículas, nanopartículas, liposomas o micelas. El agente que va a ser suministrado por las partículas, liposomas o micelas podrá estar en forma de un gas, líquido o sólido y el agente podrá ser un polinucleótido, proteína, péptido o molécula de bajo peso molecular. Los compuestos lipídicos aminoalcohólicos se podrán combinar con otros compuestos lipídicos aminoalcohólicos, polímeros (sintéticos o naturales), tensioactivos, colesterol, carbohidratos, proteínas, lípidos, etc. Para formar las partículas. Estas partículas se podrán combinar a continuación opcionalmente con un excipiente farmacéutico para formar una composición farmacéutica.

La patente de los EE.UU. con n.º de publicación 20110293703 también proporciona métodos para preparar los compuestos lipídicos aminoalcohólicos. Se permite que uno o más equivalentes de una amina reaccionen con uno o

más equivalentes de un compuesto con un epóxido terminal en condiciones adecuadas para formar un compuesto lipídico aminoalcohólico adecuado para su uso en los métodos y composiciones reivindicados de la presente invención. En ciertas formas de realización, todos los grupos amino de la amina se hacen reaccionar totalmente con el compuesto con un epóxido terminal para formar aminas terciarias. En otras formas de realización, todos los grupos amino de la amina se hacen reaccionar de manera no total con el compuesto con un epóxido terminal para formar aminas terciarias y de esta manera se obtienen como resultado aminas primarias o secundarias en el compuesto lipídico aminoalcohólico. Estas aminas primarias o secundarias se dejan tal cual o se podrán hacer reaccionar con otro electrófilo tal como un compuesto con epóxido terminal diferente. Como apreciará el experto en la técnica, la reacción de una amina con una cantidad que es menor que un exceso del compuesto con un epóxido terminal dará como resultado varios compuestos lipídicos aminoalcohólicos diferentes con diversos números de colas. Ciertas aminas podrán estar totalmente funcionalizadas con dos colas que son compuestos derivados de epóxidos mientras que otras moléculas no estarán completamente funcionalizadas con colas que son compuestos derivados de epóxidos. Por ejemplo, una diamina o poliamina podrá incluir una, dos, tres o cuatro colas que son compuestos derivados de epóxidos unidos a varios restos amino de la molécula lo que da como resultado aminas primarias, secundarias y terciarias. En ciertas formas de realización, todos los grupos amino no están totalmente funcionalizados. En ciertas formas de realización, se utilizan dos compuestos con un epóxido terminal del mismo tipo. En otras formas de realización, se utilizan dos o más compuestos con un epóxido terminal diferentes. La síntesis de los compuestos lipídicos aminoalcohólicos se realiza con o sin disolvente y la síntesis se podrá realizar a temperaturas más elevadas que están comprendidas entre 30 y 100 ºC, preferentemente a aproximadamente 50-90 ºC. Los compuestos lipídicos aminoalcohólicos preparados se podrán purificar opcionalmente. Por ejemplo, la mezcla de compuestos lipídicos aminoalcohólicos se podrá purificar para generar un compuesto lipídico aminoalcohólico con un número concreto de colas que son un compuesto derivado de un epóxido. O la mezcla se podrá purificar para generar un estereo- o regioisómero concreto. Los compuestos lipídicos aminoalcohólicos también se podrán alquilar utilizando un haluro de alquilo (por ejemplo, yoduro de metilo) u otro agente alquilante y/o se podrán acilar.

La patente de los EE.UU. con n.º de publicación 20110293703 también proporciona bibliotecas de compuestos lipídicos aminoalcohólicos preparados con los métodos de la invención. Estos compuestos lipídicos aminoalcohólicos se podrán preparar y/o cribar utilizando técnicas ultrarrápidas que conllevan manipuladores de líquidos, robots, placas de microvaloración, computadoras, etc. En ciertas formas de realización, los compuestos lipídicos aminoalcohólicos se criban para determinar su capacidad de transfectar la célula con polinucleótidos u otros agentes (por ejemplo, proteínas, péptidos, moléculas de bajo peso molecular).

La patente de los EE.UU. con n.º de publicación 20130302401 se refiere a una clase de poli(beta-aminoalcoholes) (PBAA) que se ha preparado utilizando polimerización combinatoria. Los PBAA de la invención se podrán utilizar en aplicaciones biomédicas y biotecnológicas tales como recubrimientos (tales como recubrimientos de películas o películas multicapa para implantes o dispositivos médicos), aditivos, materiales, excipientes, agentes para evitar la bioincrustación, agentes para la miniaturización de patrones y agentes de encapsulación celular. Cuando se utilizan como recubrimientos de superficies, estos PBAA suscitan diferentes niveles de inflamación, tanto *in vitro* como *in vivo*, dependiendo de sus estructuras químicas. La gran diversidad química de esta clase de materiales ha permitido a los autores identificar recubrimientos poliméricos que inhiben la activación de macrófagos *in vitro*. Además, estos recubrimientos reducen el reclutamiento de células inflamatorias y reducen la fibrosis, tras la implantación subcutánea de micropartículas de poliestireno carboxilado. Estos polímeros se podrán utilizar para formar cápsulas de complejos polielectrolíticos para la encapsulación celular. La invención también podrá tener otras aplicaciones biológicas tales como recubrimientos antimicrobianos, suministro de ADN o ARNip y modificación tisular con células madre. Los contenidos de la patente de los EE.UU. con n.º de publicación 20130302401 se podrán aplicar al sistema CRISPR-Cas usado en la composición y métodos reivindicados de la presente invención. En algunas formas de realización, se pueden usar partículas basadas en azucares, por ejemplo, GalNAc, como se describe en la presente y en referencia a WO2014118272 y Nair, JK y col., 2014, Journal of the American Chemical Society 136 (49), 16958-16961) y los contenidos de la misma, en especial con respecto a la administración que se aplica a todas las partículas a menos que sea evidente otra cosa.

En otra forma de realización, se contemplan nanopartículas lipídicas (LNP, por sus siglas en inglés). En particular, se podrá aplicar un ARN de interferencia pequeño de antitranstiretina encapsulado en nanopartículas lipídicas (véase, por ejemplo, Coelho y col., N Engl J Med 2013;369:819-29), y dicho sistema se puede adaptar y aplicar al sistema CRISPR-Cas usado en los métodos y composiciones reivindicados de la presente invención. Se contemplan dosis comprendidas entre aproximadamente 0,01 y aproximadamente 1 mg por kg de peso corporal administradas por vía intravenosa. Se contemplan medicaciones para reducir el riesgo de reacciones relacionadas con la infusión, por ejemplo, se contemplan dexametasona, acetaminofeno, difenhidramina o cetirizina y ranitidina. También se contemplan dosis múltiples de aproximadamente 0,3 mg por kilogramo cada 4 semanas durante cinco dosis.

Se ha demostrado que las LNP son sumamente eficaces para suministrar ARNip al hígado (véase, por ejemplo, Tabernero y col., Cancer Discovery, abril de 2013, Vol. 3, n.º 4, páginas 363-470) y por lo tanto se contemplan para suministrar ARN que codifica para CRISPR Cas hacia el ácido nucleico al hígado. Se podrá contemplar una dosificación de aproximadamente cuatro dosis de 6 mg/kg de la LNP cada dos semanas. Tabernero *y col.* han demostrado que se observó una regresión tumoral después de los 2 primeros ciclos de LNP con una dosificación de 0,7 mg/kg, y que al final de los 6 ciclos el paciente había logrado una respuesta parcial con una regresión completa de la metástasis de los nódulos linfáticos y una reducción sustancial del volumen de los tumores hepáticos. Se obtuvo

una respuesta completa tras 40 dosis en este paciente, que ha permanecido en remisión y completado el tratamiento tras recibir dosis a lo largo de 26 meses. Dos pacientes con RCC y sitios de enfermedad extrahepáticos que incluían el riñón, pulmón y nódulos linfáticos que estaban evolucionando tras la terapia inicial con inhibidores de la ruta VEGF experimentaron una estabilización de la enfermedad en todas las zonas durante aproximadamente 8 a 12 meses y un paciente con PNET y metástasis hepáticas continuó en la extensión del estudio durante 18 meses (36 dosis) con una enfermedad estabilizada.

Sin embargo, debe tenerse en cuenta la carga de la LNP. Los lípidos catiónicos se combinaron con lípidos cargados negativamente para inducir estructuras que no son bicapas para facilitar el suministro intracelular. Debido a que las LNP cargadas se eliminan rápidamente de la circulación tras la inyección intravenosa, se desarrollaron lípidos catiónicos ionizables con valores de pKa por debajo de 7 (véase, por ejemplo, Rosin *y col.*, Molecular Therapy, vol. 19, n.º 12, páginas 1286-2200, diciembre de 2011). Se pueden cargar polímeros cargados negativamente tales como ARN en los LNP con valores de pH bajos (por ejemplo, pH 4) donde los lípidos ionizables muestran una carga positiva. Sin embargo, a valores de pH fisiológicos, las LNP exhiben una carga superficial baja compatible con tiempos de circulación más prolongados. Cuatro especies de lípidos catiónicos ionizables han recibido una atención especial, concretamente 1,2-dilineoil-3-dimetilamoniopropano (DLinDAP), 1,2-dilinoleiloxi-3-*N,N*-dimetilaminopropano (DLinDMA), 1,2-dilinoleiloxiceto-*N,N*-dimetil-3-aminopropano (DLinKDMA) y 1,2-dilinoleil-4-(2-dimetilaminoetil)-[1,3]-dioxolano (DLinKC2-DMA). Se ha demostrado que sistemas LNP ARNip que contienen estos lípidos muestran unas propiedades de silenciamiento génico notablemente diferentes en hepatocitos *in vivo*, con potencias que varían de acuerdo con la serie DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP empleando un modelo de silenciamiento del gen del Factor VII (véase, por ejemplo, Rosin *y col.*, *Molecular Therapy*, vol. 19, n.º 12, páginas 1286-2200, diciembre de 2011). Puede contemplarse una dosificación de 1 µg/ml de LNP o ARN de CRISPR-Cas en, o asociada con la LNP, especialmente para una formulación que contiene DLinKC2-DMA.

La preparación de las LNP y la encapsulación de CRISPR Cas se podrán utilizar y/o adaptar de Rosin y col., Molecular Therapy, vol. 19, n.º 12, páginas 1286-2200, diciembre de 2011). Los lípidos catiónicos 1,2-dilineoil-3-dimetilamoniopropano (DLinDAP), 1,2-dilinoleiloxi-3-*N,N*-dimetilaminopropano (DLinDMA), 1,2-dilinoleiloxiceto-*N,N*-dimetil-3-aminopropano (DLinK-DMA), 1,2-dilinoleil-4-(2-dimetilaminoetil)-[1,3]-dioxolano (DLinKC2-DMA), (3-o-[2''-(metoxipolietilenglicol 2000)succinoíl]-1,2-dimiristoil-*sn*-glicol (PEG-S-DMG) y *R*-3-[(ω-metoxi-polietilenglicol)2000)carbamoil]-1,2-dimiristiloxlpropil-3-amina (PEG-C-DOMG) se podrán adquirir de Tekmira Pharmaceuticals (Vancouver, Canadá) o sintetizar. El colesterol se podrá adquirir de Sigma (San Luis, MO). Se podrá encapsular el ARN específico para CRISPR-Cas en las LNP que contienen DLinDAP, DLinDMA, DLinK-DMA y DLinKC2-DMA (lípidos catiónicos:DSPC:COL:PEGS-DMG o PEG-C-DOMG con relaciones molares de 40:10:40:10). Cuando sea necesario, se podrá incorporar un 0,2% de SP-DiOC18 (Invitrogen, Burlington, Canadá) para evaluar la captación celular, suministro intracelular y biodistribución. La encapsulación se podrá realizar disolviendo mezclas lípidicas que comprenden lípidos catiónicos:DSPC:colesterol:PEG-c-DOMG (con relaciones molares de 40:10:40:10) en etanol hasta alcanzar una concentración lipídica final de 10 mmol/L. Esta solución etanólica de lípidos se podrá añadir gota a gota a una solución de 50 mmol/L de citrato, pH 4,0 para formar vesículas multilaminares para producir una concentración final de un 30% de etanol vol/vol. Se podrán formar vesículas unilaminares grandes tras la extrusión de vesículas multilaminares a través de dos filtros de policarbonato Nuclepore de 80 nm apilados utilizando el Extrusor (Northern Lipids, Vancouver, Canadá). Se podrá lograr la encapsulación añadiendo ARN disuelto con una concentración de 2 mg/mL en una solución de 50 mmol/L de citrato, pH 4,0 que contenga un 30% de etanol vol/vol gota a gota a las vesículas unilaminares grandes preformadas extruídas e incubando a 31 ºC durante 30 minutos con una agitación constante hasta alcanzar una relación ponderal ARN/lípidos final de 0,06/1 p/p. La eliminación del etanol y la neutralización del regulador de pH de formulación se realizaron por diálisis frente a una solución salina de pH regulado con fosfato (PBS), pH 7,4 durante 16 horas utilizando membranas de diálisis de celulosa regenerada Spectra/Por 2. La distribución de tamaños de las nanopartículas se podrá determinar mediante dispersión dinámica de la luz utilizando un clasificador de las partículas por tamaño NICOMP 370, los modos vesícula/intensidad y el ajuste Gaussiano (Nicomp Particle Sizing, Santa Barbara, CA). El tamaño de partícula para los tres sistemas de LNP puede ser de ~70 nm de diámetro. Se puede determinar la eficacia de encapsulación del ARN mediante la eliminación del ARN libre utilizando columnas VivaPureD Minh (Sartorius Stedim Biotech) de muestras recogidas antes y después de la diálisis. El ARN encapsulado se puede extraer de las nanopartículas eluidas y cuantificarse a 260 nm. Se determinó la relación ARN a lípido por medida del contenido de colesterol en vesículas utilizando el ensayo enzimático del colesterol E de Wako Chemicals USA (Richmond, VA). Junto con la discusión de la presente sobre las LNP y los lípidos PEG, los liposomas PEGilados o las LNP son igualmente adecuados para el suministro de un sistema CRISPR-Cas o los componentes del mismo.

Se podrá utilizar y/o adaptar la preparación de LNP grandes de Rosin y col., Molecular Therapy, vol. 19, n.º 12, páginas 1286-2200, diciembre de 2011. Se podrá preparar una solución de premezcla lipídica (concentración lipídica total de 20,4 mg/mL) en etanol que contenga DLinKC2-DMA, DSPC y colesterol con relaciones molares de 50:10:38.5. Se podrá añadir acetato de sodio a la premezcla lipídica con una relación molar de 0,75:1 (acetato de sodio:DLinKC2-DMA). Los lípidos se podrán hidratar posteriormente combinando la mezcla con 1,85 volúmenes de regulador de pH citrato (10 mmol/L, pH 3,0) agitando vigorosamente, para dar como resultado la formación de liposomas espontánea en regulador de pH acuoso que contenga un 35% de etanol. La solución con liposomas se podrá incubar a 37 ºC para permitir un incremento dependiente del tiempo del tamaño de las partículas. Se podrán retirar alícuotas en diversos momentos durante la incubación para estudiar cambios en el tamaño de los liposomas mediante dispersión dinámica

de la luz (Zetasizer Nano ZS, Malvern Instruments, Worcestershire, Reino Unido). Una vez que se ha logrado el tamaño de partícula deseado, se podrá añadir una solución de PEG lípidos acuosa (solución madre = 10 mg/mL de PEG-DMG en etanol al 35% (vol/vol) a la mezcla de liposomas para obtener una concentración molar de PEG final de un 3,5% de los lípidos totales. Tras la adición del PEG-lípidos, los liposomas deberían su tamaño, y detener de manera eficaz un crecimiento adicional. A continuación, se podrá añadir ARN a los liposomas vacíos con una relación de ARN respecto de los lípidos totales de aproximadamente 1:10 (p:p), seguido por una incubación durante 30 minutos a 37 ºC para formar LNP cargadas. Posteriormente, la mezcla se podrá dializar durante toda la noche en PBS y filtrar con un filtro de jeringa de 0,45 µm.

También se contemplan construcciones de ácido nucleico esférico (SNA™) y otras nanopartículas (especialmente nanopartículas de oro) como medios para suministrar el sistema CRISPR-Cas a los blancos previstos. Se dispone de datos significativos que muestran que las construcciones de ácido nucleico esférico de AuraSense Therapeutics (SNA™), basados en nanopartículas de oro funcionalizadas con ácido nucleico, que son útiles.

La bibliografía que se puede emplear junto con las divulgaciones de la presente incluye: Cutler y col., J. Am. Chem. Soc. 2011 133:9254-9257, Hao *y col.*, Small. 2011 7:3158-3162, Zhang *y col.*, ACS Nano. 2011 5:6962-6970, Cutler y col., J. Am. Chem. Soc. 2012 134:1376-1391, Young y col., Nano Lett. 2012 12:3867-71, Zheng y col., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang y col., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi y col., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen y col., Sci. Transl. Med. 5, 209ra152 (2013) y Mirkin, y col., Small, 10:186-192.

Se pueden construir nanopartículas autoensamblables con ARN con polietilenimina (PEI) que está PEGilada con un ligando peptídico Arg-Gly-Asp (RGD) unido al extremo distal del polietilenglicol (PEG). Este sistema se ha utilizado, por ejemplo, como medio para dirigirse a la neovasculatura tumoral que expresa integrinas y suministrar ARNip que inhibe la expresión del receptor-2 del factor de crecimiento endotelial vascular (VEGF R2) y conseguir de esta forma la angiogénesis tumoral (véase, por ejemplo, Schiffelers y col., Nucleic Acids Research, 2004, Vol. 32, No. 19). Se podrán preparar *nanoplexos* mezclando volúmenes iguales de soluciones acuosas de un polímero catiónico y un ácido nucleico para proporcionar un exceso molar neto de nitrógeno ionizable (polímero) respecto al fosfato (ácido nucleico) en el intervalo de 2 a 6. Las interacciones electrostáticas entre polímeros catiónicos y ácido nucleico dieron como resultado la formación de *poliplexos* con una distribución de tamaños de partículas promedio de aproximadamente 100 nm, denominadas pues en la presente nanoplexos. Se prevé una dosificación comprendida entre aproximadamente 100 y 200 mg de CRISPR-Cas para el suministro en las nanopartículas autoensamblables de Schiffelers y col.

Los nanoplexos de Bartlett y col. (PNAS, 25 de septiembre de 2007, vol. 104, n.º 39) también se podrán aplicar a la presente invención. Los *nanoplexos* de Bartlett *y col.* se preparan mezclando volúmenes iguales de soluciones acuosas de un polímero catiónico y un ácido nucleico para proporcionar un exceso molar neto de nitrógeno ionizable (polímero) respecto al fosfato (ácido nucleico) en el intervalo de 2 a 6. Las interacciones electrostáticas entre polímeros catiónicos y ácido nucleico dieron como resultado la formación de *poliplexos* con una distribución de tamaños de partículas promedio de aproximadamente 100 nm, denominadas pues en la presente *nanoplexos*. El DOTA-ARNip de Bartlett *y col.* Se sintetizó de la siguiente manera: Se encargó el monoéster de la *N*-hidroxisuccinimida y el ácido 1,4,7,10-tetraazaciclododecano-1,4,7,10-tetraacético (DOTA-ésterNHS) de Macrocyclics (Dallas, TX). Se añadió la hebra sentido de ARN modificada en la amina con un exceso molar de 100 veces de DOTA-éster-NHS en regulador de pH de carbonato (pH 9) a un tubo de microcentrífuga. Los contenidos se hicieron reaccionar agitando durante 4 horas a temperatura ambiente. El conjugado de DOTA-ARN sentido se precipitó en etanol, se resuspendió en agua y se hibridó con la hebra antisentido no modificada para generar DOTA-ARNip. Se pretrataron todos los líquidos con Chelex-100 (Bio-Rad, Hercules, CA) para eliminar todas las trazas de contaminantes metálicos. Se podrán formar nanopartículas de ARNip dirigidas y no dirigidas al Tf utilizando policationes que contienen ciclodextrina. Normalmente, se formaron nanopartículas en agua con una relación de carga de 3 (+/-) y una concentración de ARNip de 0,5 g/litro. Un uno por ciento de las moléculas de adamantano-PEG de la superficie de las nanopartículas dirigidas se modificaron con Tf (adamantano-PEG-Tf). Se suspendieron las nanopartículas en una solución portadora de glucosa al 5% (p/v) inyectable.

Davis y col. (Nature, Vol 464, 15 de abril de 2010) llevan a cabo un ensayo clínico con ARN que utiliza un sistema de suministro con nanopartículas dirigidas (número de registro del ensayo clínico NCT00689065). A los pacientes con distintos tipos de cáncer sólido resistente a las terapias asistenciales habituales se les administran dosis de las nanopartículas dirigidas en los días 1, 3, 8 y 10 de un ciclo de 21 días mediante una infusión intravenosa de 30 min. Las nanopartículas consisten en a sistema de suministro sintético que contiene: (1) un polímero lineal basado en la ciclodextrina (CDP, por sus siglas en inglés), (2) un ligando cuya diana sea una proteína de tipo transferrina (TF) humana presentado en el exterior de la nanopartícula para unirse a los receptores de TF (TFR) en la superficie de las células cancerosas, (3) un polímero hidrofílico (polietilenglicol (PEG) utilizado para promover la estabilidad de las nanopartículas en fluidos biológicos) y (4) un ARNip diseñado para reducir la expresión de RRM2 (la secuencia utilizada en la clínica se denominó previamente siR2B+5). Hace tiempo que existe constancia del aumento de TFR en las células malignas y RRM2 es una diana contra el cáncer conocida. En estudios multidosis en primates no humanos se ha demostrado que estas nanopartículas (versión clínica denominada CALAA-01) se toleran bien. Aunque se ha administrado ARNip mediante suministro con liposomas a un único paciente con leucemia mieloide crónica, el ensayo clínico de Davis y col. es el ensayo inicial en seres humanos para suministrar de manera sistémica ARNip con un

sistema de suministro dirigido y para tratar pacientes con un cáncer sólido. Para averiguar si el sistema de suministro dirigido puede proporcionar el suministro eficaz de ARNip funcional a los tumores humanos, Davis *y col.*, estudiaron biopsias de tres pacientes de tres cohortes de dosificación diferentes; los pacientes A, B y C, todos los cuales tenían un melanoma metastásico y recibieron dosis CALAA-01 de 18, 24 y 30 mg m^{-2} de ARNip, respectivamente. También se podrán contemplar dosis similares para el sistema CRISPR Cas usado en los métodos y sistemas reivindicados de la presente invención. El suministro de la invención se podrá lograr con nanopartículas que contengan un polímero lineal basado en la ciclodextrina (CDP), un ligando cuyo diana es un tipo de proteína transferrina (TF) humana presentada en el exterior de las nanopartículas para unirse a los receptores de la TF (TFR) en la superficie de las células cancerosas y/o un polímero hidrofílico (por ejemplo, polietilenglicol (PEG) utilizado para promover la estabilidad de las nanopartículas en los fluidos biológicos).

En términos de la presente invención, es preferible tener uno o más componentes del complejo CRISPR, por ejemplo, la enzima CRISPR o ARNm, o ARN guía que se suministra usando nanopartículas o envolturas lipídicas. Otros sistemas o vectores de suministro se pueden usar junto con los aspectos de la nanopartícula de la invención.

En general, una "nanopartícula" se refiere a cualquier partícula que tiene un diámetro de menos de 1000 nm. En ciertas formas de realización preferidas, las nanopartículas de la invención tienen una dimensión mayor (por ejemplo, el diámetro) de 500 nm o menos. En otras formas de realización preferidas, las nanopartículas de la invención tienen una dimensión mayor que varía entre 25 nm y 200 nm. En otras formas de realización preferidas, las nanopartículas de la invención tienen una dimensión mayor de 100 nm o menos. En otras formas de realización preferidas, las nanopartículas de la invención tienen una dimensión mayor que varía entre 35 nm y 60 nm.

Las nanopartículas abarcadas en la presente invención pueden proporcionarse en diferentes formas, por ejemplo, como nanopartículas sólidas (por ejemplo, metales tales como plata, oro, hierro, titanio), no metales, sólidos basados en lípidos, polímeros), suspensiones de nanopartículas o combinaciones de estas. Se pueden preparar nanopartículas metálicas, dieléctricas y semiconductoras, así como estructuras híbridas (por ejemplo, nanopartículas de tipo núcleo-coraza). Las partículas preparadas con materiales semiconductores también pueden etiquetarse con puntos cuánticos si son suficientemente pequeñas (normalmente, por debajo de 10 nm) de tal manera que se produzca la cuantización de los niveles electrónicos de energía. Dichas partículas a nanoescala se usan en aplicaciones biomédicas como portadores de fármaco o agentes de formación de imágenes y se pueden adaptar para fines similares en la presente invención.

Se han fabricado nanopartículas semisólidas y blandas, y están comprendidas en el alcance de la presente invención. Un prototipo de nanopartícula de naturaleza semisólida es el liposoma. En la actualidad se usan clínicamente diversos tipos de nanopartículas liposómicas como sistemas de suministro para fármacos y vacunas anticancerosas. Las nanopartículas con una mitad hidrofílica y la otra mitad hidrofóbica se denominan partículas Jano y son particularmente eficaces para estabilizar emulsiones. Se pueden autoensamblar en interfases de agua/aceite y actúan como tensioactivos sólidos.

La patente de Estados Unidos n.º 8.709.843 proporciona un sistema de suministro de fármacos para el suministro dirigido de partículas que contienen un agente terapéutico a tejidos, células, y compartimentos intracelulares. Los métodos y composiciones reivindicados de la invención pueden comprender partículas dirigidas que comprenden un polímero conjugado con un tensioactivo, un polímero hidrofílico o un lípido.

La patente de Estados Unidos n.º 6.007.845 proporciona partículas que tienen un núcleo de copolímero multibloques que se forma enlazando covalentemente un compuesto multifuncional con uno o más polímeros hidrofóbicos y uno o más polímeros hidrofílicos y que contienen un material biológicamente activo.

La patente de Estados Unidos n.º 5.855.913 proporciona una composición particulada que tiene partículas aerodinámicamente ligeras que tienen una densidad aparente menor de 0,4 g/cm^3 con un diámetro promedio de entre 5 µm y 30 µm, que incorpora un tensioactivo sobre la superficie de estas para el suministro de fármaco al sistema pulmonar.

La patente de Estados Unidos n.º 5.985.309 proporciona partículas que incorporan un tensioactivo y/o un complejo hidrofílico o hidrofóbico de un agente terapéutico o diagnóstico cargado positiva o negativamente y una molécula cargada de carga opuesta para el suministro al sistema pulmonar.

La patente de Estados Unidos n.º 5.543.158 proporciona partículas inyectables biodegradables que tienen un núcleo sólido biodegradable que contiene un material biológicamente activo y restos de poli(alquilenglicol) sobre la superficie.

El documento WO2012135025 (publicado también como US20120251560) describe polímeros de polietilenimina conjugados (PEI) y aza-macrociclos conjugados (denominados en su conjunto "lipómero conjugado" o "lipómeros"). Puede contemplarse que dichos lipómeros conjugados pueden utilizarse en el contexto del sistema CRISPR-Cas para conseguir perturbaciones genómicas *in vitro*, *ex vivo* e *in vivo* para modificar la expresión génica, incluyendo la modulación de la expresión de las proteínas.

En una forma de realización, la nanopartícula puede ser un polímero lipídico manipulado con epóxido, de forma ventajosa 7C1 (véase, por ejemplo, James E. Dahlman y Carmen Barnes y col. Nature Nanotechnology (2014)

publicado online el 11 de mayo de 2014, doi:10.1038/nnano,2014.84). Se sintetizó C71 haciendo reaccionar los lípidos terminados en el epóxido C15 con PEI600 en una relación molar 14:1, y se formuló con C14PEG2000 para producir nanopartículas (diámetro entre 35 y 60 nm) que eran estables en solución PBS durante al menos 40 días.

Se puede utilizar un polímero lipídico manipulado con epóxido para suministrar el sistema CRISPR-Cas en los métodos y composiciones reivindicados de la presente invención a células pulmonares, cardiovasculares o renales, sin embargo, una persona experta en la materia puede adaptar el sistema para suministrar a otros órganos diana. Se contempla una variación de la dosificación desde aproximadamente 0,05 a aproximadamente 0,6 mg/kg. Se contemplan también dosificaciones de aproximadamente varios día o semanas, con una dosificación total de aproximadamente 2 mg/kg

Exosomas

Los exosomas son nanovesículas endógenas que trasportan los ARN y las proteínas, y que pueden suministrar ARN al cerebro y otros órganos diana. Para reducir la inmunogenicidad, Alvarez-Erviti y col. (2011, Nat Biotechnol 29: 341) utilizaron células dendríticas autoderivadas para la producción de exosomas. Se logró el direccionamiento al cerebro modificando las células dendríticas para que expresaran Lamp2b, una proteína de membrana exosómica, fusionada con el péptido RVG específico de las neuronas. Los exosomas purificados se cargaron con ARN exógeno mediante electroporación. La inyección intravenosa de exosomas dirigidos con RVG suministró ARNip GAPDH específicamente a las neuronas, microglía, oligodendrocitos en el cerebro, lo que dio como resultado una atenuación génica específica. La preexposición a los exosomas RVG no atenuó la atenuación génica y no se observó una captación no específica en otros tejidos. Se demostró el potencial terapéutico del suministro de ARNip mediado por exosomas mediante la fuerte atenuación de la proteína (62%) y del ARNm (60%) de BACE1, una diana terapéutica en la enfermedad de Alzheimer.

Para obtener un conjunto de exosomas inmunológicamente inertes, Alvarez-Erviti y col. recolectaron médula ósea de ratones endogámicos C57BL/6 con un haplotipo homogéneo del complejo mayor de histocompatibilidad (MHC, por sus siglas en inglés). Ya que las células dendríticas inmaduras producen grandes cantidades de exosomas desprovistos de activadores de los linfocitos T tales como MHC-II y CD86, Alvarez-Erviti *y col.* seleccionaron células dendríticas con el factor estimulador de colonias de macrófagos/granulocitos (GM-GSF, por sus siglas en inglés) durante 7 d. Se purificaron los exosomas del sobrenadante del cultivo al día siguiente utilizando protocolos sobradamente establecidos de ultracentrifugación. Los exosomas producidos fueron físicamente homogéneos, con una distribución de tamaños que tenía un máximo en 80 nm de diámetro tal como lo determinó el análisis de rastreo de nanopartículas (NTA, por sus siglas en inglés) y la microscopía electrónica. Alvarez-Erviti *y col.* obtuvieron 6-12 µg de exosomas (medidos basándose en la concentración de proteína) por 10^6 células.

A continuación, Alvarez-Erviti y col. estudiaron la posibilidad de cargar exosomas modificados con material de carga exógeno utilizando protocolos de electroporación adaptados para las aplicaciones a nanoescala. Ya que la electroporación para las partículas con membrana a escala nanométrica no está bien caracterizada, se utilizó ARN marcado con Cy5 no específico para la optimización empírica del protocolo de electroporación. Se determinó mediante un ensayo la cantidad de ARN encapsulado tras la ultracentrifugación y lisis de los exosomas. La electroporación a 400 V y 125 µF dio como resultado la mayor retención de ARN y se utilizó para todos los experimentos posteriores.

Alvarez-Erviti y col. administraron 150 µg de cada ARNip de BACE1 encapsulado en 150 µg de exosomas RVG a ratones C57BL/6 normales y se comparó con la eficacia de atenuación génica en cuatro controles: ratones no tratados, ratones a los que se inyectó únicamente exosomas RVG, ratones a los que se inyectó ARNip de BACE1 complejado con un reactivo liposómico catiónico *in vivo* y ratones a los que se inyectó ARNip de BACE1 complejado con RVG-9R, el péptido RVG conjugado con 9 D-argininas que se unen electrostáticamente al ARNip. Se analizaron muestras tisulares corticales 3 d después de la administración y se observó una atenuación significativa de la proteína (45%, $P < 0,05$, frente a 62%, $P < 0,01$) tanto en ratones tratados con ARNip-RVG-9R como tratados con exosomas con ARNipRVG, lo que fue el resultado de un descenso significativo en niveles de ARNm de BACE1 (66% [+ o -] 15%, $P < 0,001$ y 61% [+ o -] 13% respectivamente, $P < 0,01$). Además, los solicitantes demostraron un descenso significativo (55%, $P < 0,05$) en los niveles totales de [beta]-amiloide 1-42, un componente principal de las placas de amiloide en la patología de Alzheimer, en los animales tratados con exosomas RVG. El descenso observado fue superior al descenso de β-amiloide 1-40 demostrado en ratones normales tras la inyección intraventricular de inhibidores de BACE1. Alvarez-Erviti *y col.* llevaron a cabo una amplificación rápida en 5' de los extremos del ADNc (RACE) en el producto de clivaje de BACE1, que proporcionó evidencia de la atenuación génica mediada por iARN por parte del ARNip.

Finalmente, Alvarez-Erviti y col. estudiaron si los exosomas con ARN-RVG inducían respuestas inmunitarias *in vivo* evaluando las concentraciones séricas de IL-6, IP-10, TNFα e IFN-α. Tras el tratamiento de los exosomas no se registraron cambios significativos en todas las citocinas similares al tratamiento con el reactivo de transfección de ARNip a diferencia de ARNip-RVG-9R, que estimuló notablemente la secreción de IL-6, y confirmó de esta manera el perfil inmunológicamente inerte del tratamiento con exosomas. Dado que los exosomas encapsulan únicamente un 20% del ARNip, el suministro con exosomas con RVG parece ser más eficaz que el suministro con RVG-9R ya que se logró una atenuación génica del ARNm comparable y una atenuación génica de la proteína mayor con una cantidad de ARNip cinco veces menor sin el nivel correspondiente de estimulación inmunitaria. Este experimento demostró el

potencial terapéutico de la tecnología de exosomas con RVG, que es potencialmente adecuado para el silenciamiento a largo plazo de genes relacionados con enfermedades neurodegenerativas. El sistema de suministro con exosomas de Alvarez-Erviti y col., se podrá aplicar al suministro del sistema CRISPR-Cas usado en los métodos y composiciones reivindicados de la presente invención en blancos terapéuticas, especialmente en enfermedades neurodegenerativas. Para la presente invención se podrá contemplar una dosificación comprendida entre aproximadamente 100 y 1000 mg de sistema CRISPR-Cas encapsulado en una cantidad de exosomas con RVG comprendida entre aproximadamente 100 y 1000 mg.

El-Andaloussi y col. (Nature Protocols 7,2112–2126(2012)) describen cómo exosomas obtenidos a partir de células cultivadas se pueden aprovechar para el suministro de ARN *in vitro* e *in vivo*. Este protocolo describe en primer lugar la generación de exosomas dirigidos mediante la transfección con un vector de expresión que comprende una proteína exosómica fusionada con un ligando peptídico. A continuación, El-Andaloussi y col. explican cómo purificar y caracterizar exosomas a partir del sobrenadante celular transfectado. A continuación, El-Andaloussi y col. detallan los pasos cruciales para cargar ARN en los exosomas. Finalmente, El-Andaloussi y col. describen brevemente cómo utilizar exosomas para suministrar de manera eficaz ARN *in vitro* e *in vivo* en el cerebro del ratón. También se proporcionan ejemplos de los resultados previstos en los que se evalúa el suministro de ARN mediado por exosomas mediante ensayos funcionales y obtención de imágenes. Todo el protocolo requiere ~3 semanas. El suministro o la administración de acuerdo con la invención se podrá realizar utilizando exosomas producidos a partir de células dendríticas autoderivadas. Con respecto de las divulgaciones de la presente, esta se puede emplear en la práctica de la invención.

En otra forma de realización, se contemplan los exosomas plasmáticos de Wahlgren y col. (Nucleic Acids Research, 2012, Vol. 40, n.º 17 e130). Los exosomas son vesículas de tamaño nano (tamaño de 30-90 nm) producidos por muchos tipos celulares, incluidas las células dendríticas (DC, por sus siglas en inglés), linfocitos B, linfocitos T, mastocitos, células epiteliales y células tumorales. Estas vesículas se forman por gemación hacia el interior de endosomas tardíos y se liberan a continuación al entorno extracelular tras la fusión con la membrana plasmática. Debido a que los exosomas trasportan naturalmente ARN entre las células, esta propiedad puede ser útil en la terapia génica, y se puede emplear a partir de esta divulgación en la práctica de la presente invención.

Se pueden preparar exosomas a partir del plasma por centrifugación de la capa leucocitaria a 900 g durante 20 min para aislar el plasma y a continuación se recolectan los sobrenadantes celulares, se centrifugan a 300 g durante 10 min para eliminar células y a 16 500 g durante 30 min seguido por filtración a través de un filtro de 0,22 mm. Se obtiene el sedimento con los exosomas mediante ultracentrifugación a 120.000g durante 70 min. La transfección química del ARNip al interior de los exosomas se lleva a cabo de acuerdo con las instrucciones del fabricante con el conjunto de elementos Starter para iARN de Seres humanos/Ratones (Quiagen, Hilden, Alemania). Se añade ARNip a 100 ml de PBS con una concentración final de 2 mmol/ml. Después de añadir el reactivo de transfección HiPerFect, se incuba la mezcla durante 10 min a t. amb. Con el fin de eliminar el exceso de micelas, se vuelven a aislar los exosomas utilizando perlitas de látex con aldehído/sulfato. La transfección química de exosomas con CRISPR-Cas se podrá llevar a cabo de manera similar al ARNip. Los exosomas se podrán cocultivar con monocitos y linfocitos aislados de sangre periférica de donantes sanos. Por lo tanto, se podrá contemplar que exosomas que contienen CRISPR-Cas se podrán introducir en monocitos y linfocitos de un ser humano y se podrán reintroducir autólogamente en dicho ser humano. En consecuencia, el suministro o administración de acuerdo con la invención se podrá realizar utilizando exosomas plasmáticos.

Liposomas

El suministro o la administración para los métodos y composiciones reivindicados de acuerdo con la invención se podrá realizar con liposomas. Los liposomas son estructuras vesiculares esféricas compuestas de una bicapa lipídica uni o multilaminar que rodea compartimentos acuosos internos y una bicapa fosfolipídica lipófila externa relativamente impermeable. Los liposomas han atraído una atención considerable como portadores para el suministro de fármacos ya que son biocompatibles, atóxicos, pueden suministrar tanto moléculas farmacológicas hidrófilas como lipófilas, protegen su material de carga de la degradación por parte de enzimas plasmáticas y transportan su carga a través de membranas biológicas y de la barrera hematoencefálica (BHE) (para una revisión véase, por ejemplo, Spuch y Navarro, *Journal of Drug Delivery*, vol. 2011, ID del artículo 469679, 12 páginas, 2011. doi:10.1155/2011/469679).

Se pueden obtener liposomas a partir de varios tipos de lípidos diferentes; sin embargo, los fosfolípidos son los que se utilizan más habitualmente para generar liposomas como portadores de fármacos. Aunque la formación de liposomas es espontánea cuando una película lipídica se mezcla con una solución acuosa, también se puede acelerar aplicando fuerza en forma de agitación utilizando un homogeneizador, sonicador o un aparato de extrusión (para una revisión véase, por ejemplo, Spuch y Navarro, *Journal of Drug Delivery*, vol. 2011, ID del artículo 469679, 12 páginas, 2011. doi:10.1155/2011/469679).

Se podrán añadir varios aditivos diferentes a los liposomas para modificar su estructura y propiedades. Por ejemplo, se podrán añadir colesterol o esfingomielina a la mezcla de liposomas con el fin de ayudar a estabilizar la estructura liposomal y para prevenir el escape del material de carga interno liposomal. Además, los liposomas se preparan a partir de fosfatidilcolina de huevo hidrogenada o fosfato dicetílico, colesterol y fosfatidilcolina de huevo y sus tamaños vesiculares medio se ajustan entre aproximadamente 50 y 100 nm (para una revisión véase, por ejemplo, Spuch y

Navarro, Journal of Drug Delivery, vol. 2011, ID del artículo 469679, 12 páginas, 2011. doi:10.1155/2011/469679).

La formulación de liposomas convencional comprende principalmente fosfolípidos y lípidos naturales tales como 1,2-diestearoril-sn-glicero-3-fosfatidilcolina (DSPC, por sus siglas en inglés), esfingomielina, fosfatidilcolinas de huevo y monosialogangliósido. Debido a que esta formulación está constituida únicamente por fosfolípidos, las formulaciones liposomales se han topado con muchos obstáculos, siendo uno de ellos la inestabilidad en plasma. Se han realizado varios intentos de superar estos obstáculos, específicamente en la manipulación de la membrana lipídica. Uno de estos intentos se ha centrado en la manipulación del colesterol. La adición de colesterol a las formulaciones convencionales reduce la liberación rápida del compuesto bioactivo encapsulado en el plasma o la 1,2-dioleoil-*sn*-glicero-3-fosfoetanolamina (DOPE) incrementa la estabilidad (para una revisión véase, por ejemplo, Spuch y Navarro, *Journal of Drug Delivery*, vol. 2011, ID del artículo 469679, 12 páginas, 2011. doi:10.1155/2011/469679).

En una forma de realización particularmente conveniente, son deseables los liposomas de tipo caballo de Troya (también conocidos como caballos de Troya moleculares) y se podrán acceder a los protocolos en http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long. Estas partículas permiten el suministro de un transgén a todo el cerebro tras una inyección intravascular. Sin querer ceñirse a ninguna teoría, se cree que las partículas lipídicas neutras con anticuerpos específicos conjugados a la superficie permiten cruzar la barrera hematoencefálica mediante endocitosis. Fuera del alcance de la invención, el solicitante propone utilizar liposomas de tipo caballo de Troya para suministrar la familia CRISPR de nucleasas al cerebro mediante una inyección intravascular, lo que permitiría animales transgénicos en la totalidad de su cerebro sin requerir manipulación de embriones. Se podrán contemplar aproximadamente 1-5 g de ADN o ARN para la administración *in vivo* en liposomas.

En otra forma de realización, el sistema CRISPR-Cas o los componentes del mismo se podrán administrar en liposomas, tales como una partícula ácido nucleico-lípido estable (SNALP, por sus siglas en inglés) (véase, por ejemplo, Morrissey y col., Nature Biotechnology, vol. 23, n.º 8, agosto de 2005). Se contemplan inyecciones intravenosas diarias de aproximadamente 1, 3 o 5 mg/kg/día de CRISPR-Cas específico dirigido en una SNALP. El tratamiento diario podrá prolongarse durante aproximadamente tres días y a continuación semanalmente durante aproximadamente cinco semanas. En otra forma de realización, también se contempla un CRISPR-Cas específico encapsulado (SNALP) administrado mediante inyección intravenosa con dosis de aproximadamente 1 o 2,5 mg/kg (véase, por ejemplo, Zimmerman y col., Nature Letters, Vol. 441, 4 de mayo de 2006). La formulación SNALP puede contener los lípidos 3-N-[(w metoxipoli(etilenglicol) 2000)carbamoil]-1,2-dimiristiloxipropilamina (PEG-C-DMA), 1,2-dilinoleiloxi-N,N-dimetil-3-aminopropano (DLinDMA), 1,2-diestearoil-sn-glicero-3-fosfocolina (DSPC) y colesterol con una relación molar porcentual de 2:40:10:48 (véase, por ejemplo, Zimmerman y col., Nature Letters, Vol. 441, 4 de mayo de 2006).

En otra forma de realización, se ha demostrado que las partículas de ácido nucleico-lípido estables (SNALP) son moléculas de suministro eficaces para tumores hepáticos derivados de HepG2 sumamente vascularizados pero no para los tumores hepáticos derivados de HCT-116 poco vascularizados (véase, por ejemplo, Li, Gene Therapy (2012) 19, 775–780). Se podrán preparar liposomas de SNALP formulando D-Lin-DMA y PEG-C-DMA con diestearoilfosfatidilcolina (DSPC), colesterol y ARNip utilizando una relación 25:1 de lípido/ARNip y una relación molar de 48/40/10/2 de colesterol/D-Lin-DMA/DSPC/PEG-C-DMA. Los liposomas de SNALP resultantes tienen un tamaño de aproximadamente 80-100 nm.

En otra forma de realización más, una SNALP podrá comprender colesterol sintético (Sigma-Aldrich, San Luis, MO, EE.UU.), dipalmitoilfosfatidilcolina (Avanti Polar Lipids, Alabaster, AL, EE.UU.), 3-N-[(w-metoxipoli(etilenglicol)2000)carbamoil]-1,2-dimirestiloxipropilamina y 1,2-dilinoleiloxi-3-N,N-dimetilaminopropano catiónico (véase, por ejemplo, Geisbert y col., Lancet 2010; 375: 1896-905). Se podrá contemplar una dosificación de aproximadamente 2 mg/kg de CRISPR-Cas total por dosis administrada, por ejemplo, como una inyección intravenosa rápida.

En otra forma de realización más, una SNALP podrá comprender colesterol sintético (Sigma-Aldrich), 1,2-diestearoil-sn-glicero-3-fosfocolina (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA y 1,2-dilinoleiloxi-3-(N,N-dimetil)aminopropano (DLinDMA) (véase, por ejemplo, Judge, J. Clin. Invest. 119:661-673 (2009)). Las formulaciones utilizadas para los estudios *in vivo* podrán comprender una relación másica de lípidos/ARN final de aproximadamente 9:1.

Barros y Gollob de Alnylam Pharmaceuticals (véase, por ejemplo, Advanced Drug Delivery Reviews 64 (2012) 1730–1737) han revisado el perfil de seguridad de las nanomedicinas con iARN. La partícula de ácido nucleico y lípido estable (SNALP) comprende cuatro lípidos diferentes - un lípido ionizable (DLinDMA) que es catiónico a un pH bajo, un lípido neutro cooperador, colesterol y un polietilenglicol (PEG) difundible-lípido. La partícula tiene un diámetro de aproximadamente 80 nm y el pH fisiológico tiene una carga neutra. Durante la formulación, el lípido ionizable sirve para condensar el lípido con el ARN aniónico durante la formación de la partícula. Cuando se carga positivamente en condiciones endosómicas cada vez más ácidas, el lípido ionizable también media en la fusión de la SNALP con la membrana endosómica y hace posible la liberación del ARN en el citoplasma. El PEG-lípido estabiliza la partícula y reduce la agregación durante la formulación y posteriormente proporciona un exterior hidrofílico neutro que mejora las propiedades farmacocinéticas.

Hasta la fecha, se han iniciado dos programas clínicos utilizando formulaciones de SNALP con ARN. Tekmira Pharmaceuticals completó recientemente un estudio de fase I de dosis única de SNALP-ApoB en voluntarios adultos

con un nivel de colesterol LDL elevado. ApoB se expresa predominantemente en el hígado y el yeyuno y es esencial para el ensamblaje y secreción de VLDL y LDL. Diecisiete sujetos recibieron una dosis única de SNALP-ApoB (aumento escalonado de la dosis en 7 niveles de dosis). No se dispone de evidencia de hepatotoxicidad (prevista debido a la toxicidad limitante de la dosis potencial según los estudios preclínicos). Uno (de los dos) sujetos con la dosis más alta experimentaron síntomas similares a los de la gripe coherentes con una estimulación del sistema inmunitario y se tomó la decisión de finalizar el ensayo.

Alnylam Pharmaceuticals ha avanzado de manera similar ALN-TTR01, que emplea la tecnología de SNALP descrita anteriormente y actúa sobre la producción de hepatocitos tanto en TTR no modificada como la mutante para tratar la amiloidosis TTR (ATTR, por sus siglas en inglés). Se han descrito tres síndromes de ATTR: la polineuropatía amiloidótica familiar (FAP, por sus siglas en inglés) y la cardiomiopatía amiloidótica familiar (FAC, por sus siglas en inglés) - ambas provocadas por mutaciones dominantes autosómicas en la TTR; y la amiloidosis sistémica senil (SSA, por sus siglas en inglés) provocada por la TTR no modificada. Recientemente se ha completado un ensayo controlado con placebo de fase I con aumento escalonado de una dosis única de ALN-TTR01 en pacientes con ATTR. Se administró ALN-TTR01 como una infusión IV de 15 min a 31 pacientes (23 con el fármaco de estudio y 8 con el placebo) con una dosis comprendida en el intervalo de 0,01 a 1,0 mg/kg (basándose en el ARNip). El tratamiento se toleró bien sin aumentos significativos en las pruebas de la función hepática. Se observaron reacciones relacionadas con la infusión en 3 de los 23 pacientes con ≥0,4 mg/kg; todos respondieron a una ralentización de la velocidad de infusión y todos continuaron en el estudio. Se observaron elevaciones transitorias y mínimas de las citocinas séricas IL-6, IP-10 e IL-1ra en dos pacientes con la dosis más alta de 1 mg/kg (tal como se había previsto a partir de los estudios preclínicos y en NHP). Al bajar la TTR sérica, se observó el efecto farmacodinámico esperado de ALN-TTR01 con 1 mg/kg.

En otra forma de realización más, se puede generar una SNALP solubilizando un lípido catiónico, DSPC, colesterol y PEG-lípido por ejemplo en etanol, por ejemplo con una relación molar de 40:10:40:10, respectivamente (véase Semple y col., Nature Biotechnology, Volumen 28 Número 2 febrero de 2010, págs. 172-177). Se añadió la mezcla lipídica a un regulador de pH acuoso (citrato 50 mM, pH 4) mientras se agitaba hasta alcanzar una concentración final de etanol y lípidos de un 30% (vol/vol) y 6,1 mg/mL, respectivamente, y se permitió que se equilibrara a 22 ºC durante 2 min antes de la extrusión. Se extruyeron los lípidos hidratados a través de dos filtros apilados con un tamaño de poro de 80 nm (Nuclepore) a 22 ºC utilizando un Extrusor Lipex (Northern Lipids) hasta obtener vesículas con un diámetro de 70-90 nm, tal como se determinó mediante análisis de dispersión dinámica de la luz. Esto requirió, en general, 1-3 pases. El ARNip (solubilizado en una solución acuosa de citrato 50 mM, pH 4, que contenía un 30% de etanol) se añadió a las vesículas preequilibradas (35 ºC) con una velocidad de ~5 mL/min mientras se agitaba. Después de alcanzar la relación ARNip/lípido objetivo final de 0,06 (p/p), se incubó la mezcla durante 30 min más a 35 ºC para permitir la reorganización de las vesículas y la encapsulación del ARNip. A continuación, el etanol se eliminó y el regulador de pH externo se reemplazó con PBS (NaCl 155 mM, Na_2HPO_4 3 mM, KH_2PO_4 1 mM, pH 7,5) ya sea por diálisis o diafiltración con flujo tangencial. Se encapsuló el ARNip en SNALP utilizando un proceso metodológico de dilución en etapas controlado. Los constituyentes lipídicos de KC2-SNALP fueron DLin-KC2-DMA (lípido catiónico), dipalmitoilfosfatidilcolina (DPPC; Avanti Polar Lipids), colesterol sintético (Sigma) y PEG-C-DMA utilizados con una relación molar de 57.1:7.1:34,3:1,4. Tras la formación de las partículas con el material de carga, se dializaron las SNALP frente a PBS y se esterilizaron por filtración a través de un filtro de 0,2 µm antes de su uso. Los tamaños de partículas medios fueron de 75-85 nm y se encapsuló un 90-95% del ARNip dentro de las partículas lipídicas. La relación final de ARNip/lípido en las formulaciones utilizadas para las pruebas *in vivo* fue de ~0,15 (p/p). Los sistemas de LNP-ARNip que contenían ARNip contra el Factor VII se diluyeron hasta alcanzar las concentraciones apropiadas en PBS estéril inmediatamente antes de su uso y se administraron las formulaciones por vía intravenosa a través de la vena de la cola lateral en un volumen total de 10 mL/kg. Este método y estos sistemas de suministro se podrán extrapolar al sistema CRISPR-Cas de la presente invención.

Otros lípidos

Se podrán utilizar otros lípidos catiónicos, tales como el amino lípido 2,2-dilinoleil-4-dimetilaminoetil-[1,3]-dioxolano (DLin-KC2-DMA) para encapsular a CRIPSR-Cas o componentes del mismo o molécula(s) de ácido nucleico de manera similar al ARNip (véase, por ejemplo, Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 –8533), y por tanto, se puede emplear en la práctica de la invención. Se podrá contemplar una vesícula preformada con la siguiente composición lipídica: amino lípido, diestearoilfosfatidilcolina (DSPC), colesterol y (*R*)-2,3-bis(octadeciloxi)propil-1-(metoxipoli(etilenglicol)2000)propilcarbamato (PEG-lípido) con una relación molar de 40/10/40/10, respectivamente, y una relación ARNip contra FVII/lípidos totales de aproximadamente 0,05 (p/p). Para garantizar una distribución estrecha del tamaño de las partículas comprendida en el intervalo de 70-90 nm y un índice de polidispersidad por debajo de 0,11-0,04 (n=56), las partículas se podrán extrudir hasta tres veces a través de membranas de 80 nm antes de añadir el ARN guía. Se podrán utilizar partículas que contengan un amino lípido 16 sumamente potente, donde la relación molar de los cuatro componentes lipídicos 16, DSPC, colesterol y PEG-lípido (50/10/38,5/1,5) se podrá optimizar en mayor grado para potenciar la actividad *in vivo*.

Michael S D Kormann y col. ("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: *Nature Biotechnology*, Volumen:29, Páginas: 154–157 (2011) describen el uso de envolturas lipídicas para suministrar ARN. También se prefiere la utilización de envolturas lipídicas en la presente invención.

En otra forma de realización, los lípidos se pueden formular con el sistema CRISPR-Cas usado en los métodos y composiciones reivindicados de la presente invención o componente(s) del mismo o molécula(s) de ácido nucleico que codifica(n) el mismo para formar nanopartículas lipídicas (LNP, por las siglas en inglés de Lipid NanoParticles). Los lípidos incluyen, de manera no taxativa, DLin-KC2-DMA4, C12-200 y los colípidos diesteroilfosfatidilcolina, colesterol y PEG-DMG, y se podrán formular con CRISPR-Cas en lugar de ARNip (véase, por ejemplo, Novobrantseva, Molecular Therapy–Nucleic Acids (2012) 1, e4; doi:10.1038/mtna,2011,3) utilizando un procedimiento de formación de vesículas espontáneo. La relación molar de los componentes podrá ser aproximadamente 50/10/38,5/1,5 (DLin-KC2-DMA o C12-200/diesteroilfosfatidilcolina/colesterol/PEG-DMG). La proporción en peso final lípido:ARNip puede ser de ~12:1 y 9:1 en el caso de las nanopartículas lipídicas (LNP) DLin-KC2-DMA y C12-200, respectivamente. Las formulaciones podrán tener diámetros de partículas promedio de ~80 nm con una eficacia de atrapamiento de >90%. Se podrá contemplar una dosis de 3 mg/kg.

Tekmira tiene una cartera de aproximadamente 95 familias de patentes, en los EE.UU. y otros países, que tratan sobre diversos aspectos de la LNP y formulaciones con LNP (véase, por ejemplo, las patentes de los EE.UU. con n.º 7.982.027; 7.799.565; 8.058.069; 8.283.333; 7.901.708; 7.745.651; 7.803.397; 8.101.741; 8.188.263; 7.915.399; 8.236.943 y 7.838.658 y patentes europeas n.º 1766035; 1519714; 1781593 y 1664316), todas las cuales pueden usarse y/o adaptarse a la presente invención.

El sistema CRISPR-Cas o sus componentes o molécula(s) de ácido nucleico que codifican el anterior se podrán suministrar encapsuladas en microesferas de PLGA como las que se describen adicionalmente en las solicitudes publicadas en los EE.UU. 20130252281 y 20130245107 y 20130244279 (adjudicada a Moderna Therapeutics) que se relacionan con aspectos de la formulación de composiciones que comprenden moléculas de ácido nucleico manipulado que podrán codificar una proteína, un precursor proteico o una forma parcial o totalmente procesada de la proteína o precursor proteico. La formulación podrá tener una relación molar de 50:10:38,5:1,5-3,0 (lípido catiónico:lípido fusogénico:colesterol:PEG lípido). El PEG lípido se podrá seleccionar entre, de manera no taxativa, PEG-c-DOMG, PEG-DMG. El lípido fusogénico podrá ser DSPC. Véase también a Schrum *y col.*, Delivery and Formulation of Engineered Nucleic Acids, solicitud publicada de los EE.UU. 20120251618.

La tecnología de Nanomerics, aborda los obstáculos referentes a la biodisponibilidad para una gama amplia de tratamientos, que incluyen tratamientos con ácido nucleico (plásmidos, ARNip, ARNmi), péptidos y fármacos hidrofóbicos con un bajo peso molecular. Las vías de administración específicas para las que esta tecnología ha demostrado ventajas claras incluyen la vía oral, transporte a través de la barrera hematoencefálica, suministro a tumores sólidos, así como también al ojo. Véase, por ejemplo, Mazza *y col.*, 2013, *ACS Nano*. 26 de febrero de 2013;7(2):1016-26; Uchegbu y Siew, 2013, J Pharm Sci. 102(2):305-10 y Lalatsa y col., 2012, J Control Release. 20 de julio de 2012; 161(2):523-36.

La patente de los EE.UU. con n.º de publicación 20050019923 describe dendrímeros catiónicos para el suministro de moléculas bioactivas tales como moléculas polinucleotídicas, péptidos y polipéptido y/o agentes farmacéuticos, al cuerpo de un mamífero. Los dendrímeros son adecuados para dirigir el suministro de moléculas bioactivas a, por ejemplo, el hígado, bazo, pulmón, riñón o corazón (o incluso el cerebro). Los dendrímeros son macromoléculas sintéticas tridimensionales que se preparan por etapas a partir de unidades monoméricas ramificadas simples, donde su naturaleza y funcionalidad se pueden controlar y variar fácilmente. Los dendrímeros se sintetizan a partir de la adición repetida de unidades básicas a un núcleo multifuncional (estrategia divergente para la síntesis) o hacia un núcleo multifuncional (estrategia convergente para la síntesis) y cada adición de una coraza tridimensional de unidades básicas conduce a la formación de una generación más elevada de dendrímeros. Los dendrímeros de polipropilenimina comienzan con un núcleo de diaminobutano al que se añaden el doble de grupos amino mediante una adición doble de Michael de acrilonitrilo a las aminas primarias seguido por la hidrogenación de nitrilos. Esto da como resultado la duplicación de los grupos amino. Los dendrímeros de polipropilenimina contienen un 100% de nitrógenos protonables y hasta 64 grupos amino terminal (generación 5, 64 DBA). Los grupos protonables son normalmente grupos amino que son capaces de aceptar protones a un pH neutro. El uso de dendrímeros como agentes de suministro de genes se ha centrado principalmente en el uso de la poliamidoamina y compuestos que contienen fósforo con una mezcla de amina/amida o N--$P(O_2)S$ como unidades de conjugación respectivamente sin que se haya publicado ningún trabajo sobre el uso de dendrímeros de polipropilenimina con una generación más baja para el suministro de genes. También se han estudiado dendrímeros de polipropilenimina como sistemas de liberación controlada sensibles al pH para el suministro de fármacos y para la encapsulación de moléculas externas cuando se modifican químicamente mediante grupos de tipo aminoácido periféricos. La citotoxicidad y la interacción de los dendrímeros de polipropilenimina con ADN, así como también la eficacia de transfección de DAB 64 también se han estudiado.

La patente de los EE.UU. con n.º de publicación 20050019923 se basa en la observación de que, a diferencia de informes anteriores, los dendrímeros catiónicos tales como dendrímeros de polipropilenimina, presentan propiedades adecuadas tales como un direccionamiento específico y una toxicidad baja, para su uso en el suministro dirigido de moléculas bioactivas, tal como el material genético. Además, los derivados de dendrímeros catiónicos también presentan propiedades adecuadas para el suministro dirigido de moléculas bioactivas. Véase también a, Bioactive Polymers, solicitud de los EE.UU. Publicada 20080267903, que divulga "diversos polímeros, que incluyen polímeros de poliamina catiónica y polímeros dendriméricos de los que se demuestra que poseen actividad antiproliferativa y, por lo tanto, podrán ser útiles para el tratamiento de afecciones caracterizadas por una proliferación celular no deseada

tales como neoplasias y tumores, trastornos inflamatorios (que incluyen los trastornos autoinmunitarios), psoriasis y aterosclerosis. Los polímeros se podrán utilizar solo como agentes activos o como vehículos para el suministro de otros agentes terapéuticos tales como moléculas farmacológicas o ácidos nucleicos para la terapia génica. En tales casos, la propia actividad antitumoral intrínseca de los polímeros podrá complementar la actividad del agente que se va a suministrar". Se pueden emplear las divulgaciones de estas publicaciones de patentes conjuntamente con las divulgaciones de la presente para el suministro de sistema(s) CRISPR Cas o sus componentes o moléculas de ácidos nucleicos que codifican los anteriores

Proteínas supercargadas

Las proteínas supercargadas son una clase de proteínas modificadas o de origen natural que tienen una carga teórica neta negativa o positiva inusualmente alta y pueden emplearse en el suministro de sistemas CRISPR-Cas o sus componentes o moléculas de ácidos nucleicos que codifican los anteriores. Tanto las proteínas cargadas supernegativamente como superpositivamente exhiben una capacidad notable de soportar la agregación inducida térmica o químicamente. Las proteínas cargadas superpositivamente también son capaces de penetrar en células de mamíferos. La asociación de material de carga con estas proteínas, tales como el ADN plasmídico, ARN, u otras proteínas, puede posibilitar el suministro funcional de estas macromoléculas en células de mamífero tanto *in vitro* como *in vivo*. El laboratorio de David Liu publicó la creación y caracterización de proteínas supercargadas en 2007 (Lawrence *y col.*, 2007, *Journal of the American Chemical Society* 129, 10110–10112).

El suministro no viral de ARN y ADN plasmídico a células de mamíferos es valioso tanto para aplicaciones terapéuticas como en investigación (Akinc y col., 2010, Nat. Biotech. 26, 561–569). La proteína +36 GFP purificada (u otra proteína cargada superpositivamente) se mezcla con los ARN en el medio exento de suero apropiado y se permite que compleje antes de la adición de células. La inclusión de suero en esta etapa inhibe la formación de los complejos proteína supercargada-ARN y reduce la eficacia del tratamiento. Se ha descubierto que el siguiente protocolo es eficaz para una variedad de líneas celulares (McNaughton y col., 2009, Proc. Natl. Acad. Sci. USA 106, 6111–6116). Sin embargo, deberán realizarse experimentos piloto en los que se varíe la dosis de proteína y ARN para optimizar el procedimiento para líneas celulares específicas.

(1) Un día antes del tratamiento, colocar en placas 1 × 10^5 células por pocillo en una placa de 48 pocillos.

(2) El día del tratamiento, diluir la proteína +36 GFP purificada en medio exento de suero hasta una concentración final de 200 nM. Añadir ARN hasta una concentración final de 50 nM. Agitar con un vórtice para mezclar e incubar a temperatura ambiente durante 10 min.

(3) Durante la incubación, aspirar el medio de las células y lavar una vez con PBS.

(4) Tras la incubación de +36 GFP y ARN, añadir los complejos proteína-ARN a las células.

(5) Incubar las células con complejos a 37 ºC durante 4 h.

(6) Tras la incubación, aspirar el medio y lavar tres veces con 20 U/ml de heparina PBS. Incubar las células con medio que contenga suero durante 48 h más o durante más tiempo dependiendo del ensayo para determinar la actividad.

(7) Analizar las células mediante inmunotransferencia, qPCR, ensayo fenotípico u otro método apropiado.

El laboratorio de David Liu ha descubierto además que +36 GFP es un reactivo de suministro plasmídico eficaz con una gama de células. Ya que el ADN plasmídico es un material de carga mayor que el ARNip, se requiere proporcionalmente más proteína +36 GFP para complejar de manera eficaz los plásmidos. Para el suministro plasmídico eficaz, los solicitantes han desarrollado una variante de +36 GFP que porta un identificador peptídico HA2 C-terminal, un conocido péptido desorganizador de endosomas obtenido a partir de la proteína hemaglutinina del virus influenza. El siguiente protocolo ha sido eficaz en varias células, pero al igual que anteriormente, se aconseja que las dosis de ADN plasmídico y proteína supercargada se optimicen para líneas celulares específicas y aplicaciones de suministro.

(1) Un día antes del tratamiento, sembrar en placas 1 × 10^5 por pocillo en una placa de 48 pocillos.

(2) El día del tratamiento, diluir la proteína Þ36 GFP purificada en medio exento de suero hasta una concentración final de 2 mM. Añadir 1 mg de ADN plasmídico. Agitar con un vórtice para mezclar e incubar a temperatura ambiente durante 10 min.

(3) Durante la incubación, aspirar el medio de las células y lavar una vez con PBS.

(4) Tras la incubación de Þ36 GFP y ADN plasmídico, añadir cuidadosamente los complejos proteína-ADN a las células.

(5) Incubar las células con complejos a 37 ºC durante 4 h.

(6) Tras la incubación, aspirar el medio y lavar con PBS. Incubar las células en medio que contenga suero e incubar

durante 24-48 h más.

(7) Analizar el suministro del plásmido (por ejemplo, mediante expresión génica dirigida por el plásmido) según sea apropiado.

Véase también a, por ejemplo, McNaughton y col., Proc. Natl. Acad. Sci. USA 106, 6111-6116 (2009); Cronican y col., ACS Chemical Biology 5, 747-752 (2010); Cronican y col., Chemistry & Biology 18, 833-838 (2011); Thompson y col., Methods in Enzymology 503, 293-319 (2012); Thompson, D.B., y col., Chemistry & Biology 19 (7), 831-843 (2012). Los métodos de las proteínas supercargadas se podrán utilizar y/o adaptar para el suministro del sistema CRISPR-Cas usado en los métodos y composiciones reivindicados de la presente invención. Se pueden emplear estos sistemas del Dr. Lui y los documentos de la presente junto con las divulgaciones de la presente en el suministro de sistema(s) CRISPR-Cas o componente(s) de estos o molécula(s) de ácidos nucleicos que codifican los anteriores.

Péptidos penetrantes en células (CPP)

En otra forma de realización más, se contemplan péptidos penetrantes en células (CPP) para la administración del sistema CRISPR Cas. Los CPP son péptidos cortos que facilitan la captación celular de varias cargas moleculares (desde partículas de tamaño nanométrico hasta moléculas químicas pequeñas y grandes fragmentos de ADN). El término "carga", tal como se usa en el presente documento incluye, pero sin limitación, el grupo que consiste en agentes terapéuticos, sondas diagnósticas, péptidos, ácidos nucleicos, oligonucleótidos antisentido, plásmidos, proteínas, nanopartículas, liposomas, cromóforos, moléculas pequeñas y materiales radioactivos. En aspectos de los métodos y composiciones reivindicados de la invención, la carga también puede comprender cualquier componente del sistema CRISPR Cas o el sistema CRISPR Cas funcional completo. Los aspectos de la presente invención proporcionan además métodos para administrar una carga deseada en un sujeto que comprenden: (a) preparar un complejo que comprende el péptido penetrante en la célula de la presente invención y una carga deseada y (b) administrar el complejo a un sujeto por vía oral, intraarticular, intraperitoneal, intratecal, 193intraarterial, intranasal, intraparenquimal, subcutánea, intramuscular, intravenosa, dérmica, intratecal o tópica. La carga se asocia con los péptidos bien mediante ligadura química a través de enlaces covalentes o mediante interacciones no covalentes.

La función de los CPP es administrar la carga a las células, un proceso que tiene lugar comúnmente mediante endocitosis con la carga administrada a los endosomas de células vivas de mamífero. Los péptidos penetrantes en células son de diferentes tamaños, secuencias de aminoácidos y cargas, pero todos los CPP tienen una característica distinta, que es la capacidad para translocarse en la membrana plasmática y facilitar la administración de diversas cargas moleculares al citoplasma o a un orgánulo. La translocación de CPP puede clasificarse en tres principales mecanismos de entrada: penetración directa en la membrana, entrada mediada por endocitosis y translocación mediante la formación de una estructura transitoria. Los CPP han encontrado numerosas aplicaciones en medicina como agentes de suministro de fármacos en el tratamiento de diferentes enfermedades, incluyendo el cáncer e inhibidores virales, así como agentes de contraste para el marcado de células. Los ejemplos del último incluyen actuar como transportador para GFP, agentes de contraste de IRM o puntos cuánticos. Los CPP tienen un gran potencial como vectores de administración *in vitro* e *in vivo* para su uso en investigación y medicina. Los CPP tienen típicamente una composición de aminoácidos que bien contiene una abundancia relativa mayor de aminoácidos cargados positivamente, tales como lisina o arginina o tiene secuencias que contienen un patrón alterno de aminoácidos polares/cargados y aminoácidos no polares/hidrofóbicos. Estos dos tipos de estructuras se citan como policatiónicas y anfipáticas, respectivamente. Una tercera clase de CPP son los péptidos hidrofóbicos que solo contienen restos apolares, con una carga neta baja o tienen grupos aminoácidos hidrofóbicos que son cruciales para la captación celular. Uno de los CPP descubiertos inicialmente fue el activador transcripcional trans-activante (Tat) del virus de la inmunodeficiencia humana 1 (VIH-1) que se observó que era captado eficazmente desde el medio circundante por numerosos tipos celulares en cultivo. Desde entonces, se ha expandido considerablemente el número de CPP conocidos y se han generado análogos sintéticos de molécula pequeña con propiedades de transducción de proteínas más eficaces. Los CPP incluyen, pero sin limitación, Penetratin, Tat (48-60), Transportan y (R-AhX-R4) (Ahx= aminohexanoilo).

La Patente de los EE.UU. 8.372.951 proporciona un CPP procedente de proteína catiónica de eosinófilos (ECP) que muestra una alta eficiencia de penetración en la célula y baja toxicidad. También se proporcionan aspectos de administración del CPP con su carga en un sujeto vertebrado. Los aspectos adicionales de los CPP y su administración se describen en las Patentes de Estados Unidos 8,575,305; 8.614.194 y 8,044,019. Los CPP pueden usarse para administrar el sistema CRISPR-Cas o componentes del mismo. El hecho de que puedan emplearse CPP para administrar el sistema CRISPR-Cas o componentes del mismo también se proporciona en el manuscrito "Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA", por Suresh Ramakrishna, Abu-Bonsrah Kwaku Dad, Jagadish Beloor, y col. Genome Res. 2 de abril de 2014. [Publicación electrónica nunca impresa], donde se demuestra que el tratamiento con proteínas Cas9 recombinante conjugada a CPP y ARN guía complejado a CPP da lugar a disrupciones génicas endógenas en líneas celulares humanas. En el artículo, la proteína Cas9 se conjugó a CPP a través de un enlace tioéter, mientras que el ARN guía se complejó al CPP, formando partículas condensadas cargadas positivamente. Se demostró que el tratamiento simultáneo y secuencial de células humanas, incluyendo células madre embrionarias, fibroblastos dérmicos, células HEK293T, células HeLa y células de carcinoma embrionario, con la Cas9 modificada y ARN guía dio lugar a disrupciones génicas eficaces con mutaciones fuera del diana reducidas en relación a las transfecciones con plásmido.

Dispositivos implantables

En otra forma de realización, se contemplan también dispositivos implantables para el suministro de sistemas CRISPR-Cas o componentes(s) de estos o molécula(s) de ácidos nucleicos que codifican los anteriores. Por ejemplo, la publicación de patente de los EE.UU. 20110195123 divulga un dispositivo médico implantable el cual eluye un fármaco de manera localizada y en un periodo prolongado, que incluye varios tipos de dispositivos de este tipo, y se proporcionan modos de tratamiento de la implementación y métodos de la implantación. El dispositivo comprende un sustrato polimérico tal como una matriz, por ejemplo, que se utiliza como el cuerpo del dispositivo y fármacos y, en algunos casos, materiales que sirven de esqueleto adicionales tales como metales o polímeros adicionales y materiales para potenciar la visibilidad y obtención de imágenes. Puede ser ventajoso proporcionar un dispositivo de suministro implantable que proporcione una liberación local y durante un periodo prolongado, donde el fármaco se libera directamente a la matriz extracelular (ECM, por sus siglas en inglés) del área con la enfermedad tal como el tumor, inflamación, degeneración o con objetivos sintomáticos o a células del músculo liso herido o para la prevención. Un tipo de fármaco es el ARN que se ha descrito anteriormente, y este sistema se podrá utilizar y/o adaptar al sistema CRISPR-Cas usado en los métodos y composiciones reivindicados de la presente invención. Los modos de implantación en algunas formas de realización son los procedimientos de implantación existentes que se han desarrollado y se utilizan en la actualidad para otros tratamientos que incluyen la braquirradioterapia y la punción biópsica. En tales casos, las dimensiones del nuevo implante descrito en esta invención son similares a las del implante original. Normalmente, se implantan unos pocos dispositivos durante el mismo procedimiento del tratamiento.

La Publicación de Patente de los EE.UU. 20110195123 proporciona un sistema de suministro de fármaco implantable o insertable que incluye sistemas aplicables a una cavidad tal como la cavidad abdominal y/o cualquier otro tipo de administración en la que el sistema de suministro del fármaco no está anclado ni unido, que comprende un sustrato polimérico bioestable y/o degradable y/o bioabsorbible que podrá ser, por ejemplo, opcionalmente una matriz. Cabe señalar que el término "inserción" también incluye la implantación. El sistema de suministro del fármaco se implementa preferentemente como un "Loder" tal y como se describe en la publicación de patente de los EE.UU. 20110195123.

El polímero o pluralidad de polímeros son biocompatibles, incorporan un agente y/o una pluralidad de agentes, lo que hace posible la liberación del agente con una velocidad controlada, donde el volumen total del sustrato polimérico, tal como una matriz, por ejemplo, en algunas formas de realización es opcional y preferentemente no superior a un volumen máximo que permite que se alcance un nivel terapéutico del agente. Como un ejemplo no limitante, un volumen de este tipo está comprendido preferentemente en el intervalo entre 0,1 m^3 y 1000 mm^3, tal y como lo requiere el volumen para alojar el agente. El Loder podrá ser opcionalmente más grande, por ejemplo, cuando tiene incorporado un dispositivo cuyo tamaño está determinado por la funcionalidad, por ejemplo, y de manera no taxativa, la articulación de la rodilla, un anillo intrauterino o cervicouterino y similares.

El sistema de suministro de fármacos (para suministrar la composición) se diseña en algunas formas de realización para emplear preferentemente polímeros degradables, donde el principal mecanismo de liberación es la erosión de todo el volumen; o en algunas formas de realización se utilizan polímeros no degradables o de degradación lenta, donde el principal sistema de liberación es la difusión más que la erosión de todo el volumen, de modo que la parte externa funciona como una membrana y su parte interna funciona como un depósito del fármaco, el cual prácticamente no se ve afectado por el entorno durante un periodo prolongado (por ejemplo, desde aproximadamente una semana hasta aproximadamente unos pocos meses). También se podrán utilizar opcionalmente combinaciones de polímeros diferentes con mecanismos de liberación diferentes. Preferentemente, el gradiente de concentración en la superficie se mantiene eficazmente constante durante un periodo de tiempo significativo del periodo total de liberación del fármaco y, por lo tanto, la velocidad de difusión es eficazmente constante (denominada difusión de "modo cero"). Por el término "constante" se entiende una velocidad de difusión que se mantiene preferentemente por encima del umbral inferior de la eficacia terapéutica pero que aún podrá presentar opcionalmente una fluctuación y/o variación brusca inicial, por ejemplo, un incremento y disminución hasta cierto grado. La velocidad de difusión se mantiene de esta manera preferentemente durante un periodo prolongado y se puede considerar constante hasta un cierto nivel para optimizar el periodo terapéuticamente eficaz, por ejemplo, el periodo de silenciamiento eficaz.

Opcional y preferentemente, el sistema de suministro de fármacos se diseña para proteger el agente terapéutico nucleotídico de la degradación, ya sea de naturaleza química o debida al ataque de enzimas y otros factores del cuerpo del sujeto.

El sistema de suministro de fármacos de la publicación de patente de Estados Unidos 20110195123 se asocia opcionalmente con aparatos de detección y/o activación operados en la implantación del dispositivo y/o después de esta, mediante métodos de activación y/o aceleración/desaceleración no invasivos, y/o mínimamente invasivos, por ejemplo, que incluyen opcionalmente, pero sin limitación, métodos o dispositivos de enfriamiento y calentamiento térmico, haces de rayos láser y ultrasónicos, que incluyen los ultrasonidos focalizados y/o RF (radiofrecuencia).

De acuerdo con algunas formas de realización de la publicación de patente de los EE.UU. 20110195123, el sitio para el suministro localizado podrá incluir opcionalmente sitios diana caracterizados por una proliferación elevada anómala de células y una supresión de la apoptosis que incluyen tumores, la inflamación crónica y/o activa y la infección que incluye los estados patológicos autoimunitarios, tejido que se está degenerando que incluye el tejido muscular y nervioso, dolor crónico, sitios degenerativos y ubicación de fracturas óseas y otras ubicaciones de heridas para

potenciar la regeneración del tejido y músculo liso y estriado cardiaco lesionado.

El sitio para la implantación de la composición, o sitio diana, presenta preferentemente un radio, área y/o volumen que es lo suficientemente pequeño para el suministro localizado dirigido. Por ejemplo, el sitio diana tiene opcionalmente un diámetro comprendido en el intervalo de aproximadamente 0,1 mm a aproximadamente 5 cm.

Preferentemente, se selecciona la ubicación del sitio diana para una eficacia terapéutica máxima. Por ejemplo, la composición del sistema de suministro del fármaco (opcionalmente con un dispositivo para la implantación tal como se ha descrito anteriormente) se implanta opcional y preferentemente en las proximidades del entorno tumoral o en este, o en el suministro sanguíneo asociado con este.

Por ejemplo, la composición (opcionalmente con el dispositivo) se implanta opcionalmente dentro o en las proximidades del páncreas, próstata, mama, hígado, mediante el pezón, dentro del sistema vascular y así sucesivamente.

La localización del diana se selecciona opcionalmente entre el grupo constituido por (únicamente a modo de ejemplos no taxativos, ya que opcionalmente cualquier sitio dentro del cuerpo podrá ser adecuado para implantar un Loder): 1. cerebro en sitios degenerativos como en la enfermedad de Parkinson o Alzheimer en los núcleos basales, materia blanca y gris; 2. columna vertebral como en el caso de la esclerosis lateral amiotrófica (ELA); 3. cuello uterino para prevenir la infección por HPV; 4. articulaciones que padecen inflamación crónica y activa; 5. dermis como en el caso de la psoriasis; 6. Sitios nerviosos sensoriales y simpáticos para un efecto analgésico; 7. implantación intraósea; 8. Sitios de infección crónica y aguda; 9. intravaginal; 10. oído interno--sistema auditivo, laberinto del oído interno, sistema vestibular; 11. intratraqueal; 12. intracardiaco; coronario, epicardíaco; 13. vejiga urinaria; 14. Sistema biliar; 15. tejido parenquimatoso que incluye de manera no taxativa el riñón, hígado, bazo; 16. nódulos linfáticos; 17. glándulas salivares; 18. Encías; 19. intraarticular (al interior de las articulaciones); 20. intraocular; 21. tejido cerebral; 22. ventrículos cerebrales; 23. cavidades, que incluyen la cavidad abdominal (por ejemplo, de manera no taxativa, para el cáncer de ovario); 24. intraesofágica y 25. intrarrectal.

Opcionalmente, la inserción del sistema (por ejemplo, un dispositivo que contiene la composición) se asocia con la inyección de material a la ECM en el sitio diana y las inmediaciones de ese sitio para afectar a la temperatura y/o pH local y/u otros factores biológicos que afectan a la difusión del fármaco y/o cinética del fármaco en la ECM del sitio diana y las inmediaciones de tal sitio.

Opcionalmente, de acuerdo con algunas formas de realización, la liberación de dicho agente se podría asociar con aparatos de detección y/o activación operados antes de la inserción y/o en esta y/o después de esta, mediante métodos de activación y/o aceleración/desaceleración no invasivos y/o mínimamente invasivos y/o de otro tipo, que incluyen métodos o dispositivos de haz de rayos láser, radiación, enfriamiento y calentamiento térmico, y ultrasónicos, que incluyen el ultrasonido focalizado y/o RF (radiofrecuencia), y activadores químicos.

De acuerdo con otras formas de realización de la publicación de patente de los EE.UU. 20110195123, el fármaco comprende preferentemente un ARN, por ejemplo, para casos localizados de cáncer en la mama, páncreas, cerebro, riñón, vejiga, pulmón y próstata como se ha descrito anteriormente. Aunque se ilustra con iARN, se pueden usar muchos fármacos para encapsularse en Loder, y se pueden usar en asociación con esta invención, siempre que dichos fármacos se puedan encapsular con el sustrato Loder, tal como por ejemplo una matriz, y este sistema se puede usar y/o adaptar para suministrar el sistema CRISPR-Cas usados en los métodos y composiciones reivindicados de la presente invención.

Como otro ejemplo de una aplicación específica, las enfermedades neurodegenerativas y musculares degenerativas se desarrollan debido a una expresión génica anómala. El suministro localizado de los ARN podrá tener propiedades terapéuticas para interferir con tal expresión génica anómala. El suministro localizado de fármacos antiapoptósicos, antiinflamatorios y antidegenerativos incluidos los fármacos con un bajo peso molecular y macromoléculas también podrá ser opcionalmente terapéutico. En tales casos, se aplica el Loder para una liberación prolongada a una velocidad constante y/o a través de un dispositivo dedicado que se implanta por separado. Todo esto se podrá utilizar y/o adaptar al sistema CRISPR-Cas usados en los métodos y composiciones reivindicados de la presente invención.

Como otro ejemplo más de la aplicación específica, los trastornos cognitivos y psiquiátricos se tratan con manipuladores génicos. La atenuación génica es una opción de tratamiento. Los Loders que suministran de manera localizada agentes a sitios del sistema nervioso central son opciones terapéuticas para trastornos cognitivos y psiquiátricos que incluyen, sin limitación, la psicosis, enfermedades bipolares, trastornos neuróticos y enfermedades conductuales. Los Loder también podrían suministrar de manera localizada fármacos, incluidos fármacos de bajo peso molecular y macromoléculas tras la implantación en sitios cerebrales específicos. Todo esto se podrá utilizar y/o adaptar al sistema CRISPR-Cas usados en los métodos y composiciones reivindicados de la presente invención.

Como otro ejemplo de la aplicación específica, el silenciamiento de mediadores inmunitarios innatos y/o adaptativos en sitios localizados posibilita la prevención del rechazo del trasplante de órganos. El suministro localizado de ARN y reactivos inmunomoduladores con el Loder implantado en el órgano trasplantado y/o sitio implantado produce la supresión inmunitaria localizada repeliendo las células inmunitarias tales como CD8 activadas contra el órgano trasplantado. Todo esto se podrá utilizar y/o adaptar al sistema CRISPR-Cas usados en los métodos y composiciones

reivindicados de la presente invención.

Como otro ejemplo de la aplicación específica, los factores de crecimiento vascular, incluidos los VEGF y la angiogenina y otros, son esenciales para la neovascularización. El suministro localizado de factores, péptidos, peptidomiméticos o la supresión de sus represores es una modalidad terapéutica importante; el silenciamiento de los represores y el suministro localizado de los factores, péptidos, macromoléculas y fármacos de bajo peso molecular que estimulan la angiogénesis con el Loder es terapéutico para la enfermedad vascular periférica, sistémica y cardíaca.

El método de inserción, tal como implantación, puede opcionalmente, estar ya siendo utilizado en otros tipos de implantación tisular y/o para inserciones y/u obtención de muestras tisulares opcionalmente sin modificaciones o, como alternativa, opcionalmente únicamente con modificaciones poco importantes de dichos métodos. Tales métodos incluyen opcionalmente, de manera no taxativa, métodos de braquirradioterapia, biopsia, endoscopia con y/o sin ultrasonido tal como la ERCP, métodos estereotácticos en el tejido cerebral, laparoscopia, incluida la implantación con un laparoscopio en las articulaciones, órganos abdominales, la pared de la vejiga y las cavidades corporales.

La tecnología de dispositivo implantable discutida en el presente documento puede emplearse con las divulgaciones del presente documento y por lo tanto mediante la presente divulgación y el conocimiento en la técnica, pueden suministrarse el sistema CRISPR-Cas o los componentes del mismo o moléculas de ácido nucleico del mismo o componentes codificantes del mismo a través de un dispositivo implantable.

Métodos de selección específica de pacientes

Un sistema de direccionamiento hacia ácidos nucleicos que se dirige a ADN, por ejemplo, a repeticiones de trinucleótidos, se puede usar para seleccionar pacientes o muestras de pacientes para determinar la presencia de dichas repeticiones. Las repeticiones pueden ser el diana del ARN del sistema de direccionamiento hacia ácidos nucleicos y, si existe unión entre el anterior y el sistema de direccionamiento hacia ácidos nucleicos, esta unión se puede detectar para indicar de esta forma que dicha repetición está presente. De esta manera, se puede usar un sistema de direccionamiento hacia ácidos nucleicos para seleccionar pacientes o muestras de pacientes para determinar la presencia de la repetición. A continuación, se pueden administrar al paciente compuesto(s) adecuado(s) para resolver la afección o se puede administrar un sistema de direccionamiento hacia ácidos nucleicos para unirse y producir la inserción, supresión o mutación y aliviar la afección.

La invención utiliza ácidos nucleicos para unirse a secuencias de ADN diana.

ARNm de la proteína efectora de CRISPR y ARN guía

El ARNm de la enzima CRISPR y el ARN guía también se podrán suministrar por separado. El ARNm de la enzima CRISPR se podrá suministrar antes que el ARN guía para dar tiempo a que se exprese la enzima CRISPR. El ARNm de la enzima CRISPR se podrá administrar 1-12 horas (preferentemente aproximadamente 2-6 horas) antes de la administración del ARN guía.

Como alternativa, el ARNm de la enzima CRISPR y el ARN guía se pueden administrar juntos. Convenientemente, se puede administrar una segunda dosis de refuerzo de ARN guía 1-12 horas (de manera preferida aproximadamente 2-6 horas) después de la administración inicial del ARNm de la proteína efectora de CRISPR + el ARN guía.

La proteína CRISPR efectora usada en los métodos y composiciones reivindicados de la presente invención, es decir una proteína efectora Cpf1 aquí se denomina a veces una enzima CRISPR. Se apreciará que la proteína efectora se basa en una enzima o se deriva de la misma, de manera que en algunas formas de realización la expresión "proteína efectora" desde luego incluye al término 'enzima'. Sin embargo, también se apreciará que, según sea necesario, en algunas formas de realización la proteína efectora puede unirse a ADN o ARN, pero no necesariamente con actividad de corte o corte de una sola hebra, incluyendo la función de la proteína efectora de Cas-muerta.

Las administraciones adicionales del ARNm de la enzima CRISPR y/o ARN guía podrán ser útiles para lograr los niveles más eficaces de modificación genómica. En algunas formas de realización, la alteración fenotípica es preferiblemente el resultado de la modificación del genoma cuando se dirige hacia una enfermedad genética, especialmente en métodos de terapia y preferiblemente donde se provee un molde de reparación para corregir o alter el fenotipo.

En algunas formas de realización, las enfermedades que pueden ser objetivos incluyen a aquellas relacionadas con defectos del corte y empalme causantes de enfermedades.

En algunas formas de realización, los blancos celulares incluyen células madre hematopoyéticas/progenitoras (CD34+); células T humanas; y células del ojo (células retinales) – por ejemplo, precursor fotorreceptor.

En algunas formas de realización los genes diana incluyen: Beta globina humana – HBB (para tratar la anemia falciforme, inclusive por estímulo de la conversión de genes (usando un gen de HBD estrechamente relacionado como molde endógeno)); CD3 (células T); y CEP920 - retina (ojo).

En algunas formas de realización, la enfermedad objetivo también incluye: cáncer; anemia falciforme (en base a una

mutación puntual); VIH; Beta talasemia; y enfermedad oftálmica u ocular– por ejemplo, amaurosis congénita de Leber (LCA, por las siglas en inglés de Leber Congenital Amaurosis) causante de defectos por corte y empalme.

En algunas formas de realización, los métodos de suministro incluyen: Administración "directa" mediada por lípido catiónico de complejo Enzima-Guía (RiboNucleoProteína) y electroporación de ADN de plásmido.

Los métodos de la invención pueden comprender además el suministro de moldes, por ejemplo, reparación de moldes, que pueden ser ODNhd u ODNhs, véase a continuación. El suministro de moldes se puede realizar mediante el suministro simultáneo o por separado de cualquiera o de todos de la enzima CRISPR o guía o de ambas y mediante el mismo mecanismo de suministro o uno diferente. En algunas formas de realización, es preferible que el molde se suministre junto con la guía, y, preferiblemente, también con la enzima CRISPR. Un ejemplo puede ser un vector AAV.

Los métodos de la invención pueden comprender además: (a) suministrar a la célula un oligodesoxinucleótido de doble hebra (ODNhd) que comprende sobreextensiones complementarias a las sobreextensiones creadas por rotura de dicha doble hebra, donde dicha ODNhd se integra al locus de interés; o – (b) suministrar a la célula un oligodeoxinucleótido de hebra simple (ODNhs), donde dicho ODNhs actúa como molde para la reparación dirigida por homología de dicha rotura de la doble hebra. Los métodos de la invención se pueden usar para la prevención o el tratamiento de la enfermedad en un individuo, opcionalmente donde dicha enfermedad es causada por un defecto en dicho locus de interés. Los métodos de la invención se pueden realizar *ex vivo* sobre una célula tomada del individuo, opcionalmente donde dicha célula se retorna al individuo.

Para minimizar la toxicidad y los efectos no específicos, será importante controlar la concentración suministrada del ARNm de la enzima CRISPR y del ARN guía. Se pueden determinar las concentraciones óptimas del ARNm de la enzima CRISPR y del ARN guía probando diferentes concentraciones en un modelo en células o en animales y utilizando una secuenciación profunda para analizar el grado de modificación en loci genómicos fuera del diana potenciales. Por ejemplo, para la secuencia guía que se dirige a 5'-GAGTCCGAGCAGAAGAAGAA-3' (SEQ ID NO: 23) en el gen EMX1 del genoma humano, se puede utilizar la secuenciación profunda para evaluar el nivel de modificación de los dos siguientes loci no específicos, 1: 5'-GAGTCCTAGCAGGAGAAGAA-3' (SEQ ID NO: 24) y 2: 5'-GAGTCTAAGCAGAAGAAGAA-3' (SEQ ID NO: 25). Debería elegirse la concentración que proporcione el nivel más elevado de modificación en el diana y a la vez minimice el nivel de modificación fuera del diana para el suministro *in vivo*.

Sistemas inducibles

En algunas formas de realización, una enzima CRISPR podrá formar un componente de un sistema inducible. La naturaleza inducible del sistema permitiría el control espaciotemporal de la edición génica o expresión génica utilizando una forma de energía. La forma de energía podrá incluir, de manera no taxativa, la radiación electromagnética, energía sonora, energía química y energía térmica. Los ejemplos de sistemas inducibles incluyen promotores inducibles por tetraciclina (Tet-On o Tet-Off), sistemas activadores de la transcripción de doble híbrido que son moléculas de bajo peso molecular (FKBP, ABA, etc.) o sistemas inducibles por la luz (Fitocromo, dominios LOV o criptocromo). En una forma de realización, la enzima CRISPR puede ser parte de un efector transcripcional inducible por la luz (LITE por las siglas en inglés de Light Inducible Transcriptional Effector) para dirigir los cambios de la actividad transcripcional de manera específica de la secuencia. Los componentes de una luz podrán incluir una enzima CRISPR, un heterodímero de un citocromo que responde a la luz (por ejemplo, de *Arabidopsis thaliana*) y un dominio de activación/represión transcripcional. Los ejemplos adicionales de proteínas de unión a ADN inducibles y métodos para su uso se proporcionan en US 61/736.465 y US 61/721.283 y WO 2014/018423 A2.

Sistemas autoinactivantes

Una vez que se han editado todas las copias de un gen en el genoma de una célula, la expresión continuada de CRISPRP/Cpf1p es esta célula ya no es más necesaria. De hecho, sería indeseable una expresión sostenida para el caso de los efectos fuera de diana en sitios genómicos no pretendidos, *etc.* Por lo tanto, sería útil una expresión limitada en el tiempo. La expresión inducible ofrece una solución, pero además, los solicitantes han manipulado un sistema CRISPR autoinactivante que se basa en el uso de una secuencia diana guía no codificante en el propio vector CRISPR. De esta manera, tras comenzar la expresión, el sistema CRISPR-Cas conducirá a su propia destrucción, pero antes de que se complete la destrucción tendrá tiempo de editar las copias genómicas del gen diana (que, con una mutación puntual normal en una célula diploide, requiere cuando menos dos ediciones). Sencillamente, el sistema CRISPR-Cas autoinactivante incluye ARN adicional (es decir, ARN guía) que dirige la secuencia de codificación para la propia enzima CRISP o que dirige una o más secuencias guía no codificantes complementarias a las secuencias singulares presentes en una o más de los siguientes:

(a) dentro del promotor, lo que impulsa la expresión de los elementos de ARN no codificantes,

(b) dentro del promotor conduciendo la expresión de la proteína efectora del gen Cpf1,

(c) dentro de los 100 pb del codón de inicio de la traducción ATG en la secuencia codificante de la proteína efectora Cpf1,

(d) dentro de la repetición terminal invertida (iTR) de un vector de suministro viral, por ejemplo, en el genoma de

AAV.

Además, este ARN puede suministrarse mediante un vector, por ejemplo, un vector separado o el mismo vector que codifica para el complejo CRISPR. Cuando se proporciona mediante un vector separado, el ARN de CRISPR que dirige la expresión de Cas puede administrase secuencial o simultáneamente. Cuando se administra secuencialmente, el ARN de CRISPR que dirige la expresión de Cas es para suministrarse después del ARN de CRISPR que se pretende para, por ejemplo, editar o modificar el gen. Este periodo puede ser un periodo de minutos (por ejemplo, 5 minutos, 10 minutos, 20 minutos, 30 minutos, 45 minutos, 60 minutos). Este periodo puede ser un periodo de horas (por ejemplo, 2 horas, 4 horas, 6 horas, 8 horas, 12 horas, 24 horas). Este periodo puede ser un periodo de días (por ejemplo, 2 días, 3 días, 4 días, 7 días). Este periodo puede ser un periodo de semanas (por ejemplo, 2 semanas, 3 semanas, 4 semanas). Este periodo puede ser un periodo de meses (por ejemplo, 2 meses, 4 meses, 8 meses, 12 meses). Este periodo puede ser un periodo de años (2 años, 3 años, 4 años). De esta manera, la enzima Cas se asocia con un primer ARNg capaz de hibridarse a un primer diana, tal como un locus genómico o loci genómicos de interés y asume la función o las funciones deseadas del sistema CRISPR-Cas (por ejemplo, la modificación génica); y posteriormente, la enzima Cas puede asociarse a continuación con el segundo ARNg capaz de hibridarse a la secuencia que comprende al menos parte del casete de Cas o CRISPR. Cuando el ARN guía se direcciona contra las secuencias que codifican para la expresión de la proteína Cas, la enzima se vuelve inactiva y el sistema se vuelve autoinactivante. De la misma manera, el ARN de CRISPR que dirige la expresión de Cas se aplica mediante, por ejemplo, liposomas, lipofección, partículas, microvesículas como se explica en la presente, se puede administrar secuencial o simultáneamente. De forma similar, se puede usar la autoinactivación para la inactivación de uno o más ARN guía usados para dirigirse a una o más dianas.

En algunos aspectos, se proporciona un ARNg singular que es capaz de hibridación a una secuencia en la dirección 3' de un codón de inicio de la enzima CRISPR, por lo cual, tras un periodo de tiempo existe una pérdida de expresión de la enzima CRISPR. En algunos aspectos, se proporcionan uno o más ARNg que son capaces de hibridación a una o más regiones codificantes o no codificantes del polinucleótido que codifica el sistema CRISPR-Cas, por lo cual, tras un periodo de tiempo existe la inactivación de uno o más, o en algunos casos todos, los sistemas CRISPR-Cas. En algunos aspectos del sistema, y no para quedar limitado por teoría alguna, la célula puede comprender una pluralidad de complejos CRISPR-Cas, en donde un primer subconjunto de complejos CRISPR comprende un primer ARN guía capaz de dirigirse a un locus genómico o a loci genómicos que se van a editar, y un segundo subconjunto de complejos CRISPR comprende al menos un segundo ARN guía capaz de dirigirse al polinucleótido que codifica el sistema CRISPR-Cas, en donde el primer subconjunto de complejos CRISPR-Cas media en la edición del locus genómico o de los loci genómicos dirigidos y el segundo subconjunto de complejos CRISPR inactiva eventualmente el sistema CRISPR-Cas, inactivando por tanto la expresión de CRISPR-Cas adicional en la célula.

De esta manera, los métodos y composiciones reivindicados de la invención pueden implicar proporcionar un sistema CRISPR-Cas que comprende uno o más vectores para el suministro a una célula eucariota, donde el vector o los vectores codifican: (i) una enzima CRISPR; (ii) un primer ARN guía capaz de hibridar con una secuencia diana en la célula; (iii) un segundo ARN guía capaz de hibridar con una o más secuencias diana en el vector que codifica para la enzima CRISPR, cuando se expresa dentro de la célula: el primer ARN guía dirige la unión específica de secuencia de un primer complejo CRISPR con la secuencia diana en la célula; el segundo ARN guía dirige la unión específica de secuencia de un segundo complejo CRISPR con la secuencia diana en el vector que codifica para la enzima CRISPR; los complejos CRISPR comprenden una enzima CRISPR unida a un ARN guía, de forma que un ARN guía puede hibrida con su secuencia diana; y el segundo complejo CRISPR inactiva al sistema CRISPR-Cas para prevenir la expresión continua de la enzima CRISPR por parte de la célula.

Las diversas secuencias codificantes (enzima CRISPR y ARN guías) se pueden incluir en un vector individual o en múltiples vectores. Por ejemplo, es posible codificar la enzima en un vector y las diversas secuencias de ARN en otro vector, o codificar la enzima y un ARN guía en un vector, y el restante ARN guía en otro vector, o cualquier otra permutación. En general, se prefiere un sistema que utiliza un total de uno o dos vectores diferentes.

Cuando se usan múltiples vectores, es posible suministrarlos en número desigual, e idealmente con un exceso de un vector que codifica el primer ARN guía con respecto al segundo ARN guía, ayudando por tanto al retraso final de la inactivación del sistema CRISPR hasta que la edición del genoma ha tenido la oportunidad de producirse.

El primer ARN guía puede dirigir cualquier secuencia de interés en el genoma como se describe en otra parte en la presente. El segundo ARN guía dirige una secuencia en el vector que codifica la enzima Cpf1 de CRISPR, e inactiva por tanto la expresión de la enzima procedente de este vector. De esta manera, la secuencia diana en el vector debe ser capaz de inactivar la expresión. Las secuencias diana apropiadas pueden estar, por ejemplo, próximas o comprendidas dentro del codón de inicio de la traducción para la secuencia que codifica para Cpf1p, en una secuencia no codificante en el promotor que impulsa la expresión de elementos de ARN no codificantes, en el promotor que impulsa la expresión del gen Cpf1p, dentro de los 100 pb del codón ATG de inicio de la traducción en la secuencia que codifica para Cas, y/o en la repetición terminal invertida (iTR) de un vector de suministro viral, por ejemplo, en el genoma de AAV. Una rotura de región bicatenaria próxima a esta región puede inducir un desplazamiento del marco en la secuencia que codifica Cas, produciendo una pérdida de expresión de la proteína. Una secuencia diana alternativa para el ARN guía "autoinactivante" tendría como objetivo editar/inactivar las regiones/secuencias reguladoras necesarias para la expresión del sistema CRISPR-Cpf1 o para la estabilidad del vector. Por ejemplo, si el

promotor de la secuencia de codificación de Cas está perturbado, entonces se puede inhibir o evitar la transcripción. De forma similar, si un vector incluye secuencias para la replicación, el mantenimiento o la estabilidad, entonces es posible dirigir estas. Por ejemplo, en un vector de AAV la secuencia diana útil está en la iTR. Otras secuencias útiles para el direccionamiento pueden ser las secuencias promotoras, los sitios de poliadenilación, etc.

Además, si los ARN guía se expresan en formato de matriz, los ARN guía "autoinactivantes" que se dirigen a ambos promotores simultáneamente darán como resultado la escisión de los nucleótidos intermedios dentro de la construcción de expresión CRISPR-Cas, lo que conduce eficazmente a su completa inactivación. De forma similar, la escisión de los nucleótidos intervinientes dará resultado cuando los ARN guía se dirigen a ambas iTR, o se dirigen a dos o más componentes de CRISPR-Cas diferentes simultáneamente. La autoinactivación tal como se explica en la presente es aplicable, en general, con sistemas CRISPR-Cas con el objetivo de proveer una regulación de CRISPR-Cas. Por ejemplo, la autoinactivación como se explica en la presente puede aplicarse a la reparación CRISPR de mutaciones, por ejemplo, trastornos de extensión, como se explica en la presente. Como resultado de esta autoinactivación, la reparación de CRISPR está solo transitoriamente activa.

La adición de nucleótidos que no se dirigen al extremo 5' (por ejemplo, 1 - 10 nucleótidos, preferentemente 1 - 5 nucleótidos) del ARN guía "autoinactivante" se puede usar para retrasar su procesamiento y/o modificar su eficacia como un medio de asegurar la edición en el locus genómico dirigido antes de la inactivación de CRISPR-Cas.

En un aspecto de la autoinactivación del sistema AAV-CRISPR-Cas, los plásmidos que expresan simultáneamente uno o más ARN guías contra secuencias genómicas de interés (por ejemplo, 1-2, 1-5, 1-10, 1-15, 1-20, 1-30) pueden establecerse con ARNg "autoinactivantes" que se dirigen a una secuencia SpCas9 en, o cerca del sitio de inicio ATG manipulado (por ejemplo, dentro de los 5 nucleótidos, dentro de los 15 nucleótidos, dentro de los 30 nucleótidos, dentro de los 50 nucleótidos, dentro de los 100 nucleótidos). Una secuencia reguladora en la región del promotor de U6 también puede ser el diana de un ARN guía. Los ARN guías conducidos por U6 se pueden diseñar en un formato de matriz de forma que se pueden liberar simultáneamente múltiples secuencias de ARN guía. Cuando se administran primero en tejidos/células diana (célula izquierda) los ARN guías comienzan a acumularse mientras que los niveles de Cas aumentan en el núcleo. La Cas se compleja con todos los ARN guías para mediar la edición de genoma y la autoinactivación de los plásmidos que tienen CRISPR-Cas.

Un aspecto de un sistema CRISPR-Cas de autoinactivación es la expresión en un formato de matriz individual o en tándem de entre 1 y hasta 4 o más secuencias guía diferentes, por ejemplo, hasta aproximadamente 20 o aproximadamente 30 secuencias guía. Cada secuencia guía autoinactivante individual puede dirigirse a una diana diferente. La mencionada puede procesarse a partir de, por ejemplo, un transcrito pol3 quimérico. Se puede usar promotores Pol3 tales como promotores U6 o H1. Promotores Pol2 tales como los mencionado a través de la presente. Las secuencias repetidas terminales invertidas (iTR) pueden flanquear al promotor de Pol3 – ARN guía(s) – promotor de Pol2 - Cas.

Un aspecto de un transcripto de arreglo en tándem es que una o más guías editan al uno o más blancos mientras que una o más guías autoinactivantes inactivan al sistema CRISPR-Cas. Por lo tanto, por ejemplo, el sistema CRISPR-Cas descrito para reparar trastornos de expansión se puede combinar directamente con el sistema CRISPR-Cas autoinactivante que se describe en la presente. Dicho sistema puede, por ejemplo, tener dos guías dirigidas contra la región diana para la reparación, así como también al menos una tercera guía dirigida a la autoinactivación del CRISPR-Cas. Se hace referencia a la Solicitud de No de Acta PCT/US2014/069897, titulada "Compositions And Methods Of Use Of Crispr-Cas Systems In Nucleotide Repeat Disorders," publicada el 12 de diciembre de 2014 como WO/2015/089351.

El ARN guía puede ser una guía control. Por ejemplo, puede diseñarse para que esté dirigido a una secuencia de ácido nucleico que codifica para la enzima CRISPR misma, como se describe en la referencia US2015232881A1. En algunas formas de realización, puede proveerse un sistema o composición solo con el ARN guía diseñado para estar dirigido a la secuencia de ácido nucleico que codifica para la enzima CRISPR. Adicionalmente, el sistema o composición puede proveerse con el ARN guía diseñado para estar dirigido a la secuencia de ácido nucleico que codifica para la enzima CRISPR, así como a la secuencia de ácido nucleico que codifica para la enzima CRISPR y, opcionalmente un segundo ARN guía y, además opcionalmente, un molde de reparación. El segundo ARN guía puede ser el objetivo primario del sistema o composición CRISPR (tal como terapéutico, diagnóstico, de noqueo, etc., como se define en la presente). En este sentido, el sistema o composición es autoinactivante. Esto se ejemplifica en relación a Cas9 en US2015232881A1 (también publicado como WO2015070083) (A1) referido en otro lugar en la presente, y puede extrapolarse a Cpf1.

Enzimas usadas en los métodos y composiciones reivindicados de la invención usadas en una estrategia de direccionamiento multiplex (en tándem)

Los inventores han mostrado que las enzimas CRISPR como se definen en la presente pueden emplear más de un ARN guía sin perder actividad. Esto permite el uso de las enzimas CRISPR, sistemas o complejos como se definen en la presente para dirigir a múltiples blancos de ADN, genes o loci génicos, con un sistema o complejo de enzima individual como se define en la presente. Los ARN guías pueden disponerse en tándem, separados opcionalmente por una secuencia nucleotídica tal como una repetición directa como se define en la presente. La posición de los diferentes ARN guías en el tándem no influye en la actividad. Se hace notar que los términos "sistema CRISPR-Cas",

"complejo CRISP-Cas" "complejo CRISPR" y "sistema CRISPR" se usan como sinónimos. Adicionalmente los términos "enzima CRISPR", "enzima Cas", o "enzima CRISPR-Cas", pueden usarse como sinónimos. En formas de realización preferidas, la mencionada enzima CRISPR, enzima CRISP-Cas o enzima Cas es Cpf1, o cualquiera de las variantes modificadas o mutadas de la misma descrita en alguna parte de la presente.

En un aspecto, los métodos y composiciones reivindicados de la invención implican proveer una CRISPR de origen no natural o de Tipo V como se describe en la presente, tal como sin limitación Cpf1 como se describe en otra parte en la presente, usada para el direccionamiento en tándem o multiplex. Se debe entender que cualquier enzima CRISPR (o CRISPR-Cas o Cas), complejo, o sistema de acuerdo con la invención como se describe en otra parte en la presente puede usarse en la mencionada estrategia. Cualquiera de los métodos, productos, composiciones y usos como se describe en otra parte en la presente son igualmente aplicables con la estrategia de direccionamiento multiplex o en tándem detallada adicionalmente más adelante. Para mayor orientación, se proveen los siguientes aspectos y formas de realización particulares.

En un aspecto, la invención provee el uso de una enzima Cpf1, complejo o sistema como se define en la presente para el direccionamiento a múltiples loci génicos. En una forma de realización, esto puede establecerse con el uso de secuencias de ARN guía (ARNg) múltiples (en tándem o multiplexados).

En un aspecto, la invención provee los métodos reivindicados en la presente para usar uno o más elementos de una enzima Cpf1, complejo o sistema como se define en la presente para el direccionamiento en tándem o multiplexado, en donde el mencionado sistema CRISP comprende múltiples secuencias de ARN guía. Preferiblemente, las mencionadas secuencias de ARNg están separadas por una secuencia nucleotídica, tal como una repetición directa como se define en otra parte en la presente.

La enzima Cpf1, sistema o complejo como se define en la presente provee un medio eficaz para modificar múltiples polinucleótidos diana. La enzima Cpf1, sistema o complejo como se define en la presente tiene una variedad amplia de utilidades que incluyen modificar (por ejemplo, eliminar, insertar, translocar, inactivar, activar) uno o más polinucleótidos diana en una multiplicidad de tipos celulares. Como tal la enzima Cpf1, sistema o complejo como se define en la presente para su uso en los métodos y composiciones reivindicados de la invención tiene un espectro amplio de aplicaciones en, por ejemplo, terapia génica, selección de fármacos, diagnóstico de enfermedades, y pronóstico, incluyendo direccionamiento a múltiples loci génicos en un sistema CRISPR individual.

En un aspecto, los métodos y composiciones reivindicados de la invención implican proveer una enzima Cpf1, sistema o complejo como se define en la presente, es decir un complejo Cpf1 CRISPR-Cas que tiene una proteína Cpf1 que tiene por lo menos un dominio de desestabilización asociado al mismo, y múltiples ARN guías direccionados a múltiples moléculas de ácido nucleico tales como moléculas de ADN, mediante el cual cada uno de los múltiples ARN guías están dirigidos específicamente a su correspondiente molécula de ácido nucleico, por ejemplo, molécula de ADN. Cada diana molécula de ácido nucleico, por ejemplo, molécula de ADN puede codificar para un producto genético o abarca un locus génico. El uso de múltiples ARN guías permite entonces el direccionamiento a múltiples loci génicos o múltiples genes. En algunas formas de realización la enzima Cpf1 puede clivar la molécula de ADN que codifica para el producto genético. En algunas formas de realización la expresión del producto genético está alterada. La proteína Cpf1 y los ARN guías no existen juntos naturalmente. La invención comprende los ARN guías que comprenden secuencias guía dispuestas en tándem. La invención además comprende secuencias codificantes para que la proteína Cpf1 sea optimizada en codones para la expresión en una célula eucariótica. En una forma de realización preferida la célula eucariótica es una célula de mamífero, una célula vegetal o una célula de levadura y en una forma de realización preferida adicional la célula de mamífero es una célula humana. La expresión del producto genético puede disminuirse. La enzima Cpf1 puede formar parte de un sistema o complejo CRISPR, que además comprende ARN guías dispuestos en tándem (ARNg) que comprende una serie de 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 25, 25, 30, o más de 30 secuencias guía, cada una con capacidad de hibridar específicamente con una secuencia diana en un locus genómico de interés en una célula. En algunas formas de realización, el sistema o complejo Cpf1 CRISPR funcional se une a las múltiples secuencias diana. En algunas formas de realización, el sistema o complejo CRISPR funcional puede editar las múltiples secuencias diana, por ejemplo, las secuencias diana pueden comprender un locus genómico, y en algunas formas de realización puede haber una alteración en la expresión del gen. En algunas formas de realización, el sistema o complejo CRISPR funcional puede comprender dominios funcionales adicionales. En algunas formas de realización, los métodos y composiciones reivindicados de la invención implican alterar o modificar la expresión de múltiples productos genéticos. El método puede comprender la introducción *in vitro* o *ex vivo* a una célula que contiene los mencionados ácidos nucleicos diana, por ejemplo, moléculas de ADN, o que contienen y expresan ácido nucleico diana, por ejemplo, moléculas de ADN; por ejemplo, los ácidos nucleicos diana puede codificar productos genéticos o se proveen par la expresión de productos genéticos (por ejemplo, secuencias regulatorias).

En formas de realización preferidas la enzima CRISPR usada para el direccionamiento multiplexado es Cpf1, o el sistema o complejo CRISPR comprende Cpf1. En algunas formas de realización, la enzima CRISPR usada para el direccionamiento multiplexado es AsCpf1, o el sistema o complejo CRISPR usado para el direccionamiento multiplexado comprende una AsCpf1. En algunas formas de realización, la enzima CRISPR es una LbCpf1, o el sistema o complejo CRISPR comprende LbCpf1. En algunas formas de realización, la enzima Cpf1 usada para el direccionamiento multiplexado cliva ambas hebras de ADN para producir una rotura de hebra doble (DSB). En algunas formas de realización, la enzima CRISPR usada para el direccionamiento multiplexado es una nickasa. En algunas

formas de realización, la enzima Cpf1 usada para el direccionamiento multiplexado es una nickasa dual. En algunas formas de realización, la enzima Cpf1 usada para el direccionamiento multiplexado es una enzima Cpf1 tal como una enzima DD Cpf1 como se define en otra parte en la presente.

En algunas formas de realización generales, la enzima Cpf1 usada para el direccionamiento multiplexado está asociada con uno o más dominios funcionales. En algunas formas de realización más específicas, la enzima CRISPR usada para el direccionamiento multiplexado es una deadCpf1 como se define en otra parte en la presente.

En un aspecto, los métodos y composiciones reivindicados de la presente invención implican proveer un medio para administrar la enzima Cpf1, sistema o complejo para usar en el direccionamiento múltiple como se define en la presente o los polinucleótidos definidos en la presente. Los ejemplos no limitantes del medio de administración mencionado son por ejemplo partícula(s) que administra(n) componente(s) del complejo, vector(es) que comprenden el o los polinucleótidos descritos en la presente (por ejemplo, que codifican para la enzima CRISPR, proveyendo los nucleótidos que codifican para el complejo CRISPR). En algunas formas de realización, el vector puede ser un plásmido o un vector viral tal como AAV, o lentivirus. La transfección transitoria con plásmidos, por ejemplo, en células HEK puede ser ventajoso, especialmente debido a las limitaciones de tamaño de los AAV y que si bien Cpf1 encaja dentro de los AAV, uno puede alcanzar un límite superior con ARN guías adicionales.

También se provee un modelo que expresa constitutivamente la enzima Cpf1, complejo o sistema como se usa en la presente para usar en el direccionamiento multiplexado. El organismo no humano puede ser transgénico y puede haber sido transfectado con los vectores presentes o puede ser la descendencia de un organismo no humano transfectado de esa manera. En un aspecto adicional, los métodos y composiciones reivindicados de la presente invención implican proveer composiciones que comprenden la enzima CRISPR, sistema y complejo como se define en la presente o los polinucleótidos o vectores descritos en la presente. También se proveen sistemas o complejos Cpf1 CRISPR que comprenden múltiples ARN guías, preferiblemente en un formato dispuesto en tándem. Los diferentes ARN guías pueden separarse por secuencias nucleotídicas tales como repeticiones directas.

También se provee un método para tratar a un sujeto, por ejemplo, un sujeto que lo requiere, que comprende inducir la edición de genes por transformación del sujeto con el polinucleótido que codifica para el sistema o complejo Cpf1 CRISPR o cualquiera de los polinucleótidos o vectores descritos en la presente y administración de los mismos al sujeto. También puede proveerse un molde de reparación adecuado, por ejemplo provisto por un vector que comprende el mencionado molde de reparación. También se provee un método para tratar a un sujeto, por ejemplo, un sujeto que lo requiere, que comprende inducir la activación o represión transcripcional de múltiples loci genéticos diana por transformación del sujeto con los polinucleótidos o vectores descritos en la presente, en donde el mencionado polinucleótido o vector codifica para o comprende la enzima Cpf1, complejo o sistema que comprende múltiples ARN guía, preferiblemente dispuestos en tándem. Cuando cualquier tratamiento se realiza *ex vivo*, por ejemplo, en un cultivo celular, entonces se apreciará que el término "sujeto" puede reemplazarse por la frase "célula o cultivo celular".

También se proveen composiciones que comprenden enzima, complejo o sistema de Cpf1 que comprenden múltiples ARN guías, preferiblemente dispuestos en tándem, o el polinucleótido o vector que codifica para o que comprende la mencionada enzima, complejo o sistema de Cpf1 que comprende múltiples ARN guías, preferiblemente dispuestos en tándem, para usar en los métodos de tratamiento como se define en otra parte en la presente. Puede proveerse un conjunto de partes que incluye las mencionadas composiciones. También se provee el uso de la mencionada composición en la fabricación de un medicamento para dichos métodos de tratamiento. La presente invención también provee el uso de un sistema Cpf1 CRISPR en la selección, por ejemplo, selección de ganancia de función. Las células que se fuerzan artificialmente para sobreexpresar un gen tienen la capacidad de disminuir la expresión del gen en el tiempo (estableciendo nuevamente el equilibrio) por ejemplo por bucles de retroalimentación negativos. Al momento en que comienza la selección el gen no regulado podría haberse reducido nuevamente. El uso de un activador de Cpf1 inducible permite inducir la transcripción justo antes de la selección y minimizar de esta manera la posibilidad de hallazgos negativos falsos. En consecuencia, con el uso de la presente invención en la selección, por ejemplo, selección de ganancia de función, puede minimizarse la posibilidad de resultados negativos falsos.

En un aspecto, los métodos y composiciones reivindicados de la invención implican proveer un sistema CRISPR diseñado, de origen no natural que comprende una proteína Cpf1 y múltiples ARN guías que cada uno hace diana específicamente en una molécula de ADN que codifica para un producto genético en una célula, en donde los múltiples ARN guías están dirigidos cada uno a su molécula de ADN especifica que codifica para el producto genético y la proteína Cpf1 cliva la molécula de ADN diana que codifica para el producto genético, por lo cual se altera la expresion del producto genético; y, en donde la proteína CRISPR y los ARN guías no se presentan juntos naturalmente. La invención comprende los múltiples ARN guías que comprenden múltiples secuencias guía, preferiblemente separadas por una secuencia nucleotídica tal como una repetición directa. En una forma de realización de la invención la proteína CRISPR es una proteína CRISPR-Cas tipo V o VI y en una forma de realización más preferida la proteína CRIPSR es una proteína Cpf1. La invencion ademas comprende una proteína Cpf1 con optimizacion de codones para la expresion en una célula eucariotica. En una forma de realización preferida, la célula eucariota es una célula de mamífero y en una forma de realización más preferida, la célula de mamífero es una célula de ser humano. En una forma de realización adicional de la invención, la expresión del producto genético está reducida.

En otro aspecto, los métodos y composiciones reivindicados de la invención implican proveer un sistema de vector diseñado, de origen no natural que comprende uno o más vectores que comprenden un primer elemento regulador ligado operativamente a los múltiples ARN guías del sistema Cpf1 CRISPR en donde cada uno está dirigido específicamente a una molécula de ADN que codifica para un producto genético y un segundo elemento regulador ligado operativamente que codifica para una proteína CRISPR. Ambos elementos reguladores pueden estar localizados en el mismo vector o en diferentes vectores del sistema. Los múltiples ARN guías están dirigidos a las múltiple moléculas de ADN que codifica para los múltiples productos genéticos en una célula y la proteína CRISPR puede clivar las múltiples moléculas de ADN que codifican para los productos genéticos (puede clivar una o ambas hebras o puede sustancialmente no tener actividad nucleasa), por lo cual se altera la expresión de los múltiples productos genéticos; y, en donde la proteína CRISPR y los múltiples ARN guías no se presentan juntos naturalmente. En una forma de realización preferida la proteína CRISPR es la proteína Cpf1, opcionalmente con optimización de codones para la expresión en una célula eucariótica. En una forma de realización preferida la célula eucariótica es una célula de mamífero, una célula vegetal o una célula de levadura y en una forma de realización más preferida la célula de mamífero es una célula humana. En otra forma de realización de la invención, la expresión de cada uno de los múltiples productos genéticos está alterada, preferiblemente disminuida.

En un aspecto, los los métodos y composiciones reivindicados de la invención implican proveer un sistema vector que comprende uno o más vectores. En algunas formas de realización, el sistema comprende: (a) un primer elemento regulador ligado operativamente a una secuencia de repetición directa y uno o más sitios de inserción para insertar una o más secuencias guía corriente arriba o abajo (según sea aplicable) de la secuencia de repetición directa, en donde cuando se expresa, la una o más secuencias guía dirigen la unión específica de secuencia del complejo CRISPR a la una o más secuencias diana en una célula eucariótica, en donde el complejo CRISPR comprende una enzima Cpf1 complejada con la una o más secuencias guía que están hibridadas con la una o más secuencias diana; y (b) un segundo elemento regulador ligado operativamente a una secuencia que codifica para una enzima que codifica para la mencionada enzima Cpf1, preferiblemente que comprende por lo menos una secuencia de localizacion nuclear y/o por lo menos una NES; en donde los componentes (a) y (b) están localizados en el mismo o diferentes vectores del sistema. En algunas formas de realización, el componente (a) comprende además dos o más secuencias guía ligadas operativamente con el primer elemento regulador, donde cada una de las dos o más secuencias guía, cuando se expresan, dirigen la unión específica respecto a la secuencia de un complejo Cpf1 CRISPR con una secuencia diana diferente en una célula eucariota. En algunas formas de realización, el complejo CRISPR comprende una o más secuencias de localización nuclear y/o una o más NES de suficiente fuerza para dirigir la acumulación del mencionado complejo Cpf1 CRISPR en una cantidad que puede detectarse dentro o fuera del n[ucleo de una célula eucariótica. En algunas formas de realización, el primer elemento regulador es un promotor de la polimerasa III. En algunas formas de realización, el segundo elemento regulador es un promotor de la polimerasa II. En algunas formas de realización, cada una de las secuencias gu[ia tiene por lo menos 16, 17, 18, 19, 20, 25 nucleótidos, o entre 16 y 30, o entre 16 y 25, o entre 16 y 20 nucleótidos de longitud.

Los vectores de expresión recombinantes pueden comprender los polinucleótidos que codifican para la enzima, sistema o complejo de Cpf1 para usar en el direccionamiento múltiple como se define en la presente en una forma adecuada para la expresión del ácido nucleico en una célula huésped, que significa que los vectores de expresión recombinantes incluyen uno o más elementos reguladors, que pueden seleccionarse en base a las células huéspedes a usar para la expresión, que están ligados operativamente a la secuencia de ácido nucleico a expresar. En un vector de expresión recombinante, “ligado operativamente” se pretende que signifique que la secuencia nucleotídica de interés está ligada al elemento o los elementos reguladores de modo que permita la expresión de la secuencia nucleotídica (por ejemplo, en un sistema de transcripción/traducción in vitro o en una célula huésped cuando se introduce el vector en la célula huésped).

En algunas formas de realización, se transfecta una célula huésped transitoriamente o no transitoriamente con uno o más vectores que comprenden los polinucleótidos que codifican para la enzima, sistema o complejo de Cpf1 para usar en el direccionamiento múltiple como se define en la presente. En algunas formas de realización, se transfecta una célula tal como ocurre de manera natural en un sujeto. En algunas formas de realización, la célula que se transfecta se toma de un sujeto. En algunas formas de realización, la célula se obtiene a partir de células que se toman de un sujeto, tal como una línea celular. En el arte se conoce una amplia variedad de líneas celulares para el cultivo de tejido y se ejemplifican en otra parte en la presente. Estas líneas celulares se pueden adquirir de diferentes proveedores conocidos por los expertos en la técnica (véase, por ejemplo, la Colección Americana de Cultivos Tipo (ATCC, por sus siglas en inglés) (Manassus, Va.)). En algunas formas de realización, se usa una célula transfectada con uno o más vectores que comprenden los polinucleótidos que codifican para la enzima, sistema o complejo de Cpf1 para usar en direccionamiento múltiple como se define en la presente para establecer una nueva línea celular que comprende una o más secuencias derivadas de vector. En algunas formas de realización, se usa una célula transfectada transitoriamente con los componentes de un sistema o complejo Cpf1 CRISPR para usar en el direccionamiento múltiple como se describe en la presente (tal como por transfeccion transitoria de uno o más vectores, o transfección con ARN), y modificada por la actividad de un sistema o complejo Cpf1 CRISPR, para establecer una nueva línea celular que comprende células que contienen la modificacion pero que carecen de cualquier otra secuencia exógena. En algunas formas de realización, las células transfectadas transitoriamente or no transitoriamente con uno o más vectores que comprenden los polinucleótidos que codifican para la enzima, sistema o complejo de Cpf1 para usar en el direccionamiento múltiple como se define en la presente, o líneas celulares derivadas de dichas células se usan

para evaluar uno o más compuestos de ensayo.

El término "elemento regulador" es tal como se define en otra parte en la presente.

Los vectores ventajosos incluyen lentivirus y virus adenoasociados, y pueden seleccionarse tipos de dichos vectores para el direccionamiento particular a tipos de células.

En un aspecto, los métodos y composiciones reivindicados de la invención pueden proveer una célula huésped eucariótica que comprende (a) un primer elemento regulador ligado operativamente a una secuencia de repetición directa y uno o más sitios de inserción para insertar una o más secuencias de ARN guía corriente arriba o abajo (según sea aplicable) de la secuencia de repetición directa, en donde cuando se expresa, la o las secuencias guías dirigen unión específica de secuencia del complejo Cpf1 CRISPR a la o las secuencias diana respectivas en una célula eucariótica, en donde el complejo Cpf1 CRISPR comprende una enzima Cpf1 complejada con la una o más secuencias guía que está hibridada con la o las respectivas secuencias diana; y/o (b) un segundo elemento regulador ligado operativamente a una secuencia que codifica para una enzima que codifica para la mencionada enzima Cpf1 que comprende preferiblemente por lo menos una secuencia de localización nuclear y/o NES. En algunas formas de realización, la célula huésped comprende los componentes (a) y (b). En algunas formas de realización, el componente (a), el componente (b), o los componentes (a) y (b) están integrados de manera estable en el genoma de la célula eucariota huésped. En algunas formas de realización, el componente (a) además comprende dos o más secuencias guía ligada operativamente al primer elemento regulador, y opcionalmente separadas por una repetición directa, en donde cuando se expresan, cada una de las dos o más secuencias guía dirigen la union específica de secuencia de un complejo Cpf1 CRISPR a una secuencia diana diferente en una célula eucariótica. En algunas formas de realización, la enzima Cpf1 comprende una o más secuencias de localización nuclear y/o secuencias de exportación nuclear o NES de suficiente potencia para dirigir la acumulacion de la mencionada enzima CRISPR en una cantidad que puede detectarse dentro y/o fuera del núcleo de una célula eucariótica.

En algunas formas de realización, la enzima Cpf1 es una enzima del sistema CRISPR tipo V. En algunas formas de realización, la enzima Cpf1 es una enzima Cpf1. En algunas formas de realización, la enzima Cpf1 deriva de Francisella tularensis 1, Francisella tularensis subsp. novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, o Porphyromonas macacae Cpf1, y puede incluir otras alteraciones o mutaciones de la Cpf1 como se define en otra parte en la presente, y puede ser una Cpf1 quimérica. En algunas formas de realización, la enzima Cpf1 tiene codones optimizados para la expresión en una célula eucariota. En algunas formas de realización, la enzima CRISPR dirige el clivaje de una o dos hebras en la ubicación de la secuencia diana. En algunas formas de realización, el primer elemento regulador es un promotor de la polimerasa III. En algunas formas de realización, el segundo elemento regulador es un promotor de la polimerasa II. En algunas formas de realización, la una o más secuencias guía tienen (cada una) por lo menos 16, 17, 18, 19, 20, 25 nucleótidos, o entre 16 y 30, o entre 16 y 25, o entre 16 y 20 nucleótidos de longitud. Cuando se usan múltiples ARN guías, preferiblemente están separados por una secuencia de repetición directa. Se describe, pero no se reivindica en la presente, un organismo eucariota no humano; tal como un organismo eucariota multicelular, que comprende una célula eucariota huésped según cualquiera de las formas de realización descritas, un organismo eucariota, tal como un organismo eucariota multicelular, que comprende una célula eucariota huésped como se ha descrito anteriormente. También se describe en la presente un organismo que puede ser un animal, tal como un mamífero. Por tanto, el organismo puede ser un artrópodo tal como un insecto. El organismo también puede ser una planta. Además, el organismo puede ser un hongo.

También se describe en la presente, pero no se reivindica, un conjunto de elementos que comprende uno o más de los componentes descritos en la presente. En algunas formas de realización, el conjunto de elementos comprende un sistema vectorial e instrucciones para utilizar el conjunto de elementos. En algunas formas de realización, el sistema vector comprende (a) un primer elemento regulador ligado operativamente a una secuencia de repetición directa y uno o más sitios de inserción para insertar una o más secuencias guía corriente arriba o abajo (según sea aplicable) de la secuencia de repetición directa, en donde cuando se expresa, la secuencia guía dirige la unión específica de secuencia de un complejo Cpf1 CRISPR a una secuencia diana en una célula eucariótica, en donde el complejo Cpf1 CRISPR comprende una enzima Cpf1 complejada con la secuencia guía que está hibridada con la secuencia diana; y/o (b) un segundo elemento regulador ligado operativamente a una secuencia que codifica para una enzima que codifica para la mencionada enzima Cpf1 que comprende una secuencia de localización nuclear. En algunas formas de realización, el conjunto de elementos comprende los componentes (a) y (b) ubicados en el mismo vector o en vectores diferentes del sistema. En algunas formas de realización, el componente (a) comprende además dos o más secuencias guía ligadas operativamente con el primer elemento regulador, donde cada una de las dos o más secuencias guía, cuando se expresan, dirigen la unión específica respecto a la secuencia de un complejo CRISPR con una secuencia diana diferente en una célula eucariota. En algunas formas de realización, la enzima CRISPR comprende una o más secuencias de localización nuclear con una potencia suficiente para impulsar la acumulación de dicha enzima CRISPR en una cantidad detectable en el núcleo de una célula eucariota. En algunas formas de realización, la enzima CRISPR es una enzima de sistema CRISPR tipo V. En algunas formas de realización, la enzima CRISPR es una enzima Cpf1. En algunas formas de realización, la enzima Cpf1 deriva de Francisella tularensis 1, Francisella tularensis subsp.

novicida, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, o Porphyromonas macacae Cpf1 (por ejemplo, modificada para tener o que esté asociada con por lo menos un DD), y puede incluir una alteración o mutación adicional de la Cpf1, y puede ser una Cpf1 quimérica. En algunas formas de realización, la enzima DD-CRISPR tiene codones optimizados para la expresión en una célula eucariota. En algunas formas de realización, la enzima DD-CRISPR dirige el clivaje de una o dos hebras en la ubicación de la secuencia diana. En algunas formas de realización, la enzima DD-CRISPR carece o sustancialmente carece de actividad de clibaje de hebra de ADN (por ejemplo, no más de 5% de actividad nucleasa en comparacion a una enzima de tipo salvaje o enzima que no tiene la mutación o alteración que disminuye la actividad nucleasa). En algunas formas de realización, el primer elemento regulador es un promotor de la polimerasa III. En algunas formas de realización, el segundo elemento regulador es un promotor de la polimerasa II. En algunas formas de realización, la secuencia guía is por lo menos 16, 17, 18, 19, 20, 25 nucleótidos, o entre 16 y 30, o entre 16 y 25, o entre 16 y 20 nucleótidos de longitud.

En un aspecto, los métodos reivindicados pueden implicar modificar los múltiples polinucleótidos diana en una célula huésped tal como una célula eucariótica. En algunas formas de realización, el método comprende permitir que un complejo Cpf1CRISPR se usan múltiples polinucleótidos diana, por ejemplo, para producir el clivaje de los mencionados múltiples polinucleótidos diana, modificando de esta manera los múltiples polinucleótidos diana, en donde el complejo Cpf1CRISPR comprende una enzima Cpf1 complejada con múltiples secuencias guía cada una de ellas hibridada a una secuencia diana específica en el mencionado polinucleótido diana, en donde las mencionadas múltiples secuencias guía están ligadas a una secuencia de repetición directa. En algunas formas de realización, el mencionado clivaje comprende clivar una o dos hebras en la posición de cada una de las secuencia diana por la mencionada enzima Cpf1. En algunas formas de realización, el mencionado clivaje genera una transcripción disminuida de los múltiples genes diana. En algunas formas de realización, el método además comprende reparar uno o más de los mencionados polinucleótidos diana clivados por recombinación homóloga con un polinucleótido molde exógeno, en donde la mencionada reparacion produce una mutación que comprende una inserción, eliminación o sustitución de uno o más nucleótidos de uno o más de los mencionados polinucleótidos diana. En algunas formas de realización, la mencionada mutación genera uno o más cambios de aminoácidos en una proteína expresada a partir de un gen que comprende una o más de las secuencias diana. En algunas formas de realización, el método además comprende proveer uno o más vectores a la mencionada célula eucariótica, en donde el uno o más vectores dirigen la expresión de uno o más de: la enzima Cpf1 y las múltiples secuencias de ARN guía ligadas a una secuencia de repetición directa. En algunas formas de realización, dichos vectores se suministran a la célula eucariota en un sujeto. En algunas formas de realización, dicha modificación tiene lugar en dicha célula eucariota en un cultivo celular. En algunas formas de realización, el método comprende además aislar dicha célula eucariota de un sujeto antes de dicha modificación. En algunas formas de realización, el método comprende además devolver dicha célula eucariota y/o células obtenidas a partir de ella a dicho sujeto.

En un aspecto, los métodos reivindicados pueden implicar un método para modificar la expresión de múltiples polinucleótidos en una célula eucariótica. En algunas formas de realización, el método comprende permitir que un complejo Cpf1 CRISPR se una a múltiples polinucleótidos tal que la mencionada unión produzca una expresión aumentada o disminuida de los mencionados polinucleótidos; en donde el complejo Cpf1 CRISPR comprende una enzima Cpf1 complejada con múltiples secuencias guía cada una hibridada específicamente a su propia secuencia diana en el mencionado polinucleótido, en donde las mencionadas secuencias guía están unidas a una secuencia de repetición directa. En algunas formas de realización, el método además comprende administrar uno o más vectores a las mencionadas células eucarióticas, en donde el uno o más vectores dirige la expresión de uno o más de: la enzima Cpf1 y las múltiples secuencias guía unidas a las secuencias de repetición directa.

En un aspecto, los métodos y composiciones reivindicados de la invención pueden implicar proveer un polinucleótido recombinante que comprende múltiples secuencias ARN guía corriente arriba o abajo (según sea aplicable) de una secuencia de repetición directa, en donde cada una de las secuencias guía cuando se expresa dirige la unión específica de secuencia de un complejo Cpf1CRISPR a su correspondiente secuencia diana presente en una célula eucariótica. En algunas formas de realización, una secuencia diana es una secuencia vírica presente en una célula eucariota. En algunas formas de realización, una secuencia diana es un protooncogén o un oncogén.

Los aspectos de los métodos y composiciones reivindicados de la invención abarcan proporcionar una composición de origen no natural o diseñada que puede comprender un ARN guía (ARNg) que comprende una secuencia guía con capacidad de hibridar con una secuencia diana en un locus genómico de interés en una célula y una enzima Cpf1 como se define en la presente que puede comprender por lo menos una o más secuencias de localización nuclear.

Un aspecto de los métodos reivindicados de la invención modificar un locus genómico de interés para cambiar la expresión genética en una célula por introducción en la célula de cualquiera de las composiciones descritas en la presente.

Un aspecto de la invención es que los elementos precedentes están comprendidos en una composición individual o comprendidos en composiciones individuales. Estas composiciones pueden aplicarse ventajosamente a un huésped

para que genere un efecto funcional a nivel genómico.

Como se usa en la presente, el término "ARN guía" o "ARNg" tiene el significado que se usa en otra parte en la presente y comprende cualquier secuencia polinucleotídica que tiene suficiente complementariedad con una secuencia de ácido nucleico diana para hibridar con la secuencia diana de ácido nucleico y dirigir la unión específica de secuencia de un complejo de direccionamiento a ácido nucleico a la secuencia de ácido nucleico diana. Cada ARNg puede diseñarse para que incluya múltiples sitios de reconocimiento de unión (por ejemplo, aptámeros) específicos para la misma proteína adaptadora o diferente. Cada ARNg puede diseñarse para que se una a la región promotora -1000 - +1 ácidos nucleicos corriente arriba del sitio de inicio de la transcripción (es decir TSS), preferiblemente -200 ácidos nucleicos. Este posicionamiento mejora los dominios funcionales que afecta a la activación del gen (por ejemplo, activadores de la transcripción) o inhibición del gen (por ejemplo, represores de la transcripción). El ARNg modificado puede ser uno o más de los ARNgs modificados dirigidos a uno o más de los loci diana (por ejemplo, por lo menos 1 ARNg, por lo menos 2 ARNg, por lo menos 5 ARNg, por lo menos 10 ARNg, por lo menos 20 ARNg, por lo menos 30 g RNA, por lo menos 50 ARNg) comprendidos en una composición. Las mencionadas múltiples secuencias de ARNg pueden disponerse en tándem y preferiblemente están separadas por una repetición directa.

Por lo tanto, ARNg, la enzima CRISPR como se define en la presente puede estar cada uno comprendido individualmente en una composición y administrarse a un huésped individual o colectivamente. Como alternativa, estos componentes pueden proveerse en una composición individual para la administración a un huésped. La administración a un huésped puede realizarse mediante vectores virales conocidos para la persona con experiencia o se describe en la presente para la administración a un huésped (por ejemplo, vector lentiviral, vector adenoviral, vector AAV). Como se explica en la presente, el uso de diferentes marcadores de selección (por ejemplo, para selección de ARNg lentiviral) y concentración de ARNg (por ejemplo, dependiendo de si se usan múltiple ARNgs) puede ser ventajoso para producir un efecto mejorado. En base a este concepto, son apropiadas diferentes variaciones para producir un evento en un locus genómico, que incluye clivaje del ADN, activación de un gen, o desactivación de un gen. Con el uso de las composiciones provistas, la persona con experiencia en el arte puede dirigir ventajoso y específicamente un locus simple o múltiple con el mismo o diferentes dominios funcionales para producir uno o más eventos de locus genómico. Las composiciones pueden aplicarse en una variedad amplia de métodos para la selección en bibliotecas en células y para el modelado funcional in vivo (por ejemplo, activación de un gen de lincARN e identificación de función; modelado de ganancia de función; modelado de pérdida de función; el uso de las composiciones de la invención para establecer líneas celulares y animales transgénicos no humanos para propósitos de optimización y selección).

También se describe, pero no se reivindica en la presente, el uso de las composiciones de la presente invención para establecer y utilizar células/animales transgénicos condiciones o inducibles para CRISPR; véase, por ejemplo, Platt y col., Cell (2014), 159(2): 440-455, o las publicaciones de patente PCT citadas en la presente, tal como WO 2014/093622 (PCT/US2013/074667). Por ejemplo, células o animales tales como animales no humanos, por ejemplo, vertebrados o mamíferos, tales como roedores, por ejemplo, ratones, ratas, u otros animales de laboratorio o de campo, por ejemplo, gatos, perros, ovejas, etc., pueden ser "knock-in" en donde el animal expresa Cpf1 condicionalmente o de manera inducible similar a Platt y col. La célula o animal diana comprende por lo tanto la enzima CRISRP (por ejemplo, Cpf1) condicionalmente o de manera inducible (por ejemplo, en la forma de construcciones dependientes de Cre), de la expresión de un vector introducido en la célula diana, en donde el vector expresa induciendo o generando la condición de la expresión de la enzima CRISRP (por ejemplo, Cpf1) en la célula diana. Por aplicación de la descripción y composiciones como se definen en la presente con el método conocido para crear un complejo CRISPR, los eventos genómicos inducibles también son un aspecto de la presente invención. Los ejemplos de dichos eventos inducibles han sido descritos en otra parte en la presente.

En algunas formas de realización, la alteración fenotípica preferiblemente es el resultado de una modificación en el genoma cuando está direccionado a una enfermedad genética, especialmente en métodos de terapia y preferiblemente cuando se provee un molde de reparación para corregir o alterar el fenotipo.

En algunas formas de realización las enfermedades a las que puede estar dirigido incluyen aquellas relacionadas con defectos en corte y empalme que provocan enfermedad.

En algunas formas de realización, los blancos celulares incluyen células madre/progenitoras hematopoyéticas (CD34+); células T humanas; y células del ojo (células de retina) – por ejemplo precursoras de fotorreceptores.

En algunas formas de realización los blancos genéticos incluyen: globina beta humana – HBB (para tratar la anemia de células falciformes, que incluye la estimulación de la conversión del gen (usando un gen de HBD cercanamente relacionado como un molde endógeno)); CD3 (células T); y CEP920 - retina (ojo).

En algunas formas de realización las enfermedades diana también incluyen: cáncer; anemia de células falciformes (en base en una mutación puntual); HBV, HIV; Beta-Talasemia; y enfermedad oftálmica u ocular – por ejemplo defecto de corte y empalme que provoca la amaurosis congénita de Leber (LCA).

En algunas formas de realización los métodos de administración incluyen: administración "directa" mediada por lípidos catiónicos del complejo enzima-guía (RiboNucleoProtein) y electroporación de ADN plasmídico.

Los métodos, productos y usos descritos en la presente se pueden usar para fines no terapéuticos. Además, se puede

aplicar cualquiera de los métodos descritos en la presente in vitro y ex vivo.

En un aspecto, se provee una composición de origen no natural o diseñada que comprende:

I. dos o más secuencias polinucleotídicas del sistema CRISPR-Cas que comprenden

(a) una primera secuencia guía con capacidad de hibridar con una primera secuencia diana en un locus polinucleotídico,

(b) una segunda secuencia guía con capacidad de hibridar con una segunda secuencia diana en un locus polinucleotídico,

(c) una secuencia de repetición directa,

y

II. una enzima Cpf1 o una segunda secuencia polinucleotídica que codifica para la misma,

en donde cuando se transcribe, la primera y la segunda secuencias guías dirigen la unión específica de la secuencia de un primer y segundo complejo Cpf1 CRISPR a la primera y segunda secuencias diana respectivamente, en donde el primer complejo CRISPR comprende la enzima Cpf1 complejada con la primera secuencia guía que puede hibridar con la primera secuencia diana, en donde el segundo complejo CRISPR comprende la enzima Cpf1 complejada con la segunda secuencia guía que puede hibridar con la segunda secuencia diana, y en donde la primera secuencia guía dirige el clivaje de una hebra del dúplex de ADN cerca de la primera secuencia diana y la segunda secuencia guía dirige el clivaje de la otra hebra cerca de la segunda secuencia diana induciendo una interrupción en la hebra doble, modificando de esta manera el organismo no humano o no animal. Similarmente, las composiciones que comprenden más de dos ARN guías pueden ser pensadas por ejemplo cada uno específico por un diana, y dispuestos en tándem en la composición o sistema o complejo CRISPR como se describe en la presente.

En otra forma de realización, la Cpf1 se administra a la célula como una proteína. En otra forma de realización y particularmente preferida, la Cpf1 se administra a la célula como una proteína o como una secuencia nucleotídica que codifica para la misma. La administración a la célula como una proteína puede incluir la administración de un complejo de ribonucleoproteína (RNP), en donde la proteína está complejada con las múltiples guías.

También se describe, pero no se reivindica específicamente en la presente, células huésped y líneas celulares modificadas por lo que comprenden las composiciones, sistemas o enzimas modificadas de la presente invención, que incluye células madre, y progenie de las mismas.

También se describe, pero no se reivindica en la presente, métodos de terapia celular, en donde, por ejemplo, se muestrea una célula individual o una población de células se muestrea o cultiva, en donde esa célula o células está o ha sido modificada ex vivo como se describe en la presente, y luego se reintroduce (células muestreadas) o se introduce (células cultivadas) en el organismo. También se describen particularmente a este respecto las células madre, tanto células madre no humanas embrionarias como células madre no humanas pluripotentes o totipotentes inducidas. Pero, por supuesto, también se describen métodos *in vivo*, aunque no se reivindican en la presente.

Los métodos según las reivindicaciones además pueden comprender la administración de moldes, tal como moldes de reparación, que pueden ser ODNhd u ODNhs, véase más adelante. La administración de moldes puede ser contemporánea o separada de la administración de cualquiera o todos de enzima CRISPR o ARN guías y mediante el mismo o diferente mecanismo de administración. En algunas formas de realización, se prefiere que el molde se administre junto con los ARN guías y, preferiblemente, también la enzima CRISPR. Un ejemplo puede ser un vector AAV en el cual la enzima CRISPR es AsCpf1 o LbCpf1.

Los métodos según las reivindicaciones además pueden comprender: (a) administración a la célula de un oligodesoxinucleótido de hebra doble (ODNhd) que comprende sobreextensiones complementarias con las sobreextensiones creadas por la mencionada interrupción de hebra doble, en donde el mencionado ODNhd se integra en el locus de interés; o –(b) administración a la célula de un oligodesoxinucleótido de hebra simple (ssODN), en donde el mencionado ODNhs actúa como un molde para la reparación dirigida por homología de la interrupción de hebra doble. Se describen, pero no se reivindican en la presente, métodos para la prevención o tratamiento de enfermedades en un individuo, opcionalmente en donde la mencionada enfermedad es provocada por un defecto en el mencionado locus de interés. Los métodos de la invención se realizan *ex vivo* en una célula tomada del individuo, opcionalmente en donde la mencionada célula es regresada al individuo.

La solicitud también describe productos obtenidos del uso de la enzima CRISPR o enzima Cas o enzima Cpf1 o enzima CRISPR-CRISPR o sistema CRISPR-Cas o sistema CRISPR-Cpf1 para usar en el direccionamiento en tándem o múltiple como se define en la presente.

Conjuntos de elementos

La solicitud también describe, pero no reivindica específicamente, conjuntos de elementos que contienen uno cualquiera o más de los elementos divulgados en los métodos y composiciones anteriores. En algunas formas de realización, el conjunto de elementos comprende un sistema vector como se describe en la presente e instrucciones para usar el conjunto de elementos. Se podrán proveer los elementos individualmente o combinados y se podrán proveer en cualquier envase adecuado tal como un vial, botella o tubo. Los conjuntos de elementos puede incluir el ARNg y la hebra protectora no unida como se describe en la presente. Los conjuntos de elementos pueden incluir el ARNg con la hebra protectora unida por lo menos parcialmente a la secuencia guía (es decir ARNgp). Por lo tanto los conjuntos de elementos pueden incluir el ARNgp en la forma de una secuencia nucleotídica parcialmente en hebra dobe como se describe aquí. En algunas formas de realización, el conjunto de elementos incluye instrucciones en uno o más idiomas, por ejemplo, en más de un idioma. Las instrucciones pueden ser específicas para las aplicaciones y métodos descritos en la presente.

En algunas formas de realización, un conjunto de elementos comprende uno o más reactivos para su uso en un proceso que utilice uno o más de los elementos descritos en la presente. Se podrán proveer los reactivos en cualquier envase adecuado. Por ejemplo, un conjunto de elementos podrá proveer uno o más tampones de reacción o de almacenamiento. Se podrán proveer los reactivos en una forma que se pueda utilizar en un ensayo particular, o en una forma que requiera la adición de uno o más componentes diferentes antes de su uso (por ejemplo, en forma concentrada o liofilizada). Una solución amortiguadora puede ser cualquier solución amortiguadora incluida, a título enunciativo no taxativo, una solución amortiguadora de carbonato de sodio, una solución amortiguadora de bicarbonato de sodio, una solución amortiguadora de bicarbonato de sodio, una solución amortiguadora de borato, una solución amortiguadora Tris, una solución amortiguadora MOPS, una solución amortiguadora HEPES y combinaciones de estas. En algunas formas de realización, la solución amortiguadora es alcalina. En algunas formas de realización, la solución amortiguadora tiene un pH entre aproximadamente 7 y aproximadamente 10. En algunas formas de realización, el conjunto de elementos comprende uno o más oligonucleótidos correspondientes a la secuencia guía para la insercion en un vector tal que se una operativamente a la secuencia guía y a un elemento regulador. En algunas formas de realización, el conjunto de elementos comprende un polinucleótido molde para la recombinación homóloga. En algunas formas de realización, el conjunto de elementos comprende uno o más de los vectores y/o uno o más de los polinucleótidos descritos en la presente. El conjunto de elementos podrá permitir convenientemente que se provean todos los elementos de los sistemas descritos en la presente.

En un aspecto, los métodos reivindicados de la invención pueden implicar métodos para utilizar uno o más elementos de un sistema CRISPR. El complejo CRISPR de la invención provee un medio eficaz para modificar un polinucleótido diana. El complejo CRISPR tiene una amplia variedad de utilidades que incluyen modificar (por ejemplo, eliminar, insertar, translocar, inactivar, activar) un polinucleótido diana en una multiplicidad de tipos celulares. En este sentido, el complejo CRISPR tiene un amplio espectro de aplicaciones en, por ejemplo, terapia génica, selección de fármacos, diagnóstico y pronóstico de enfermedades. Un complejo CRISPR ejemplificativo comprende una proteína efectora CRISPR complejada con una secuencia guía hibridada con una secuencia diana con el polinucleótido diana. En determinadas formas de realización, una secuencia de repetición directa está ligada a la secuencia guía.

En una forma de realización, este método reivindicado implica clivar un polinucleótido diana. El método comprende modificar un polinucleótido diana utilizando un complejo CRISPR que se una al polinucleótido diana y efectúe el clivaje de dicho polinucleótido diana. Normalmente, el complejo CRISPR de la invención, cuando se introduce en una célula, crea una rotura (por ejemplo, una rotura mono- o bicatenaria) en la secuencia genómica. Por ejemplo, se puede utilizar el método para clivar un gen ligado a una enfermedad en una célula.

La rotura creada por el complejo CRISPR se puede reparar mediante procesos de reparación tal como la ruta de unión de extremos no homólogos propensa a error (NHEJ) o la reparación dirigida por homología de alta fidelidad (HDR). Durante este proceso de reparación, se puede introducir un molde de polinucleótidos exógeno en la secuencia genómica. En algunos métodos, el proceso de HDR se usa para modificar la secuencia del genoma. Por ejemplo, se introduce en una célula un molde de polinucleótidos exógeno que comprende una secuencia que se va a integrar flanqueada por una secuencia en dirección 5' y una secuencia en dirección 3'. Las secuencias en dirección 5' y en dirección 3' comparten similitud secuencial con cualquier lado de un sitio de integración del cromosoma.

Cuando se desee, un polinucleótido donante puede ser ADN, por ejemplo, un plásmido de ADN, un cromosoma artificial bacteriano (BAC, por sus siglas en inglés), un cromosoma artificial de levaduras (YAC, por sus siglas en inglés), un vector vírico, un fragmento lineal de ADN, un fragmento de PCR, un ácido nucleico puro o un ácido nucleico complejado con un vehículo de suministro tal como un liposoma o un poloxámero.

El molde de polinucleótidos exógeno comprende una secuencia que se va a integrar (por ejemplo, un gen mutado). La secuencia para la integración podrá ser una secuencia endógena o exógena para la célula. Los ejemplos de una secuencia que se va a integrar incluyen polinucleótidos que codifican una proteína o ARN no codificante (por ejemplo, un microARN). Por lo tanto, la secuencia para la integración podrá estar ligada operativamente a una secuencia o secuencias de control apropiadas. Como alternativa, la secuencia que se va a integrar podrá proveer una función reguladora.

Las secuencias en dirección 5' y en dirección 3' en el molde de polinucleótidos exógeno se seleccionan para promover la recombinación entre la secuencia cromosómica de interés y el polinucleótido donante. La secuencia en dirección 5'

es una secuencia de ácido nucleico que comparte similitud secuencial con la secuencia genómica en dirección 5' respecto al sitio diana para la integración. De manera similar, la secuencia en dirección 3' es una secuencia de ácido nucleico que comparte similitud secuencial con la secuencia cromosómica en dirección 3' respecto al sitio diana de la integración. Las secuencias en dirección 5' y en dirección 3' en el molde de polinucleótidos exógeno puede tener un 75%, 80%, 85%, 90%, 95% o un 100% de identidad secuencial con la secuencia genómica diana. Preferiblemente, las secuencias en dirección 5' y en dirección 3' en el molde de polinucleótidos exógeno tiene aproximadamente un 95%, 96%, 97%, 98%, 99% o un 100% de identidad secuencial con la secuencia genómica diana. En algunos métodos, las secuencias en dirección 5' y en dirección 3' en el molde de polinucleótidos exógeno tienen aproximadamente un 99% o un 100% de identidad secuencial con la secuencia genómica diana.

Una secuencia en dirección 5' o en dirección 3' podrá comprender de aproximadamente 20 pb a aproximadamente 2500 pb, por ejemplo, aproximadamente 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400 o 2500 pb. En algunos métodos, la secuencia en dirección 5' o 3' ilustrativa tiene de aproximadamente 200 pb a aproximadamente 2000 pb, de aproximadamente 600 pb a aproximadamente 1000 pb o, más particularmente, de aproximadamente 700 pb a aproximadamente 1000 pb.

En algunos métodos, el molde de polinucleótidos exógeno podrá comprender además un marcador. Un marcador de este tipo podrá facilitar la selección para detectar las integraciones objetivo. Los ejemplos de marcadores adecuados incluyen sitios de restricción, proteínas fluorescentes o marcadores seleccionables. Se puede construir el molde de polinucleótidos exógeno previsto en formas de realización particulares de la invención utilizando técnicas recombinantes (véase, por ejemplo, Sambrook y col., 2001 y Ausubel y col., 1996).

En un método ilustrativo para modificar un polinucleótido diana integrando un molde de polinucleótidos exógeno, se introduce una rotura bicatenaria en la secuencia genómica mediante el complejo CRISPR, se repara la rotura mediante recombinación homóloga de un molde de polinucleótidos exógeno de modo que se integre el molde en el genoma. La presencia de una rotura bicatenaria facilita la integración del molde.

En otras formas de realización, los métodos reivindicados de la invención implican modificar la expresión de un polinucleótido en una célula eucariota. El método comprende incrementar o disminuir la expresión de un polinucleótido diana utilizando un complejo CRISPR que se une al polinucleótido.

En algunos métodos, se puede inactivar un polinucleótido diana para efectuar la modificación de la expresión en una célula. Por ejemplo, tras la unión de un complejo CRISPR a la secuencia diana en una célula, el polinucleótido diana se inactiva de modo que la secuencia no se transcriba, la proteína codificada no se produce o la secuencia no funciona como lo hace la secuencia no modificada. Por ejemplo, una proteína o una secuencia codificante de microARN se podrán inactivar de modo que no se produzca la proteína.

En algunos métodos, se puede inactivar una secuencia control de modo que ya no actúe como una secuencia de control. Tal y como se utiliza en la presente una "secuencia de control" se refiere a cualquier secuencia de ácido nucleico que efectúa la transcripción, traducción o accesibilidad de una secuencia de ácido nucleico. Los ejemplos de una secuencia de control incluyen un promotor, un terminador de la transcripción y un potenciador son secuencias de control. La secuencia diana inactivada podrá incluir una mutación de tipo eliminación (es decir, la eliminación de uno o más nucleótidos), una mutación de tipo inserción (es decir, la inserción de uno o más nucleótidos) o una mutación sin sentido (es decir, una sustitución de un único nucleótido por otro nucleótido de modo que se introduzca un codón de parada). En algunos métodos, la inactivación de una secuencia diana produce un "noqueo" de la secuencia diana.

Métodos ejemplificativos de uso del sistema CRISPR Cas

La solicitud provee una composición de origen natural o diseñada, o uno o más polinucleótidos que codifican para componentes de la mencionada composición, o vector o sistemas de administración que comprenden uno o más polinucleótidos que codifican para componentes de la mencionada composición para usar en una modificación de una célula diana *in vivo*, *ex vivo* o *in vitro* y, puede realizarse de una manera que altera la célula tal que una vez modificada la progenie o la línea celular de la célula modificada con CRISPR retiene el fenotipo alterado. En un aspecto que no es parte de la invención, las células modificadas y la progenie pueden ser parte de un organismo multicelular no humano tal como una planta o animal con aplicación *ex vivo* o *in vivo* del sistema CRISPR a los tipos de células deseados. Se describe en el presente documento, pero no se reivindica, un método terapéutico de tratamiento. El método terapéutico de tratamiento puede comprender la edición de un gen o genoma, o terapia génica.

Uso de enzima Cpf1 CRISPR inactivado para métodos de detección tal como FISH

En un aspecto, los métodos y composiciones reivindicados pueden implicar el uso de un sistema CRISPR-Cas diseñado, de origen no natural que comprende una proteína Cas catalíticamente inactivada descrita en la presente, preferiblemente una Cpf1 inactivada (dCpf1), y el uso de este sistema en métodos de detección tales como hibridación in situ por fluorescencia (FISH). La dCpf1 que carece de capacidad para producir interrupciones de doble hebra en el ADN puede fusionarse con un marcador, tal como una proteína fluorescente, tal como la proteína fluorescente verde potenciada (eEGFP) y coexpresarse con pequeños ARN guías para dirigir repeticiones pericéntricas, céntricas y teloméricas in vivo. El sistema dCpf1 puede usarse para visualizar las secuencias repetitivas y genes individuales en el genoma humano. Dichas aplicaciones nuevas de los sistemas CRISPR-cas con dCpf1 marcada pueden ser

importantes para obtener imágenes de células y estudiar la arquitectura nuclear funcional, especialmente en casos con un volumen de núcleo pequeño o estructuras 3-D complejas. (Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155(7):1479-91. doi: 10.1016/j.cell.2013.12.001.)

Modificación de un diana con el sistema o complejo CRISPR Cas (por ejemplo, complejo Cpf1-ARN)

En un aspecto, los métodos reivindicados pueden implicar modificar un polinucleótido diana en una célula eucariota que podrán ser in vivo, ex vivo o in vitro. En algunas realizaciones, el método comprende obtener muestras de una célula o población de células de un ser humano o animal no humano y modificar la célula o las células. El cultivo podrá tener lugar en cualquier etapa ex vivo. Se describe, pero no se reivindica, que la célula o las células incluso podrán reintroducirse al animal no humano o la planta. En el caso de las células reintroducidas, se prefiere particularmente que las células sean células madre.

En algunas realizaciones, el método reivindicado comprende permitir que un complejo CRISPR se una al polinucleótido diana para efectuar la escisión de dicho polinucleótido diana y modificar de este modo el polinucleótido diana, donde el complejo CRISPR comprende una enzima CRISPR complejada con una secuencia guía hibridada o que puede hibridarse con una secuencia diana en dicho polinucleótido diana.

En un aspecto, el método reivindicado de la invención comprende modificar la expresión de un polinucleótido en una célula eucariota. En algunas realizaciones, el método comprende permitir que un complejo CRISPR se una al polinucleótido de modo que dicha unión dé como resultado un aumento o un descenso de la expresión de dicho polinucleótido; en donde el complejo CRISPR comprende una enzima CRISPR complejada con una secuencia guía hibridada o que puede hibridarse con una secuencia diana en dicho polinucleótido. Para los métodos de modificación de un polinucleótido diana son válidas consideraciones y condiciones similares a las mencionadas anteriormente. De hecho, estas opciones de obtención de muestras, cultivo y reintroducción son válidas en todos los aspectos de la presente invención.

De hecho, en cualquier aspecto de la invención, el complejo CRISPR puede comprender una enzima CRISPR complejada con una secuencia guía hibridada o que puede hibridar con una secuencia diana. Para los métodos de modificación de un polinucleótido diana son válidas consideraciones y condiciones similares a las mencionadas anteriormente.

Por lo tanto, cualquiera de las enzimas CRISPR de origen no natural descritas en la presente comprende por lo menos una modificación y por ende la enzima tiene determinadas capacidades mejoradas. En particular, cualquiera de las enzimas tiene la capacidad de formar un complejo CRISPR con un ARN guía. Cuando se forma dicho complejo, el ARN guía tiene la capacidad de unirse a una secuencia polinucleotídica diana y la enzima tiene la capacidad de modificar un *locus* diana. Adicionalmente, la enzima en el complejo CRISPR ha reducido la capacidad de modificar uno o más *loci* fuera de diana en comparación a una enzima no modificada.

Adicionalmente, las enzimas CRISPR modificadas descritas en la presente abarcan enzimas por las cuales en el complejo CRISPR la enzima tiene capacidad aumentada para modificar el uno o más *loci* diana en comparación a una enzima no modificada. Dicha función puede ser provista por separado o provista en combinación con la función mencionada precedentemente de capacidad reducida para modificar uno o más *loci* fuera de diana. Cualquiera de las enzimas mencionadas puede proveerse con cualquiera de las otras modificaciones a la enzima CRISPR como se describe en la presente, tal como en combinación con cualquier actividad provista por uno o más dominios funcionales heterólogos asociados, cualquier otra mutación para reducir la actividad nucleasa y similares.

En formas de realización ventajosas de los métodos y composiciones reivindicados de la invención, la enzima CRISPR modificada se provee con capacidad reducida para modificar uno o más *loci fuera de diana* en comparación con una enzima no modificada y capacidad aumentada para modificar el uno o más *loci* diana en comparación con una enzima no modificada. En combinación con otras modificaciones de la enzima, puede obtenerse una especificidad significativamente aumentada. Por ejemplo, se provee una combinación de dichas formas de realización ventajosas con una o más mutaciones adicionales en donde la una o más mutaciones adicionales están en uno o más dominios catalíticamente activos. Dichas mutaciones catalíticas adicionales pueden conferir funcionalidad de nickasa como se describe en detalle en otra parte en la presente. En dichas enzimas, la especificidad aumentada puede lograrse debido a una especificidad mejorada en términos de la actividad de la enzima.

Pueden hacerse modificaciones para reducir los efectos fuera de diana y/o potenciar los efectos en diana como se describió precedentemente en residuos de aminoácidos localizados en una región/fosa cargada positivamente situada entre los dominios RuvC-III y HNH. Se apreciará que cualquiera de los efectos funcionales descritos precedentemente puede obtenerse por modificación de aminoácidos en la fosa mencionada precedentemente pero también por modificación de aminoácidos adyacentes a o por fuera de esa fosa.

Las funcionalidades adicionales que pueden diseñarse en enzimas CRISPR modificadas como se describe en la presente incluyen las siguientes. 1. enzimas CRISPR modificadas que alteran las interacciones ADN:proteína sin afectar la estructura terciaria o secundaria. Esto incluye residuos que están en contacto con cualquier parte del dúplex ARN:ADN. 2. enzimas CRISPR modificadas que debilitan las interacciones intraproteína que tienen Cpf1 en

conformación esencial para el corte con nucleasa en respuesta a la unión al ADN (en diana o fuera de diana). Por ejemplo: una modificación que inhibe levemente, pero aún permite, la conformación de nucleasa del dominio HNH (posicionado en el fosfato escindible). 3. enzimas CRISPR modificadas que fortalecen las interacciones intraproteína que tienen Cpf1 en una conformación que inhibe la actividad nucleasa en respuesta a la unión a ADN (en diana o fuera de los blancos). Por ejemplo: una modificación que estabiliza el dominio HNH en una conformación fuera del fosfato escindible. Cualquier mejora funcional adicional puede ser provista en combinación con cualquier otra modificación de la enzima CRISPR como se describe en detalle en otro lugar en la presente.

Puede hacerse cualquiera de las funcionalidades mejoradas descritas en la presente a cualquier enzima CRISPR, tal como una enzima Cpf1. Sin embargo, se apreciará que puede diseñarse cualquiera de las funcionalidades descritas en la presente en enzimas Cpf1 a partir de ortólogos, que incluye enzimas quiméricas que comprenden fragmentos de ortólogos múltiples.

Ácidos nucleicos, aminoácidos y proteínas, secuencias regulatorias, vectores, etc.

Los métodos y composiciones reivindicados de la invención utilizan ácidos nucleicos para unirse a secuencias de ADN diana. Esto es conveniente ya que los ácidos nucleicos son mucho más fáciles y más baratos de producir que las proteínas y se puede variar la especificidad de acuerdo con la longitud del fragmento donde se desea la homología. Por ejemplo, no se requiere el posicionamiento 3-D complejo de múltiples dedos. Los términos "polinucleótido", "nucleótido", "secuencia nucleotídica", "ácido nucleico" y "oligonucleótido" se utilizan como sinónimos. Se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sean desoxirribonucleótidos o ribonucleótidos o análogos de estos. Los polinucleótidos podrán tener cualquier estructura tridimensional y podrán realizar cualquier función, conocida o desconocida. Los siguientes son ejemplos no limitantes de polinucleótidos: regiones codificantes o no codificantes de un gen o fragmento génico, loci (locus) definidos a partir del análisis de la unión, exones, intrones, ARN mensajero (ARNm), ARN de transferencia, ARN ribosómico, ARN interferente pequeño (ARNip), ARN de horquilla pequeña (ARNhp), micro-ARN (miARN), ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácido nucleico y cebadores. El término también abarca estructuras de tipo ácido nucleico con esqueletos sintéticos, véase, por ejemplo, Eckstein, 1991; Baserga y col., 1992; Milligan, 1993; WO 97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; y Samstag, 1996. Un polinucleótido puede comprender uno o más nucleótidos modificados tales como nucleótidos metilados y análogos de nucleótidos. Si están presentes, las modificaciones en la estructura nucleotídica se podrán practicar antes o después del ensamblaje del polímero. Se puede interrumpir la secuencia nucleotídica con componentes no nucleotídicos. Se puede modificar un polinucleótido adicionalmente después de la polimerización, por ejemplo, conjugándolo con un componente marcador. Tal y como se utiliza en la presente, el término "de tipo no modificado" es un término de la técnica cuyo significado conocen los expertos y se refiere a la forma típica de un organismo, cepa, gen o característica tal y como ocurre en la naturaleza y que se distingue de formas mutantes o variantes. Un "tipo salvaje" puede ser un valor inicial. Tal y como se utiliza en la presente, se debería sobreentender que el término "variante" se refiere a la exhibición de cualidades que tienen un patrón que se desvía del que ocurre en la naturaleza. Los términos "de origen no natural" y "diseñado" se utilizan como sinónimos e indican la participación del hombre. Los términos, cuando se refieren a moléculas de ácido nucleico o polipéptidos se refieren a que la molécula de ácido nucleico o el polipéptido están por lo menos sustancialmente libres de por lo menos otro componente con el que se asocian de manera natural en la naturaleza y tal y como se observan en la naturaleza. El término "complementariedad" se refiere a la capacidad de un ácido nucleico para formar uno o más puentes de hidrógeno con otra secuencia de ácido nucleico ya sea mediante el emparejamiento tradicional de bases de Watson-Crick o mediante otros tipos no tradicionales. Un porcentaje de complementariedad indica el porcentaje de residuos en una molécula de ácido nucleico que pueden formar puentes de hidrógeno (por ejemplo, emparejamiento de bases de Watson-Crick) con una segunda secuencia de ácido nucleico (por ejemplo, 5, 6, 7, 8, 9, 10 de 10 representan una complementariedad de un 50%, 60%, 70%, 80%, 90% y 100%). "Perfectamente complementario" significa que todos los residuos contiguos de una secuencia de ácido nucleico se unirán mediante puentes de hidrógeno con el mismo número de residuos contiguos en una segunda secuencia de ácido nucleico. "Sustancialmente complementario" tal y como se utiliza en la presente se refiere a un grado de complementariedad que es por lo menos de un 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% o 100% a lo largo de una región de 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 o más nucleótidos o se refiere a dos ácidos nucleicos que se hibridan entre sí en condiciones rigurosas. Tal y como se utiliza en la presente, las "condiciones rigurosas" para la hibridación se refieren a condiciones en las que un ácido nucleico que es complementario a una secuencia diana se hibrida predominantemente con la secuencia diana y sustancialmente no se hibrida con secuencias que no son diana. Las condiciones rigurosas normalmente dependen de la secuencia y varían dependiendo de una serie de factores. En general, cuanto más larga sea la secuencia, mayor será la temperatura a la que la secuencia se hibride específicamente con su secuencia diana. Algunos ejemplos no limitantes de condiciones rigurosas se describen detalladamente en Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Parte I, Segundo capítulo "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y. Cuando se hace referencia a una secuencia polinucleotídica, entonces también se contemplan secuencias complementarias o parcialmente complementarias. Preferiblemente estas tienen la capacidad de hibridarse con la secuencia de referencia en condiciones sumamente rigurosas. En general, con el fin de maximizar la tasa de hibridación, se seleccionan condiciones de hibridación con una rigurosidad relativamente baja: de aproximadamente entre 20 y 25 ºC más bajas que el punto de fusión térmica (T_m). La T_m es la temperatura a la cual

un 50% de la secuencia diana específica se hibrida con una sonda perfectamente complementaria en solución con una fuerza iónica y pH definidos. En general, cuando se necesita por lo menos aproximadamente un 85% de complementariedad nucleotídica de las secuencias hibridadas, se seleccionan condiciones de lavado sumamente rigurosas y que son de aproximadamente entre 5 y 15 °C más bajas que la T_m. Cuando se necesita por lo menos aproximadamente un 70% de complementariedad nucleotídica de las secuencias hibridadas, se seleccionan condiciones de lavado moderadamente rigurosas y que son de aproximadamente entre 15 y 30 °C más bajas que la T_m. Las condiciones de lavado sumamente permisivas (rigurosidad muy baja) podrán ser de incluso 50 °C por debajo de la T_m, lo que permite un nivel elevado de emparejamientos erróneos entre las secuencias hibridadas. Los expertos en la técnica reconocerán que también se pueden alterar otros parámetros físicos y químicos en las etapas de hibridación y lavado para modificar la respuesta de una señal de hibridación detectable que procede de un nivel específico de homología entre las secuencias diana y sonda. Las condiciones muy rigurosas preferidas comprenden la incubación en formamida al 50%, 5×SSC, y SDS al 1% a 42 °C, o incubación en 5×SSC y SDS al 1% a 65 °C, con lavado en 0,2×SSC y SDS al 0,1% a 65 °C. "Hibridación" se refiere a una reacción en la que uno o más polinucleótidos reaccionan para formar un complejo que se estabiliza mediante un enlace de hidrógeno entre las bases de los residuos de nucleótidos. La formación de puentes de hidrógeno podrá ocurrir mediante el emparejamiento de bases de Watson y Crick, unión de Hoogstein o de cualquier otra manera específica de la secuencia. El complejo podrá comprender dos hebras que forman una estructura dúplex, tres o más hebras que forman un complejo con múltiples hebras o una hebra sencilla autohibridante o cualquier combinación de estos. Una reacción de hibridación podrá constituir un paso de un proceso más extenso, tal como la iniciación de la PCR o el clivaje de un polinucleótido por una enzima. Una secuencia capaz de hibridarse con una secuencia dada se denomina la "complementaria" de la secuencia dada. Tal y como se utiliza en la presente, el término "locus genómico" o "locus" (plural loci) es la ubicación específica de un gen o secuencia de ADN en un cromosoma. Un "gen" se refiere a fragmentos de ADN o ARN que codifican un polipéptido o una cadena de ARN que ejerce un papel funcional en un organismo y, por lo tanto, es la unidad molecular heredable en organismos vivos. A efectos de esta invención, se puede considerar que los genes incluyen regiones que regulan la producción del producto genético, sin que importe si tales secuencias reguladoras son o no adyacentes a las secuencias codificantes y/o transcritas. En consecuencia, un gen incluye, sin carácter necesariamente limitante, secuencias promotoras, terminadores, secuencias reguladoras de la traducción tales como sitios de unión al ribosoma y sitios de entrada ribosómicos internos, potenciadores, silenciadores, aislantes, elementos frontera, orígenes de la replicación, sitios de unión a la matriz y regiones controladoras del locus. Tal y como se utilizan en la presente, la "expresión de un locus genómico" o la "expresión génica" es el proceso por el cual la información de un gen se utiliza en la síntesis de un producto genético funcional. Los productos de la expresión génica son a menudo proteínas, pero en los genes que codifican compuestos que no son proteínas tales como genes de ARNr o genes de ARNt, el producto es un ARN funcional. Todas las formas de vida conocidas utilizan el proceso de expresión génica - eucariotas (incluidos organismos multicelulares), procariotas (bacterias y arqueas) y virus para generar productos funcionales para sobrevivir. Tal y como se utiliza en la presente, la "expresión" de un gen o ácido nucleico engloba no solamente la expresión génica celular sino también la transcripción y traducción de uno o más ácidos nucleicos en sistemas de clonación y en cualquier otro contexto. Tal como se utiliza en la presente, la "expresión" también se refiere al proceso por el cual se transcribe un polinucleótido a partir de un molde de ADN (tal como para obtener ARNm u otro transcrito de ARN) y/o el proceso por el cual un transcrito de ARNm se traduce posteriormente para obtener péptidos, polipéptidos o proteínas. Los polipéptidos codificados y transcritos se podrán denominar de manera colectiva "producto genético". Si el polinucleótido se obtiene a partir de ADN genómico, la expresión podrá incluir el corte y empalme de ARNm en una célula eucariota. Los términos "polipéptido", "péptido" y "proteína" se utilizan como sinónimos en la presente para referirse a polímeros de aminoácidos de cualquier longitud. El polímero podrá ser lineal o ramificado y podrá comprender aminoácidos modificados y podrá estar interrumpido por componentes que no son aminoácidos. Los términos también abarcan un polímero de aminoácidos que se ha modificado; por ejemplo, mediante formación de enlaces disulfuro, glicosilación, lipidación, acetilación, fosforilación o cualquier otra manipulación tal como conjugación con un componente marcador. Tal y como se utiliza en la presente, el término "aminoácido" incluye aminoácidos naturales y/o no naturales o sintéticos que incluyen la glicina y los isómeros ópticos D o L, y análogos de aminoácidos y peptidomiméticos. Tal y como se utiliza en la presente, el término "dominio" o la expresión "dominio proteico" se refieren a una parte de una secuencia proteica que podrá existir y funcionar independientemente del resto de la cadena proteica. Tal y como se describe en los aspectos de la invención, la identidad secuencial está relacionada con la homología secuencial. Las comparaciones de la homología se podrán llevar a cabo a simple vista o, más comúnmente, con la ayuda de programas de comparación de secuencias a los que se puede acceder fácilmente. Estos programas computarizados comercializados podrán calcular la homología porcentual (%) entre dos o más secuencias y también podrán calcular la identidad secuencial compartida por dos o más secuencias de aminoácidos o ácidos nucleicos.

En aspectos de la invención el término "ARN guía", se refiere a la secuencia polinucleotídica que comprende una secuencia ARNcr putativa o identificada o una secuencia guía.

Tal y como se utiliza en la presente, la expresión "de tipo salvaje " es un término de la técnica cuyo significado conocen los expertos y se refiere a la forma típica de un organismo, cepa, gen o característica tal y como ocurre en la naturaleza y que se distingue de formas mutantes o variantes. Un "de tipo salvaje" puede ser un valor inicial.

Tal y como se utiliza en la presente, se debería sobreentender que el término "variante" se refiere a la exhibición de cualidades que tienen un patrón que se desvía del que ocurre en la naturaleza.

Los términos "de origen no natural" o "diseñado" se utilizan como sinónimos e indican la participación del hombre. Los términos, cuando se refieren a moléculas de ácido nucleico o polipéptidos se refieren a que la molécula de ácido nucleico o el polipéptido están por lo menos sustancialmente libres de por lo menos otro componente con el que se asocian de manera natural en la naturaleza y tal y como se observan en la naturaleza. En todos los aspectos y formas de realización, incluyan estos términos o no, se entenderá que, preferiblemente, pueden ser opcionales y por lo tanto incluidos preferiblemente o no incluidos no preferiblemente. Adicionalmente, los términos "de origen no natural" y "diseñado" pueden usarse como sinónimos y por lo tanto pueden usarse solos o en combinación y uno u otro puede reemplazar la mención de ambos juntos. En particular, "diseñado" se prefiere en lugar de "de origen no natural" o "de origen no natural y/o diseñado".

Se podrán generar homologías secuenciales mediante cualquiera de numerosos programas computarizados conocidos en la técnica, por ejemplo, BLAST o FASTA, etc. Un programa computarizado adecuado para llevar a cabo una alineación de este tipo es el paquete GCG Wisconsin Bestfit (Universidad de Wisconsin, EE.UU.; Devereux y col., 1984, Nucleic Acids Research 12:387). Los ejemplos de otros programas de computación que pueden realizar las comparaciones secuenciales incluyen, a título enunciativo no taxativo, el paquete BLAST (véase Ausubel y col., 1999 ibid – Capítulo 18), FASTA (Atschul y col., 1990, J. Mol. Biol., 403-410) y el paquete GENEWORKS de herramientas de comparación. Tanto BLAST como FASTA están disponibles para búsquedas con conexión y sin conexión a internet (véase Ausubel y col., 1999 ibid, páginas 7-58 a 7-60). Sin embargo, se prefiere utilizar el programa GCG Bestfit. La homología secuencial porcentual (%) se puede calcular en secuencias contiguas, es decir, se alinea una secuencia con otra secuencia y cada aminoácido o nucleótido de una secuencia se compara directamente con el aminoácido o nucleótido correspondiente de la otra secuencia, un residuo de cada vez. Esto se denomina alineación "sin huecos". Normalmente, se realizan alineaciones sin huecos solamente en números de residuos relativamente bajos. Aunque este es un método muy simple y consistente, falla a la hora de tener en cuenta que, por ejemplo, en pares de secuencias que por lo demás son idénticas, una inserción o eliminación podrá provocar que los siguientes residuos aminoacídicos dejen de estar alineados lo que puede conllevar potencialmente una gran reducción en el % de homología cuando se realiza una alineación global. En consecuencia, la mayoría de los métodos para la comparación secuencial están diseñados para producir alineaciones óptimas que tengan en cuenta posibles inserciones y eliminaciones sin penalizar indebidamente la homología global o calificación de la identidad. Esto se logra insertando "no coincidencias" en la alineación secuencial para tratar de maximizar la homología o identidad local. Sin embargo, estos métodos más complejos asignan "penalizaciones por no coincidencia" a cada no coincidencia que aparece en la alineación de modo que, para el mismo número de aminoácidos idénticos, una alineación secuencial con tan pocos huecos como sea posible - que refleje una gran semejanza entre las dos secuencias comparadas - podrá lograr una calificación más elevada que una con muchas no coincidencias. Normalmente se utilizan "costos de no coincidencia de afinidad" que imponen un costo relativamente elevado a la existencia de una no coincidencia y una penalización más pequeña para cada residuo posterior a la no coincidencia. Este es el sistema de calificación de no coincidencias utilizado más comúnmente. Las penalizaciones de no coincidencias elevadas podrán, obviamente, producir alineaciones optimizadas con menos no coincidencias. La mayoría de los programas de alineación permiten que se modifiquen las penalizaciones por no coincidencia. Sin embargo, se prefiere utilizar los valores predeterminados cuando se utilice un programa de computación de este tipo para las comparaciones secuenciales. Por ejemplo, cuando se utilice el paquete GCG Wisconsin Bestfit la penalización de no coincidencia predeterminada para las secuencias de aminoácidos es -12 para una no coincidencia y -4 para cada extensión. El cálculo del % de homología máximo, por lo tanto, requiere en primer lugar la producción de una alineación óptima que tenga en cuenta las penalizaciones por no coincidencias. Un programa computarizado adecuado para llevar a cabo una alineación de este tipo es el paquete GCG Wisconsin Bestfit (Devereux y col., 1984 *Nuc. Acids Research* 12 p387). Los ejemplos de otros softwares que pueden realizar las comparaciones secuenciales incluyen, a título enunciativo no taxativo, el paquete BLAST (véase Ausubel y col., 1999 Short Protocols in Molecular Biology, 4ª Ed. – Capítulo 18), FASTA (Altschul y col., 1990 J. Mol. Biol. 403-410) y las herramientas de comparación del paquete GENEWORKS. Tanto BLAST como FASTA están disponibles para búsquedas con conexión y sin conexión a internet (véase Ausubel y col., 1999 Short Protocols in Molecular Biology, páginas 7-58 a 7-60). Sin embargo, para algunas aplicaciones, se prefiere utilizar el programa GCG Bestfit. Una nueva herramienta, denominada BLAST 2 Sequences también está disponible para comparar secuencias de nucleótidos y proteínas (véase *FEMS Microbiol Lett.* 1999 174(2): 247-50; *FEMS Microbiol Lett.* 1999 177(1): 187-8 y la página web del Centro Nacional para la Información Biotecnológica en la página web de los Institutos Nacionales de Salud). Aunque el % de homología final se puede medir en función de la identidad, el propio proceso de alineación normalmente no se basa en una comparación de parejas de todo o nada. En su lugar, se utiliza en general una matriz de puntuación de la similitud con escala que asigna calificaciones a cada comparación por parejas en función de la similitud química o la distancia evolutiva. Como un ejemplo de esto una matriz que se utiliza normalmente es la matriz BLOSUM62 - la matriz predeterminada del grupo de programas BLAST. Los programas GCG Wisconsin utilizan generalmente los valores predeterminados públicos o una tabla de comparación de símbolos personalizada, si se suministra (véase el manual del usuario para más detalles). Para algunas aplicaciones, se prefiere utilizar los valores predeterminados públicos para el paquete GCG o, en el caso de otro programa de computación, la matriz predeterminada tal como BLOSUM62. Como alternativa, las homologías porcentuales se podrán calcular utilizando la característica de alineación múltiple en DNASIS™ (Hitachi Software), que se basa en un algoritmo análogo a CLUSTAL (Higgins DG & Sharp PM (1988), *Gene* 73(1), 237-244). Una vez que el software ha producido una alineación óptima, es posible calcular el % de homología, preferentemente el % de identidad secuencial. El programa de computación normalmente hace esto como parte de la comparación secuencial y genera un resultado numérico. Las secuencias también podrán tener eliminaciones, inserciones o sustituciones de residuos aminoacídicos que producen un cambio

silencioso y dan como resultado una sustancia funcionalmente equivalente. Se pueden realizar sustituciones de aminoácidos deliberadas en función de la similitud de las propiedades de los aminoácidos (tales como la polaridad, carga, solubilidad, hidrofobicidad, hidrofilicidad y/o la naturaleza anfipática de los residuos) y, por lo tanto, es útil agrupar los aminoácidos en grupos funcionales. Los aminoácidos se pueden agrupar en función únicamente de las propiedades de las cadenas laterales. Sin embargo, es más útil incluir también los datos de las mutaciones. Los conjuntos de aminoácidos que se obtienen de esta manera es probable que estén conservados por razones estructurales. Estos conjuntos se podrán describir en forma de un diagrama de Venn (Livingstone C.D. y Barton G.J. (1993) "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation" *Comput. Appl. Biosci.* 9: 745-756) (Taylor W.R. (1986) "The classification of amino acid conservation" *J. Theor. Biol.* 119; 205-218). Se podrán realizar sustituciones conservadoras, por ejemplo, de acuerdo con la siguiente tabla que describe un diagrama de Venn de aceptación generalizada para agrupar aminoácidos.

Conjunto		Subconjunto	
Hidrofóbico	F W Y H K M I L V A G C	Aromático	F W Y H
		Alifático	I L V
Polar	W Y H K R E D C S T N Q	Cargado	H K R E D
		Cargado positivamente	H K R
		Cargado negativamente	E D
Pequeño	V C A G S P T N D	Diminuto	A G S

Los términos "sujeto", "individuo" y "paciente" se utilizan indistintamente en la presente para referirse a un vertebrado, preferentemente un mamífero, más preferentemente un ser humano. Los mamíferos incluyen, a título enunciativo no taxativo, múridos, simios, seres humanos, animales de granja, animales para deportes y mascotas. También se engloban los tejidos, las células y la progenie de una entidad biológica obtenida *in vivo* o cultivada *in vitro*.

Los términos "agente terapéutico", "agente con capacidad terapéutica" o "agente de tratamiento" se utilizan indistintamente y se refieren a una molécula o compuesto que confiere algún efecto beneficioso tras la administración a un sujeto. El efecto beneficioso incluye la posibilidad de determinaciones diagnósticas; la mejora de una enfermedad, síntoma, trastorno o afección patológica; reducir o prevenir el inicio de una enfermedad, síntoma, trastorno o afección; y en general contrarrestar una enfermedad, síntoma, trastorno o afección patológica.

Tal y como se emplean en la presente, "tratamiento" o "tratar" o "paliar" o "mejorar" se utilizan indistintamente. Estos términos se refieren a una estrategia para obtener resultados beneficiosos o deseados que incluyen, a título enunciativo no taxativo, un beneficio terapéutico y/o un beneficio profiláctico. El beneficio terapéutico se refiere a cualquier mejora relevante terapéuticamente en una o más enfermedades, afecciones o síntomas que se están tratando, o su efecto sobre ellos. Para un beneficio profiláctico, se podrán administrar las composiciones a un sujeto en riesgo de desarrollar una enfermedad, afección o síntoma particulares o a un sujeto que refiera uno o más síntomas fisiológicos de una enfermedad, incluso aunque la enfermedad, afección o síntoma aún no se hayan manifestado.

El término "cantidad eficaz" o "cantidad terapéuticamente eficaz" se refiere a la cantidad de un agente que es suficiente para obtener un efecto beneficioso o los resultados deseados. La cantidad terapéuticamente eficaz podrá variar dependiendo de uno o más de los siguientes: el sujeto y estado patológico que se está tratando, el peso y edad del sujeto, la gravedad del estado patológico, el modo de administración y similares que el experto en la técnica podrá determinar fácilmente. La expresión también se aplica a una dosis que provea una imagen para la detección mediante cualquiera de los métodos de obtención de imágenes descritos en la presente. La dosis específica podrá variar dependiendo de uno o más de los siguientes: el agente particular escogido, la pauta posológica que se ha de seguir, si se administra o no combinado con otros compuestos, el momento de la administración, el tejido del que se va a obtener la imagen y el sistema de suministro físico que lo porta.

Varios aspectos de los métodos y composiciones reivindicados de la invención se refieren al uso de sistemas vectoriales que comprenden uno o más vectores, o a vectores como tal. Se pueden diseñar vectores para la expresión de transcritos de CRISPR (por ejemplo, proteínas, enzimas o transcritos de ácido nucleico) en células procariotas o eucariotas. Por ejemplo, se pueden expresar transcritos de CRISPR en células bacterianas tales como *Escherichia coli*, células de insecto (utilizando vectores de expresión de baculovirus), células de levadura o células de mamíferos. Las células huésped adecuadas se analizan en Goeddel, EXPRESIÓN GENÉTICA TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Como alternativa, el vector de expresión recombinante se puede transcribir y traducir *in vitro*, por ejemplo, utilizando secuencias reguladoras del promotor T7 y la polimerasa T7.

Las formas de realización de los métodos y composiciones reivindicados de la invención usan secuencias (tanto de

polinucleótidos como de polipéptidos) que podrán comprender una sustitución homóloga (tanto sustitución como reemplazo se utilizan en la presente para denotar el intercambio de un residuo aminoacídico o nucleótido existente con un residuo o nucleótido alternativo) que podrá ocurrir, es decir, la sustitución de igual por igual en el caso de los aminoácidos tal como básico por básico, ácido por ácido, polar por polar, etc. La sustitución no homóloga también podrá ocurrir, es decir, de una clase de residuo a otro o como alternativa que conlleve la inclusión de aminoácidos no naturales tales como ornitina (denominado en lo sucesivo Z), ácido diaminobutírico ornitina (denominado en lo sucesivo B), norleucina ornitina (denominado en lo sucesivo O), piriilalanina, tienilalanina, naftilalanina y fenilglicina. Las secuencias aminoacídicas variantes podrán incluir grupos separadores adecuados que se podrán insertar entre dos residuos aminoacídicos de la secuencia que incluyen grupos alquilo tales como grupos metilo, etilo o propilo además de separadores aminoacídicos tales como los residuos de glicina o β-alanina. Los expertos en la técnica podrán conocer bien una forma adicional de variación, que conlleva la presencia de uno o más residuos aminoacídicos en forma peptoide. Para evitar dudas, se utiliza "la forma peptoide" para referirse a residuos aminoacídicos variantes donde el grupo sustituyente del carbono está en el átomo de nitrógeno del residuo más que en el carbono α. En la técnica existe constancia de procesos para preparar péptidos en forma peptoide, por ejemplo, Simon RJ y col., *PNAS* (1992) 89(20), 9367-9371 y Horwell DC, *Trends Biotechnol.* (1995) 13(4), 132-134.

Modelado de homología: Los residuos correspondientes en otros ortólogos de Cpf1 pueden identificarse por los métodos de Zhang y col., 2012 (Nature; 490(7421): 556-60) y Chen y col., 2015 (PLoS Comput Biol; 11(5): e1004248)- un método computacional de interacción proteína-proteína (PPI) para predecir interacciones mediadas por interfaces dominio-motivo. El PrePPI (Predicting PPI), un método para predicción PPI en base a estructura, combina evidencia estructural con evidencia no estructural usando una estructura de estadística Bayesiana. El método involucra tomar un par de proteínas de consulta y usar alineamiento estructural para identificar representantes estructurales que se correspondan con sus estructuras determinadas experimentalmente o modelos de homología. El alineamiento estructural se usa adicionalmente para identificar vecinos cercanos y remotos considerando las relaciones geométricas globales y locales. Cuando dos vecinos de los representantes estructurales forman un complejo informado en el Banco de Datos de Proteínas, esto define un molde para modelar la interacción entre las dos proteínas de búsqueda. Los modelos del complejo se crean por superposición de las estructuras representativas en su vecino estructural correspondiente en el molde. Esta estrategia se describe adicionalmente en Dey y col., 2013 (Prot Sci; 22: 359-66).

A efectos de esta invención, el término amplificación se refiere a cualquier método que emplea un cebador y una polimerasa capaces de replicar una secuencia diana con una fidelidad razonable. La amplificación se puede llevar a cabo mediante polimerasas de ADN naturales o recombinantes tales como TaqGold™, ADN polimerasa T7, fragmento Klenow de la ADN polimerasa de *E. coli* y transcriptasa inversa. Un método de amplificación preferido es la PCR.

En ciertos aspectos, los métodos y composiciones reivindicados de la invención implican el uso de vectores. Tal como se usa en el presente documento, un "vector" es una herramienta que permite o facilita la transferencia de una entidad de un entorno a otro. Es un replicón, tal como un plásmido, fago o cósmido en el cual se puede insertar otro segmento de ADN de modo que provoque la replicación del segmento insertado. En general, un vector es capaz de replicarse cuando se asocia con los elementos de control apropiados. En general, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que está unido. Los vectores incluyen, a título enunciativo no taxativo, moléculas de ácido nucleico que son monocatenarias, bicatenarias o parcialmente bicatenarias; moléculas de ácido nucleico que comprenden uno o más extremos libres o sin extremos libres (por ejemplo, circulares); moléculas de ácido nucleico que comprenden ADN, ARN o ambos; y otras variedades de polinucleótidos conocidos en la técnica. Un tipo de vector es un vector vírico, donde están presentes en el vector secuencias de ADN o ARN de origen vírico para empaquetarlo en un virus (por ejemplo, retrovirus, retrovirus de replicación defectuosa, adenovirus, adenovirus de replicación defectuosa y virus adenoasociados (AAV)). Los vectores víricos también incluyen polinucleótidos portados por un virus para la transfección de una célula huésped. Ciertos vectores son capaces de una replicación autónoma en la célula huésped en la que se introducen (por ejemplo, vectores bacterianos que tienen un origen bacteriano de replicación y vectores episómicos de mamíferos). Otros vectores (por ejemplo, vectores no episómicos de mamíferos) se integran en el genoma de la célula huésped tras la introducción en la célula huésped y, de esta manera, se replican junto con el genoma huésped. Además, ciertos vectores son capaces de dirigir la expresión de genes a los que están ligados operativamente. Este tipo de vectores se denominan en la presente "vectores de expresión". Los vectores de expresión comunes útiles en las técnicas de ADN recombinante se encuentran a menudo en forma de plásmidos.

Los vectores de expresión recombinantes pueden comprender un ácido nucleico para su uso en los métodos y composiciones reivindicados de la invención en una forma adecuada para la expresión del ácido nucleico en una célula huésped, lo que significa que los vectores de expresión recombinantes incluyen uno o más elementos reguladores que se podrán seleccionar en función de las células huésped que se van a utilizar para la expresión, que están ligados operativamente a la secuencia de ácido nucleico que se va a expresar. En un vector de expresión recombinante, "ligado operativamente" se pretende que signifique que la secuencia nucleotídica de interés está ligada al elemento o los elementos reguladores de modo que permita la expresión de la secuencia nucleotídica (por ejemplo, en un sistema de transcripción/traducción *in vitro* o en una célula huésped cuando se introduce el vector en la célula huésped). En lo que respecta a los métodos de recombinación y clonación, cabe mencionar la solicitud de patente de los EE.UU. 10/815,730, publicada el 2 de septiembre de 2004 como US 2004-0171156 A1.

Los aspectos de los métodos y composiciones reivindicados de la invención se relacionan con el uso de vectores

bicistrónicos para el ARN guía y enzimas CRISPR (opcionalmente modificadas o mutadas) (por ejemplo, Cpf1). Se prefieren vectores de expresión bicistrónicos para ARN guía y enzimas CRISPR (opcionalmente modificadas o mutadas) (por ejemplo, Cpf1). En general y particularmente en esta forma de realización las enzimas CRISPR (opcionalmente modificadas o mutadas) preferiblemente están dirigidas por el promotor CBh. El ARN podrá estar impulsado preferentemente por un promotor Pol III, tal como un promotor U6. Idealmente se combinan los dos.

En algunas formas de realización, se provee un bucle en el ARN guía. Este puede ser un bucle tipo tallo o un bucle tetra. El bucle es preferiblemente GAAA pero no está limitado a esta secuencia ni, por supuesto, a tener una longitud de solo 4 pb. Ciertamente, las secuencias que forman bucles preferidas para su uso en estructuras de tipo horquilla tienen una longitud de cuatro nucleótidos y, más preferiblemente, tienen la secuencia GAAA. Sin embargo, se podrán utilizar secuencias bucle más largas o más cortas al igual que secuencias alternativas. Las secuencias incluyen preferiblemente un triplete de nucleótidos (por ejemplo, AAA) y un nucleótido adicional (por ejemplo, C o G). Los ejemplos de secuencias que forman bucles incluyen CAAA y AAAG. Al llevar a la práctica cualquiera de los métodos divulgados en la presente, se puede introducir un vector adecuado en una célula o un embrión no humano mediante uno o más métodos conocidos en la técnica incluidos, sin limitación, microinyección, electroporación, sonoporación, biolística, transfección mediada por fosfato de calcio, transfección catiónica, transfección con liposomas, transfección con dendrímeros, transfección por choque térmico, transfección por nucleofección, magnetofección, lipofección, impalafección, transfección óptica, captación de ácidos nucleicos potenciada mediante un agente exclusivo y suministro mediante liposomas, inmunoliposomas, virosomas o viriones artificiales. En algunos métodos, se introduce el vector en un embrión no humano por microinyección. El vector o los vectores se podrán microinyectar en el núcleo o en el citoplasma del embrión no humano. En algunos métodos, el vector o los vectores se podrán introducir en una célula por nucleofección.

La expresión "elemento regulador" se pretende que incluya promotores, potenciadores, sitios internos de entrada al ribosoma (IRES, por sus siglas en inglés) y otros elementos de control de la expresión (por ejemplo, señales de terminación de la transcripción tales como señales de poliadenilación y secuencias poliU). Este tipo de elementos reguladores se describen, por ejemplo, en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Los elementos reguladores incluyen aquellos que dirigen la expresión constitutiva de una secuencia nucleotídica en muchos tipos de células huésped y aquellos que dirigen la expresión de la secuencia nucleotídica únicamente en ciertas células huésped (por ejemplo, secuencias reguladoras específicas del tejido). Un promotor específico del tejido podrá dirigir la expresión principalmente en un tejido deseado de interés tal como músculo, neurona, hueso, piel, sangre, órganos específicos (por ejemplo, hígado, páncreas) o tipos celulares particulares (por ejemplo, linfocitos). Los elementos reguladores también podrán dirigir la expresión de manera dependiente del tiempo, tal como de manera dependiente del ciclo celular o dependiente de la etapa de desarrollo, que también podrá ser o no ser específica del tipo celular o del tejido. En algunas formas de realización, un vector comprende uno o más promotores de la pol III (por ejemplo, 1, 2, 3, 4, 5 o más promotores de la pol III), uno o más promotores de la pol II (por ejemplo, 1, 2, 3, 4, 5 o más promotores de la pol II), uno o más promotores de la pol I (por ejemplo, 1, 2, 3, 4, 5 o más promotores de la pol I) o combinaciones de estos. Los ejemplos de los promotores de la pol III incluyen, a título enunciativo no taxativo, los promotores H1 y U6. Los ejemplos de los promotores de la pol II incluyen, a título enunciativo no taxativo, el promotor LTR retrovírico del virus del sarcoma de Rous (RSV, por sus siglas en inglés) (opcionalmente con el potenciador de RSV), el promotor del citomegalovirus (CMV) (opcionalmente con el potenciador de CMV) [véase, por ejemplo, Boshart y col., *Cell*, 41:521-530 (1985)], el promotor de SV40, el promotor de la dihidrofolato-reductasa, el promotor de la β-actina, el promotor de la fosfoglicerol-quinasa (PGK, por sus siglas en inglés) y el promotor EF1α. También están abarcados en el término "elemento regulador" los elementos potenciadores, tales como WPRE; potenciadores de CMV; el segmento R-U5' en LTR de HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); potenciador de SV40; y la secuencia intrónica entre los exones 2 y 3 de la β-globina de conejo (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). Los expertos en la técnica comprenderán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que se va a transformar, el nivel de expresión deseado, etc. Se puede introducir un vector en las células huésped para producir de esta manera transcritos, proteínas o péptidos, incluidos péptidos o proteínas de fusión, codificados por ácidos nucleicos como los descritos en la presente (por ejemplo, enzimas, proteínas, transcritos de grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR), formas mutantes de estos, proteínas de fusión de estos, etc.). En lo que se refiere a las secuencias reguladoras, cabe mencionar la solicitud de patente de los EE.UU. 10/491,026. En lo que se refiere a los promotores, cabe mencionar la publicación PCT WO 2011/028929 y la solicitud de los EE.UU. 12/511 940.

Se pueden diseñar vectores para la expresión de transcritos de CRISPR (por ejemplo, proteínas, enzimas o transcritos de ácido nucleico) en células procariotas o eucariotas. Por ejemplo, se pueden expresar transcritos de CRISPR en células bacterianas tales como *Escherichia coli*, células de insecto (utilizando vectores de expresión de baculovirus), células de levadura o células de mamíferos. Las células huésped adecuadas se analizan en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Como alternativa, el vector de expresión recombinante se puede transcribir y traducir *in vitro*, por ejemplo, utilizando secuencias reguladoras del promotor T7 y la polimerasa T7.

Los vectores se podrán introducir y propagar en un procariota o célula procariota. En algunas formas de realización, se utiliza un procariota para amplificar copias de un vector que se va a introducir en una célula eucariota o como un vector intermediario para la producción de un vector que se va a introducir en una célula eucariota (por ejemplo,

amplificando un plásmido como parte de un sistema de empaquetamiento de un vector vírico). En algunas formas de realización, se utiliza un procariota para amplificar copias de un vector y expresar uno o más ácidos nucleicos, por ejemplo, para proveer una fuente de una o más proteínas para el suministro a una célula huésped u organismo huésped. La expresión de proteínas en procariotas se lleva a cabo muy a menudo en *Escherichia coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de las proteínas de fusión o de proteínas que no son de fusión. Los vectores de fusión añaden varios aminoácidos a una proteína que se ha codificado allí, tal como al extremo amino de la proteína recombinante. Tales vectores de fusión podrán cumplir uno o más propósitos, tales como: (i) incrementar la expresión de la proteína recombinante; (ii) incrementar la solubilidad de la proteína recombinante; y (iii) cooperar en la purificación de la proteína recombinante actuando como un ligando en la purificación por afinidad. A menudo, en los vectores de expresión de fusión, se introduce el sitio de escisión proteolítica en la unión del resto de fusión y la proteína recombinante para permitir la separación de la proteína recombinante y el resto de fusión después de la purificación de la proteína de fusión. Tales enzimas, y sus secuencias de reconocimiento afines, incluyen el Factor Xa, trombina y enteroquinasa. Los ejemplos de vectores de expresión de fusión incluyen pGEX (Pharmacia Biotech Inc; Smith y Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) y pRIT5 (Pharmacia, Piscataway, N.J.) que fusionan glutatión S-transferasa (GST), la proteína de unión a la maltosa E o la proteína A, respectivamente, a una proteína recombinante diana. Los ejemplos de vectores de expresión de *E. coli* inducibles adecuados que no son de fusión adecuados incluyen pTrc (Amrann y col., (1988) *Gene* 69:301-315) y pET 11d (Studier y col., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)(60-89). En algunas formas de realización, un vector es un vector de expresión de levaduras. Los ejemplos de vectores de expresión en levaduras Saccharomyces cerivisae incluyen pYepSec1 (Baldari, y col., 1987. EMBO J. 6: 229-234), pMFa (Kuijan y Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz y col., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) y picZ (InVitrogen Corp, San Diego, Calif.). En algunas formas de realización, un vector impulsa la expresión proteica en células de insecto utilizando vectores de expresión de baculovirus. Los vectores de Baculovirus disponibles para la expresión de proteínas en células de insecto cultivadas (por ejemplo, células SF9) incluyen las series pAc (Smith, y col., 1983. Mol. Cell. Biol. 3: 2156-2165) y las series pVL (Lucklow y Summers, 1989. Virology 170: 31-39).

En algunas formas de realización, un vector es capaz de impulsar la expresión de una o más secuencias en células de mamífero utilizando un vector de expresión de mamíferos. Los ejemplos de vectores de expresión mamíferos incluyen pCDM8 (Seed, 1987. Nature 329: 840) y pMT2PC (Kaufman, y col., 1987. EMBO J. 6: 187-195). Cuando se utilizan en células de mamíferos, normalmente uno o más elementos reguladores son los que proveen las funciones de control del vector de expresión. Por ejemplo, los promotores utilizados normalmente se obtienen a partir del polioma, adenovirus 2, citomegalovirus, virus del simio 40 y otros descritos en la presente y conocidos en la técnica. Para consultar otros sistemas de expresión adecuados tanto para células procariotas como eucariotas véase, por ejemplo, los Capítulos 16 y 17 de Sambrook y col., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

En algunas formas de realización, el vector de expresión de mamíferos recombinante es capaz de dirigir la expresión del ácido nucleico preferiblemente en un tipo celular concreto (por ejemplo, se utilizan elementos reguladores específicos del tejido para expresar el ácido nucleico). En el arte existe constancia de elementos reguladores específicos del tejido. Los ejemplos no limitantes de promotores específicos del tejido adecuados incluyen el promotor de la albúmina (específico del hígado; Pinkert, y col., 1987. *Genes Dev.*1:268-277), promotores con especificidad linfoide (Calame y Eaton, 1988. *Adv. Immunol.* 43:235-275), en particular promotores de receptores de linfocitos T (Winoto y Baltimore, 1989. *EMBO J.* 8:729-733) e inmunoglobulinas (Baneiji, y col., 1983. *Cell* 33:729-740; Queen y Baltimore, 1983. *Cell* 33:741-748), promotores específicos de las neuronas (por ejemplo, el promotor del neurofilamento; Byrne y Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86:5473-5477), promotores específicos del páncreas (Edlund, y col., 1985. *Science* 230:912-916) y promotores específicos de las glándulas mamarias (por ejemplo, promotor del suero de la leche, patente de los EE.UU. con n.º 4.873.316 y la solicitud europea con n.º de publicación 264.166). También se engloban los promotores regulados por el desarrollo, por ejemplo, los promotores hox murinos (Kessel y Gruss, 1990. *Science* 249:374-379) y el promotor de la α-fetoproteína (Campes y Tilghman, 1989. *Genes Dev.* 3:537-546). En lo que se refiere a estos vectores procariotas y eucariotas, cabe mencionar la patente de los EE.UU. 6.750.059. Otras formas de realización de la invención se podrán relacionar con el uso de vectores víricos, respecto a los cuales cabe mencionar la solicitud de patente de los EE.UU. 13/092 085. En la técnica existe constancia de elementos reguladores específicos del tejido y en lo que se refiere a estos, cabe mencionar la patente de los EE.UU. 7,776,321. En algunas formas de realización, un elemento regulador está ligado operativamente a uno o más elementos de un sistema CRISPR de modo que impulse la expresión del elemento o elementos del sistema CRISPR. En general, las CRISPR (grupos de repeticiones palindrómicas cortas en intervalos regulares), también conocidas como SPIDR (siglas en inglés de repeticiones directas con espaciadores intercalados), constituyen una familia de loci de ADN que son específicos normalmente para una especie bacteriana particular. El locus CRISPR comprende una clase diferente de repeticiones de secuencias cortas espaciadas entre sí (SSR, por sus siglas en inglés) que se reconoció en E. coli (Ishino y col., J. Bacteriol., 169:5429-5433 [1987]; y Nakata y col., J. Bacteriol., 171:3553-3556 [1989]) y en genes asociados. Se han identificado SSR separadas entre sí similares en Haloferax mediterranei, Streptococcus pyogenes, Anabaena y Mycobacterium tuberculosis (véase, Groenen y col., Mol. Microbiol., 10:1057-1065 [1993]; Hoe y col., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl y col., Biochim. Biophys. Acta 1307:26-30 [1996]; y Mojica y col., Mol. Microbiol., 17:85-93 [1995]). Los loci CRISPR difieren normalmente de otras SSR en la estructura de las repeticiones, que se han denominado repeticiones cortas separadas regularmente (SRSR, por sus

siglas en inglés) (Janssen y col., OMICS J. Integ. Biol., 6:23-33 [2002]; y Mojica y col., Mol. Microbiol., 36:244-246 [2000]). En general, las repeticiones son elementos cortos que se presentan en agrupaciones que están espaciadas regularmente por secuencias intrónicas singulares con una longitud sustancialmente constante (Mojica y col., [2000], precedentemente). Aunque las secuencias repetidas están sumamente conservadas entre cepas, el número de las repeticiones espaciadas entre sí y las secuencias de las regiones espaciadoras difiere normalmente de cepa a cepa (van Embden y col., J. Bacteriol., 182:2393-2401 [2000]). Se han identificado loci CRISPR en más de 40 procariotas (véase, por ejemplo, Jansen y col Microbiol., 43:1565-1575 [2002]; y Mojica y col., [2005]) que incluyen, a título enunciativo no taxativo, Aeropyrum, Pyrobaculum, Sulfolobus, Archaeoglobus, Halocarcula, Methanobacterium, Methanococcus, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thermoplasma, Corynebacterium, Mycobacterium, Streptomyces, Aquifex, Porphyromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myxococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema y Thermotoga.

En general, "sistema de direccionamiento a ácido nucleico" como se usa en la presente solicitud se refiere colectivamente a transcriptos y otros elementos involucrados en la expresión de o en el direccionamiento de la actividad de genes asociados con CRISPR de direccionamiento a ácido nucleico ("Cas") (también referido en la presente como una proteína efectora), que incluye secuencias que codifican para una proteína (efectora) Cas de direccionamiento a ácido nucleico y un ARN guía u otras secuencias y transcriptos de un locus CRISPR de direccionamiento a ácido nucleico. En algunas realizaciones, uno o más elementos de un sistema de direccionamiento a ácido nucleico se obtienen a partir de un sistema CRISPR de direccionamiento a ácido nucleico de tipo V/Tipo VI. En algunas realizaciones, uno o más elementos de un sistema de direccionamiento a ácido nucleico se obtienen a partir de un organismo particular que comprende un sistema CRISPR endógeno de direccionamiento a ácido nucleico. En general, un sistema de direccionamiento a ácido nucleico se caracteriza por elementos que promueven la formación de un complejo de direccionamiento a ácido nucleico en el sitio de una secuencia diana. En el contexto de la formación de un complejo de direccionamiento a ácido nucleico, la "secuencia diana" se refiere a una secuencia respecto a la cual se ha diseñado una secuencia guía complementaria, donde la hibridación entre una secuencia diana y un ARN guía promueve la formación de un complejo dirigido a ADN o ARN. No se requiere necesariamente una complementariedad completa, siempre que haya una complementariedad suficiente para provocar la hibridación y promover la formación de un complejo de direccionamiento a ácido nucleico. Una secuencia diana puede comprender polinucleótidos de ARN. En algunas realizaciones, una secuencia diana está ubicada en el núcleo o citoplasma de una célula. En algunas realizaciones, la secuencia diana podrá estar en el interior de una organela de una célula eucariota, por ejemplo, la mitocondria o el cloroplasto. Una secuencia o molde que se puede utilizar para la recombinación en el locus diana que comprende las secuencias diana se denomina "molde de edición" o "ARN de edición" o "secuencia de edición". En algunos aspectos de la invención, se puede denominar molde de edición a un polinucleótido de molde exógeno. En un aspecto de la invención, la recombinación es recombinación homóloga.

Normalmente, en el contexto de un sistema de direccionamiento a ácido nucleico endógeno, la formación de un complejo de direccionamiento a ácido nucleico (que comprende un ARN guía que se hibrida con una secuencia diana y forma complejo con una o más proteínas efectoras de direccionamiento a ácido nucleico) da como resultado el clivaje de una o ambas hebras del ARN en la secuencia diana o cerca de esta (por ejemplo, en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50 o más pares de bases de esta). En algunas formas de realización, se introducen uno o más vectores que impulsan la expresión de uno o más elementos de un sistema de direccionamiento a ácido nucleico en una célula hospedadora de modo que la expresión de los elementos del sistema de direccionamiento a ácido nucleico dirija la formación de un complejo de direccionamiento a ácido nucleico en uno o más sitios diana. Por ejemplo, una proteína efectora y un ARN guía de direccionamiento a ácido nucleico podría estar cada uno operativamente ligado a elementos regulatorios por separado en vectores por separado. Como alternativa, pueden combinarse dos o más de los elementos expresados de los mismos o diferentes elementos regulatorios, en un vector individual, con uno o más vectores adicionales que proveen cualquier componente del sistema de direccionamiento a ácido nucleico no incluido en el primer vector. Los elementos del sistema de direccionamiento a ácido nucleico que se combinan en un vector individual pueden disponerse en cualquier orientación tal como un elemento localizado 5' respecto a ("corriente arriba" de) o 3' respecto a ("corriente abajo" de) un segundo elemento. La secuencia codificante de un elemento podrá estar ubicada en la misma o en una hebra opuesta a la secuencia codificante de un segundo elemento, y orientada en una dirección igual u opuesta. En algunas formas de realización, un promotor único impulsa la expresión de un transcrito que codifica una proteína efectora de direccionamiento a ácido nucleico y un ARN guía integrado en una o más secuencias intrónicas (por ejemplo, cada una en un intrón diferente, dos o más en por lo menos un intrón o todas en un único intrón). En algunas formas de realización, la proteína efectora y ARN guía de direccionamiento a ácido nucleico están operativamente ligados a y expresados a partir del mismo promotor.

En general, una secuencia guía es cualquier secuencia de polinucleótidos que tiene una complementariedad suficiente con una secuencia polinucleotídica diana para hibridarse con la secuencia diana y dirigir la unión específica respecto a la secuencia de un complejo de direccionamiento a ácido nucleico con la secuencia diana. En algunas formas de realización, el grado de complementariedad entre una secuencia guía y su secuencia diana correspondiente, cuando se alinean de manera óptima utilizando un algoritmo de alineación adecuado, es aproximadamente o más de aproximadamente un 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97,5%, 99% o más. Se puede determinar la alineación

óptima utilizando cualquier algoritmo adecuado para alinear secuencias, cuyos ejemplos no limitantes incluyen el algoritmo de Smith-Waterman, el algoritmo de Needleman-Wunsh, los algoritmos basados en la transformada de Burrows-Wheeler (por ejemplo, el alineador de Burrows Wheeler), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies), ELAND (Illumina, San Diego, CA), SOAP (disponible en soap.genomics.org.cn), y Maq (disponible en maq.sourceforge.net). En algunas formas de realización, una secuencia guía tiene una longitud de aproximadamente o más de aproximadamente 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 o más nucleótidos. En algunas formas de realización, una secuencia guía tiene una longitud inferior a aproximadamente 75, 50, 45, 40, 35, 30, 25, 20, 15, 12 o menos nucleótidos. La capacidad de una secuencia guía para dirigir una unión específica de la secuencia de un complejo de direccionamiento a ácido nucleico a una secuencia diana se puede evaluar mediante cualquier ensayo adecuado. Por ejemplo, se pueden proveer los componentes de un sistema de direccionamiento a ácido nucleico suficientes para formar un complejo de direccionamiento a ácido nucleico, incluida la secuencia guía que se va a probar, a una célula hospedadora que tenga la secuencia diana correspondiente, tal como por transfección con vectores que codifican los componentes de la secuencia CRISPR de direccionamiento a ácido nucleico, y a continuación evaluar el clivaje preferente dentro de la secuencia diana o en la vecindad de ella, tal como mediante el ensayo Surveyor tal y como se describe en la presente. De manera similar, se puede evaluar el clivaje de una secuencia polinucleotídica diana en un tubo de ensayo proveyendo la secuencia diana, los componentes de un complejo de direccionamiento a ácido nucleico, incluida la secuencia guía que se va a probar y una secuencia guía de control diferente de la secuencia guía de prueba, y comparando la unión o la tasa de clivaje en la secuencia diana entre las reacciones de la secuencia guía de control y de prueba. Son posibles otros ensayos como comprenderán los expertos en la técnica.

Se puede seleccionar una secuencia guía para que tenga como diana cualquier secuencia diana. En algunas formas de realización, la secuencia diana es una secuencia dentro de un transcripto de un gen o ARNm.

En algunas formas de realización, una secuencia diana es una secuencia dentro de un genoma de una célula.

En algunas formas de realización, se selecciona una secuencia guía para reducir el grado de estructura secundaria dentro de la secuencia guía. Se puede determinar la estructura secundaria mediante cualquier algoritmo de plegamiento de polinucleótidos adecuado. Algunos problemas están basados en el cálculo de la mínima energía libre de Gibbs. Un ejemplo de un algoritmo de este tipo es mFold, tal como lo describen Zuker y Stiegler (*Nucleic Acids Res.* 9 (1981), 133-148). Otro ejemplo de un algoritmo de plegamiento es el servidor web con conexión a internet RNAfold, desarrollado en el Instituto de Química Teórica de la Universidad de Viena, que utiliza el algoritmo de predicción con estructura centroide (véase, por ejemplo, A.R. Gruber y col., 2008, Cell 106(1): 23-24; y PA Carr y GM Church, 2009, Nature Biotechnology 27(12): 1151-62). Se puede acceder a más algoritmos en la solicitud de los EE.UU. con n.º de acta TBA (número de expediente 44790.11.2022; referencia amplia BI-2013/004A).

En algunas formas de realización, también se provee un molde de recombinación. Un molde de recombinación puede ser un componente de otro vector como se describe en la presente, contenido en un vector separado, o provisto como un polinucleótido separado. En algunas formas de realización, se diseña un molde de recombinación para que sirva como molde de la recombinación homóloga, tal como dentro o cerca de una secuencia diana con corte de hebra simple o clivada por una proteína efectora dirigida a ácido nucleico como una parte de un complejo dirigido a ácido nucleico. Un polinucleótido molde puede tener cualquier longitud adecuada, tal como aproximadamente o más de aproximadamente 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, o más nucleótidos de longitud. En algunas formas de realización, el polinucleótido molde es complementario con una porción de un polinucleótido que comprende la secuencia diana. Cuando se alinea de manera óptima, un polinucleótido molde podría superponerse con uno o más nucleótidos de una secuencia diana (por ejemplo, aproximadamente o más de aproximadamente 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 o más nucleótidos). En algunas formas de realización, cuando una secuencia molde y una polinucleótido que comprende una secuencia diana están óptimamente alineados, el nucleótido más cercano del polinucleótido molde está dentro de aproximadamente 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, o más nucleótidos de la secuencia diana.

En algunas formas de realización, la proteína efectora de direccionamiento a ácido nucleico es parte de una proteína de fusión que comprende uno o más dominios proteicos heterólogos (por ejemplo, aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más dominios además de la proteína efectora de direccionamiento a ácido nucleico). En algunas formas de realización, la proteína efectora CRISPR es parte de una proteína de fusión que comprende uno o más dominios proteicos heterólogos (por ejemplo, aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más dominios además de la enzima CRISPR). Una proteína de fusión de una enzima CRISPR podrá comprender cualquier secuencia proteica adicional y, opcionalmente, una secuencia conectora entre dos dominios cualesquiera. Los ejemplos de dominios proteicos que se podrán fusionar con una enzima CRISPR incluyen, a título enunciativo no taxativo, epítopos de identificación, secuencias génicas indicadoras y dominios proteicos que tienen una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor liberador de la transcripción, actividad de modificación de histonas, actividad de clivaje del ARN y actividad de unión a ácido nucleico. Los ejemplos no limitantes de epítopos de identificación incluyen los identificadores de histidina (His), identificadores V5, identificadores FLAG, identificadores de la hemaglutinina de influenza (HA), identificadores Myc, identificadores VSV-G e identificadores de tiorredoxina (Trx). Los ejemplos de genes indicadores incluyen, a título enunciativo no taxativo, glutatión-S-transferasa (GST), peroxidasa de rábano picante (HRP), cloranfenicol

acetiltransferasa (CAT) beta-galactosidasa, beta-glucuronidasa, luciferasa, proteína fluorescente amarilla (GFP), HcRed, DsRed, proteína fluorescente cian (CFP), proteína fluorescente amarilla (YPF) y proteínas autofluorescentes incluida la proteína fluorescente azul (BFP). Se puede fusionar una enzima CRISPR a una secuencia génica que codifique una proteína o un fragmento de una proteína que se une a moléculas de ADN o que se une a otras moléculas celulares incluidas, a título enunciativo no taxativo, la proteína de unión a maltosa (MBP), identificador S, fusiones del dominio de unión a ADN Lex A (DBD), fusiones del dominio de unión a ADN GAL4 y fusiones de la proteína BP16 del virus del herpes simple (HSV). Dominios adicionales que podrán formar parte de una proteína de fusión que comprende una enzima CRISPR se describen en US20110059502. En algunas formas de realización, se utiliza una enzima CRISPR marcada para identificar la ubicación de una secuencia diana.

En algunas formas de realización, una enzima CRISPR podrá formar un componente de un sistema inducible. La naturaleza inducible del sistema permitiría el control espaciotemporal de la edición génica o expresión génica utilizando una forma de energía. La forma de energía podrá incluir, a título enunciativo no taxativo, la radiación electromagnética, energía sonora, energía química y energía térmica. Los ejemplos de un sistema inducible incluyen promotores inducibles por tetraciclina (Tet-On o Tet-Off), sistemas activadores de la transcripción de doble híbrido que son moléculas de bajo peso molecular (FKBP, ABA, etc.) o sistemas inducibles por la luz (Fitocromo, dominios LOV o criptocromo). En una forma de realización, la enzima CRISPR podrá ser una parte de un efector transcripcional inducible por la luz (LITE, por sus siglas en inglés) para dirigir cambios en la actividad transcripcional de una manera específica de la secuencia. Los componentes de una luz podrán incluir una enzima CRISPR, un heterodímero de un citocromo que responde a la luz (por ejemplo, de *Arabidopsis thaliana*) y un dominio de activación/represión transcripcional. Otros ejemplos de proteínas de unión a ADN inducibles y métodos para su uso se proveen en US 61/736465 y US 61/721.283 y WO 2014/018423 y US8889418, US8895308, US20140186919, US20140242700, US20140273234, US20140335620, WO2014093635.

Suministro

En algunos aspectos, los métodos reivindicados de la invención comprenden suministrar uno o más polinucleótidos, tal como uno o más vectores como se describen en la presente, uno o más transcritos de estos y/o una o proteínas transcritas a partir de ellos, a una célula huésped. También se describe, pero no se reivindica específicamente en la presente, células producidas mediante tales métodos y organismos no humanos (tales como animales, plantas u hongos) que comprenden tales células o producidos a partir de ellas. En algunas formas de realización de los métodos reivindicados, se provee a una célula una proteína efectora de direccionamiento a ácido nucleico combinada con (y opcionalmente complejada con) un ARN guía. Se pueden utilizar métodos de transferencia génica con virus y sin virus convencionales para introducir ácidos nucleicos en células de mamíferos o tejidos diana. Se pueden utilizar tales métodos para administrar ácidos nucleicos que codifican componentes de un sistema de direccionamiento a ácido nucleico a células en cultivo, o en un organismo huésped. Los sistemas de suministro con vectores no víricos incluyen plásmidos de ADN, ARN (por ejemplo, un transcrito de un vector descrito en la presente), ácido nucleico desnudo y ácido nucleico complejado con un vehículo de suministro, tal como un liposoma. Los sistemas de suministro vectoriales víricos incluyen virus de ADN y ARN, que tienen genomas episómicos o integrados tras el suministro a la célula. Para una revisión de los procedimientos de la terapia génica véase Anderson, Science 256:808-813 (1992); Nabel y Felgner, TIBTECH 11:211-217 (1993); Mitani y Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer y Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada y col., en Current Topics in Microbiology and Immunology Doerfler y Böhm (eds) (1995); y Yu y col., Gene Therapy 1:13-26 (1994).

Los métodos de administración no víricos de ácidos nucleicos incluyen la lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, conjugados de policatión o lípido:ácido nucleico, ADN puro, viriones artificiales y captación de ADN potenciada por un agente. La lipofección se describe en por ejemplo, las Patentes de los EE.UU. n.º 5.049.386, 4.946.787; y 4.897.355) y los reactivos de lipofección se venden de manera comercial (por ejemplo, Transfectam™ y Lipofectin™). Los lípidos catiónicos y neutros que son adecuados para la lipofección de reconocimiento de receptor eficaz de polinucleótidos inlcuyen aquellos de Felgner, WO 91/17424; WO 91/16024. La administración puede ser a células (por ejemplo administración *in vitro* o *ex vivo*) o tejidos diana (por ejemplo administración *in vivo*). La preparación de complejos lípido:ácido nucleico, incluidos los liposomas dirigidos tales como los complejos de inmunolípidos, es muy conocida por el experto en la técnica (véase, por ejemplo, Crystal, *Science* 270:404-410 (1995); Blaese y col., *Cancer Gene Ther.* 2:291-297 (1995); Behr y col., Bioconjugate Chem. 5:382-389 (1994); Remy y col., Bioconjugate Chem. 5:647-654 (1994); Gao y col., Gene Therapy 2:710-722 (1995); Ahmad y col., Cancer Res. 52:4817-4820 (1992); Patente de los EE.UU. n.º 4.186.183, 4.217.344, 4.235.871, 4.261.975, 4.485.054, 4.501.728, 4.774.085, 4.837.028 y 4.946.787).

El uso de sistemas con ARN o ADN vírico para el suministro de ácidos nucleicos aprovecha los procesos sumamente evolucionados para dirigir un virus a células específicas en el cuerpo y para introducir la carga dañina viral en el núcleo. Los vectores víricos se pueden administrar directamente a pacientes (in vivo) o se pueden utilizar para tratar células in vitro y las células modificadas se podrán administrar opcionalmente a pacientes (ex vivo). Los sistemas con virus convencionales podrían incluir vectores retrovíricos, lentivíricos, adenovíricos, adenoasociados y del virus del herpes simple para la transferencia génica. La integración en el genoma huésped es posible con métodos de transferencia génica con retrovirus, lentivirus y virus adenoasociados y a menudo dan como resultado la expresión a largo plazo del

transgén insertado. Además, se han observado eficacias de transducción elevadas en muchos tipos celulares y tejidos diana.

Se puede alterar el tropismo de un retrovirus incorporando proteínas de la envoltura foráneas para expandir la población diana potencial de células diana. Los vectores lentivíricos son vectores retrovíricos que son capaces de transducir o infectar células que no se dividen y normalmente producen títulos víricos elevados. La selección de un sistema de transferencia génico retrovírico podría depender, por lo tanto, del tejido diana. Los vectores retrovíricos comprenden repeticiones terminales largas que actúan en *cis* con capacidad de empaquetar hasta 6-10 kb de una secuencia foránea. Los LTR que actúan en *cis* mínimos son suficientes para la replicación y empaquetamiento de los vectores, que se utilizan a continuación para integrar el gen terapéutico en la célula diana para proveer una expresión del transgén permanente. Los vectores retrovíricos utilizados ampliamente incluyen los basados en el virus de la leucemia murina (MuLV), el virus de la leucemia de monos gibones (GaLV), virus de la inmunodeficiencia del simio (VIS), virus de la inmunodeficiencia humana (VIH) y combinaciones de estos (véase, por ejemplo, Buchscher y col., J. Virol. 66:2731-2739 (1992); Johann y col., J. Virol. 66:1635-1640 (1992); Sommnerfelt y col., Virol. 176:58-59 (1990); Wilson y col., J. Virol. 63:2374-2378 (1989); Miller y col., J. Virol. 65:2220-2224 (1991); PCT/US94/05700). En las aplicaciones en las que se prefiere la expresión transitoria se podrán utilizar sistemas derivados de adenovirus. Los vectores derivados de adenovirus son capaces de una eficacia de transducción muy elevada en muchos tipos celulares y no requieren división celular. Con este tipo de vectores se han obtenido niveles de expresión y títulos elevados. Este vector se puede producir en grandes cantidades en un sistema relativamente simple. Los vectores derivados de virus adenoasociados ("AAV") también se podrán utilizar para transducir células con ácidos nucleico diana, por ejemplo, en la producción in vitro de ácidos nucleicos y péptidos y para los procedimientos de terapia génica in vivo y ex vivo (véase, por ejemplo, West y col., Virology 160:38-47 (1987); Patente de los EE.UU. n.º 4.797.368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). La construcción de vectores de AAV recombinantes se describe en varias publicaciones, que incluyen la Patente de los EE.UU. n.º 5.173.414; Tratschin y col., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, y col., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); y Samulski y col., J. Virol. 63:03822-3828 (1989).

Opciones para ADN/ARN o ADN/ADN o ARN/ARN o proteína/ARN

En algunas formas de realización, los componentes del sistema CRISPR pueden administrarse en diferentes formas, tal como combinaciones de ADN/ARN o ARN/ARN o proteína/ARN. Por ejemplo, la Cpf1 puede administrarse como un polinucleótido que codifica para un ADN o un polinucleótido que codifica para un ARN o como una proteína. La guía puede administrarse como un polinucleótido que codifica para un ADN o un ARN. Se contemplan todas las combinaciones posibles, incluyendo formas mezcladas de administración.

En algunas formas de realización, todas las combinaciones mencionadas (ADN/ARN o ADN/ADN o ARN/ARN o proteína/ARN).

En alguna forma de realización, cuando se administra Cpf1 en forma de proteína, es posible preensamblarla con una o más guías.

Nanoovillos

Adicionalmente, el sistema CRISPR puede administrarse usando nanoovillos, por ejemplo, como se describe en Sun W y col, *Cocoon-like self-degradable DNA nanoclew for anticancer drug delivery.,* J Am Chem Soc. 2014 Oct 22;136(42):14722-5. doi: 10.1021/ja5088024. Epub 2014 Oct 13; o en Sun W y col, *Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing.,* Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. doi: 10.1002/anie.201506030. Epub 2015 Aug 27.

La práctica de la presente invención emplea, a menos que se indique de otra manera, técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica y ADN recombinante, que se encuentran dentro de la experiencia del arte. Véase Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2da edición (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, y col. eds., (1987)); las series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames y G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, y ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

Modelos de condiciones genéticas y epigenéticas

También se describen, pero no son parte de la invención, métodos que pueden usarse para crear una planta, un animal no humano o célula que puede usarse para modelar y/o estudiar las condiciones géneticas o epigéneticas de interés, tal como mediante un modelo de mutaciones de interés o un modelo de enfermedad. Tal y como se utiliza en la presente, "enfermedad" se refiere a una enfermedad, trastorno o indicio en un sujeto. Por ejemplo, se puede utilizar dicho método para crear un animal no humano o célula que comprenda una modificación en una o más secuencias de ácido nucleico asociadas con una enfermedad, o una planta, animal o célula en la que está alterada la expresión de una o más secuencias de ácido nucleico asociadas con una enfermedad. Una secuencia de ácido nucleico de este tipo podrá codificar una secuencia proteica asociada con una enfermedad o podrá ser una secuencia de control asociada con una enfermedad. En consecuencia, se entiende que en estos métodos una planta, sujeto, paciente,

organismo o célula puede ser un sujeto, paciente, organismo o célula no humano. Así, también se describe, pero no se reivindica específicamente en la presente, una planta, animal o célula producidos por los métodos de la presente o una progenie de estos. La progenie podrá ser un clon de la planta o animal producidos o podrá ser el resultado de la reproducción sexual por cruzamiento con otros individuos de la misma especie para lograr la introgresión de rasgos deseables adicionales en su descendencia. La célula puede ser *in vivo* o *ex vivo* en los casos de organismos multicelulares, particularmente no humanos o plantas. En el caso en el que la célula esté en cultivo, se puede establecer una línea celular si se cumplen las condiciones de cultivo apropiadas y preferiblemente si la célula se adapta de manera adecuada para este fin (por ejemplo, una célula madre). También se describen las líneas celulares bacterianas producidas mediante la invención. Así, pues también se describen líneas celulares.

En algunos métodos, se puede utilizar el modelo de la enfermedad para estudiar los efectos de mutaciones en el animal o célula y el desarrollo y/o evolución de la enfermedad utilizando medidas utilizadas comúnmente en el estudio de la enfermedad. Como alternativa, un modelo de una enfermedad de este tipo es útil para estudiar el efecto de un compuesto farmacéuticamente activo en la enfermedad.

En algunos métodos, se puede utilizar el modelo de la enfermedad para evaluar la eficacia de una estrategia de terapia génica potencial. Es decir, se puede modificar un gen o polinucleótido asociado con una enfermedad de modo que se inhiba o reduzca el desarrollo y/o evolución de la enfermedad. En particular, el método comprende modificar un gen o polinucleótido asociado con una enfermedad de modo que se produzca una proteína alterada y, como resultado, el animal o célula tenga una respuesta alterada. En consecuencia, en algunos métodos, se puede comparar un animal modificado genéticamente con un animal predispuesto al desarrollo de la enfermedad de modo que se pueda evaluar el efecto de la terapia génica.

También se describe, pero no se reivindica en la presente, un método para desarrollar un agente biológicamente activo que module un evento de señalización celular asociado con un gen ligado con una enfermedad. El método comprende poner en contacto un compuesto de prueba con una célula que comprenda uno o más vectores que impulsen la expresión de una o más enzimas CRISPR, y una secuencia de repetición directa ligada a una secuencia guía; y detectar un cambio en una lectura que sea indicativa de una reducción o un aumento de un evento de señalización celular asociado con, por ejemplo, una mutación en un gen ligado a una enfermedad contenido en la célula.

Se puede construir un modelo en una célula o un modelo en animales combinado con el método de la invención para explorar un cambio en la función celular. Se puede utilizar un modelo de este tipo para estudiar los efectos de una secuencia genómica modificada por los métodos reivindicados que hacen uso del complejo CRISPR en una función celular de interés. Por ejemplo, se puede utilizar un modelo de una función celular para estudiar el efecto de una secuencia genómica modificada en la señalización intracelular o la señalización extracelular. Como alternativa, se puede utilizar un modelo de una función celular para estudiar el efecto de una secuencia genómica modificada en la percepción sensorial. En algunos modelos de este tipo, una o más secuencias genómicas asociadas con una ruta bioquímica de señalización del modelo están modificadas.

Se han estudiado específicamente varios modelos de enfermedades. Estos incluyen los genes de riesgo del autismo de novo CHD8, KATNAL2 y SCN2A; y el gen del autismo sindrómico (síndrome Angelman) UBE3A. Obviamente, se prefieren estos genes y los modelos de autismo resultantes, pero sirven para mostrar la amplia aplicabilidad de la invención en genes y sus modelos correspondientes. Se puede determinar una expresión alterada de una o más secuencias genómicas asociadas con una ruta bioquímica de señalización mediante un ensayo la diferencia en los niveles de ARNm de los genes correspondientes entre la célula modelo de estudio y una célula de control, cuando se ponen en contacto con un agente candidato. Como alternativa, se determina la expresión diferencial de las secuencias asociadas con una ruta bioquímica de señalización detectando una diferencia en el nivel del polipéptido codificado o producto genético.

Para determinar mediante un ensayo la alteración inducida por un agente en el nivel de los transcritos de ARNm o polinucleótidos correspondientes, se extrae en primer lugar el ácido nucleico contenido en una muestra de acuerdo con métodos habituales en la técnica. Por ejemplo, puede aislarse ARNm usando varias enzimas líticas o soluciones químicas de acuerdo con los procedimientos expuestos en Sambrook y col. (1989) o extraerse mediante resinas de unión a ácido nucleico siguiendo las instrucciones adjuntas provistas por los fabricantes. El ARNm contenido en la muestra de ácido nucleico extraída se detecta a continuación mediante procedimientos de amplificación o ensayos de hibridación convencionales (por ejemplo, un análisis mediante la técnica de Northern) de acuerdo con métodos ampliamente conocidos en la técnica o que se basan en los métodos ejemplificados en la presente.

A efectos de esta invención, el término amplificación se refiere a cualquier método que emplea un cebador y una polimerasa capaces de replicar una secuencia diana con una fidelidad razonable. La amplificación se puede llevar a cabo mediante polimerasas de ADN naturales o recombinantes tales como TaqGold™, ADN polimerasa T7, fragmento Klenow de la ADN polimerasa de *E. coli* y transcriptasa inversa. Un método de amplificación preferido es la PCR. En particular, el ARN aislado se puede someter a un ensayo de transcripción inversa que esté acoplado con una reacción en cadena de la polimerasa cuantitativa (RT-PCR) con el fin de cuantificar el nivel de expresión de una secuencia asociada con una ruta bioquímica de señalización.

La detección del nivel de expresión génica se puede realizar en tiempo real en un ensayo de amplificación. En un aspecto, los productos amplificados se pueden visualizar directamente con agentes de unión a ADN fluorescentes que

incluyen, a título enunciativo no taxativo, intercalantes de ADN y agentes de unión al surco del ADN. Debido a que la cantidad de intercalantes incorporada en las moléculas de ADN bicatenario es normalmente proporcional a la cantidad de los productos de ADN amplificados, se puede determinar convenientemente la cantidad de los productos amplificados cuantificando la fluorescencia del indicador intercalado utilizando sistemas ópticos convencionales de la técnica. El indicador de unión al ADN adecuado para esta aplicación incluye verde SYBR, azul SYBR, DAPI, yoduro de propidio, Hoeste, oro SYBR, bromuro de etidio, acridinas, proflavina, naranja de acridina, acriflavina, fluorocumanina, elipticina, daunomicina, cloroquina, distamicina D, cromomicina, homidio, mitramicina, polipiridilos de rutenio, antramicina y similares.

En otro aspecto, se pueden emplear otras marcas fluorescentes tales como sondas específicas de la secuencia en la reacción de amplificación para facilitar la detección y cuantificación de los productos amplificados. La amplificación cuantitativa con sonda se basa en la detección específica de la secuencia de un producto amplificado deseado. Utiliza sondas específicas de la diana fluorescentes (por ejemplo, sondas TaqMan®), lo que da como resultado un aumento de la especificidad y sensibilidad. Los métodos para realizar una amplificación cuantitativa con sonda están muy consolidados en la técnica y se exponen en la patente de los EE.UU. con n.º 5.210.015.

En otro aspecto más, se pueden realizar ensayos de hibridación convencionales que utilizan sondas de hibridación que comparten una homología secuencial con secuencias asociadas con una ruta bioquímica de señalización. Normalmente, se permite que las sondas formen complejos estables con las secuencias asociadas con una ruta bioquímica de señalización contenida dentro de la muestra biológica obtenida a partir del sujeto de estudio en una reacción de hibridación. El experto en la técnica comprenderá que cuando el antisentido se utiliza como el ácido nucleico sonda, los polinucleótidos diana que se proveen en la muestra se escogen para que sean complementarios con las secuencias de los ácidos nucleicos antisentido. Por el contrario, cuando la sonda de nucleótidos es un ácido nucleico sentido, el polinucleótido diana se selecciona para que sea complementario a las secuencias de ácido nucleico sentido.

La hibridación se puede realizar en condiciones de rigurosidad variable. Las condiciones de hibridación adecuadas para llevar a la práctica la presente invención son tales que la interacción de reconocimiento entre la sonda y las secuencias asociadas con una ruta bioquímica de señalización es tanto lo suficientemente específica como lo suficientemente estable. Las condiciones que incrementan la rigurosidad de una reacción de hibridación son ampliamente conocidas y se han publicado en la técnica. Véase, por ejemplo, (Sambrook, y col., (1989); Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, segunda edición). El ensayo de hibridación se puede elaborar utilizando sondas inmovilizadas en cualquier soporte sólido que incluye, a título enunciativo no taxativo, nitrocelulosa, vidrio, silicio y diversas matrices génicas. Un ensayo de hibridación preferido se lleva a cabo en genochips de densidad elevada tal y como se describe en la patente de Estados Unidos n.º 5.445.934.

Para una detección conveniente de los complejos sonda-diana formados durante el ensayo de hibridación, se conjugan las sondas de nucleótidos con una marca detectable. Las marcas detectables adecuadas para su uso en la presente invención incluyen cualquier composición detectable por medios fotoquímicos, bioquímicos, espectroscópicos, inmunoquímicos, eléctricos, ópticos o químicos. En la técnica existe constancia de una gran variedad de marcas detectables apropiadas, que incluyen marcas fluorescentes o quimioluminiscentes, marcas de isótopos radioactivos, enzimáticas o con otros ligandos. En las formas de realización preferidas, probablemente se deseará emplear una marca fluorescente o un identificador enzimático tal como digoxigenina, β-galactosidasa, ureasa, fosfatasa alcalina o peroxidasa, complejo avidina/biotina.

Los métodos de detección utilizados para detectar o cuantificar la intensidad de hibridación dependerán normalmente de la marca seleccionada anteriormente. Por ejemplo, las radiomarcas se podrán detectar utilizando una película fotográfica o un lector de luminiscencia fotoestimulada (*phosphoimager*). Se podrán detectar y cuantificar los marcadores fluorescentes utilizando un fotodetector para detectar la luz emitida. Las marcas enzimáticas se detectan normalmente proveyendo a la enzima un sustrato y midiendo el producto de la reacción producido por la acción de la enzima sobre el sustrato; y finalmente las marcas colorimétricas se detectan simplemente visualizando la marca coloreada.

Un cambio inducido por un agente en la expresión de secuencias asociadas con una ruta bioquímica de señalización también se puede determinar examinando los productos genéticos correspondientes. La determinación del nivel proteico normalmente conlleva a) poner en contacto la proteína contenida en la muestra biológica con un agente que se una específicamente a una proteína asociada con una ruta bioquímica de señalización; y (b) identificar cualquier complejo agente:proteína formado de esta manera. En un aspecto de esta realización, el agente que se une específicamente a la proteína asociada con la ruta bioquímica de señalización es un anticuerpo, preferiblemente un anticuerpo monoclonal.

La reacción se lleva a cabo poniendo en contacto el agente con una muestra de las proteínas asociadas con una ruta bioquímica de señalización obtenida a partir de las muestras de prueba en condiciones que permitan que se forme un complejo entre el agente y las proteínas asociadas con una ruta de bioquímica de señalización. La formación del complejo se puede detectar directa o indirectamente de acuerdo con procedimientos habituales en la técnica. En el método de detección directo, se suministran los agentes con una marca detectable y los agentes que no hayan reaccionado se podrán separar del complejo; y de esta manera la cantidad de marca restante indica la cantidad de

complejo formado. Para un método de este tipo, es preferible seleccionar marcas que permanezcan unidas a los agentes incluso durante condiciones de lavado rigurosas. Es preferible que la marca no interfiera con la reacción de unión. Como alternativa, un procedimiento de detección indirecto podrá utilizar un agente que contenga una marca introducida química o enzimáticamente. Una marca deseable generalmente no interfiere con la unión o la estabilidad del complejo agente:polipéptido resultante. Sin embargo, la marca normalmente se diseña para que sea accesible a un anticuerpo para una unión eficaz y generar así pues una señal detectable.

En la técnica existe constancia de una amplia variedad de marcas adecuadas para detectar niveles de proteínas. Los ejemplos no limitantes incluyen radioisótopos, enzimas, metales coloidales, compuestos fluorescentes, compuestos bioluminiscentes y compuestos quimioluminiscentes.

La cantidad de complejos agente:polipéptido formados durante la reacción de unión se puede cuantificar mediante ensayos cuantitativos habituales. Tal y como se ha ilustrado anteriormente, se puede medir la formación de un complejo agente:polipéptido directamente mediante la cantidad de marca que permanece en el sitio de unión. Como alternativa, se estudia la proteína asociada con una ruta bioquímica de señalización para determinar su capacidad para competir con un análogo marcado por los sitios de unión en el agente específico. En este ensayo competitivo, la cantidad de marca capturada es inversamente proporcional a la cantidad de secuencias proteicas asociadas con una ruta bioquímica de señalización presentes en una muestra de prueba.

En la técnica se dispone de varias técnicas para el análisis proteico basadas en los principios generales presentados de manera esquemática anteriormente. Estos incluyen, a título enunciativo no taxativo, radioinmunoensayos, ELISA (ensayos inmunorradiométricos enzimáticos), inmunoensayos de tipo “sándwich”, ensayos inmunorradiométricos, inmunoensayos in situ (que utilizan, por ejemplo, oro coloidal, marcas con radioisótopos o enzimas), análisis por inmunoelectrotransferencia, ensayos de inmunoprecipitación, ensayos inmunofluorescentes y SDS-PAGE.

Los anticuerpos que reconocen o se unen específicamente a proteínas asociadas con una ruta bioquímica de señalización son preferibles para llevar a cabo los análisis proteicos mencionados anteriormente. Cuando se desee, se podrán utilizar anticuerpos que reconozcan un tipo específico de modificaciones postraduccionales (por ejemplo, modificaciones inducibles en la ruta bioquímica de señalización). Las modificaciones postraduccionales incluyen, a título enunciativo no taxativo, glicosilación, lipidación, acetilación y fosforilación. Estos anticuerpos se podrán adquirir de proveedores comerciales. Por ejemplo, los anticuerpos anti-fosfotirosina que reconocen específicamente proteínas fosforiladas en la tirosina se pueden adquirir de varios proveedores que incluyen Invitrogen y Perkin Elmer. Los anticuerpos anti-fosfotirosina son particularmente útiles para detectar proteínas que están fosforiladas de manera diferencial en sus residuos de tirosina como respuesta al estrés del RE. Tales proteínas incluyen, a título enunciativo no taxativo, el factor 2 alfa de la iniciación de la traducción eucariota (eIF-2α). Como alternativa, se pueden generar estos anticuerpos utilizando tecnologías convencionales con anticuerpos monoclonales o policlonales inmunizando un animal huésped o una célula productora de anticuerpos con una proteína diana que exhiba la modificación postraduccional deseada.

Al llevar a la práctica el método en cuestión, puede ser deseable distinguir el patrón de expresión de una proteína asociada con una ruta bioquímica de señalización en diferentes tejidos corporales, en diferentes tipos celulares y/o en diferentes estructuras subcelulares. Estos estudios se pueden realizar con la utilización de anticuerpos específicos del tejido, específicos de la célula o específicos de la estructura subcelular que sean capaces de unirse a marcadores proteicos que se expresen de manera preferencial en ciertos tejidos, tipos celulares o estructuras subcelulares.

Una expresión alterada de un gen asociado con una ruta bioquímica de señalización también se puede determinar examinando un cambio en la actividad del producto genético respecto a una célula de control. El ensayo para determinar un cambio inducido por un agente en la actividad de una proteína asociada con una ruta bioquímica de señalización dependerá de la actividad biológica y/o la ruta de transducción de la señal que se esté estudiando. Por ejemplo, cuando la proteína es una quinasa, se puede determinar un cambio en su capacidad para fosforilar el sustrato o los sustratos posteriores mediante varios ensayos conocidos en la técnica. Los ensayos representativos incluyen, a título enunciativo no taxativo, inmunotransferencia e inmunoprecipitación con anticuerpos tales como anticuerpos anti-fosfotirosina que reconocen las proteínas fosforiladas. Además, se puede detectar la actividad quinasa mediante ensayos quimioluminiscentes ultrarrápidos tales como el ensayo AlphaScreen™ (se puede adquirir de Perkin Elmer) y eTag™ (Chan-Hui, y col. (2003) Clinical Immunology 111: 162-174).

Cuando la proteína asociada con una ruta bioquímica de señalización es parte de una cascada de señalización que conlleva una fluctuación de estado del pH intracelular, las moléculas sensibles al pH tales como indicadores de pH fluorescentes se pueden utilizar como las moléculas indicadoras. En otro ejemplo, cuando la proteína asociada con una ruta bioquímica de señalización es un canal iónico, se pueden monitorizar las fluctuaciones en el potencial de membrana y/o concentración iónica intracelular. Varios conjuntos de elementos comerciales y dispositivos ultrarrápidos son particularmente adecuados para una selección rápida y robusta que detecte moduladores de canales iónicos. Los instrumentos representativos incluyen FLIPRTM (Molecular Devices, Inc.) y VIPR (Aurora Biosciences). Estos instrumentos son capaces de detectar reacciones simultáneamente en más de 1000 pocillos de muestra de una microplaca y de proveer medidas en tiempo real y datos funcionales en un segundo o incluso un milisegundo.

Al llevar a la práctica cualquiera de los métodos divulgados en la presente, se puede introducir un vector adecuado en una célula o un embrión no humano mediante uno o más métodos conocidos en la técnica incluidos, sin limitación,

microinyección, electroporación, sonoporación, biolística, transfección mediada por fosfato de calcio, transfección catiónica, transfección con liposomas, transfección con dendrímeros, transfección por choque térmico, transfección por nucleofección, magnetofección, lipofección, impalafección, transfección óptica, captación de ácidos nucleicos potenciada mediante un agente exclusivo y suministro mediante liposomas, inmunoliposomas, virosomas o viriones artificiales. En algunos métodos, se introduce el vector en un embrión no humano por microinyección. El vector o los vectores se podrán microinyectar en el núcleo o en el citoplasma del embrión no humano. En algunos métodos, el vector o los vectores se podrán introducir en una célula por nucleofección.

El polinucleótido diana de un complejo CRISPR puede ser cualquier polinucleótido endógeno o exógeno respecto a la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que resida en el núcleo de una célula eucariota. El polinucleótido diana puede ser una secuencia que codifique un producto genético (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, un polinucleótido regulador o ADN no codificante).

Los ejemplos de polinucleótidos diana incluyen una secuencia asociada con una ruta bioquímica de señalización, por ejemplo, un gen o un polinucleótido asociado con una ruta bioquímica de señalización. Los ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado con una enfermedad. Un gen o polinucleótido "asociados con una enfermedad" se refiere a cualquier gen o polinucleótido que genere productos de transcripción o traducción con un nivel anómalo o en una forma anómala en células que se obtienen a partir de tejidos afectados por una enfermedad en comparación con tejidos o células de un control libre de la enfermedad. Podrá ser un gen que se exprese con un nivel anómalamente alto; podrá ser un gen que se exprese con un nivel anómalamente bajo, donde la expresión alterada se correlacione con la presencia y/o evolución de la enfermedad. Un gen asociado con una enfermedad también se refiere a un gen que posee una o más mutaciones o una variación genética que es responsable directamente de un gen o genes que son responsables por la etiología de una enfermedad o que está en un desequilibro de ligamiento con ellos. Los productos transcritos o traducidos podrán ser conocidos o desconocidos y podrán tener un nivel normal o anómalo.

El polinucleótido diana de un complejo CRISPR puede ser cualquier polinucleótido endógeno o exógeno respecto a la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que resida en el núcleo de una célula eucariota. El polinucleótido diana puede ser una secuencia que codifique un producto genético (por ejemplo, una proteína) o una secuencia no codificante (por ejemplo, un polinucleótido regulador o ADN no codificante). Sin querer ceñirse a ninguna teoría, se cree que la secuencia diana debería estar asociada con un PAM (motivo adyacente a un protoespaciador); es decir, una secuencia corta reconocida por el complejo CRISPR. Los requisitos de secuencia y longitud precisos para el PAM difieren dependiendo de la enzima CRISPR utilizada, pero los PAM tienen normalmente secuencias de 2-5 pares de bases adyacentes a un protoseparador (es decir, la secuencia diana). Se proveen ejemplos de las secuencias PAM en la sección de ejemplos siguiente y el experto en la técnica será capaz de identificar secuencias PAM adicionales para su uso con una enzima CRISPR concreta. Además, el diseño del dominio de interacción con PAM (PI) puede permitir la programación de la especificidad de PAM, mejorar la fidelidad de reconocimiento del sitio diana, y aumentar la versatilidad de la plataforma de diseño de genoma de Cas, por ejemplo, Cas9. Las proteínas Cas, tales como las proteínas Cas9 pueden diseñarse para alterar su especificidad por PAM, por ejemplo, como se describe en Kleinstiver BP y col. *Engineered CRISPR-Cas9 nucleases with altered PAM specificities*. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592.

El polinucleótido diana de un complejo CRISPR puede incluir un número de genes y polinucleótidos asociados con enfermedades, así como genes y polinucleótidos asociados con vías bioquímicas de señalización como se lista en las solicitudes de patentes provisionales de los EE.UU. 61/736.527 y 61/748.427 como se hace referencia en general en BI-2011/008/WSGR n.º de expediente. 44063-701.101 y BI-2011/008/WSGR n.º de expediente 44063-701.102 respectivamente con el título SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION presentada el 12 de diciembre de 2012 y el 2 de enero de 2013, respectivamente, y la solicitud PCT PCT/US2013/074667, con el título DELIVERY, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION AND THERAPEUTIC APPLICATIONS, presentada el 12 de diciembre de 2013.

Los ejemplos de polinucleótidos diana incluyen una secuencia asociada con una ruta bioquímica de señalización, por ejemplo, un gen o un polinucleótido asociado con una ruta bioquímica de señalización. Los ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado con una enfermedad. Un gen o polinucleótido "asociados con una enfermedad" se refiere a cualquier gen o polinucleótido que genere productos de transcripción o traducción con un nivel anómalo o en una forma anómala en células que se obtienen a partir de tejidos afectados por una enfermedad en comparación con tejidos o células de un control libre de la enfermedad. Podrá ser un gen que se exprese con un nivel anómalamente alto; podrá ser un gen que se exprese con un nivel anómalamente bajo, donde la expresión alterada se correlacione con la presencia y/o evolución de la enfermedad. Un gen asociado con una enfermedad también se refiere a un gen que posee una o más mutaciones o una variación genética que es responsable directamente de un gen o genes que son responsables por la etiología de una enfermedad o que está en un desequilibro de ligamiento con ellos. Los productos transcritos o traducidos podrán ser conocidos o desconocidos y podrán tener un nivel normal o anómalo.

Selección de noqueo amplio del genoma

Las proteínas y sistemas de CRISPR descritos en la presente pueden usarse para realizar selecciones genómicas funcionales eficientes y eficaces en costos. Dichas selecciones pueden utilizar bibliotecas amplias del genoma en base a la proteína efectora CRISPR. Dichas selecciones y bibliotecas pueden proveerse para determinar la función de genes, qué genes de vías celulares están involucrados, y cómo cualquier alteración en la expresión génica puede producir un proceso biológico particular. Una ventaja de los métodos de la presente es que el sistema CRISPR evita la unión fuera del diana y los consiguientes efectos secundarios. Esto se logra utilizando sistemas dispuestos para tener un grado elevado de especificidad secuencial por el ADN diana. En formas de realización preferidas de la invención, los complejos de proteína efectora CRISPR son complejos proteína efectora Cpf1.

Como se describe en la presente, una biblioteca amplia del genoma puede comprender una pluralidad de ARN guías de Cpf1, como se describe en la presente, que comprenden secuencia guías que tienen la capacidad de dirigir una pluralidad de secuencias diana en una pluralidad de *loci* genómicos en una población de células eucarióticas. La población de células puede ser una población de células madre embrionarias (ES). La secuencia diana en el *locus* genómico puede ser una secuencia no codificante. La secuencia no codificante puede ser un intrón, secuencia regulatoria, sitio de corte y empalme, 3' UTR, 5' UTR, o señal de poliadenilación. La función génica de uno o más productos génicos puede alterarse por el mencionado direccionamiento. El direccionamiento puede producir un noqueo de la función del gen. El direccionamiento de un producto génico puede comprender más de un ARN guía. Un producto génico puede ser dirigido por 2, 3, 4, 5, 6, 7, 8, 9, o 10 ARN guías, preferentemente 3 a 4 por gen. Las modificaciones fuera de diana pueden minimizarse por explotación de cortes de hebra doble escalonados generados por los complejos proteína efectora Cpf1 o por utilización de métodos análogos a aquellos usados en los sistemas CRISPR-Cas9 (véase, por ejemplo, DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013)). El direccionamiento puede ser de aproximadamente 100 o más secuencias. El direccionamiento puede ser de aproximadamente 1000 o más secuencias. El direccionamiento puede ser de aproximadamente 20.000 o más secuencias. El direccionamiento puede ser del genoma entero. El direccionamiento puede ser de un panel de secuencias diana enfocado en una vía relevante o deseada. El direccionamiento puede ser una vía inmunológica. El direccionamiento puede ser una vía de división celular.

Como se describe, pero no se reivindica específicamente en la presente, una biblioteca amplia del genoma que puede comprender una pluralidad de ARN guías de Cpf1 que puede comprender secuencias guías que tienen la capacidad de dirigir una pluralidad de secuencias diana en una pluralidad de *loci* genómicos, en donde dicho direccionamiento resulta en un noqueo/knockdown de la función génica. Esta biblioteca puede comprender potencialmente ARN guías que dirigen a cada uno de los genes en el genoma de un organismo.

Con respecto a las bibliotecas descritas, el organismo o sujeto es una eucariota (que incluye mamíferos que incluyen seres humanos) o una eucariota que no es de ser humano o un animal no humano o un mamífero no humano. En algunas formas de realización, el organismo o sujeto es un animal no humano y podrá ser un artrópodo, por ejemplo, un insecto o podrá ser un nematodo. En algunos métodos de la invención, el organismo o sujeto es una planta. En algunos métodos de la invención, el organismo o sujeto es un mamífero o un mamífero no humano. Un mamífero no humano podrá ser, por ejemplo, un roedor (preferiblemente un ratón o una rata), un ungulado o un primate. En algunos métodos, el organismo o sujeto es un alga, que incluye microalgas, o es un hongo.

El noqueo/knockdown de la función génica puede comprender: introducir en cada célula en la población de células un sistema vector de uno o más vectores que comprenden un sistema de proteína efectora Cpf1 diseñado, de origen no natural que comprende I. una proteína efectora Cpf1, y II. uno o más ARN guías, en donde los componentes I y II pueden estar en el mismo o diferentes vectores del sistema, integrando los componentes I y II en cada célula, en donde la secuencia guía direcciona un gen único en cada célula, en donde la proteína efectora Cpf1 está operativamente ligada a un elemento regulatorio, en donde cuando se transcribe, el ARN guía que comprende la secuencia guía dirige la unión específica de secuencia del sistema de proteína efectora Cpf1 a una secuencia diana que se corresponde con los loci genómicos del gen único, induciendo el clivaje de los *loci* genómicos por la proteína efectora Cpf1, y confirmando diferentes mutaciones de noqueo/knockdown en una pluralidad de genes únicos en cada célula de la población de células generando de esta manera una biblioteca celular de noqueo/knockdown de genes. La solicitud también proporciona que la población de células es una población de células eucarióticas, y en una forma de realización preferida, la población de células es una población de células madre embriónicas (ES).

El uno o más vectores pueden ser vectores plasmídicos. The El vector puede ser un vector individual que comprende una proteína efectora Cpf1, un ARNgs, y opcionalmente, un marcador de selección en las células diana. Sin estar ligado a la teoría, la capacidad de suministrar simultáneamente una proteína efectora Cpf1 y ARNgs mediante un vector individual permite la aplicación a cualquier tipo de célula de interés, sin la necesidad de generar primero las líneas celulares que expresen la proteína efectora Cpf1. El elemento regulatorio puede ser un promotor inducible. El promotor inducible puede ser un promotor inducible por doxiciclina. En algunos métodos de la invención, la expresión de la secuencia guía está controlada por el promotor T7 y está impulsada por la expresión de la polimerasa T7. La confirmación de diferentes mutaciones de noqueo/knockdown puede ser por secuenciamiento del exoma completo. La mutación de noqueo/knockdown puede obtenerse en 100 o más genes únicos. La mutación de noqueo/knockdown puede obtenerse en 1000 o más genes únicos. La mutación de noqueo/knockdown puede obtenerse en 20.000 o más genes únicos. La mutación de noqueo/knockdown puede obtenerse en el genoma completo. El noqueo/knockdown de

la función génica puede lograrse en una pluralidad de genes únicos que funcionan en una vía o condición fisiológica particular. La vía o condición puede ser una vía o condición inmunológica. La vía o condición puede ser una vía o condición de división celular.

También se describe, pero no se reivindica específicamente en la presente, conjuntos de elementos que comprenden las bibliotecas amplias del genoma mencionadas en la presente. El conjunto de elementos puede comprender un recipiente único que comprende vectores o plásmidos que comprenden la biblioteca descrita en la presente. El conjunto de elementos también puede comprender un panel que comprende una selección de ARN guías de sistema de proteína efectora Cpf1 únicas que comprenden secuencia guías de la biblioteca descrita en la presente, en donde la selección es indicativa de una condición fisiológica particular. La solicitud comprende que el direccionamiento es de aproximadamente 100 o más secuencias, aproximadamente 1000 o más secuencias o aproximadamente 20.000 o más secuencias o el genoma completo. Adicionalmente, un panel de secuencias diana puede enfocarse en una vía relevante o deseable, tal como una vía inmunológica o división celular.

En un aspecto adicional de los métodos y composiciones reivindicados de la invención, la proteína efectora Cpf1 puede comprender una o más mutaciones y puede usarse como una proteína de unión a ADN genérica con o sin fusión a un dominio funcional. Las mutaciones podrán ser mutaciones introducidas artificialmente o mutaciones de ganancia o pérdida funcional. Las mutaciones han sido caracterizadas como se describe en la presente. En un aspecto de la invención, el dominio funcional puede ser un dominio de activación transcripcional, el cual puede ser VP64. En otros aspectos de la invención, el dominio funcional puede ser un dominio represor transcripcional, que podrá ser KRAB o SID4X. Otros aspectos de la invención se refieren a la proteína efectora Cpf1 mutada fusionada a dominios que incluyen, a título enunciativo no taxativo, un activador de la transcripción, represor, una recombinasa, una transposasa, un remodelador de histonas, una desmetilasa, una ADN metil-transferasa, un criptocromo, un dominio inducible/controlable lumínicamente o un dominio inducible/controlable químicamente. Algunos métodos de la invención podrán incluir inducir la expresión de los genes diana. En una forma de realización, la inducción de la expresión por direccionamiento a una pluralidad de secuencias diana en una pluralidad de *loci* genómicos en una población de células eucarióticas se hace con el uso de un dominio funcional.

En la práctica de la presente invención que utiliza complejos de la proteína efectora Cpf1 son útiles los métodos usados en los sistemas CRISPR-Cas9 y se hace referencia a:

Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013). [Epub previo a la impresión]; Publicado en su forma editada final como: Science. 2014 Jan 3; 343(6166): 84–87.

Shalem y col. han descrito una nueva manera de interrogar la función génica en una escala genómica amplia. Sus estudios mostraron que el suministro de una biblioteca de CRISPR-Cas9 inactivado génicamente a escala genómica (GeCKO) dirigida a 18,080 genes con 64,751 secuencias guía únicas permitió la selección que comprende una selección negativa y positiva en células humanas. En primer lugar, los autores han mostrado el uso de la biblioteca GeCKO para identificar genes esenciales para la viabilidad celular en células cancerosas y citoblastos pluripotentes. En primer lugar, en un modelo de melanoma, los autores seleccionaron los genes cuya pérdida está implicada en la resistencia a vemurafenib, un agente terapéutico que inhibe la proteína quinasa mutante BRAF. Sus estudios han mostrado que los candidatos mejor clasificados incluyeron los genes anteriormente validados NF1 y MED12 así como los novedosos blancos NF2, CUL3, TADA2B y TADA1. Los autores observaron un alto nivel de consistencia entre los ARN guías independientes que dirigían el mismo gen y una elevada tasa de confirmación de blancos, y de esta manera demostraron lo prometedor de la selección con Cas9 a escala genómica.

También se hace referencia a la publicación de patente de los EE.UU. número US20140357530; y Publicación de Patente PCT WO2014093701. También se hace referencia a la Publicación de Prensa de NIH de Oct. 22, 2015 con el título, “Researchers identify potential alternative to CRISPR-Cas genome editing tools: New Cas enzymes shed light on evolution of CRISPR-Cas systems”.

Alteración funcional y selección

En otro aspecto, los métodos y composiciones reivindicados de la presente invención pueden incluir la evaluación funcional y selección de genes. El uso del sistema CRISPR de la presente invención para suministrar precisamente dominios funcionales, para activar o reprimir genes o para alterar el estado epigenético por alteración precisa del sitio de metilación en un *locus* específico de interés, puede ser con uno o más ARN guías aplicado a una célula única o población de células o con una biblioteca aplicada al genoma en un conjunto de células *ex vivo* o *in vivo* que comprende la administración o expresión de una biblioteca que comprende una pluralidad de ARN guías (ARNgs) y en donde la selección además comprende el uso de una proteína efectora Cpf1, en donde el complejo CRISPR que comprende la proteína efectora Cpf1 es modificado para que comprenda un dominio funcional heterólogo. Se describe, pero no se reivindica en la presente, un método para explorar un genoma que comprende la administración a un huésped o la expresión en un huésped in vivo de una biblioteca. En un aspecto, la invención provee un método tal como se describe en la presente que comprende además un activador administrado al huésped o expresado en el huésped. En un aspecto la invención provee un método como se describe en la presente en donde el activador está unido a una proteína efectora Cpf1. En un aspecto, la invención provee un método tal como se discute en el presente documento en donde el activador se acopla al extremo N-terminal o al extremo C-terminal de la proteína efectora Cpf1. En un

aspecto, la invención provee un método tal como se divulga en la presente en donde el activador está acoplado a un bucle de ARNsg. En un aspecto, la invención provee un método tal como se describe en la presente que comprende además un represor administrado al huésped o expresado en el huésped. En un aspecto, la invención provee un método tal como se describe en el presente documento, en donde la exploración comprende afectar a y detectar la activación génica, la inhibición génica o el clivaje en el locus.

En un aspecto, la invención provee una actividad eficaz en la diana y minimiza la actividad fuera de la diana. En un aspecto, la invención provee el clivaje eficaz en diana por la proteína efectora Cpf1 y minimiza el civaje fuera de diana por la proteína efectora Cpf1. En un aspecto, la invención provee la unión específica de guía de la proteína efectora Cpf1 en el locus genético sin clivaje de ADN. Por consiguiente, en un aspecto, la invención provee regulación genética específica de diana. En un aspecto, la invención provee la union específica de guía de la proteína efectora Cpf1 en el locus genético sin clivaje de ADN. En consecuencia, en un aspecto, la invención provee el clivaje en un locus genético y regulación genética en un locus genético diferente usando una proteína efectora Cpf1 individual. En un aspecto, la invención provee activación y/o inhibición y/o clivaje ortogonal de múltiples blancos usando una o más de una proteína efectora y/o enzima Cpf1.

En un aspecto, la invención provee un método tal como se describe en el presente documento, en donde el huésped es una célula eucariota. En un aspecto, la invención provee un método tal como se describe en el presente documento, en donde el huésped es una célula de mamífero. En un aspecto, la invención provee un método tal como se describe en el presente documento, en donde el huésped es una célula eucariota no humana. En un aspecto la invención provee un método como se describe en la presente, en donde el eucariota no humano es un mamifero no humano. En un aspecto la invención provee un método como se describe en la presente, en donde el mamifero no humano es un ratón. Un aspecto de la invención provee un método como se describe en la presente que comprende la administración de complejos de la proteína efectora Cpf1 o uno o más componentes de la misma o una o más moléculas de ácido nucleico que codifican para la misma, en donde la o las mencionadas moléculas de ácido nucleico están ligadas operativamente a una o más secuencias regultorias y expresadas. En un aspecto, la invención provee un método tal como se describe en la presente en donde la expresión es mediante un lentivirus, un adenovirus o un AAV. En un aspecto, la invención provee un método tal como se describe en la presente en donde el suministro es mediante una partícula, una nanopartícula, un lípido o péptido que penetra en una célula (CPP).

En un aspecto, los métodos y composiciones reivindicados de la invención hacen uso de un par de complejos CRISPR, comprendiendo cada uno una proteína efectora Cpf1, comprendiendo cada uno un ARN guía (ARNsg) que comprende una secuencia guía con capacidad de hibridar con una secuencia diana en un locus genómico de interés en una célula, en donde por lo menos un bucle de cada ARNsg está modificado mediante la inserción de distintas secuencias de ARN que se unen a una o más proteínas adaptadoras y en donde la proteína adaptadora se asocia con uno o más dominios funcionales, en donde cada ARNsg de cada complejo de proteína efectora Cpf1 comprende un dominio funcional que tiene una actividad de clivaje de ADN. En un aspecto, la invención provee complejos de la proteína efectora Cpf1 apareados tal como se describe en el presente documento, en donde la actividad de clivaje de ADN se debe a una nucleasa Fok1.

También se describe, pero no se reivindica especialmente en la presente, un método para cortar una secuencia diana en un locus genómico de interés que comprende el suministro a una célula de los complejos de la proteína efectora Cpf1 o componentes del mismo o moléculas de ácido nucleico que codifican a los mismos, en donde dichas moléculas de ácido nucleico están ligadas operativamente a secuencias reguladoras y se expresan in vivo. En un aspecto, la invención provee un método tal como se describe en la presente en donde el suministro es mediante un lentivirus, un adenovirus o un AAV. En un aspecto, la invención provee un método tal como se describe en la presente o complejos de proteína efectora Cpf1 apareados tal como se describe en la presente en donde la secuencia diana para un primer complejo del par se encuentra en una primera hebra de ADN bicatenario y la secuencia diana para un segundo complejo del par se encuentra en una segunda hebra de ADN bicatenario. En un aspecto, la invención provee un método tal como se describe en la presente o complejos de proteína efectora Cpf1 apareados tal como se describe en la presente en donde las secuencias diana de los complejos primero y segundo se encuentran próximas entre sí de tal forma que el ADN se corta de un modo tal que facilita la reparación dirigida por homología. En un aspecto, un método del presente documento puede incluir además introducir en una célula ADN molde. En un aspecto un método de la presente o complejos de la proteína efectora Cpf1 apareada de la presente pueden involucrar que cada complejo de proteína efectora Cpf1 tiene una enzima efectora Cpf1 que está mutada tal que no tiene más de aproximadamente 5% de la actividad nucleasa de la enzima efectora Cpf1 que no está mutada.

También se describe, pero no se reivindica en la presente, una biblioteca, método o complejo tal como se discuten en la presente en donde el ARNsg se modifica para tener por lo menos un bucle funcional no codificante, por ejemplo, en donde el por lo menos un bucle funcional no codificante es represor; por ejemplo, en donde el por lo menos un bucle funcional no codificante comprende Alu.

En un aspecto, el método reivindicado de la invención comprende alterar o modificar la expresión de un producto genético. El mencionado método puede comprender introducir una célula que contiene y expresa una molécula de ADN que codifica para el producto genético un sistema CRISPR diseñado, de origen no natural que comprende una proteína efectora Cpf1 y ARN guía dirigido contra la molécula de ADN, por el cual el ARN guía está dirigido a la molécula de ADN que codifica para el producto genético y la proteína efectora Cpf1 cliva la molécula de ADN que

codifica para el producto genético, por lo cual se altera la expresión del producto genético; y, en donde la proteína efectora Cpf1 y el ARN guía no se presentan juntos naturalmente. La invención comprende el ARN guía que comprende una secuencia guía ligada a una secuencia de repetición directa. La invención además comprende la proteína efectora Cpf1 que tiene optimización de codones para la expresión en una célula eucariótica. En una forma de realización preferida, la célula eucariota es una célula de mamífero y en una forma de realización más preferida, la célula de mamífero es una célula de ser humano. En una forma de realización adicional de la invención, la expresión del producto genético está reducida.

En algunas formas de realización, uno o más dominios funcionales están asociados con la proteína efectora Cpf1. En algunas formas de realización, uno o más dominios funcionales están asociados con una proteína adaptadora, por ejemplo, como se usa con las guías modificadas de Konnerman y col. (Nature 517, 583–588, 29 de enero de 2015). En algunas formas de realización, uno o más dominios funcionales están asociados con un ARNg muerto (ARNd). En algunas formas de realización, un complejo de ARNd con proteína efectora Cpf1 activa dirige la regulación genética por un dominio funcional en un locus génico mientras que un ARNg dirige el clivaje del ADN por la proteína efectora Cpf1 activa en otro locus, por ejemplo, como se describe de manera análoga en los sistemas CRISPR-Cas9 de Dahlman y col., "Orthogonal gene control with a catalytically active Cas9 nuclease" (en prensa). En algunas formas de realización, los ARNd se seleccionan para maximizar la selectividad de regulación para un locus genético de interés en comparación con la regulación fuera de diana. En algunas formas de realización, los ARNd se seleccionan para maximizar la regulación del gen diana y minimizar el clivaje del diana.

Para los propósitos de la siguiente descripción, la referencia a un dominio funcional podría ser un dominio funcional asociado con la proteína efectora Cpf1 o un dominio funcional asociado con la proteína adaptadora.

En la práctica de la invención, los bucles del ARNg pueden ser extendidos, sin colisionar con la proteína Cpf1 por inserción de uno o más bucles diferentes de ARN o una o más secuencias diferentes que pueden reclutar proteínas adaptadoras que pueden unirse a uno o más bucles distintos de ARN o una o más secuencias distintas. Las proteínas adaptadoras pueden incluir a título enunciativo no taxativo combinaciones de proteína de unión a ARN ortogonal /aptámero que existen en una diversidad de proteínas de recubrimiento de bacteriofagos. Una lista de las mencionadas proteínas de recubrimiento incluyen, a título enunciativo no taxativo: Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, φCb5, φCb8r, φCb12r, φCb23r, 7s y PRR1. Estas proteínas adaptadoras o protenías de unión a ARN ortogonal pueden reclutar adicionalmente proteínas efectoras o fusiones que comprenden uno o más dominios funcionales. En algunas formas de realización, el dominio funcional puede seleccionarse del grupo que consiste en: dominio transposasa, dominio integrasa, dominio recombinasa, dominio resolvasa, dominio invertasa, dominio proteasa, dominio ADN metiltransferasa, dominio ADN hidroxilmetilasa, dominio ADN desmetilasa, dominio histona acetilasa, dominio histona desacetilasas, dominio nucleasa, dominio represor, dominio activador, dominio de señal de localización nuclear, dominio de proteína regulatoria de la transcripción (o reclutamiento de complejo de transcripción), dominio asociado con actividad de incorporación celular, dominio de unión a ácido nucleico, dominio de presentación de anticuerpo, enzimas modificadoras de histona, reclutador de enzimas modificadoras de histonas; inhibidor de enzimas modificadoras de histonas, histona metiltransferasa, histona desmetilasa, histona quinasa, histona fosfatasa, histona ribosilasa, histona desribosilasa, histona ubiquitinasa, histona desubiquitinasa, histona biotinasa e proteasa de cola de histona. En algunas formas de realización preferidas, el dominio funcional es un dominio de activación transcripcional, tal como, a título enunciativo no taxativo, VP64, p65, MyoD1, HSF1, RTA, SET7/9 o una histona acetiltransferasa. En algunas formas de realización, el dominio funcional es un dominio de represión transcripcional, preferiblemente KRAB. En algunas formas de realización, el dominio de represión transcripcional es SID, o concatémeros de SID (por ejemplo, SID4X). En algunas formas de realización, el dominio funcional es un dominio de modificación epigenética, de modo que se provee una enzima de modificación epigenética. En algunas formas de realización, el dominio funcional es un dominio de activación, que podrá ser el dominio de activación P65.

En algunas formas de realización, el uno o más dominios funcionales es una NLS (secuencia de localización nuclear) o una NES (señal de exportación nuclear). En algunas formas de realización, el uno o más dominios funcionales es un dominio de activación transcripcional que comprende VP64, p65, MyoD1, HSF1, RTA, SET7/9 y una histona acetiltransferasa. Otras referencias en la presente a dominios de activación (o activadores) respecto a aquellos asociados con la enzima CRISPR incluyen cualquier dominio de activación transcripcional y específicamente VP64, p65, MyoD1, HSF1, RTA, SET7/9 o una histona acetiltransferasa.

En algunas formas de realización, el uno o más dominios funcionales es un dominio represor transcripcional. En algunas formas de realización, el dominio represor transcripcional es un dominio KRAB. En algunas formas de realización, el dominio represor transcripcional es un dominio NuE, dominio NcoR, dominio SID o un dominio SID4X.

En algunas formas de realización, el uno o más dominios funcionales tienen una o más actividades que comprenden actividad metilasa, actividad desmetilasa, actividad de activación transcripcional, actividad de represión de la transcripción, actividad de factor de liberación de la transcripción, actividad de modificación de histona, actividad de clivaje de ARN, actividad de clivaje de ADN, actividad de integración del ADN o actividad de unión a ácido nucleico.

Los dominios de modificación de histona también se prefieren en algunas formas de realización. Los dominios de modificación de histonas ejemplificativos se describen más adelante. Como dominios funcionales de la presente

también se prefieren dominios de transposasa, dominios de maquinaria de HR (Recombinación Homóloga), dominios de recombinasa, y/o dominios de integrasa. En algunas formas de realización, la actividad de integración de ADN incluye dominios de maquinaria de HR, dominios de integrasa, dominios de recombinasa y/o dominios de transposasa. En algunas formas de realización se prefieren histona acetiltransferasas.

En algunas formas de realización, la actividad de clivaje de ADN es debido a una nucleasa. En algunas formas de realización, la nucleasa comprende una nucleasa Fok1. Véase, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014), se relaciona con nucleasas FokI guiadas por ARN dimérico que reconocen secuencias extendidas y pueden editar genes endógenos con alta eficiencia en células humanas.

En algunas formas de realización, el uno o más dominios funcionales está unido a la proteína efectora Cpf1 tal que con la unión al ARNgs y diana, el dominio funcional se encuentra en una orientación espacial que permite que el dominio funcional funcione en su función atribuida.

En algunas formas de realización, el uno o más dominios funcionales está unido a la proteína adaptadora tal que con la unión de la proteína efectora Cpf1 al ARNgs y al diana, el dominio funcional está en una orientación espacial que permite que el dominio funcional funcione en su función atribuida.

En un aspecto la invención provee una composición como se describe en la presente en donde el uno o más dominios funcionales está unido a la proteína efectora Cpf1 o proteína adaptadora mediante un conector, opcionalmente un conector GlySer, como se describe en la presente.

La represión transcripcional endógena con frecuencia está mediada por enzimas que modifican la cromatina tales como histona metiltransferasas (HMT) y desacetilasas (HDAC). Los dominios efectores de histona represores son conocidos y más adelante se provee una lista de ejemplos. En la tabla de ejemplos, se da preferencia a proteínas y truncados funcionales de pequeño tamaño para facilitar el empaquetamiento viral eficaz (por ejemplo, con AAV). En general, sin embargo, los dominios pueden incluir inhibidores de HDAC, histona metiltransferasas (HMT), e histona acetiltransferasa (HAT), así como proteínas reclutadoras de HDAC y HMT. El dominio funcional puede ser o incluir, en algunas formas de realización, dominios efectores de HDAC, dominios efectores reclutadores de HDAC, dominios efectores de Histona Metiltransferasa (HMT), dominios efectores reclutadores de Histona Metiltransferasa (HMT), o dominios efectores inhibidores de Histona acetiltransferasa.

Dominios efectores de HDAC

Subtipo/ Complejo	Nombre	Sustrato (si es conocido)	Modificación (si es conocido)	Organismo	Tamaño completo (aa)	Truncado seleccionado (aa)	Tamaño final (aa)	Dominio catalítico
HDAC I	**HDAC8**	-	-	*X. laevis*	325	1-325	325	1-272: HDAC
HDAC I	**RPD3**	-	-	*S. cerevisiae*	433	19-340	322 (Vannier)	19-331: HDAC
HDAC IV	**MesoLo4**	-	-	*M. loti*	300	1-300 (Gregoretti)	300	-
HDAC IV	**HDAC11**	-	-	*H. sapiens*	347	1-347 (Gao)	347	14-326: HDAC
HE2	**HDT1**	-	-	*A. thaliana*	245	1-211 (Wu)	211	-
SIRT I	**SIRT3**	H3K9Ac H4K16Ac H3K56Ac	-	*H. sapiens*	399	143-399 (Scher)	257	126-382: SIRT
SIRT I	**HST2**	-	-	*C. albicans*	331	1-331 (Hnisz)	331	-
SIRT I	**CobB**	-	-	*E. coli (K12)*	242	1-242 (Landry)	242	-
SIRT I	**HST2**	-	-	*S. cerevisiae*	357	8-298 (Wilson)	291	-
SIRT III	**SIRT5**	H4K8Ac H4K16Ac	-	*H. sapiens*	310	37-310 (Gertz)	274	41-308: SIRT
SIRT III	**Sir2A**	-	-	*P. falciparum*	273	1-273 (Zhu)	273	19-273: SIRT
SIRT IV	**SIRT6**	H3K9Ac H3K56Ac	-	*H. sapiens*	355	1-289 (Tennen)	289	35-274: SIRT

En consecuencia, los dominios represores de la presente invención pueden seleccionarse de inhibidores de histona metiltransferasas (HMTs), histona desacetilasas (HDACs), histona acetiltransferasa (HAT), así como proteínas reclutadoras de HDAC y HMT.

El dominio HDAC puede ser cualquiera de aquellos en la tabla precedente, es decir: HDAC8, RPD3, MesoLo4, HDAC11, HDT1, SIRT3, HST2, CobB, HST2, SIRT5, Sir2A, o SIRT6.

En alguna forma de realización, el dominio funcional puede ser un dominio efector reclutador de HDAC. Los ejemplos preferidos incluyen aquellos de la Tabla más adelante, es decir MeCP2, MBD2b, Sin3a, NcoR, SALL1, RCOR1. NcoR se ejemplifica en los presentes Ejemplos y, a pesar que se prefiere, se prevé que otros en la clase también serán útiles.

Tabla de dominios efectores reclutadores de HDAC

Subtipo/ Complejo	Nombre	Sustrato (si es conocido)	Modificación (si es conocido)	Organismo	Tamaño completo (aa)	Truncado seleccionado (aa)	Tamaño final (aa)	Dominio catalítico
Sin3a	**MeCP2**	-	-	*R. norvegicus*	492	207-492 (Nan)	286	-
Sin3a	**MBD2b**	-	-	*H. sapiens*	262	45-262 (Boeke)	218	-
Sin3a	**Sin3a**	-	-	*H. sapiens*	1273	524-851 (Laherty)	328	627-829: interacción con HDAC1
NcoR	**NcoR**	-	-	*H. sapiens*	2440	420-488 (Zhang)	69	-
NuRD	**SALL1**	-	-	*M. musculus*	1322	1-93 (Lauberth)	93	-
CoREST	**RCOR1**	-	-	*H. sapiens*	482	81-300 (Gu, Ouyang)	220	-

En alguna forma de realización, el dominio funcional puede ser un dominio efector Metiltransferasa (HMT). Los ejemplos preferidos incluyen aquellos en la Tabla más adelante, es decir NUE, vSET, EHMT2/G9A, SUV39H1, dim-5, KYP, SUVR4, SET4, SET1, SETD8, y TgSET8. NUE se ejemplifica en los presentes Ejemplos y, a pesar que se prefiere, se prevé que otros en la clase también sean útiles.

Tabla de dominios efectores de Histona Metiltransferasa (HMT)

Subtipo/ Complejo	Nombre	Sustrato (si es conocido)	Modificación (si es conocido)	Organismo	Tamaño completo (aa)	Truncado seleccionado (aa)	Tamaño final (aa)	Dominio catalítico
SET	**NUE**	H2B, H3, H4	-	*C. trachomatis*	219	1-219 (Pennini)	219	-
SET	**vSET**	-	H3K27me3	*P. bursaria chlorella virus*	119	1-119 (Mujtaba)	119	4-112: SET2
SUV39 familia	**EHMT2/G9A**	H1.4K2, H3K9, H3K27	H3K9me1/2, H1K25me1	*M. musculus*	1263	969-1263 (Tachibana)	295	1025-1233: preSET, SET, postSET
SUV39	**SUV39H1**	-	H3K9me2/3	*H. sapiens*	412	79-412 (Snowden)	334	172-412: preSET, SET, postSET
Suvar3-9	**dim-5**	-	H3K9me3	*N. crassa*	331	1-331 (Rathert)	331	77-331: preSET, SET, postSET
Suvar3-9 (SUVH subfamilia)	**KYP**	-	H3K9me1/2	*A. thaliana*	624	336-601	267 (Jackson)	-
Suvar3-9 (SUVR subfamilia)	**SUVR4**	H3K9me1	H3K9me2/3	*A. thaliana*	492	180-492	313 (Thorstensen)	192-462: preSET, SET, postSET
Suvar4-20	**SET4**	-	H4K20me3	*C. elegans*	288	1-288 (Vielle)	288	-
SET8	**SET1**	-	H4K20me1	*C. elegans*	242	1-242 (Vielle)	242	-
SET8	**SETD8**	-	H4K20me1	*H. sapiens*	393	185-393	209 (Couture)	256-382: SET
SET8	**TgSET8**	-	H4K20me1/2/3	*T. gondii*	1893	1590-1893 (Sautel)	304	1749-1884: SET

En alguna forma de realización, el dominio funcional puede ser un dominio efector reclutador de Histona Metiltransferasa (HMT). Los ejemplos preferidos incluyen aquellos en la Tabla a continuación, es decir Hp1a, PHF19, y NIPP1.

Tabla de dominios efectores reclutadores de Histona Metiltransferasa (HMT)

Subtipo/ Complejo	Nombre	Sustrato (si es conocido)	Modificación (si es conocido)	Organismo	Tamaño completo (aa)	Truncado seleccionado (aa)	Tamaño final (aa)	Dominio catalítico
-	**Hp1a**	-	H3K9me3	*M. musculus*	191	73-191	119 (Hatmaway)	121-179: chromoshadow
-	**PHF19**	-	H3K27me3	*H. sapiens*	580	(1-250) + GGSG conector + (500-580)	335 (Ballaré)	163-250: PHD2
-	**NIPP1**	-	H3K27me3	*H. sapiens*	351	1-329 (Jin)	329	310-329: EED

En alguna forma de realización, el dominio funcional puede ser un dominio efector inhibidor de Histona acetiltransferasa. Los ejemplos preferidos incluyen SET/TAF-1β que se lista en la Tabla más adelante.

Tabla de dominios efectores inhibidores de Histona acetiltransferasa

Subtipo/ Complejo	Nombre	Sustrato (si es conocido)	Modificación (si es conocido)	Organism o	Tamaño completo (aa)	Truncado seleccionado (aa)	Tamaño final (aa)	Dominio catalítico
-	**SET/TAF-1β**	-	-	*M. musculus*	289	1-289 (Cervoni)	289	-

También se prefiere hacer diana contra elementos de control (regulatorios) endógenos (tales como promotores y silenciadores) adicionalmente a un promotor o elementos proximales al promotor. Por lo tanto, la invención también puede usarse para hacer diana en elementos de control endógeno (que incluyen potenciadores y silenciadores) adicionalmente a hacer diana en el promotor. Estos elementos de control pueden localizarse corriente arriba y corriente abajo del sitio de inicio de la transcripción (TSS), comenzando desde 200 pb del TSS hasta más allá de 100kb. El direccionamiento a elementos de control conocidos puede usarse para activar o reprimir el gen de interés. En algunos casos, un elemento de control simple puede influir en la transcripción de múltiples genes diana. El direccionamiento de un elemento de control simple podría por lo tanto usarse para controlar la transcripción de genes múltiples simultáneamente.

El direccionamiento a elementos de control putativos por otra parte (por ejemplo, por revestimiento de la región del elemento de control putativo así como 200bp hasta 100kB alrededor del elemento) puede usarse como un medio para verificar dichos elementos (por medida de la transcripción del gen de interés) o para detectar elementos de control novedosos (por ejemplo por revestimiento de 100kb corriente arriba y corriente abajo del TSS del gen de interés). Adicionalmente, el direccionamiento de elementos de control putativos puede ser útil en el contexto de entender causas genéticas de enfermedades. Muchas mutaciones y variantes de SNP comunes asociados con fenotipos de enfermedades están localizados fuera de las regiones codificantes. El direccionamiento a dichas regiones con sistemas de activación o de represión descritos en la presente pueden seguirse por lectura de la transcripción de a) un conjunto de blancos putativos (por ejemplo, un conjunto de genes localizados en una proximidad cercana al elemento de control) o b) lectura del transcriptoma completo por ejemplo por ARNseq o microarreglo. Esto permitiría la identificación de genes posiblemente candidatos involucrados en el fenotipo de la enfermedad. Dichos genes candidatos podrían ser útiles como novedosos blancos farmacológicos.

Los inhibidores de histona acetiltransferasa (HAT) se mencionan en la presente. Sin embargo, una alternativa en algunas formas de realización es que el uno o más dominios funcionales comprendan una acetiltransferasa, preferentemente una histona acetiltransferasa. Estos son útiles en el campo de la epigenómica, por ejemplo en métodos para investigar el epigenoma. Los métodos para investigar el epigenoma pueden incluir, por ejemplo, hacer diana en secuencias epigenómicas. El hacer diana en secuencias epigenómicas puede incluir que la guía sea dirigida a una secuencia diana epigenómica. El diana de secuencia epigenómica puede incluir, en algunas formas de realización, una secuencia de promotor, de silenciador o de un potenciador.

El uso de un dominio funcional unido a una proteína efectora Cpf1 como se describe en la presente, preferentemente una proteína efectora Cpf1 muerta, más preferentemente una proteína efectora FnCpf1 muerta, para hacer diana en secuencias epigenómicas puede usarse para activar o reprimir promotores, silenciadores o potenciadores.

Los ejemplos de acetiltransferasas son conocidos pero pueden incluir, en algunas formas de realización, histona acetiltransferasas. En algunas formas de realización, la histona acetiltransferasa puede comprender el centro catalítico de la acetiltransferasa p300 (Gerbasch & Reddy, Nature Biotech 6th April 2015).

En algunas formas de realización preferidas, el dominio funcional está unido a una proteína efectora Cpf1 muerta a un diana y activar secuencias epigenómicas tales como promotores o potenciadores. Una o más de las guías dirigidas a dichos promotores o potenciadores también pueden proveerse para dirigirse la unión de la enzima CRISPR a dichos promotores o potenciadores.

El término "asociado con" se usa en la presente en relación a la asociación del dominio funcional a la proteína efectora Cpf1 o la proteína adaptadora. Se usa respecto a cómo una molécula se "asocia" respecto a otra, por ejemplo entre una proteína adaptadora y un dominio funcional, o entre la proteína efectora Cpf1 y un dominio funcional. En el caso de dichas interacciones proteína-proteína, esta asociación puede ser vista en términos de reconocimiento en el sentido en que un anticuerpo reconoce un epítopo. Como alternativa, una proteína puede asociarse con otra proteína mediante una fusión de las dos, por ejemplo una subunidad estando fusionada con otra subunidad. La fusión típicamente se produce por agregado de la secuencia de aminoácidos de una a la otra, por ejemplo mediante corte y empalme a la vez de las secuencias de nucleótidos que codifican para cada proteína o subunidad. Como alternativa, esto puede ser visto esencialmente como la unión entre dos moléculas o enlace directo, tal como una proteína de fusión. En cualquier evento, la proteína de fusión puede incluir un conector entre las dos subunidades de interés (es decir entre la enzima

y el dominio funcional o entre la proteína adaptadora y el dominio funcional). Por lo tanto, en algunas formas de realización, la proteína efectora Cpf1 o proteína adaptadora está asociada con un dominio funcional por unión al mismo. En otras formas de realización, la proteína efectora Cpf1 o proteína adaptadora está asociada con un dominio funcional porque los otros dos están fusionados entre sí, opcionalmente mediante un conector intermediario.

La unión de un dominio funcional o proteína de fusión puede ser a través de un conector, por ejemplo, un conector flexible glicina-serina (GlyGlyGlySer) o $(GGGS)_3$ o un conector rígido alfa-hélice tal como (Ala(GluAlaAlaAlaLys)Ala). Los conectores como $(GGGGS)_3$ se usan preferiblemente en la presente para separar dominios proteicos o peptídicos. Se prefiere $(GGGGS)_3$ porque es un conector relativamente largo (15 aminoácidos). Los residuos glicina son los más flexibles y los residuos serina potencian la posibilidad de que el conector esté en el exterior de la proteína. Como alternativas puede usarse preferiblemente $(GGGGS)_6$, $(GGGGS)_9$ o $(GGGGS)_{12}$. Otras alternativas preferidas son $(GGGGS)_1$, $(GGGGS)_2$, $(GGGGS)_4$, $(GGGGS)_5$, $(GGGGS)_7$, $(GGGGS)_8$, $(GGGGS)_{10}$, o $(GGGGS)_{11}$. Hay disponibles conectores alternativos, pero se cree que los conectores muy flexibles funcionan mejor para permitir la oportunidad máxima para que las 2 partes de la Cpf1 se junten y de esta manera se reconstituya la actividad Cpf1. Una alternativa es que la NLS de nucleoplasmina pueda usarse como un conector. Por ejemplo, también puede usarse un conector entre la Cpf1 y cualquier dominio funcional. Nuevamente, puede usarse un conector $(GGGGS)_3$ aquí (o las versiones de 6, 9, o 12 repeticiones del mismo) o la NLS de nucleoplasmina como conector entre Cpf1 y el dominio funcional.

Mutagénesis saturante

El o los sistemas de proteína efectora Cpf1 descritos en la presente pueden usarse para realizar mutagénesis de barrido saturante o profunda de loci genómicos junto con un fenotipo celular—por ejemplo, para determinar las características mínimas críticas y vulnerabilidades discretas de elementos funcionales requeridos para la expresión génica, resistencia de fármacos, y reversión de enfermedad. Por mutagénesis saturante o profunda se entiende que cada una o esencialmente cada base de ADN es cortada en los loci genómicos. Puede introducirse una biblioteca de ARN guías de la proteína efectora Cpf1 en una población de células. La biblioteca puede introducirse, tal que cada célula reciba un ARN guía simple (ARNgs). En el caso en el que la biblioteca se introduce por transducción de un vector viral, como se describe en la presente, se usa una multiplicidad de infección (MOI) baja. La biblioteca puede incluir ARNgs dirigidos a cada secuencia corriente arriba de una secuencia (motivo adyacente protoespaciador) (PAM) en un locus genómico. La biblioteca puede incluir por lo menos 100 secuencias genómicas no superpuestas corriente arriba de una secuencia PAM para cada 1000 pares de bases en el locus genómico. La biblioteca puede incluir secuencias de direccionamiento a ARNgs corriente arriba de por lo menos una secuencia PAM diferente. Los sistemas de proteína efectora Cpf1 pueden incluir más de una proteína Cpf1. Puede usarse cualquier proteína efectora Cpf1 como se describe en la presente, incluyendo proteínas efectoras Cpf1 ortólogas o diseñadas que reconocen diferentes secuencias PAM. La frecuencia de los sitios fuera de diana para un ARNgs puede ser menor que 500. Las puntuaciones fuera de diana pueden generarse para seleccionar ARNgs con la menor cantidad de sitios fuera de diana. Cualquier fenotipo determinado a ser asociado con corte en un sitio diana del ARNgs puede confirmarse con el uso de ARNgs dirigido al mismo sitio en un experimento único. La validación de un sitio diana también puede realizarse con el uso de una proteína efectora Cpf1 modificada, como se describe en la presente, y dos ARNgs dirigidos al sitio genómico de interés. Si estar ligado a la teoría, un sitio diana es una coincidencia verdadera si se observa el cambio en el fenotipo en experimentos de validación.

Los loci genómicos pueden incluir por lo menos una región genómica continua. La por lo menos una región genómica continua puede comprender hasta el genoma completo. La por lo menos una región genómica continua puede comprender un elemento funcional del genoma. El elemento funcional puede estar dentro de una región no codificante, gen codificante, región intrónica, promotor, o potenciador. La por lo menos una región genómica continua puede comprender por lo menos 1 kb, preferentemente por lo menos 50 kb de ADN genómico. La por lo menos una región genómica continua puede comprender un sitio de unión de factor de transcripción. La por lo menos una región genómica continua puede comprender una región de hipersensibilidad a DNAsa I. La por lo menos una región genómica continua puede comprender un potenciador o elemento de represión de transcripción. La por lo menos una región genómica continua puede comprender un sitio enriquecido para una signatura epigenética. La por lo menos una región de ADN genómico continua puede comprender un aislante epigenético. La por lo menos una región genómica continua puede comprender dos o más regiones genómicas continuas que interaccionan físicamente. Las regiones genómicas que interaccionan pueden determinarse por "tecnología 4C". La tecnología 4C permite la selección del genoma completo de una manera no sesgada para segmentos de ADN que interaccionan físicamente con un fragmento de ADN de elección, como se describe en Zhao y col. ((2006) Nat Genet 38, 1341-7) y en la Patente de los EE.UU. 8.642.295. La signatura epigenética puede ser acetilación de histona, metilación de histona, ubiquitinación de histona, fosforilación de histona, metilación de ADN, o una falta de los mismos.

El o los sistemas de proteína efectora Cpf1 para la mutagénesis por barrido saturante o profunda puede usarse en una población de células. El o los sistemas de proteína efectora Cpf1 pueden usarse en células eucarióticas, que incluyen a título enunciativo no taxativo a células de mamífero y vegetales. La población de células puede ser células procarióticas. La población de células eucarióticas puede ser una población de células madre embriónicas (ES), células neuronales, células epiteliales, células inmunológicas, células endócrinas, células musculares, eritrocitos, linfocitos, células vegetales, o células de levaduras.

En un aspecto, los métodos y composiciones reivindicados de la presente invención se pueden usar en un método

para la selección de los elementos funcionales asociados con un cambio en un fenotipo. La biblioteca puede introducirse en una población de células que están adaptadas para contener una proteína efectora Cpf1. Las células pueden separarse en por lo menos dos grupos en base al fenotipo. El fenotipo puede ser la expresión de un gen, crecimiento celular, o viabilidad celular. Se determina la representación relativa de los ARN guías presentes en cada grupo, mientras que los sitios genómicos asociados con el cambio en el fenotipo se determinan por representación de los ARN guías presentes en cada grupo. El cambio en el fenotipo puede ser un cambio en la expresión de un gen de interés. Al gen de interés se le puede aumentar la expresión, disminuir la expresión o noquearse. Las células pueden separarse en un grupo de expresión alta y un grupo de expresión baja. La población de células puede incluir una construcción reportera que se usa para determinar el fenotipo. La construcción reportera puede incluir un marcador que puede detectarse. Las células pueden separarse con el uso del marcador que puede detectarse.

En otro aspecto, los métodos y composiciones reivindicados de la presente invención se pueden usan en un método para la selección de sitios genómicos asociados con resistencia a un compuesto químico. El compuesto químico puede ser un fármaco o pesticida. La biblioteca puede introducirse en una población de células que están adaptadas para contener una proteína efectora Cpf1, en donde cada célula de la población contiene no más de un ARN guía; la población de células son tratadas con el compuesto químico; y la representación de ARN guías se determina después del tratamiento con el compuesto químico en un punto de tiempo posterior en comparación con un punto de tiempo más temprano, de manera que se determinan los sitios genómicos asociados con la resistencia al compuesto químico por enriquecimiento de los ARN guías. La representación de los ARNgs puede determinarse por métodos de secuenciamiento profundo.

En la práctica de la presente invención que utiliza los complejos de la proteína efectora Cpf1 son útiles los métodos usados en los sistemas CRISPR-Cas9 y se hace referencia al artículo con el título BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Canver, M.C., Smith,E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., Luc, S., Kurita, R., Nakamura, Y., Fujiwara, Y., Maeda, T., Yuan, G., Zhang, F., Orkin, S.H., & Bauer, D.E. DOI:10.1038/nature15521, publicado en internet el 16 de septiembre de 2015, el artículo se describe brevemente a continuación:

Canver y col. involucran novedosas bibliotecas de ARN guía de CRISPR-Cas9 agrupados para realizar mutagénesis saturante in situ de los potenciadores eritroides BCL11A de humano y ratón previamente identificados como un potenciador asociado con el nivel de hemoglobina fetal (HbF) y cuyo ortólogo de ratón es necesario para la expresión de BCL11A eritroide. Esta estrategia reveló las características mínimas críticas y las vulnerabilidades discretas de estos potenciadores. Mediante la edición de progenitores humanos primarios y transgénesis en ratón, los autores validaron el potenciador eritroide BCL11A como un diana para la reinducción de HbF. Los autores generaron un mapa de potenciadores detallado que informa la edición de genoma terapéutica.

Método para usar los sistemas Cpf1 para modificar una célula u organismo

Los métodos y composiciones reivindicados de la invención en algunas formas de realización comprenden modificar una célula. La célula puede ser una célula eucariota o una célula procariota. La célula puede ser una célula de mamífero. La célula de mamífero puede ser cualquier célula de primate no humano, bovino, porcino, roedor o ratón. La célula puede ser una célula eucariótica no mamífera tal como de ave, pescado o camarón. La célula también puede ser una célula vegetal. La célula vegetal puede ser de una planta de cultivo tal como mandioca, maíz, sorgo, trigo, o arroz. La célula vegetal también puede ser un alga, árbol o vegetal. La modificación introducida a la célula por la presente invención puede ser tal que la célula y progenie de la célula están alteradas para una producción mejorada de productos biológicos tales como un anticuerpo, almidón, alcohol u otro producto celular deseado. La modificación introducida por la célula por la presente invención puede ser tal que la célula y la progenie de la célula incluya una alteración que cambia el producto biológico producido.

El sistema previsto para su uso en los métodos y composiciones reivindicados puede comprender uno o más vectores diferentes. En un aspecto de la invención, la proteína Cas con codones optimizados para la expresión en un tipo de célula deseada, preferiblemente una célula eucariótica, preferiblemente una célula de mamífero o una célula humana.

Normalmente se utilizan células empaquetadoras para formar partículas de virus que tienen la capacidad de infectar una célula huésped. Este tipo de células incluyen las células 293, que empaquetan adenovirus, y las células ψ2 o células PA317, que empaquetan retrovirus. Los vectores víricos que se utilizan en la terapia génica se generan normalmente mediante la producción de una línea celular que empaqueta un vector con ácido nucleico para obtener una partícula vírica. Los vectores contienen normalmente las secuencias víricas mínimas requeridas para el empaquetamiento e integración posterior en un huésped, reemplazándose otras secuencias víricas por un casete de expresión para el polinucleótido o los polinucleótidos que se van a expresar. Las funciones víricas ausentes se suministran normalmente por un mecanismo en *trans* desde la línea celular empaquetadora. Por ejemplo, los vectores derivados de AAV utilizados en la terapia génica normalmente poseen tan solo secuencias ITR procedentes del genoma de AAV necesarias para el empaquetamiento e integración en el genoma huésped. El ADN vírico se empaqueta en una línea celular, que contiene un plásmido cooperador que codifica otros genes de AAV, concretamente rep y cap, pero que carece de las secuencias ITR. También puede infectarse la línea celular con adenovirus como auxiliar. El virus cooperador promueve la replicación del vector derivado de AAV y la expresión de los genes de AAV procedentes del plásmido cooperador. El plásmido cooperador no se empaqueta en cantidades

significativas debido a la ausencia de secuencias ITR. La contaminación con adenovirus se puede reducir mediante, por ejemplo, tratamiento térmico al cual los adenovirus son más sensibles que los AAV.

Administración

La invención implica las composiciones y métodos reivindicados que hacen uso de por lo menos un componente del complejo CRISPR, por ejemplo, ARN administrado a través de por lo menos un complejo de nanopartícula. En algunos aspectos, los métodos reivindicados de la invención comprenden suministrar uno o más polinucleótidos, tal como uno o más vectores como se describen en la presente, uno o más transcritos de estos y/o una o proteínas transcritas a partir de ellos, a una célula huésped. También se describen, pero no se reivindican específicamente en la presente, células producidas mediante tales métodos y animales no humanos que comprenden tales células o producidos a partir de ellas. En algunas formas de realización, se suministra a una célula una enzima CRISPR combinada con (y opcionalmente complejada con) una secuencia guía. Se pueden utilizar métodos de transferencia génica con virus y sin virus convencionales para introducir ácidos nucleicos en células de mamíferos o tejidos diana. Se pueden utilizar tales métodos para administrar ácidos nucleicos que codifican componentes de un sistema CRISPR a células en cultivo o en un organismo huésped. Los sistemas de suministro con vectores no víricos incluyen plásmidos de ADN, ARN (por ejemplo, un transcrito de un vector descrito en la presente), ácido nucleico desnudo y ácido nucleico complejado con un vehículo de suministro, tal como un liposoma. Los sistemas de suministro vectoriales víricos incluyen virus de ADN y ARN, que tienen genomas episómicos o integrados tras el suministro a la célula. Para una revisión de los procedimientos de la terapia génica véase Anderson, Science 256:808-813 (1992); Nabel y Felgner, TIBTECH 11:211-217 (1993); Mitani y Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer y Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada y col., en Current Topics in Microbiology and Immunology Doerfler y Böhm (eds) (1995); y Yu y col., Gene Therapy 1:13-26 (1994).

Los métodos de administración no víricos de ácidos nucleicos incluyen la lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, conjugados de policatión o lípido:ácido nucleico, ADN puro, viriones artificiales y captación de ADN potenciada por un agente. La lipofección se describe en por ejemplo, las Patentes de los EE.UU. n.º 5.049.386, 4.946.787; y 4.897.355) y los reactivos de lipofección se venden de manera comercial (por ejemplo, Transfectam™ y Lipofectin™). Los lípidos catiónicos y neutros que son adecuados para la lipofección de reconocimiento de receptor eficaz de polinucleótidos inlcuyen aquellos de Felgner, WO 91/17424; WO 91/16024. La administración puede ser a células (por ejemplo administración *in vitro* o *ex vivo*).

La preparación de complejos lípido:ácido nucleico, incluidos los liposomas dirigidos tales como los complejos de inmunolípidos, es muy conocida por el experto en la técnica (véase, por ejemplo, Crystal, Science 270:404-410 (1995); Blaese y col., Cancer Gene Ther. 2:291-297 (1995); Behr y col., Bioconjugate Chem. 5:382-389 (1994); Remy y col., Bioconjugate Chem. 5:647-654 (1994); Gao y col., Gene Therapy 2:710-722 (1995); Ahmad y col., Cancer Res. 52:4817-4820 (1992); Patentes de los EE.UU. n.º 4.186.183, 4.217.344, 4.235.871, 4.261.975, 4.485.054, 4.501.728, 4.774.085, 4.837.028 y 4.946.787).

El uso de sistemas con ARN o ADN vírico para el suministro de ácidos nucleicos aprovecha los procesos sumamente evolucionados para dirigir un virus a células específicas en el cuerpo y para introducir la carga dañina viral en el núcleo. Los vectores víricos se pueden administrar directamente a pacientes (in vivo, no reivindicado en la presente) o se pueden utilizar para tratar células in vitro y las células modificadas se podrán administrar opcionalmente a pacientes (ex vivo). Los sistemas con virus convencionales podrían incluir vectores retrovíricos, lentivíricos, adenovíricos, adenoasociados y del virus del herpes simple para la transferencia génica. La integración en el genoma huésped es posible con métodos de transferencia génica con retrovirus, lentivirus y virus adenoasociados y a menudo dan como resultado la expresión a largo plazo del transgén insertado. Además, se han observado eficacias de transducción elevadas en muchos tipos celulares y tejidos diana diferentes.

Se puede alterar el tropismo de un retrovirus incorporando proteínas de la envoltura foráneas para expandir la población diana potencial de células diana. Los vectores lentivíricos son vectores retrovíricos que son capaces de transducir o infectar células que no se dividen y normalmente producen títulos víricos elevados. La selección de un sistema de transferencia génico retrovírico podría depender, por lo tanto, del tejido diana. Los vectores retrovíricos comprenden repeticiones terminales largas que actúan en cis con capacidad de empaquetar hasta 6-10 kb de una secuencia foránea. Los LTR que actúan en cis mínimos son suficientes para la replicación y empaquetamiento de los vectores, que se utilizan a continuación para integrar el gen terapéutico en la célula diana para proveer una expresión del transgén permanente. Los vectores retrovirales usados ampliamente incluyen aquellos en base al virus de leucemia murina (MuLV), virus de leucemia de mono gibón (GaLV), virus de inmunodificiencia de simio (SIV), virus de inmuno deficiencia humana (HIV), y combinaciones de los mismos (véase, por ejemplo, Buchscher y col., J. Virol. 66:2731-2739 (1992); Johann y col., J. Virol. 66:1635-1640 (1992); Sommnerfelt y col., Virol. 176:58-59 (1990); Wilson y col., J. Virol. 63:2374-2378 (1989); Miller y col., J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

En otra forma de realización, se contemplan partículas de vectores retrovíricos pseudotipados con envoltura de vesiculovirus cocal (véase, por ejemplo, la patente de los EE.UU. con n.º de publicación 20120164118 adjudicada al Centro de Investigación sobre el Cáncer Fred Hutchinson). El virus cocal pertenece al género de vesiculovirus y es el agente causante de la estomatitis vesicular en mamíferos. El virus cocal se aisló originalmente de garrapatas en

Trinidad (Jonkers y col., Am. J. Vet. Res. 25:236-242 (1964)), y se han identificado infecciones en Trinidad, Brazil, y Argentina de insectos, ganado y caballos. Muchos de los vesiculovirus que infectan mamíferos se han aislado de artrópodos infectados de manera natural lo que sugieren que se transmiten mediante vectores. Los anticuerpos contra vesiculovirus son comunes entre personas que viven en áreas rurales donde los virus son endémicos y adquiridos en laboratorio; las infecciones en seres humanos normalmente conllevan síntomas similares a la gripe. La glucoproteína de la envoltura del virus Cocal comparte un 71,5% de identidad a nivel aminoacídico con el VSV-G Indiana y la comparación filogenética del gen de la envoltura de vesiculovirus muestra que el virus Cocal es serológicamente distinto, aunque estrechamente relacionado con ellas, de las cepas de VSV-G Indiana entre los vesiculovirus. Jonkers y col., Am. J. Vet. Res. 25:236-242 (1964) y Travassos da Rosa y col., Am. J. Tropical Med. & Hygiene 33:999-1006 (1984). Las partículas del vector retrovírico pseudotipado con la envoltura del vesiculovirus Cocal podrán incluir, por ejemplo, partículas de vectores lentivíricos, alfarretrovíricos, betarretrovíricos, gammarretrovíricos, deltarretrovíricos y epsilonretrovíricos que podrán comprender la Gag, Pol y/o una o más proteínas accesorias retrovíricas y una proteína de envoltura del vesiculovirus Cocal. Dentro de determinados aspectos de estas formas de realización, las proteínas Gag, Pol y accesorias son lentivirales y/o gammaretrovirales. La invención provee AAV que contiene o consiste esencialmente de una molécula de ácido nucleico exógena que codifica para un sistema CRISPR, por ejemplo, una pluralidad de casetes que comprenden o consisten en un primer casete que comprende o consiste esencialmente en un promotor, una molécula de ácido nucleico que codifica para una proteína asociada a CRISPR (Cas) (proteínas putativas de nucleasa o helicasa), por ejemplo, Cpf1 y un terminador, y dos, o más, ventajosamente hasta el límite de tamaño de empaquetamiento del vector, por ejemplo, en total (incluyendo el primer casete) cinco, casetes que comprenden o consisten esencialmente en un promotor, molécula de ácido nucleico que codifica para ARN guía (ARNg) y un terminador (por ejemplo, cada casete representado esquemáticamente como Promotor-ARNg1-terminador, Promotor-ARNg2-terminador ... Promotor-ARNg(N)-terminador (donde N es un número que se puede insertar que está en el límite superior del límite del tamaño de empaquetamiento del vector) o dos o más AAVr individuales, donde cada uno contiene uno o más de un casete de un sistema CRISPR, por ejemplo, un primer AAVr que contiene el primer casete que comprende o está constituido esencialmente por un promotor, una molécula de ácido nucleico que codifica Cas, por ejemplo, Cas9 y un terminador, y un segundo AAVr que contiene varios, cuatro, casetes que comprenden o están constituidos esencialmente por un promotor, una molécula de ácido nucleico que codifica un ARN guía (ARNg) y un terminador (por ejemplo, cada casete se representa esquemáticamente como Promotor-ARNg1-terminador, Promotor-ARNg2-terminador... Promotor-ARNg(N)-terminador (donde N es un número que se puede insertar que está en el límite superior del límite del tamaño de empaquetamiento del vector). Ya que AAVr es un virus con ADN, las moléculas de ácido nucleico en la presente discusión que trata sobre AAV o AAVr son convenientemente ADN. El promotor es convenientemente en algunas formas de realización el promotor de la sinapsina I humana (hSyn). Aquella persona con experiencia en el arte conoce métodos adicionales para la administración de ácidos nucleicos a las células. Véase, por ejemplo, US20030087817.

En algunas formas de realización, se transfecta una célula huésped de manera transitoria o no transitoria con uno o más vectores descritos en la presente. En algunas formas de realización, se transfecta una célula tal como ocurre de manera natural en un sujeto. En algunas formas de realización, la célula que se transfecta se toma de un sujeto. En algunas formas de realización, la célula se obtiene a partir de células que se toman de un sujeto, tal como una línea celular. Los ejemplos de líneas celulares incluyen, a título enunciativo no taxativo, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, células epiteliales de riñón de mono BS-C-1, fibroblastos embrionarios de ratón BALB/ 3T3, 3T3 suizo, 3T3-L1, fibroblastos fetales embrionarios 132-d5; fibroblastos de ratón 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, células BCP-1, BEAS-2B, bEnd.3, BHK-21, BR293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, células JY, células K562, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, líneas celulares OPCN / OPCT, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, células Saos-2, Sf-9, SkBr3, T2, T-47D, T84, línea celular THP1, U373, U87, U937, VCaP, células Vero, WM39, WT-49, X63, YAC-1, YAR y variedades transgénicas de estos. Estas líneas celulares se pueden adquirir de varios proveedores que conocen los expertos en la técnica (véase, por ejemplo, la Colección de Cultivos Tipo Americana (ATCC, por sus siglas en inglés) (Manassus, Va.)). En algunas formas de realización, se utiliza una célula transfectada con uno o más vectores descritos en la presente para establecer una nueva línea celular que comprenda una o más secuencias derivadas de un vector. En algunas formas de realización, se utiliza una célula transfectada de manera transitoria con los componentes de un sistema CRISPR tal y como se describen en la presente (por ejemplo, mediante la transfección transitoria de uno o más vectores o la transfección con ARN) y modificada mediante la actividad de un complejo CRISPR para establecer una nueva línea celular que comprenda células que contengan la modificación pero que carezcan de cualquier otra secuencia exógena. En algunas formas de realización, se utilizan células transfectadas de manera transitoria o no transitoria con uno o más vectores descritos en la presente, o se utilizan líneas celulares obtenidas a partir de tales células para evaluar uno o más compuestos de prueba.

En algunas formas de realización, se utilizan uno o más vectores descritos en la presente para producir un animal no humano transgénico o una planta transgénica. En algunas formas de realización, el animal transgénico es un mamífero, tal como un ratón, rata o conejo. En la técnica existe constancia de métodos para producir plantas y animales transgénicos y generalmente comienzan con un método de transfección celular, tal como se describe en la presente. En otra forma de realización, se puede contemplar un dispositivo para el suministro de fluidos con una matriz de agujas (véase, por ejemplo, la patente de EE.UU. con n.º de publicación 20110230839 adjudicada al Centro de Investigación sobre el Cáncer Fred Hutchinson) para el suministro de CRISPR Cas al tejido sólido. Un dispositivo de la patente de los EE.UU. con n.º de publicación 20110230839 para el suministro de un fluido a un tejido sólido podrá comprender un conjunto de agujas dispuestas en una matriz; un conjunto de depósitos, cada uno en comunicación fluida con un elemento respectivo del conjunto de agujas; y un conjunto de actuadores acoplados operativamente con los elementos respectivos del conjunto de depósitos y configurado para controlar la presión fluida dentro del depósito. En ciertas formas de realización, cada elemento del conjunto de actuadores podrá comprender un elemento de un conjunto de émbolos, donde un primer extremo de cada elemento del conjunto de émbolos se aloja en un elemento respectivo del conjunto de depósitos y en ciertas formas de realización adicionales los émbolos del conjunto de émbolos se acoplan operativamente entre sí en los segundos extremos respectivos de modo que se desplacen hacia abajo simultáneamente. Ciertas formas de realización adicionales más podrán comprender un empujador de émbolos configurado para desplazar hacia abajo todos los elementos del conjunto de émbolos con una velocidad variable de manera selectiva. En otras formas de realización, cada elemento del conjunto de actuadores podrá comprender un elemento del conjunto de líneas de transmisión fluida que tenga un primer y segundo extremos, donde un primer extremo de cada elemento del conjunto de líneas de transmisión fluida se acopla con un elemento respectivo del conjunto de depósitos. En otras formas de realización, el dispositivo podrá comprender una fuente de presión fluida y cada elemento del conjunto de actuadores comprenderá un acoplamiento fluido entre la fuente de presión fluida y un elemento respectivo del conjunto de depósitos. En formas de realización adicionales, la fuente de presión fluida podrá comprender por lo menos uno de: compresor, acumulador de vacío, bomba peristáltica, cilindro maestro, bomba microfluídica y válvula. En otra forma de realización, cada elemento del conjunto de agujas podrá comprender un conjunto de conexiones distribuidas a lo largo de su longitud.

En un aspecto, los métodos y composiciones reivindicados de la invención pueden comprender modificar un polinucleótido diana en una célula eucariota. En algunas formas de realización, el método puede comprender permitir que se una un complejo direccionado a ácido nucleico al polinucleótido diana para efectuar el clivaje del mencionado polinucleótido diana, modificando de este modo el polinucleótido diana, en donde el complejo direccionado a ácido nucleico comprende una proteína efectora direccionada a ácido nucleico complejada con un ARN guía hibridado a una secuencia diana en el mencionado polinucleótido diana.

En un aspecto, los métodos y composiciones reivinficados de la invención comprenden modificar la expresión de un polinucleótido en una célula eucariota. En algunas formas de realización, el método comprende permitir que un complejo direccionado a ácido nucleico se una al polinucleótido de forma tal que dicha unión de como resultado la expresión aumentada o reducida de dicho polinucleótido; en donde el complejo direccionado a ácido nucleico comprende una proteína efectora direccionada a ácido nucleico complejada con un ARN guía a una secuencia diana en el mencionado polinucleótido.

Los componentes del complejo CRISPR pueden administrarse por conjugación o asociación con porciones de transporte (adaptado por ejemplo de estrategias descritas en las Patentes de los EE.UU. n.º 8.106.022; 8.313.772). Las estrategias de administración de ácido nucleico pueden usarse por ejemplo para mejorar la administración de ARN guía, o ARN mensajeros o ADN codificantes que codifican para componentes del complejo CRISPR. Por ejemplo, los ARN pueden incorporar nucleótidos de ARN modificados para mejorar la estabilidad, reducir la inmunoestimulación, y/o mejorar la especificidad (véase Deleavey, Glen F. y col., 2012, Chemistry & Biology, Volume 19, Issue 8, 937 – 954; Zalipsky, 1995, Advanced Drug Delivery Reviews 16: 157-182; Caliceti and Veronese, 2003, Advanced Drug Delivery Reviews 55: 1261-1277). Se han descrito varias construcciones que pueden usarse para modificar ácidos nucleicos, tal como los ARNgs, para una administración más eficaz, tal como modificaciones de esqueleto fosfodiéster neutralizante de carga reversible que pueden adaptarse para modificar los ARNgs para que sean más hidrofóbicos y no aniónicos, mejorando de esa manera la entrada a la célula (Meade BR y col., 2014, Nature Biotechnology 32,1256–1261). En otras formas de realización alternativas, los motivos de ARN seleccionados pueden ser útiles para mediar la transfección celular (Magalhães M., y col., Molecular Therapy (2012); 20 3, 616–624). Similarmente, los aptámeros pueden adaptarse para la administración de componentes del complejo CRISPR, por ejemplo, por agregado de aptámeros a ARNgs (Tan W. y col., 2011, Trends in Biotechnology, December 2011, Vol. 29, No. 12).

En algunas formas de realización, la conjugación de N-acetilgalactosamina ternaria (GalNAc) a componentes oligonucleotídicos puede usarse para mejorar la administración, por ejemplo, administración a tipos celulares seleccionados, por ejemplo hepatocitos (véase WO2014118272; Nair, JK y col., 2014, Journal of the American Chemical Society 136 (49), 16958-16961). Esto puede considerarse que es una partícula en base a azúcar y en la presente se proveen detalles adicionales de otros sistemas y/o formulaciones de administración de partículas. Por lo tanto, el GalNAc puede considerarse que es una partícula en el sentido de las otras partículas descritas en la presente, tal que los usos generales y otras consideraciones, por ejemplo, administración de las mencionadas partículas, se aplican también para las partículas de GalNAc. Puede usarse por ejemplo una estrategia de conjugación en fase solución para unir agrupamientos de GalNAc ternaria (peso molecular ~2000) activados como PFP (pentafluorofenil) ésteres sobre oligonucleótidos modificados con 5′-hexilamino (5′-HA ASOs, peso molecular ~8000 Da; Østergaard y

col., Bioconjugate Chem., 2015, 26 (8), pp 1451-1455). Similarmente, se han descrito polímeros de poli(acrilato) para la administración *in vivo* de ácido nucleico (véase el documento de patente WO2013158141). En otras formas de realización alternativas, la premezcla de nanopartículas de CRISPR (o complejos proteicos) con proteínas de suero de origen natural puede usarse con el objetivo de mejorar la administración (Akinc A y col., 2010, Molecular Therapy vol. 18 no. 7, 1357-1364).

Las técnicas de selección se encuentran disponibles para identificar la administración de potenciadores, por ejemplo, mediante selección de bibliotecas químicas (Gilleron J. y col., 2015, Nucl. Acids Res. 43 (16): 7984-8001). También se han descrito estrategias para evaluar la eficacia de vehículos de administración, tal como nanopartículas de lípidos, que puede emplearse para identificar vehículos de administración de componentes de CRISPR (véase Sahay G. y col., 2013, Nature Biotechnology 31, 653-658).

En algunas formas de realización, la administración de componentes de la proteína CRISPR puede facilitarse con el agregado de péptidos funcionales a la proteína, tales como péptidos que cambian la hidrofobicidad de la proteína, por ejemplo, para mejorar la funcionalidad *in vivo*. Las proteínas del componente CRISPR pueden modificarse similarmente para facilitar las reacciones químicas posteriores. Por ejemplo, pueden agregarse aminoácidos a una proteína que tiene un grupo que es sometido a química de marcaje (Nikić I. y col., 2015, Nature Protocols 10,780–791). En formas de realización de este tipo, el grupo químico de marcaje puede usarse entonces para agregar una variedad amplia de estructuras alternativas, tal como poli(etilenglicol) para estabilidad, péptidos penetrantes de la célula, aptámeros de ARN, lípidos, o hidratos de carbono tal como GalNAc. En otras alternativas, una proteína componente de CRISPR puede modificarse para adaptar la proteína para la entrada a la célula (véase Svensen y col., 2012, Trends in Pharmacological Sciences, Vol. 33, No. 4), por ejemplo, por agregado de péptidos penetrantes a la proteína (véase Kauffman, W. Berkeley y col., 2015, Trends in Biochemical Sciences, Volume 40, Issue 12, 749 – 764; Koren y Torchilin, 2012, Trends in Molecular Medicine, Vol. 18, No. 7). En otra forma de realización alternativa, los pacientes o sujetos pueden pretratarse con compuestos o formulaciones que facilitan la administración tardía de componentes CRISPR.

Los complejos de proteína efectora Cpf1 pueden usarse en plantas

El o los sistemas de proteína efectora Cpf1 (por ejemplo, simples o multiplexados) pueden usarse junto con los recientes avances en la genómica de cultivos. Los sistemas descritos en la presente pueden usarse para realizar consultas o edición o manipulación eficiente y eficaz en costos del gen o genoma de la planta—por ejemplo, para la rápida investigación y/o selección y/o consultas y/o comparación y/o manipulaciones y/o transformación de genes o genomas vegetales; por ejemplo, para crear, identificar, desarrollar, optimizar, o conferir rasgo(s) o característica(s) a la o las plantas o para transformar un genoma vegetal. Consecuentemente puede mejorarse la producción de plantas, nuevas plantas con nuevas combinaciones de rasgos o características o nuevas plantas con rasgos potenciados. El o los sistemas de proteína efectora Cpf1 pueden usarse en relación a plantas en técnicas de integración dirigida a sitio (SDI) o edición génica (GE) o cualquier cruza revertida cercana (NRB) o cruza revertida (RB). Los aspectos para utilizar los sistemas de proteína efectora Cpf1 descritos en la presente pueden ser análogos al uso del sistema CRISPR-Cas (por ejemplo, CRISPR-Cas9) en plantas, y se hace mención del sitio de internet de la Universidad de Arizona "CRISPR-PLANT" (http://www.genome.arizona.edu/crispr/) (financiado por Penn State y AGI). Las formas de realización de los métodos y composiciones reivindicados de la invención pueden usarse en la edición génica en planta o en donde se han usado previamente técnicas de ARNi o de edición de genoma similares; véase, por ejemplo, Nekrasov, "Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR-Cas system", Plant Methods 2013, 9:39 (doi:10.1186/1746-4811-9-39); Brooks, "Efficient gene editing in tomato in the first generation using the CRISPR-Cas9 system", Plant Physiology September 2014 pp 114.247577; Shan, "Targeted genome modification of crop plants using a CRISPR-Cas system", Nature Biotechnology 31, 686-688 (2013); Feng, "Efficient genome editing in plants using a CRISPR/Cas system", Cell Research (2013) 23:1229–1232. doi:10.1038/cr.2013.114; publicado en internet el 20 de agosto de 2013; Xie, "RNA-guided genome editing in plants using a CRISPR-Cas system", Mol Plant. 2013 Nov;6(6):1975-83. doi: 10.1093/mp/sst119. Epub 2013 Aug 17; Xu, "Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice", Rice 2014, 7:5 (2014), Zhou y col., "Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and Redundancy", New Phytologist (2015) (Forum) 1-4 (disponible en internet solo en www.newphytologist.com); Caliando y col., "Targeted DNA degradation using a CRISPR device stably carried in the host genome, NATURE COMMUNICATIONS 6:6989, DOI: 10.1038/ncomms6989, www.nature.com/naturecommunications DOI: 10.1038/ncomms7989; patente de Estados Unidos n.º 6.603.061 - Agrobacterium-Mediated Plant Transformation Method; patente de Estados Unidos n.º 7.868.149 - Plant Genome Sequences and Uses Thereof y el documento US 2009/0100536 – Transgenic Plants with Enhanced Agronomic Traits. En la práctica de la invención, los contenidos y divulgación de Morrell y col "Crop genomics: advances and applications", Nat Rev Genet. 2011 Dec 29;13(2):85-96. En consecuencia, la referencia en la presente a células animales también puede aplicarse, cambiando lo que se deba cambiar, a células vegetales a menos que sea evidente de otra manera; y, las enzimas en la presente que tienen efectos fuera de diana reducidos y sistemas que emplean dichas enzimas pueden usarse en aplicaciones en plantas, que incluyen aquellas mencionadas en la presente.

Aplicación del sistema Cpf1-CRISPR a plantas y levaduras

Definiciones:

En general, el término "planta" se relaciona con cualquiera de diversos organismos fotosintéticos, eucariotas, unicelular o multicelulares pertenecientes al reino Plantae que típicamente crecen por división celular, contienen cloroplastos y cuyas paredes celulares están compuestas por celulosa. El término planta abarca plantas monocotiledóneas y dicotiledóneas. Específicamente, las plantas comprenderán, en un sentido no taxativo, plantas angiospermas y gimnospermas tales como acacia, alfalfa, amaranto, manzana, damasco, alcaucil, fresno, espárrago, palta, banana, cebada, habas, remolacha, abedul, haya, moras, arándanos, brócoli, repollitos de Bruselas, repollo, canola, melón cantalupo, zanahoria, mandioca, coliflor, cedro, cereales, apio, castaño, cerezo, col china, cítricos, clementinas, trébol, café, maíz, algodón, garbanzo, pepino, ciprés, berenjena, olmo, escarola, eucalipto, hinojo, higo, abeto, geranio, uva, pomelo, maníes, tomates cherry, goma, cicuta, nogal americano, col rizada, kiwi, colinabo, alerce, lechuga, puerro, limón, lima, algarrobo, pino, ginkgo, maíz, mango, arce, melón, mijo, champiñón, mostaza, frutos secos, roble, avena, palma oleaginosa, okra, cebolla, naranja, plantas ornamentales o flores o árboles, papaya, palma, perejil, chirivía, guisante, durazno, cacahuete, pera, turba, pimiento, caqui, frijol guandú, pino, ananá, plátano, ciruela, granada, patata, zapallo, achicoria roja, rábano, colza, frambuesa, arroz, centeno, sorgo, cártamo o alazor, espinos, soja, espinaca, abeto, calabaza, frutilla, remolacha azucarera, caña de azúcar, girasol, batata, maíz dulce, mandarina, té, tabaco, tomate, árboles, triticale, céspedes, nabos, vid, nogal, berro, sandía, sorgo, ñames, tejo y calabacín. El término planta también abarca Algae, que en su mayoría son fotoautótrofos unificados primariamente por su falta de raíces, hojas y otros órganos que son característicos de las plantas superiores.

Los métodos de edición de genomas se pueden utilizar para conferir los rasgos deseados esencialmente en cualquier planta empleando el sistema Cpf1 descrito en la presente. Se puede manipular una amplia variedad de plantas y sistemas de células vegetales para las características fisiológicas y agronómicas deseadas que se describen en la presente usando las construcciones de ácido nucleico de la presente descripción y los diversos métodos de transformación mencionados previamente. En formas de realización preferidas, las plantas y células vegetales diana para la manipulación genética incluyen, pero en un sentido no limitativo, plantas monocotiledóneas y dicotiledóneas, tales como cultivos incluyendo cultivos de gramíneas (por ejemplo, trigo, maíz, arroz, mijo, cebada), cultivos frutales (por ejemplo, tomate, manzana, pera, frutilla, naranja), cultivos forrajeros (por ejemplo, alfalfa), cultivos de verduras de raíces y tubérculos (por ejemplo, zanahoria, patata, remolacha, yam), cultivos de verduras de hoja (por ejemplo, lechuga, espinaca); plantas con flores (por ejemplo, petunia, rosa, crisantemo), coníferas y pinos (por ejemplo, pino-picea, abeto); plantas empleadas en la fitorremediación (por ejemplo, plantas que acumulan metales pesados); cultivos de oleaginosas (por ejemplo, girasol, colza) y plantas usadas con fines experimentales (por ejemplo, Arabidopsis). Por lo tanto, se pueden emplear los métodos y sistemas CRISPR-Cas en un amplio rango de plantas tal como, por ejemplo, con plantas dicotiledóneas que pertenecen a los órdenes Magniolales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papeverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Poligonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, San tales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Poligalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales y Asterales; los métodos y sistemas CRISPR-Cas se pueden usar con plantas monocotiledóneas tales como las que pertenecen a los órdenes Alismatales, Hidrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Rescionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales y Orchidales, o con plantas que pertenecen a Gymnospermae, por ejemplo, aquellas que pertenecen a los órdenes Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales y Gnetales.

Los sistemas Cpf1 CRISPR y los métodos de uso que se describen en la presente se pueden usar en un amplio rango de especies de plantas, incluidas en la lista no taxativa de géneros de dicotiledóneas, monocotiledóneas o gimnospermas que se mencionan a continuación: *Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glicina, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, Litsea, Lycopersicon, Lupinus, Manihot, Majorana, Malus, Medicago, Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Senecio, Sinomenium, Stephania, Sinapis, Solanum, Theobroma, Trifolium, Trigonella, Vicia, Vinca, Vilis y Vigna; y los géneros Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, Zea, Abies, Cunninghamia, Ephedra, Picea, Pinus* y *Pseudotsuga*.

Los sistemas Cpf1 CRISPR y los métodos de uso también se pueden emplear en un amplio rango "algas" o "células de algas"; que incluyen, por ejemplo, algas seleccionadas entre varios Filas eucariotas, incluyendo Rhodophyta (algas rojas), Clorophyta (algas verdes), Phaeophyta (algas pardas), Bacillariophyta (diatomeas), Eustigmatophyta y dinoflagelados, así como el filo procariota de Cyanobacteria (algas azul-verdosas). El término "alga" incluye, por ejemplo, algas seleccionadas entre: Amphora, Anabaena, Anikstrodesmis, Botryococcus, Chaetoceros, Chlamydomonas, Clorella, Clorococcum, Cyclotella, Cylindrotheca, Dunaliella, Emiliana, Euglena, Hematococcus, Isochrysis, Monochrysis, Monoraphidium, Nannocloris, Nannnocloropsis, Navicula, Nephrocloris, Nephroselmis, Nitzschia, Nodularia, Nostoc, Oochromonas, Oocystis, Oscillartoria, Pavlova, Phaeodactylum, Playtmonas, Pleurochrysis, Porhyra, Pseudoanabaena, Pyramimonas, Stichococcus, Synechococcus, Synechocystis, Tetraselmis, Thalassiosira y Trichodesmium.

Una parte de una planta, es decir, un "tejido vegetal", puede ser tratada de acuerdo con los métodos de la presente invención para producir una planta mejorada. Un tejido vegetal también abarca células vegetales. El término "célula vegetal" según se usa en la presente se refiere a unidades individuales de una planta viva, ya sea en una planta completa intactas o en una forma aislada cultivada en cultivos tisulares in vitro, sobre un medio o agar, en suspensión en una solución amortiguadora o en un medio de crecimiento o como parte de unidades organizadas superiores tales como, por ejemplo, un tejido vegetal, un órgano vegetal o una planta completa.

Un "protoplasto" se refiere a una célula vegetal cuya pared celular protectora fue eliminada por completo o parcialmente usando, por ejemplo, medios mecánicos o enzimáticos que resulta en una unidad bioquímica competente intacta de una planta viva y que podrá volver a formar su pared celular, proliferar y regenerarse mediante crecimiento en una planta completa bajo las condiciones de crecimiento apropiadas.

El término "transformación" se refiere en general al proceso por el cual una planta huésped es modificada genéticamente mediante la introducción de ADN con Agrobacteria o cualquiera de una variedad de métodos químicos o físicos. Según se usa en la presente, el término "planta huésped" se refiere a plantas, que incluyen cualquiera entre células, tejidos, órganos o la progenie de las plantas. Hay muchos tejidos vegetales o células vegetales adecuados que pueden ser transformados e incluyen, pero en un sentido no taxativo, embriones somáticos, polen, hojas, plántulas, tallos, callos, estolones, microtubérculos y brotes. Un "tejido vegetal" también hace referencia a cualquier clon de una planta, una semilla, progenie o propágulo, generado en forma sexual o asexual, y los descendientes de cualquiera de ellos, tales como injertos o semillas.

El término "transformado", según se usa en la presente, se refiere a una célula, tejido, órgano u organismo en el cual se ha introducido una molécula de ADN extraña, tal como una construcción. La molécula de ADN introducida se puede integrar en el ADN genómico de la célula, el tejido, el órgano u organismo receptor, de modo tal que la molécula de ADN introducida es transmitida a la progenie subsiguiente. En estas formas de realización, la célula o planta "transformada" o "transgénica" también puede incluir la progenie de la célula o planta y la progenie producida en un programa de cría que emplea dicha planta transformada como progenitora en un cruzamiento y muestra un fenotipo alterado debido a la presencia de la molécula de ADN introducida. Preferiblemente, la planta transgénica es fértil y puede transmitir el ADN introducido a la progenie por medio de reproducción sexual.

El término "progenie", tal como la progenie de una planta transgénica, es aquella que nació, se generó o deriva a partir de una planta o de la planta transgénica. La molécula de ADN introducida también se puede introducir transitoriamente en la célula receptora de manera tal que la molécula de ADN introducida no será heredada por la progenie subsiguiente y por consiguiente no se la considera "transgénica". Por lo tanto, según se usa en la presente, una planta o célula vegetal "no transgénica" es una planta que no contiene un ADN extraño integrado de manera estable en su genoma.

El término "promotor vegetal", según se usa en la presente, es un promotor capaz de iniciar la transcripción en células vegetales ya sea si es originaria, o no, de una célula vegetal. Los ejemplos de promotores vegetales adecuados incluyen, pero en un sentido no taxativo, aquellos que se pueden obtener a partir de plantas, virus de plantas y bacterias tales como Agrobacterium o Rhizobium, que comprenden genes que se expresan en células vegetales.

Según se usa en la presente, una "célula fúngica" se refiere a cualquier tipo de célula eucariota perteneciente al reino de los hongos. Los Fila del reino de los hongos incluyen Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia y Neocallimastigomycota. Las células fúngicas pueden incluir levaduras, mohos y hongos filamentosos. En algunas formas de realización, la célula fúngica es una célula de levadura.

Según se usa en la presente, el término "célula de levadura" se refiere a cualquier célula fúngica pertenecientes a los Filas Ascomycota y Basidiomycota. Las células de levadura pueden incluir células de levadura de brotación, células de levadura de fisión y células de moho. Sin limitación alguna a estos organismos, muchos de los tipos de levadura usados a nivel de laboratorio e industrial forman parte del filo Ascomycota. En algunas formas de realización, la célula de levadura es una célula de S. cerervisiae, Kluyveromyces marxianus o Issatchenkia orientalis. Otras células de levadura pueden incluir, en un sentido no taxativo, Candida spp. (por ejemplo, Candida albicans), Yarrowia spp. (por ejemplo, Yarrowia lipolitica), Pichia spp. (por ejemplo, Pichia pastoris), Kluyveromyces spp. (por ejemplo, Kluyveromyces lactis y Kluyveromyces marxianus), Neurospora spp. (por ejemplo, Neurospora crassa), Fusarium spp. (por ejemplo, Fusarium oxisporum) e Issatchenkia spp. (por ejemplo, Issatchenkia orientalis, también conocida como Pichia kudriavzevii y Candida acidothermophilum). En algunas formas de realización, la célula fúngica es una célula fúngica filamentosa. Según se usa en la presente, el término "célula fúngica filamentosa" se refiere a cualquier tipo de célula fúngica que crece formando filamentos, es decir, hifas o micelios. Los ejemplos de células fúngicas filamentosas pueden incluir, en un sentido no taxativo, Aspergillus spp. (por ejemplo, Aspergillus niger), Trichoderma spp. (por ejemplo, Trichoderma reesei), Rhizopus spp. (por ejemplo, Rhizopus oryzae) y Mortierella spp. (por ejemplo, Mortierella isabellina).

En algunas formas de realización, la célula fúngica es una cepa industrial. Según se usa en la presente, una "cepa industrial" se refiere a cualquier cepa de célula fúngica usada o aislada de un proceso industrial, por ejemplo, la producción de un producto a escala comercial o industrial. La cepa industrial puede hacer referencia a una especie fúngica que típicamente se usa en un proceso industrial o puede hacer referencia a una forma aislada de una especie fúngica que también se puede emplear para fines no industriales (por ejemplo, investigación en laboratorio). Los ejemplos de procesos industriales pueden incluir una fermentación (por ejemplo, en la producción de productos

alimenticios o de bebidas), destilación, producción de biocombustible, producción de un compuesto y producción de un polipéptido. Los ejemplos de cepas industriales pueden incluir, en un sentido no taxativo, JAY270 y ATCC4124.

En algunas formas de realización, la célula fúngica es una célula poliploide. Según se usa en la presente, una célula "poliploide" se refiere a cualquier célula cuyo genoma está presente en más de una copia. Una célula poliploide se refiere a un tipo de célula que se encuentra naturalmente en un estado poliploide, o se refiere a una célula que fue inducida para existir en un estado poliploide (por ejemplo, a través de una regulación, alteración, inactivación, activación o modificación específica de la meiosis, citoquinesis o replicación de ADN). Una célula poliploide se refiere a una célula cuyo genoma completo es poliploide o se refiere a una célula que es poliploide en un locus genómico particular de interés. Sin considerar ninguna teoría, se cree que la abundancia del ARN-guía puede constituir un componente limitante en la manipulación mediante ingeniería del genoma de célula poliploides más a menudo que en células haploides, y por lo tanto los métodos que emplean los sistemas Cpf1 CRISPR descrito en la presente pueden aprovechar el uso de un determinado tipo de célula fúngica.

En algunas formas de realización, la célula fúngica es una célula diploide. Según se usa en la presente, una célula "diploide" se refiere a cualquier célula cuyo genoma está presente en dos copias. Una célula diploide se refiere a un tipo de célula que se encuentra naturalmente en un estado diploide, o se refiere a una célula que fue inducida para existir en un estado diploide (por ejemplo, a través de una regulación, alteración, inactivación, activación o modificación específica de la meiosis, citoquinesis o replicación de ADN). Por ejemplo, la cepa S228C de S. cerevisiae se puede mantener en un estado haploide o diploide. Una célula diploide se refiere a una célula cuyo genoma completo es diploide o se refiere a una célula que es diploide en un locus genómico particular de interés. En algunas formas de realización, la célula fúngica es una célula haploide. Según se usa en la presente, una célula "haploide" se refiere a cualquier célula cuyo genoma está presente en una copia. Una célula haploide se refiere a un tipo de célula que se encuentra naturalmente en un estado haploide, o se refiere a una célula que fue inducida para existir en un estado haploide (por ejemplo, a través de una regulación, alteración, inactivación, activación o modificación específica de la meiosis, citoquinesis o replicación de ADN). Por ejemplo, la cepa S228C de S. cerevisiae se puede mantener en un estado haploide o diploide. Una célula haploide se refiere a una célula cuyo genoma completo es haploide o se refiere a una célula que es haploide en un locus genómico particular de interés.

Según se usa en la presente, un "vector de expresión de levadura" se refiere a un ácido nucleico que contiene una o más secuencias que codifican un ARN y/o un polipéptido y además puede contener cualquier elemento deseado que permita controlar la expresión de dichos uno o más ácidos nucleicos, así como cualquier elementos que permita la replicación y el mantenimiento del vector de expresión en el interior de la célula de levadura. En el arte se conocen muchos vectores de expresión de levadura adecuados y características de los mismos; por ejemplo, se pueden consultar varios vectores y técnicas en Yeast Protocols, 2ª edición, Xiao, W., ed. (Humana Press, Nueva York, 2007) y Buckholz, R.G. y Gleeson, M.A. (1991) Biotechnology (NY) 9(11): 1067-72. Los vectores de levadura pueden contener, en un sentido no taxativo, una secuencia centromérica (CEN), una secuencia de replicación autónoma (ARS), un promotor, tal como un promotor de la ARN polimerasa III, ligado operativamente a una secuencia o gen de interés, un terminador, tal como un terminador de la ARN polimerasa III, un origen de replicación y un gen marcador (por ejemplo, un marcador auxotrófico, antibiótico u otros marcadores seleccionables). Los ejemplos de vectores de expresión de utilidad para levadura pueden incluir plásmidos, cromosomas artificiales de levadura, plásmidos 2µ, plásmidos de integración en levadura, plásmidos de replicación en levadura, vectores ambivalentes y plásmidos episómicos.

Integración estable de los componentes del sistema Cpf1 CRISP en el genoma de plantas y células vegetales

En formas de realización particulares, se contempla que los polinucleótidos que codifican los componentes del sistema Cpf1 CRISPR se introducen para una integración estable en el genoma de una célula vegetal. En estas formas de realización, el diseño del vector de transformación o del sistema de expresión se puede ajustar dependiendo de cuándo, dónde y bajo qué condiciones se expresa el ARN guía y/o el gen de Cpf1.

En formas de realización particulares, se contempla introducir los componentes del sistema Cpf1 CRISPR de manera estable en el ADN genómico de una célula vegetal. Adicionalmente, o como alternativa, se contempla introducir los componentes del sistema Cpf1 CRISPR para su integración estable en el ADN de un organela vegetal tal como, pero en un sentido no taxativo, un plástido, una mitocondria o un cloroplasto.

El sistema de expresión para una integración estable en el genoma de una célula vegetal puede contener uno o más de los siguientes elementos: un elemento promotor que se puede usar para expresar el ARN y/o la enzima Cpf1 en una célula vegetal; una región no traducida 5' para aumentar la expresión; un elemento de intrón para aumentar aún más la expresión en determinadas células, tales como células monocotiledóneas; un sitio de clonación múltiple para proveer sitios de restricción adecuados para la inserción de las secuencias del ARN guía y/o del gen de Cpf1 y otros elementos deseados; y una región no traducida 3' para proveer una terminación eficiente del transcripto expresado.

Los elementos del sistema de expresión se pueden encontrar en una o más construcciones de expresión que pueden ser ya sea circulares, tal como un plásmido o un vector de transformación, o no circulares tal como un ADN de hebra doble lineal.

En una forma de realización particular, un sistema de expresión Cfp1 CRISPR comprende por lo menos:

(a) una secuencia de nucleótidos que codifica un ARN guía (ARNg) que se hibrida con una secuencia diana en una planta, y en donde dicho ARN guía comprende una secuencia guía y una secuencia de repetición directa, y

(b) una secuencia de nucleótidos que codifica una proteína Cpf1,

en donde los componentes (a) o (b) están en la misma construcción o en construcciones diferentes, y con los cuales las diferentes secuencias de nucleótidos se pueden encontrar bajo el control del mismo elemento regulador, o de un elemento diferente, operable en una célula vegetal.

Las construcciones de ADN que contienen los componentes del sistema Cpf1 CRISPR y, cuando corresponda, una secuencia de molde, se pueden introducir en el genoma de una planta, de una parte de planta o de una célula vegetal mediante una variedad de técnicas convencionales. En general, el proceso comprende los pasos de seleccionar una célula huésped o un tejido huésped adecuado, introducir una o más construcciones en la célula huésped o el tejido huésped y regenerar células vegetales o plantas a partir de los mismos.

En formas de realización particulares, la construcción de ADN se puede introducir en la célula vegetal usando técnicas tales como, pero en un sentido no taxativo, electroporación, microinyección, inyección por haz de aerosol de protoplastos de células vegetales, o bien, las construcciones de ADN se pueden introducir directamente en el tejido vegetal usando métodos biolísticos, tal como bombardeo de partículas de ADN (véase también Fu y col., Transgenic Res., febrero de 2000; 9(1): 11-9). El fundamento del bombardeo de partículas es la aceleración de partículas recubiertas con uno o más genes de interés hacia las células, lo que da como resultado la penetración en el protoplasma por las partículas y típicamente una integración estable en el genoma. (véase, por ejemplo, Klein y col., Nature (1987), Klein y col., Bio/Technology (1992), Casas y col., Proc. Natl. Acad. Sci. USA (1993)).

En formas de realización particulares, las construcciones de ADN que contienen los componentes del sistema Cpf1 CRISPR se pueden introducir en la planta por transformación mediada por Agrobacterium. Las construcciones de ADN se pueden combinar con regiones flanqueadoras de ADN-T adecuadas y luego se pueden introducir en un vector huésped de Agrobacterium tumefaciens convencional. El ADN extraño se puede incorporar en el genoma de plantas mediante infección de las plantas o mediante incubación de protoplastos vegetales con bacterias Agrobacterium, que contienen uno o más plásmidos Ti (inductores de tumores). (véase, por ejemplo, Fraley y col., (1985), Rogers y col., (1987) y la Patente de los EE.UU. n.°: 5.563.055).

Promotores de plantas

Con el fin de asegurar una expresión apropiada en una célula vegetal, los componentes del sistema Cpf1 CRISPR descrito en la presente típicamente se ubican bajo el control de un promotor vegetal, es decir un promotor operable en células vegetales. Se contempla el uso de diferentes tipos de promotores.

Un promotor vegetal constitutivo es un promotor capaz de expresar el marco de lectura abierto (ORF) controlado por el mismo en todos o casi todos los tejidos vegetales durante todas o casi todas las etapas del desarrollo de la planta (conocida como una "expresión constitutiva"). Un ejemplo no taxativo de un promotor constitutivo es el promotor 35S del virus en mosaico de la coliflor. Un "promotor regulado" se refiere a un promotor que no dirige la expresión genética de manera constitutiva, sino de una manera regulada temporalmente y/o espacialmente, e incluye promotores específicos de tejidos, con preferencia por tejidos e inducibles. Hay diferentes promotores que pueden dirigir la expresión de un gen en distintos tejidos o tipos celulares, o en diferentes etapas del desarrollo, o como respuesta a diferentes condiciones ambientales. En formas de realización particulares, uno o más de los componentes de Cpf1 CRISPR se expresan bajo el control de un promotor constitutivo, tal como el promotor 35S del virus en mosaico de la coliflor, y se pueden utilizar promotores con preferencia por acciones para dirigir la expresión en determinados tipos celulares en un tejido vegetal particular, por ejemplo las células vasculares en hojas o raíces o en células específicas de semillas. Los ejemplos de promotores particulares para su uso en el sistema Cpf1 CRISPR se pueden consultar en Kawamata y col., (1997) Plant Cell Physiol 38: 792-803; Yamamoto y col., (1997) Plant J 12: 255-65; Hire y col., (1992) Plant Mol Biol 20: 207-18, Kuster y col., (1995) Plant Mol Biol 29: 759-72, y Capana y col., (1994) Plant Mol Biol 25: 681 -91.

Los ejemplos de promotores que son inducibles y que permiten un control espaciotemporal de la edición de genes o la expresión genética pueden emplear una forma de energía. La forma de energía puede incluir, pero en un sentido no taxativo, energía sonora, radiación electromagnética, energía química y/o energía térmica. Los ejemplos de sistemas inducibles incluyen promotores inducibles por tetraciclina (Tet-On o Tet-Off), sistemas de activación de la transcripción de dos híbridos de moléculas pequeñas (FKBP, ABA, etc) o sistemas inducibles por luz (fitocromos, dominios LOV o criptocromos), tal como un efector de la transcripción inducible por luz (LITE) que dirige los cambios en la actividad de transcripción de una manera específica de la secuencia. Los componentes de un sistema inducible por luz pueden incluir una enzima Cpf1 CRISPR, un heterodímero de citocromos que responde a la luz (por ejemplo, de Arabidopsis thaliana) y un dominio de activación/represión de la transcripción. Otros ejemplos de proteínas de unión a ADN inducibles y los métodos para su uso se pueden consultar en US 61/736.465 y US 61/721.283.

En formas de realización particulares, se puede lograr una expresión transitoria o inducible usando, por ejemplo, promotores regulados químicamente, es decir, donde la aplicación de una sustancia química exógena induce la expresión genética. También se puede lograr una modulación de la expresión genética con un promotor químicamente

reprimible, donde la aplicación de una sustancia química reprime la expresión genética. Los promotores inducibles por sustancias químicas incluyen, pero en un sentido no taxativo, el promotor In2-2 de maíz, activado por protectores de herbicidas de bencensulfonamida (De Veylder y col., (1997) Plant Cell Physiol 38: 568-77), el promotor GST de maíz (GST-II-27, WO93/01294), activado por compuestos electrofílicos hidrofóbicos utilizados como herbicidas de preemergencia y el promotor PR-1 de tabaco (Ono y col., (2004) Biosci Biotechnol Biochem 68: 803-7) activado por ácido salicílico. En la presente también se pueden usar los promotores regulados por antibióticos, tales como los promotores inducibles por tetraciclina y reprimibles por tetraciclina (Gatz y col., (1991) Mol Gen Genet 227: 229-37; Patentes de los EE.UU. n.°: 5.814.618 y 5.789.156).

Traslocación a organelas vegetales específicos y/o expresión en los mismos

El sistema de expresión puede comprender elementos para la traslocación a un organela vegetal específico y/o para la expresión en el mismo.

Direccionamiento a cloroplastos

En formas de realización particulares, se contempla utilizar el sistema Cpf1 CRISPR para modificar específicamente los genes de cloroplastos o para asegurar la expresión en el cloroplasto. Para tal fin, se emplean métodos de transformación en cloroplastos o compartimentalización de los componentes de Cpf1 CRISPR en el cloroplasto. Por ejemplo, la introducción de modificaciones genéticas en el genoma de plástidos puede reducir los problemas de bioseguridad, tal como el flujo de genes a través del polen.

Los métodos de transformación en cloroplasto son conocidos en el arte e incluyen bombardeo de partículas, tratamiento con PEG y microinyección. Adicionalmente, se pueden usar métodos que comprenden una traslocación de casetes de transformación desde el genoma nuclear hacia el plástico, según se describe en WO2010061186.

Como alternativa, se contempla direccionar uno o más de los componentes de Cpf1 CRISPR hacia el cloroplasto vegetal. Esto se logra mediante la incorporación en la construcción de expresión de una secuencia que codifica un péptido de tránsito a cloroplastos (CTP) o un péptido de tránsito a plástidos, ligado operativamente a la región 5' de la secuencia que codifica la proteína Cpf1. El CTP es eliminado en un paso de proceso durante la traslocación al cloroplasto. El direccionamiento a cloroplastos de proteínas expresadas es bien conocido por el especialista (véase, por ejemplo, Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology, volumen 61: 157-180). En tales formas de realización también resulta deseable direccionar el ARN guía al cloroplasto vegetal. Los métodos y las construcciones que se pueden usar para traslocar un ARN guía al cloroplasto por medio de una secuencia de localización en cloroplastos se describen, por ejemplo, en US 20040142476. Dichas variaciones de las construcciones se pueden incorporar en los sistemas de expresión de la invención para traslocar eficazmente la construcción de Cpf1-ARN guía.

Introducción de polinucleótidos que codifican el sistema CRISPR-Cpf1 en células de algas.

Las algas transgénicas (u otras plantas tal como colza) pueden ser de particular utilidad en la producción de aceites vegetales o biocombustibles tales como alcoholes (en especial metanol y etanol) u otros productos. Los mismos se pueden modificar para expresar o sobreexpresar niveles altos de aceites o de alcoholes para su uso en las industrias de aceites o de biocombustible.

En US 8945839 se describe un método para modificar especies de microalgas (células de Chlamydomonas reinhardtii) usando Cas9. El uso de herramientas similares, permite aplicar los métodos del sistema Cpf1 CRISPR descrito en la presente en especies de Chlamydomonas y otras algas. En formas de realización particulares, se introducen ambos Cpf1 y ARN guía en algas y se expresan usando un vector que expresa al Cpf1 bajo el control de un promotor constitutivo tal como de Hsp70A-Rbc S2 o de Beta2-tubulina. El ARN guía es suministrado opcionalmente usando un vector que contiene al promotor T7. Como alternativa, ambos ARNm de Cas9 y el ARN guía transcrito in vitro se pueden suministrar en las células de algas. Hay protocolos de electroporación disponibles para el especialista, tal como el protocolo estándar recomendado en el conjunto de elementos para modificación de Chlamydomonas de GeneArt.

En formas de realización particulares, la endonucleasa usada en la presente es una enzima Cpf1 de partición [split]. Las enzimas Cpf1 de corte se utilizan preferencialmente en Algas para una modificación dirigida del genoma como se describió para Cas9 en WO 2015086795. El uso del sistema Cpf1 de corte es particularmente adecuado para un método inducible de direccionamiento de genomas y evita el potencial efecto tóxico de una sobreexpresión de Cpf1 en las células de las algas. En formas de realización particulares, dichos dominios Cpf1 de corte (los dominios RuvC y HNH) se pueden introducir de manera simultánea o sucesiva en la célula de modo que dichos dominios Cpf1 de corte procesarán la secuencia de ácidos nucleicos diana en la célula de alga. El tamaño reducido de la Cpf1 de corte en comparación con la Cpf1 de tipo salvaje permite el uso de otros métodos de suministro del sistema CRISPR a las células, tal como el uso de los péptidos penetradores en células que se describen en la presente. Este método es de particular interés para generar algas modificadas genéticamente.

Introducción de polinucleótidos que codifican componentes de Cpf1 en células de levadura

En formas de realización particulares, los métodos y composiciones reivindicados de la invención implican el uso del

sistema Cpf1 CRISPR en la edición de genomas de células de levadura. Los métodos para transformar células de levadura que se pueden usar para introducir polinucleótidos que codifican componentes del sistema Cpf1 CRISPR son bien conocidos por el especialista y fueron revisados por Kawai y col., 2010, Bioeng Bugs, nov-dic de 2010; 1(6): 395-403). Los ejemplos no taxativos incluyen la transformación de células de levadura mediante tratamiento con acetato de litio (que además puede incluir un ADN transportador y tratamiento con PEG), bombardeo o mediante electroporación.

Expresión transitoria de los componentes del sistema Cpf1 CRISP en plantas y células vegetales

En formas de realización particulares, se contempla expresar transitoriamente al ARN guía y/o al gen Cpf1 en la célula vegetal. En estas formas de realización, el sistema Cpf1 CRISPR permite asegurar la modificación de un gen diana solamente cuando ambos ARN guía y la proteína Cpf1 están presentes en una célula, de modo que se puede asegurar adicionalmente la modificación genómica. Dado que la expresión de la enzima Cpf1 es transitoria, las plantas regeneradas a partir de dichas células vegetales típicamente no contienen un ADN extraño. En formas de realización particulares, la enzima Cpf1 es expresada de manera estable por la célula vegetal y la secuencia guía es expresada transitoriamente.

En formas de realización particulares, los componentes del sistema Cpf1 CRISPR se pueden introducir en las células vegetales usando una vector viral para plantas (Scholthof y col., 1996, Annu Rev Phytopathol., 1996; 34: 299-323). En otras formas de realización particulares, dicho vector viral es un vector de un virus a ADN. Por ejemplo, geminivirus (por ejemplo, virus del enrollamiento foliar de repollo, virus enano amarillo de haba, virus enano de sorgo, virus del enrollamiento foliar de tomate, virus estriado de maíz, virus del enrollamiento foliar de tabaco o virus en mosaico dorado de tomate) o nanovirus (por ejemplo, virus amarillo necrótico de poroto). En otras formas de realización particulares, dicho vector viral es un vector de un virus a ARN. Por ejemplo, un tobravirus (por ejemplo, virus del cascabel de tabaco, virus en mosaico de tabaco), un potexvirus (por ejemplo, un virus X de la patata) o un hordeivirus (por ejemplo, virus en mosaico estriado de cebada). Los genomas en replicación de los virus vegetales son vectores no integradores.

En formas de realización particulares, el vector usado para la expresión transitoria de las construcciones Cpf1 CRISPR es, por ejemplo, un vector pEAQ, que está adaptado para una expresión transitoria mediada por Agrobacterium (Sainsbury F. y col., Plant Biotechnol J., septiembre de 2009; 7(7): 682-93) en el protoplasto. El direccionamiento preciso de las localizaciones genómicas se demostró usando un vector del virus del enrollamiento foliar de repollo (CaLCuV) modificado para expresar los ARNg en plantas transgénicas estables que expresan una enzima CRISPR (Scientific Reports 5, N° Artículo: 14926 (2015), doi: 10.1038/srep14926).

En formas de realización particulares, se pueden introducir transitoriamente fragmentos de ADN de hebra doble que codifican el ARN guía y/o el gen de Cpf1 en la célula vegetal. En dichas formas de realización, los fragmentos de ADN de hebra doble introducidos se proveen en una cantidad suficiente como para modificar la célula pero no persisten una vez transcurrido un período de tiempo contemplado o después de una o más divisiones celulares. Los métodos de una transferencia directa de ADN en plantas son conocidos por el especialista (véase, por ejemplo, Davey y col., Plant Mol Biol., septiembre de 1989; 13(3): 273-85.)

En otras formas de realización, se introduce un polinucleótido de ARN que codifica la proteína Cpf1 en la célula vegetal, que luego es traducido y procesado por la célula huésped para generar la proteína en una cantidad suficiente como para modificar la célula (en la presencia de por lo menos un ARN guía) pero que no permanece allí una vez transcurrido un período de tiempo contemplado o después de una o más divisiones celulares. Los métodos para introducir ARNm en protoplastos vegetales para una expresión transitoria son conocidos por el especialista (véase, por ejemplo, en Gallie, Plant Cell Reports (1993), 13; 119-122).

También se contemplan combinaciones de los diferentes métodos descritos previamente.

Suministro de los componentes de Cpf1 CRISPR a la célula vegetal

En formas de realización particulares, resulta de interés proveer uno o más componentes del sistema Cpf1 CRISPR directamente en la célula vegetal. Esto es importante, entre otros, para la generación de plantas no transgénicas (véase más adelante). En formas de realización particulares, se prepara uno o más de los componentes Cpf1 fuera de la planta o célula vegetal y luego se suministran en la célula. Por ejemplo, en formas de realización particulares, la proteína Cpf1 se prepara in vitro antes de su introducción en la célula vegetal. La proteína Cpf1 se puede preparar mediante varios métodos conocidos por un especialista en el arte e incluyen una producción recombinante. Después de la expresión, se aísla la proteína Cpf1, se repliega nuevamente si fuera necesario, se purifica y opcionalmente es tratada para eliminar toda marca de purificación, tal como una marca His. Una vez obtenida la proteína Cpf1 cruda, parcialmente purificada, o más completamente purificada, dicha proteína se puede introducir en la célula vegetal.

En formas de realización particulares, la proteína Cpf1 se mezcla con el ARN guía dirigido al gen de interés para formar una ribonucleoproteína preensamblada.

Los componentes individuales o preensamblados de la ribonucleoproteína se pueden introducir en la célula vegetal pro electroporación, mediante bombardeo con partículas recubiertas con el producto genético asociado a Cpf1,

mediante transfección química o mediante algún otro medio de transporte a través de una membrana celular. Por ejemplo, se ha demostrado que la transfección de un protoplasto vegetal con una ribonucleoproteína CRISPR preensamblada permite asegurar una modificación dirigida del genoma vegetal (como se describe en Woo y col., Nature Biotechnology, 2015; DOI: 10.1038/nbt.3389).

En formas de realización particulares, los componentes del sistema Cpf1 CRISPR se introducen en las células vegetales usando nanopartículas. Los componentes, ya sea en la forma de una proteína o de un ácido nucleico o como una combinación de los mismos, se puede cargar o empaquetar sobre nanopartículas y luego se pueden aplicar a las plantas (tal como se describe, por ejemplo, en WO 2008042156 y en US 20130185823). En particular, las formas de realización de los métodos y composiciones reivindicados de la invención comprenden nanopartículas cargadas o empaquetadas con moléculas de ADN que codifican la proteína Cpf1. Las moléculas de ADN que codifican al ARN guía y/o un ARN guía aislado se describen en WO2015089419.

Otros medios para introducir uno o más componentes del sistema Cpf1 CRISPR en la célula vegetal es mediante el uso de péptidos penetradores de células (CPP). Por lo tanto, en particular, las formas de realización de los métodos y composiciones reivindicados de la invención comprenden el uso de composiciones que comprenden un péptido penetrador de células ligado a la proteína Cpf1. En formas de realización particulares de la presente invención, la proteína Cpf1 y/o el ARN guía se acoplan a uno o más CPP para transportarlos eficazmente al interior de los protoplastos vegetales; véase también, Ramakrishna (20140, Genome Res., junio de 2014; 24(6): 1020-7 por Cas9 en células humanas). En otras formas de realización, el gen de Cpf1 y/o el ARN guía están codificados por una o más moléculas de ADN circulares o no circulares que están acoplados a uno o más CPP para suministrarlos en un protoplasto vegetal. A continuación, los protoplastos vegetales son regenerados en células vegetales y luego en plantas. Los CPP generalmente se describen como péptidos cortos de menos de 35 aminoácidos derivados ya sea de proteínas o de secuencias quiméricas capaces de transportar biomoléculas a través de una membrana celular de una manera independiente de receptores. El CPP puede comprender péptidos catiónicos, péptidos con secuencias hidrofóbicas, péptidos anfipáticos, péptidos que tienen una secuencia rica en prolina y antimicrobiana y péptidos quiméricos o bipartitos (Pooga y Langel 2005). Los CPP pueden penetrar las membranas biológicas y como tales disparan el movimiento de varias biomoléculas a través de las membranas celulares hacia el citoplasma y mejoran su direccionamiento intracelular, y por ende facilitan la interacción de la biomolécula con el diana. Los ejemplos de CPP incluyen, entre otros: Tat, una proteína activadora de la transcripción nuclear necesaria para la replicación viral por la penetratina, tipo 1 del VIH, la secuencia del péptido señal del factor de crecimiento de fibroblastos (FGF) de Kaposi, la secuencia del péptido señal de la integrina β3; la secuencia Args del péptido de poliarginina, los transportadores moleculares ricos en guanina, el péptido sweet arrow, etc.

Uso del sistema Cpf1 CRISPR para obtener plantas no transgénicas modificadas genéticamente

En formas de realización particulares, los métodos que se describen en la presente se usan para modificar genes endógenos o para modificar su expresión sin la introducción permanente en el genoma de la planta de cualquier gen extraño, incluyendo aquellos que codifican componentes de CRISPR, para de esa manera evitar la presencia de ADN extraño en el genoma de la planta. Esto puede ser de interés ya que los requerimientos regulatorios de plantas no transgénicas son menos rigurosos.

En formas de realización particulares, esto se asegura a través de la expresión transitoria de los componentes de Cpf1 CRISPR. En formas de realización particulares, se expresa uno o más de los componentes de CRISPR en uno o más vectores virales que producen suficiente proteína Cpf1 y ARN guía como para asegurar de manera coherentemente constante la modificación de un gen de interés de acuerdo con un método descrito en la presente.

En formas de realización particulares, se asegura la expresión transitoria de las construcciones Cpf1 CRISPR en protoplastos vegetales y por lo tanto no se integran en el genoma. La ventana de expresión limitada puede ser suficiente para permitir que el sistema Cpf1 CRISPR asegure la modificación de un gen diana como se describe en la presente.

En formas de realización particulares, los diferentes componentes del sistema Cpf1 CRISPR se introducen en la célula vegetal, protoplasto o tejido vegetal ya sea por separado o en una mezcla, con la ayuda de moléculas de suministro particuladas tales como nanopartículas o moléculas de CPP como se describió precedentemente en la presente.

La expresión de los componentes de Cpf1 CRISPR puede inducir una modificación dirigida del genoma, ya sea por actividad directa de la nucleasa Cpf1 y opcionalmente por introducción de un molde de ADN o mediante la modificación de genes buscados usando el sistema Cpf1 CRISPR como se describe en la presente. Las diferentes estrategias que se describieron antes en la presente permiten una edición de los genomas buscados mediada por Cpf1 sin necesidad de introducir los componentes de Cpf1 CRISPR en el genoma vegetal. Los componentes que se introducen de manera transitoria en la célula vegetal típicamente se eliminan con el cruzamiento.

Detección de las modificaciones en los marcadores seleccionables del genoma vegetal

En formas de realización particulares, cuando el método comprende la modificación de un gen diana endógeno del genoma vegetal, se puede emplear cualquier método adecuado para determinar, después de infectar o transfectar la planta, parte de planta o célula vegetal con el sistema Cpf1 CRISPR, si ha tenido lugar el direccionamiento o la

mutagénesis dirigida del gen en el sitio diana. Cuando el método comprende la introducción de un transgen, se puede identificar y aislar la célula vegetal, callo, tejido o planta transformada mediante selección o examen del material vegetal modificado por la presencia del transgen o por características codificadas por el transgen. Se pueden usar métodos físicos y bioquímicos para identificar transformantes de plantas o células vegetales que contienen las construcciones genéticas insertadas o una modificación de ADN endógena. Estos métodos incluyen, pero en un sentido no taxativo: 1) análisis Southern o amplificación por PCR para detectar y determinar la estructura del inserto de ADN recombinante o genes endógenos modificados; 2) transferencia Northern, protección con S1 RNasa, extensión con cebadores o amplificación por transcriptasa inversa-PCR para detectar y examinar transcriptos de ARN de las construcciones genéticas; 3) ensayos enzimáticos para detectar la actividad de enzimas o ribozimas, donde dichos productos genéticos están codificados por la construcción genética o la expresión es afectada por la modificación genética; 4) electroforesis sobre gel-proteínas, técnicas de transferencia Western, inmunoprecipitación o inmunoensayos ligados a enzimas, donde los productos de la construcción genética o del gen endógeno son proteínas. También se pueden usar técnicas adicionales, tales como hibridación in situ, coloración con enzimas e inmunotinción, para detectar la presencia o expresión de la construcción recombinante o para detectar la modificación de un gen endógeno en órganos y tejidos vegetales específicos. Los métodos para conducir todos estos ensayos son bien conocidos por los especialistas en el arte.

Adicionalmente (o como alternativa), el sistema de expresión que codifica los componentes de Cpf1 CRISPR típicamente está diseñado para comprender uno o más marcadores seleccionables o detectables que proveen un medio para aislar o seleccionar eficazmente células que contienen y/o que fueron modificados por el sistema Cpf1 CRISPR en una etapa temprana y a gran escala.

En el caso de una transformación mediada por Agrobacterium, el casete marcador puede estar adyacente o entre los bordes flanqueadores de ADN-T y puede estar en un vector binario. En otra forma de realización, el casete marcador puede estar fuera del ADN-T. El casete marcador seleccionable también puede estar dentro o adyacente a los mismos bordes de ADN-T que el casete de expresión o puede estar en otra parte en un segundo ADN-T en el vector binario (por ejemplo, un sistema 2 ADN-T).

Para un bombardeo de partículas o una transformación de protoplastos, el sistema de expresión puede comprender uno o más fragmentos lineales aislados o puede formar parte de una construcción más grande que podría contener elementos de replicación bacterianos, marcadores seleccionables bacterianos u otros elementos detectables. Los casetes de expresión que comprenden los polinucleótidos que codifican la guía y/o Cpf1 se pueden unir físicamente a un casete marcador o se pueden mezclar con una segunda molécula de ácido nucleico que codifica un casete marcador. El casete marcador está compuesto por los elementos necesarios para expresar un marcador detectable o seleccionable que permita una selección eficiente de las células transformadas.

El procedimiento de selección de células basado en el marcador seleccionable dependerá de la naturaleza del gen marcador. En formas de realización particulares, se emplea un marcador seleccionable, es decir un marcador que permitirá dirigir la selección de las células basado en la expresión del marcador. Un marcador seleccionable puede conferir una selección positiva o negativa y su presencia depende, o no, de sustratos externos (Miki y col., 2004, 107(3): 193-232). Más comúnmente, se usan genes de resistencia a antibióticos o herbicidas como marcadores, con lo cual la selección se efectúa cultivando el material vegetal modificado sobre un medio que contiene una cantidad inhibidora de los antibióticos o herbicidas para los cuales el gen marcador confiere resistencia. Los ejemplos de dichos genes son genes que confieren resistencia a antibióticos tales como higromicina (hpt) y kanamicina (nptII), y genes que confieren resistencia a herbicidas tales como fosfinotricina (bar) y clorosulfurón (als),

Las plantas y células vegetales transformadas también se pueden identificar mediante selección de las actividades de un marcador visible, típicamente una enzima capaz de procesar un sustrato de color (por ejemplo, los genes de β-glucuronidasa, luciferasa, B o C1). Dichas metodologías de selección y examen son bien conocidas por los especialistas en el arte.

Cultivos y regeneración de plantas

En formas de realización particulares, las células vegetales cuyos genomas fueron modificados y que se producen o se obtienen mediante cualquiera de los métodos descritos en la presente, se pueden cultivar para regenerar una planta completa que tenga el genotipo transformado o modificado y, por lo tanto, el fenotipo deseado. Las técnicas de regeneración convencionales son bien conocidas por los especialistas en el arte. Los ejemplos particulares de dichas técnicas de regeneración se basan en la manipulación de ciertas fitohormonas en un medio de cultivo para el crecimiento de tejidos y típicamente se basan en un marcador biocida y/o herbicida que se ha introducido junto con las secuencias de nucleótido deseadas. En otras formas de realización particulares, la regeneración de plantas se obtiene a partir de protoplastos, callos vegetales, explantes, órganos, polen, embriones o partes de los mismos cultivados (véase, por ejemplo, Evans y col., (1983), Handbook of Plant Cell Culture, Klee et al (1987) Ann. Rev. of Plant Phys.).

En formas de realización particulares, las plantas transformadas o mejoradas que se describen en la presente se pueden autopolinizar para proveer semillas de las plantas homocigotas mejoradas de la invención (homocigotas para la modificación de ADN) o se pueden cruzar con plantas no transgénicas o plantas mejoradas diferentes para proveer semillas de las plantas heterocigotas. Cuando se introdujo un ADN recombinante en la célula vegetal, la planta resultante de dicho cruzamiento es una planta que es heterocigota para la molécula de ADN recombinante. Ambas plantas homocigotas y heterocigotas obtenidas mediante cruzamiento a partir de las plantas mejoradas y que comprenden la modificación genética (que puede ser un ADN recombinante) se conocen como la "progenie" en la presente. Las plantas de la progenie son plantas que descienden de la planta transgénica original y que contienen la modificación del genoma o la molécula de ADN recombinante introducida mediante los métodos provistos en la presente. Como alternativa, las plantas modificadas genéticamente se pueden obtener mediante uno de los métodos descritos supra usando la enzima Cfp1, con lo cual no se incorpora ningún ADN extraño en el genoma. La progenie de dichas plantas, obtenida mediante una cría adicional, también puede contener la modificación genética. Las crías se conducen mediante cualquiera de los métodos de cría usados comúnmente para diferentes cultivos (por ejemplo, Allard, Principles of Plant Breeding, John Wiley & Sons, NY, U. de CA, Davis, CA, 50-98 (1960).

Generación de plantas con características agronómicas mejoradas

Los sistemas CRISPR basados en Cpf1 que se proveen en la presente se pueden usar para introducir roturas dirigidas de hebra doble o de hebra simple y/o para introducir sistemas activadores y/o represores de genes y, en un sentido no limitante, se pueden usar en el direccionamiento de genes, el reemplazo de genes, una mutagénesis dirigida, supresiones o inserciones dirigidas, inversiones dirigidas y/o traslocaciones dirigidas. La coexpresión de múltiples ARN de direccionamiento dirigidos a lograr múltiples modificaciones en una sola célula, permite asegurar una modificación de genoma multiplexada. Esta tecnología se puede usar para una manipulación de gran precisión de plantas con características mejoradas, incluyendo una calidad nutricional mejorada, una mayor resistencia a enfermedades y resistencia a un estrés biótico y abiótico, y una mayor producción de productos vegetales o de compuestos heterólogos comercialmente valiosos.

En formas de realización particulares, el sistema Cpf1 CRISPR que se describe en la presente se usa para introducir una rotura de hebra doble (DSB) dirigida en una secuencia de ADN endógena. La DSB activa las vías celulares de Reparación de ADN, que se pueden implementar para lograr las modificaciones de la secuencia de ADN deseadas cerca del sitio de rotura. Esto es de interés cuando la inactivación de genes endógenos puede conferir o contribuir en una característica deseada. En formas de realización particulares, se promueve una recombinación homóloga con una secuencia de molde en el sitio de la DSB, con el fin de introducir un gen de interés.

En formas de realización particulares, el sistema Cpf1 CRISPR se puede usar como una proteína de unión a ácidos nucleicos genérica, fusionado o ligado operativamente a un dominio funcional para la activación y/o represión de genes vegetales endógenos. Los ejemplos de dominios funcionales pueden incluir, pero en un sentido no taxativo, un iniciador de la traducción, un activador de la traducción, un represor de la traducción, nucleasas, en particular ribonucleasas, un spliceosome, esferas, un dominio inducible/controlable por luz o un dominio inducible/controlable químicamente. Típicamente, en estas formas de realización, la proteína Cpf1 comprende por lo menos una mutación, de modo tal que no tiene más de un 5% de la actividad de la proteína Cpf1 que no comprende dicha por lo menos una mutación; el ARN guía comprende una secuencia guía capaz de hibridarse con una secuencia diana.

Los métodos que se describen en la presente generalmente dan como resultado la generación de "plantas mejoradas" por cuanto presentan una o más características deseables en comparación con la planta de tipo salvaje. En formas de realización particulares, las plantas, células vegetales o partes de plantas obtenidas son plantas transgénicas, que comprenden una secuencia de ADN exógena incorporada en el genoma de todas o parte de las células de la planta. En formas de realización particulares, se obtienen plantas, partes de plantas o células no transgénicas modificadas genéticamente, por cuanto no se incorpora una secuencia de ADN exógena en el genoma de cualquiera de las células vegetales de la planta. En dichas formas de realización, las plantas mejoradas no son transgénicas. Cuando solamente se asegura la modificación de un gen endógeno y no se introducen o conservan genes extraños en el genoma vegetal, los cultivos modificados genéticamente resultantes no contienen genes extraños y por lo tanto se pueden considerar básicamente como no transgénicos. Las diferentes aplicaciones del sistema Cpf1 CRISPR para la edición de genomas vegetales se describirán con más detalle a continuación:

a) introducción de uno o más genes extraños para conferir una característica agronómica de interés

Los métodos y composiciones reivindicados de la invención pueden implicar métodos de edición de genomas o de modificación de las secuencias asociadas o presentes en un locus diana de interés, en donde el método comprende introducir un complejo de proteína efectora Cpf1 en una célula vegetal, con lo cual el complejo de la proteína efectora Cpf1 funciona eficazmente en la integración de un inserto de ADN que codifica, por ejemplo, un gen extraño de interés, en el genoma de la célula vegetal. En formas de realización preferidas, la integración del inserto de ADN es facilitada por HR con un molde o molde de reparación de ADN introducido exógenamente. Típicamente, el molde o molde de reparación de ADN introducido exógenamente se suministra junto con el complejo de proteína efectora Cpf1 o un componente o un vector de polinucleótidos para la expresión de un componente del complejo.

Los sistemas Cpf1 CRISPR provistos en la presente permiten un suministro de genes dirigido. Se ha vuelto cada vez más evidente que la eficacia de expresar un gen de interés está determinada en gran medida por la ubicación de la integración en el genoma. Los métodos de la presente permiten una integración dirigida del gen extraño en una ubicación deseada en el genoma. La ubicación se puede seleccionar basado en información de eventos generados previamente o se puede seleccionar mediante los métodos divulgados en otra parte en la presente.

En formas de realización particulares, los métodos provistos en la presente incluyen (a) introducir en la célula un complejo Cpf1 CRISPR que comprende un ARN guía, que comprende una repetición directa y una secuencia guía, en donde la secuencia guía se hibrida con una secuencia diana que es endógena para la célula vegetal; (b) introducir en la célula vegetal una molécula efectora Cpf1 que forma un complejo con el ARN guía cuando la secuencia guía se hibrida con la secuencia diana e induce una rotura de hebra doble en o cerca de la secuencia a la cual está dirigida la secuencia guía; y (c) introducir en la célula una secuencia de nucleótidos que codifica un molde de reparación HDR que codifica al gen de interés y que se introduce en la ubicación de la rotura DS como resultado de HDR. En formas de realización particulares, el paso de introducción puede incluir suministrar en la célula vegetal uno o más polinucleótidos que codifican la proteína efectora Cpf1, el ARN guía y el molde de reparación. En formas de realización particulares, los polinucleótidos son suministrados en la célula mediante un virus a ADN (por ejemplo, un geminivirus) o un virus a ARN (por ejemplo, un tobravirus). En formas de realización particulares, los pasos de introducción incluyen suministrar en la célula vegetal un ADN-T que contiene una o más secuencias de polinucleótidos que codifican la proteína efectora Cpf1, el ARN guía y el molde de reparación, donde el suministro se realiza mediante Agrobacterium. La secuencia de ácidos nucleicos que codifica la proteína efectora Cpf1 se puede ligar operativamente a un promotor, tal como un promotor constitutivo (por ejemplo, un promotor 35S del virus en mosaico de la coliflor) o un promotor inducible o específico de células. En formas de realización particulares, el polinucleótido se introduce mediante bombardeo de microproyectiles. En formas de realización particulares, el método incluye además seleccionar la célula vegetal después de los pasos de introducción para determinar si se introdujo el molde de reparación, es decir, el gen de interés. En formas de realización particulares, los métodos incluyen el paso de regenerar una planta a partir de la célula vegetal. En formas de realización adicionales, los métodos incluyen el cruzamiento de la planta para obtener un linaje de plantas genéticamente deseadas. Los ejemplos de genes extraños que codifican una característica de interés se enumeran más adelante.

b) edición de genes endógenos para conferir una característica agronómica de interés

Los métodos y composiciones reivindicados pueden implicar la edición de genomas o de modificación de las secuencias asociadas con un locus diana de interés o presentes en el mismo, en donde dicho método comprende introducir un complejo de proteína efectora Cpf1 en una célula vegetal, con lo cual el complejo Cpf1 modifica la expresión de un gen endógeno de la planta. Esto se puede lograr de diferentes maneras. En formas de realización particulares, resulta deseable eliminar la expresión de un gen endógeno y se usa el complejo Cpf1 CRISPR para dirigir y clivar un gen endógeno para así modificar la expresión genética. En estas formas de realización, los métodos provistos en la presente incluyen (a) introducir en la célula vegetal un complejo Cpf1 CRISPR que comprende un ARN guía, que comprende una repetición directa y una secuencia guía, en donde la secuencia guía se hibrida con una secuencia diana en un gen de interés en el genoma de la célula vegetal; y (b) introducir en la célula una proteína efectora Cpf1 que con la unión al ARN guía comprende una secuencia guía que se hibrida con la secuencia diana, asegura una rotura de hebra doble en o cerca de la secuencia a la cual es dirigida la secuencia guía. En formas de realización particulares, el paso de introducción puede incluir suministrar en la célula vegetal uno o más polinucleótidos que codifican la proteína efectora Cpf1 y el ARN guía.

En formas de realización particulares, los polinucleótidos son suministrados en la célula mediante un virus a ADN (por ejemplo, un geminivirus) o un virus a ARN (por ejemplo, un tobravirus). En formas de realización particulares, los pasos de introducción incluyen suministrar en la célula vegetal un ADN-T que contiene una o más secuencias de polinucleótidos que codifican la proteína efectora Cpf1 y el ARN guía, donde el suministro se realiza mediante Agrobacterium. La secuencia de polinucleótidos que codifica los componentes del sistema Cpf1 CRISPR se puede ligar operativamente a un promotor, tal como un promotor constitutivo (por ejemplo, un promotor 35S del virus en mosaico de la coliflor) o un promotor inducible o específico de células. En formas de realización particulares, el polinucleótido se introduce mediante bombardeo de microproyectiles. En formas de realización particulares, el método incluye además seleccionar la célula vegetal después de los pasos de introducción para determinar si se ha modificado la expresión del gen de interés. En formas de realización particulares, los métodos incluyen el paso de regenerar una planta a partir de la célula vegetal. En formas de realización adicionales, los métodos incluyen el cruzamiento de la

planta para obtener un linaje de plantas genéticamente deseadas.

En formas de realización particulares de los métodos descritos previamente, se obtienen cultivos resistentes a enfermedades mediante mutación dirigida de genes de susceptibilidad a enfermedades o genes que codifican reguladores negativos (por ejemplo, el gen Mlo) de genes defensa de plantas. En una forma de realización particular, se generan cultivos tolerantes a herbicidas mediante sustitución dirigida de nucleótidos específicos en los genes vegetales, tales como los que codifican acetolactato sintasa (ALS) y protoporfirinógeno oxidasa (PPO). En formas de realización particulares, se generan cultivos tolerantes a sequía y sales mediante mutación dirigida de genes que codifican reguladores negativos de tolerancia a estrés abiótico, granos bajos en amilosa mediante mutación dirigida del gen Waxy, arroz u otros granos con rancidez reducida mediante mutación dirigida de los principales genes de lipasa en la capa de aleurona, etc. Más adelante se ofrece un listado más extenso de genes endógenos que codifican características de interés.

c) modulación de genes endógenos mediante el sistema Cpf1 CRISPR para conferir una característica agronómica de interés

En la presente también se proveen métodos para modular (es decir, activar o reprimir) la expresión de genes endógenos usando la proteína Cpf1 provista en la presente. Dichos métodos emplean las distintas secuencias de ARN que son dirigidas al genoma vegetal mediante el complejo Cpf1. Más particularmente, las distintas secuencias de ARN se unen a dos o más proteínas adaptadoras (por ejemplo, aptámeros) con lo cual cada proteína adaptadora está asociada con uno o más dominios funcionales y en donde por lo menos uno de dichos uno o más dominios funcionales asociadas con la proteína adaptadora presentan una o más actividades que comprenden actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad de factor de liberación de la transcripción, actividad de modificación de histonas, actividad de integración de ADN, actividad de clivaje de ARN, actividad de clivaje de ADN o actividad de unión a ácidos nucleicos; Los dominios funcionales se usan para modular la expresión de un gen vegetal endógeno para así obtener la característica deseada. Típicamente, en estas formas de realización, la proteína efectora Cpf1 contiene una o más mutaciones de modo tal que no tiene más de un 5% de la actividad nucleasa de la proteína efectora Cpf1 que no comprende dicha por lo menos una mutación.

En formas de realización particulares, los métodos provistos en la presente incluyen los pasos de (a) introducir en la célula un complejo Cpf1 CRISPR que comprende un ARN guía, que comprende una repetición directa y una secuencia guía, en donde la secuencia guía se hibrida con una secuencia diana que es endógena de la célula vegetal; (b) introducir en la célula vegetal una molécula efectora Cpf1 que forma un complejo con el ARN guía cuando la secuencia guía se hibrida con la secuencia diana; y en donde el ARN guía se modifica para comprender una secuencia de ARN distinta (aptámero) que se une a un dominio funcional y/o la proteína efectora Cpf1 se modifica por cuanto se une a un dominio funcional. En formas de realización particulares, el paso de introducción puede incluir suministrar en la célula vegetal uno o más polinucleótidos que codifican la proteína efectora Cpf1 (modificada) y el ARN guía (modificado). Los detalles de los componentes del sistema Cpf1 CRISPR para su uso en estos métodos se describen en otra parte en la presente.

En formas de realización particulares, los polinucleótidos son suministrados en la célula mediante un virus a ADN (por ejemplo, un geminivirus) o un virus a ARN (por ejemplo, un tobravirus). En formas de realización particulares, los pasos de introducción incluyen suministrar en la célula vegetal un ADN-T que contiene una o más secuencias de polinucleótidos que codifican la proteína efectora Cpf1 y el ARN guía, donde el suministro se realiza mediante Agrobacterium. La secuencia de ácidos nucleicos que codifica dichos uno o más componentes del sistema Cpf1 CRISPR se puede ligar operativamente a un promotor, tal como un promotor constitutivo (por ejemplo, un promotor 35S del virus en mosaico de la coliflor) o un promotor inducible o específico de células. En formas de realización particulares, el polinucleótido se introduce mediante bombardeo de microproyectiles. En formas de realización particulares, el método incluye además seleccionar la célula vegetal después de los pasos de introducción para determinar si se ha modificado la expresión del gen de interés. En formas de realización particulares, los métodos incluyen el paso de regenerar una planta a partir de la célula vegetal. En formas de realización adicionales, los métodos incluyen el cruzamiento de la planta para obtener un linaje de plantas genéticamente deseadas. Más adelante se ofrece un listado más extenso de genes endógenos que codifican características de interés.

<u>Uso de Cpf1 para modificar plantas poliploides</u>

Muchas plantas son poliploides, lo que significa que son portadoras de copias duplicadas de sus genomas -a veces tanto como seis, como es el caso del sorgo. Los métodos reivindicados de acuerdo con la presente invención, que emplean la proteína efectora Cpf1 CRISPR que se pueden "multiplexar" para afectar todas las copias de un gen, o para buscar como diana docenas de genes por vez. Por ejemplo, en formas de realización particulares, los métodos de la presente invención se utilizan para asegurar simultáneamente una mutación de pérdida de función en los diferentes genes responsables de suprimir las defensas contra una enfermedad. En formas de realización particulares, los métodos de la presente invención se usan para suprimir simultáneamente la expresión de las secuencias de ácidos nucleicos TaMLO-AI, TaMLO-BI y TaMLO-DI en una célula vegetal de sorgo y para regenerar una planta de sorgo a partir de la misma, con el fin de asegurar que la planta de sorgo sea resistente al mildiú pulverulento (véase también WO2015109752).

<u>Ejemplos de genes que confieren características agronómicas</u>

Según se describió previamente en la presente, en formas de realización particulares, la invención abarca el uso del sistema Cpf1 CRISPR descrito en la presente para la inserción de un ADN de interés, incluyendo uno o más genes que se pueden expresar en plantas). En formas de realización particulares adicionales, la invención abarca métodos y herramientas que emplean el sistema Cpf1 descrito en la presente para una supresión parcial o completa de uno o más genes que se pueden expresar en plantas. En otras formas de realización particulares adicionales, la invención abarca métodos y herramientas que emplean el sistema Cpf1 descrito en la presente para asegurar la modificación de uno o más genes que se expresan en plantas mediante mutación, sustitución, inserción de uno o más nucleótidos. En otras formas de realización particulares, la invención abarca el uso del sistema Cpf1 CRISPR descrito en la presente para asegurar la modificación de la expresión de uno o más genes expresados en plantas mediante la modificación específica de uno o más de los elementos reguladores que dirigen la expresión de dichos genes.

En formas de realización particulares, la invención abarca métodos que comprenden la introducción de genes exógenos y/o el direccionamiento de genes endógenos y sus elementos reguladores, tal como se enumeran a continuación:

1. Genes que confieren resistencia a plagas o enfermedades:

- Genes de resistencia a enfermedades de plantas. Se puede transformar una planta con genes de resistencia clonados para diseñar plantas resistentes a cepas patógenas específicas. Véase, por ejemplo, Jones y col., Science 266: 789 (1994) (clonación del gen Cf-9 de tomate para la resistencia a Cladosporium fulvum); Martin y col., Science 262: 1432 (1993) (gen Pto de tomate para la resistencia a Pseudomonas syringae que codifica una proteína quinasa de tomate); Mindrinos y col., Cell 78: 1089 (1994) (el gen RSP2 de Arabidopsis para la resistencia a Pseudomonas syringae).

- Genes que confieren resistencia a una plaga, tal como el nematodo quístico de soja. Véase por ejemplo, la Solicitud PCT WO 96/30517; la Solicitud PCT WO 93/19181.

- Proteínas de *Bacillus thuringiensis*; véase, por ejemplo, Geiser y col., *Gene* 48: 109 (1986).

- Lectinas, véase, por ejemplo, Van Damme y col., *Plant Molec. Biol.* 24: 25 (1994).

- Proteína de unión a vitaminas, tal como la avidina, véase la Solicitud PCT US93/06487, que divulga el uso de avidina y homólogos de avidina como larvicidas contra plagas de insectos.

- Inhibidores de enzimas, tales como inhibidores de proteasas o proteinasas o inhibidores de amilasa. Véase, por ejemplo, Abe y col., *J. Biol. Chem.* 262: 16793 (1987), Huub y col., *Plant Molec. Biol.* 21: 985 (1993)), Sumitani y col., *Biosci. Biotech. Biochem.* 57: 1243 (1993) y la Patente de los EE.UU. n.°: 5.494.813.

- Hormonas o feromonas específicas de insectos, tales como ecdisteroides o la hormona juvenil, una variante de las mismas, un mimético de las mismas o un antagonista o un agonista de las mismas. Véase, por ejemplo Hammock y col., *Nature* 344: 458 (1990).

- Péptidos o neuropéptidos específicos de insectos que, tras su expresión, alteran la fisiología de la plaga afectada. Por ejemplo Regan, *J. Biol. Chem.* 269: 9 (1994) y Pratt y col., *Biochem. Biophys. Res. Comm.* 163: 1243 (1989). Véase también la Patente de los EE.UU. n.°: 5.266.317.

- Veneno específico de insectos producido en la naturaleza por una serpiente, una avispa o cualquier otro organismo. Por ejemplo, véase Pang y col., *Gene* 116: 165 (1992).

- Enzimas responsables de una hiperacumulación de un monoterpeno, un sesquiterpeno, un esteroide, ácido hidroxámico, un derivado de fenilpropanoide u otra molécula no proteica con actividad insecticida.

- Enzimas involucradas en la modificación, incluyendo una modificación de postraducción, de una molécula biológicamente activa; por ejemplo, una enzima glicolítica, una enzima proteolítica, una enzima lipolítica, una nucleasa, una ciclasa, una transaminasa, una esterasa, una hidrolasa, una fosfatasa, una quinasa, una fosforilasa, una polimerasa, una elastasa, una quitinasa y una glucanasa, ya sea natural o sintética. Véase la Solicitud PCT WO93/02197, Kramer y col., *Insect Biochem. Molec. Biol.* 23: 691 (1993) y Kawalleck y col., *Plant Molec. Biol.* 21: 673 (1993).

- Moléculas que estimulan la transducción de señales. Por ejemplo, véase Botella y col., *Plant Molec. Biol.* 24: 757 (1994), y Griess y col., *Plant Physiol.* 104: 1467 (1994).

- Proteínas virales invasivas o una toxina compleja derivada de las mismas. Véase Beachy y col., *Ann. rev.* Phytopathol. 28: 451 (1990).

- Proteínas de arresto del desarrollo producidas en la naturaleza por un patógeno o un parásito. Véase Lamb y col., *Bio/Technology* 10: 1436 (1992) y Toubart y col., *Plant J.* 2: 367 (1992).

- Una proteína de arresto del desarrollo producida en la naturaleza por una planta. Por ejemplo, Logemann y col.,

Bio/Technology 10: 305 (1992).

- En plantas, los patógenos son a menudo específicos del huésped. Por ejemplo, algunas especies de *Fusarium* causará el marchitamiento de tomate pero solamente atacan al tomate, y otras especies de *Fusarium* solamente atacan al sorgo. Plantas que tienen defensas existentes e inducidas para resistir a la mayoría de los patógenos. Mutaciones y eventos de recombinación en generaciones de planta conducen a una variabilidad genética que da lugar a susceptibilidad, en especial dado que los patógenos se reproducen con más frecuencia que las plantas. En las plantas puede haber una resistencia que no es al huésped, por ejemplo, el huésped y el patógeno son incompatibles o puede haber una resistencia parcial contra todas las razas de un patógeno, típicamente controlada por muchos genes y/o también una resistencia completa a algunas razas de un patógeno pero no a otras razas. Dicha resistencia típicamente está controlada por unos pocos genes. El uso de los métodos y componentes del sistema CRISP-Cpf1 ofrece una nueva herramienta para inducir mutaciones específicas de aquí en adelante. Por lo tanto, se puede analizar el genoma de las fuentes de genes de resistencia y, en plantas que tienen las características o los rasgos deseados, usar el método y los componentes del sistema Cpf1 CRISPR para inducir el aumento de genes de resistencia. Los sistemas de la presente pueden hacerlo con más precisión que los agentes mutagénicos anteriores y por ende acelerar y mejorar los programas de cría de plantas.

2. Genes involucrados en las enfermedades de plantas, tales como la que se enumeran en WO 2013046247:

- Enfermedades del arroz: *Magnaporthe grisea*, *Cochliobolus miyabeanus*, *Rhizoctonia solani*, *Gibberella fujikuroi*; enfermedades del sorgo: *Erysiphe graminis, Fusarium graminaarum, F. avenaceum, F. culmorum, Microdochium nivale, Puccinia striiformis, P. graminis, P. recondita, Micronectriella nivale, Typhula* sp., *Ustilago tritici, Tilletia caries, Pseudocercosporella herpotrichoides, Mycosphaerella graminicola, Stagonospora nodorum, Pyrenophora tritici-repentis*; enfermedades de la cebada: *Erysiphe graminis, Fusarium graminaarum, F. avenaceum, F. culmorum, Microdochium nivale, Puccinia striiformis, P. graminis, P. hordei, Ustilago nuda, Rhynchosporium secalis, Pyrenophora teres, Cochliobolus sativus, Pyrenophora graminea, Rhizoctonia solani*; enfermedades del maíz: *Ustilago maydis, Cochliobolus heterostrophus, Gloeocercospora sorghi, Puccinia polisora, Cercospora zeae-maydis, Rhizoctonia solani*;

- Enfermedades de los cítricos: *Diaporthe citri, Elsinoe fawcetti, Penicillium digitatum, P. italicum, Phytophthora parasitica, Phytophthora citrophthora*; enfermedades de la manzana: *Monilinia mali, Valsa ceratosperma, Podosphaera leucotricha, Alternaria alternata* patotipo de manzana, *Venturia inaequalis, Colletotrichum acutatum, Phytophtora cactorum*;

- Enfermedades de la pera: *Venturia nashicola, V. pirina, Alternaria alternata* patotipo de pera japonés, *Gymnosporangium haraeanum, Phytophtora cactorum*;

- Enfermedades del durazno: *Monilinia fructicola, Cladosporium carpophilum, Phomopsis* sp.;

- Enfermedades de la vid: *Elsinoe ampelina, Glomerella cingulata, Uninula necator, Phakopsora ampelopsidis, Guignardia bidwellii, Plasmopara viticola*;

- Enfermedades del caqui: *Gloesporium kaki, Cercospora kaki, Mycosphaerela nawae*;

- Enfermedades de la calabaza: *Colletotrichum lagenarium, Sphaerotheca fuliginea, Mycosphaerella melonis, Fusarium oxisporum, Pseudoperonospora cubensis, Phytophthora* sp., *Pythium* sp.;

- Enfermedades del tomate: *Alternaria solani, Cladosporium fulvum, Phytophthora infestans*;

- Enfermedades de la berenjena: *Phomopsis vexans, Erysiphe cichoracearum*;

- Enfermedades de verduras Brassicaceas: *Alternaria japonica, Cercosporella brassicae, Plasmodiophora brassicae, Peronospora parasitica*;

- Enfermedades de la cebolla de verdeo: *Puccinia allii, Peronospora destructor*;

- Enfermedades de la soja: *Cercospora kikuchii, Elsinoe glicinas, Diaporthe phaseolorum* var. *sojae, Septoria glicinas, Cercospora sojina, Phakopsora pachyrhizi, Phytophthora sojae, Rhizoctonia solani, Corynespora casiicola, Sclerotinia sclerotiorum*;

- Enfermedades del frijol riñón: *Colletrichum lindemthianum*;

- Enfermedades del maní: *Cercospora personata, Cercospora arachidicola, Sclerotium rolfsii*;

- Enfermedades del guisante: *Erysiphe pisi*;

- Enfermedades de la patata: *Alternaria solani, Phytophthora infestans, Phytophthora erythroseptica, Spongospora subterranean*, f. sp. *Subterranean*;

- Enfermedades de la frutilla: *Sphaerotheca humuli, Glomerella cingulata*;
- Enfermedades del té: *Exobasidium reticulatum, Elsinoe leucospila, Pestalotiopsis* sp., *Colletotrichum theae-sinensis*;
- Enfermedades del tabaco: *Alternaria longipes, Erysiphe cichoracearum, Colletotrichum tabacum, Peronospora tabacina, Phytophthora nicotianae*;
- Enfermedades de la colza: *Sclerotinia sclerotiorum, Rhizoctonia solani*;
- Enfermedades del algodón: *Rhizoctonia solani*;
- Enfermedades de la remolacha: *Cercospora beticola, Thanatephorus cucumeris, Thanatephorus cucumeris, Aphanomyces cochlioides*;
- Enfermedades de las rosas: *Diplocarpon rosae, Sphaerotheca pannosa, Peronospora sparsa*;
- Enfermedades de crisantemos y Asteraceae: *Bremia lactuca, Septoria chrysanthemi-indici, Puccinia horiana*;
- Enfermedades de diversas plantas: *Pythium aphanidermatum, Pythium debarianum, Pythium graminicola, Pythium irregulare, Pythium ultimum, Botrytis cinerea, Sclerotinia sclerotiorum*;
- Enfermedades del rábano: *Alternaria brassicicola*;
- Enfermedades de pastos Zoysia: *Sclerotinia homeocarpa, Rhizoctonia solani*;
- Enfermedades de las bananas: *Mycosphaerella fijiensis, Mycosphaerella musicola*;
- Enfermedades del girasol: *Plasmopara halstedii*;
- Enfermedades de las semillas o enfermedades en la etapa de crecimiento inicial de varias plantas causadas por *Aspergillus* spp., *Penicillium* spp., *Fusarium* spp., *Gibberella* spp., *Tricoderma* spp., *Thielaviopsis* spp., *Rhizopus* spp., *Mucor* spp., *Corticium* spp., *Rhoma* spp., *Rhizoctonia* spp., *Diplodia* spp. o semejantes;
- Enfermedades por virus de varias plantas mediadas por *Polimixa* spp., *Olpidium* spp. o semejantes.

3. Ejemplos de genes que confieren resistencia a herbicidas:

- Resistencia a herbicidas que inhiben el punto de crecimiento o meristema, tal como una imidazolinona o una sulfonilurea, por ejemplo, según Lee y col., *EMBO J.* 7: 1241 (1988), y Miki y col., *Theor. Appl. Genet.* 80: 449 (1990), respectivamente.
- Tolerancia a glifosato (resistencia conferida, por ejemplo, por los genes mutantes de la 5-enolpiruvilshiquimato-3-fosfato sintasa (EPSP), genes de aroA y genes de la glifosato acetiltransferasa (GAT) genes, respectivamente), o resistencia a otros compuestos de fosfono tal como por genes de glufosinato (genes de la fosfinotricina acetiltransferasa (PAT) de especies de *Streptomyces*, incluyendo *Streptomyces higroscopicus* y *Streptomyces viridichromogenes*), y a los ácidos piridinoxi o fenoxipropiónico y ciclohexonas por genes que codifican inhibidores de ACCasa. Véase, por ejemplo, la Patente de los EE.UU. n.°: 4.940.835 y la Patente de los EE.UU. n.°: 6.248.876, la Patente de los EE.UU. n.°: 4.769.061, EP n.°: 0 333 033 y la Patente de los EE.UU. n.°: 4.975.374. Véase también EP N°: 0242246, DeGreef y col., *Bio/Technology* 7: 61 (1989), Marshall y col., *Theor. Appl. Genet.* 83: 435 (1992), WO 2005012515 de Castle y col. y WO 2005107437.
- Resistencia a herbicidas que inhiben la fotosíntesis, tal como una triazina (genes psbA y gs+) o un benzonitrilo (gen de nitrilasa) y la glutatión S-transferasa en Przibila y col., *Plant Cell* 3: 169 (1991), Patente de los EE.UU. n.°: 4.810.648, y Hayes y col., *Biochem. J.* 285: 173 (1992).
- Genes que codifican enzimas detoxificantes de herbicidas o una enzima glutamina sintasa mutante que es resistente a la inhibición, por ejemplo, de la Solicitud de Patente de los EE.UU. N° Acta: 11/760.602. O bien, una enzima detoxificante es una enzima que codifica una fosfinotricina acetiltransferasa (tal como la proteína bar o pat de especies de *Streptomyces*). Las fosfinotricina acetiltransferasas se describen, por ejemplo, en las Patentes de los EE.UU. n.°: 5.561.236; 5.648.477; 5.646.024; 5.273.894; 5.637.489; 5.276.268; 5.739.082; 5.908.810 y 7.112.665.
- Inhibidores de hidroxifenilpiruvatodioxigenasas (HPPD), es decir, enzimas naturales resistentes a HPPD, o genes que codifican una enzima HPPD mutada o quimérica como se describe en WO 96/38567, WO 99/24585 y WO 99/24586, WO 2009/144079, WO 2002/046387 o la Patente de los EE.UU. N°: 6,768,044.

4. Ejemplos de genes involucrados en la tolerancia al estrés abiótico:

- Transgen capaz de reducir la expresión y/o la actividad de un gen de la poli(ADP-ribosa) polimerasa (PARP) en

las células vegetales o en las plantas como se describe en WO 00/04173 o WO/2006/045633.

- Transgenes capaces de reducir la expresión y/o la actividad de los genes que codifican PARG de las plantas o las células vegetales, como se describe, por ejemplo, en WO 2004/090140.

- Transgenes que codifican una enzima funcional en plantas de la vía de síntesis de salvataje del nicotinamida adenina dinucleótido, incluyendo una nicotinamidasa, una nicotinato fosforibosiltransferasa, ácido nicotínico mononucleótido adenilo transferasa, nicotinamida adenina dinucleótido sintetasa o nicotinamida fosforibosiltransferasa como se describe, por ejemplo, en EP 04077624.7, WO 2006/133827, PCT/EP07/002,433, EP 1999263 o WO 2007/107326.

- Enzimas involucradas en la biosíntesis de carbohidratos que incluyen las que se describen, por ejemplo, en EP 0571427, WO 95/04826, EP 0719338, WO 96/15248, WO 96/19581, WO 96/27674, WO 97/11188, WO 97/26362, WO 97/32985, WO 97/42328, WO 97/44472, WO 97/45545, WO 98/27212, WO 98/40503, WO99/58688, WO 99/58690, WO 99/58654, WO 00/08184, WO 00/08185, WO 00/08175, WO 00/28052, WO 00/77229, WO 01/12782, WO 01/12826, WO 02/101059, WO 03/071860, WO 2004/056999, WO 2005/030942, WO 2005/030941, WO 2005/095632, WO 2005/095617, WO 2005/095619, WO 2005/095618, WO 2005/123927, WO 2006/018319, WO 2006/103107, WO 2006/108702, WO 2007/009823, WO 00/22140, WO 2006/063862, WO 2006/072603, WO 02/034923, EP 06090134.5, EP 06090228.5, EP 06090227.7, EP 07090007.1, EP 07090009.7, WO 01/14569, WO 02/79410, WO 03/33540, WO 2004/078983, WO 01/19975, WO 95/26407, WO 96/34968, WO 98/20145, WO 99/12950, WO 99/66050, WO 99/53072, Patente de los EE.UU. n.°: 6.734.341, WO 00/11192, WO 98/22604, WO 98/32326, WO 01/98509, WO 01/98509, WO 2005/002359, Patente de los EE.UU. N°: 5.824.790, Patente de los EE.UU. n.°: 6.013.861, WO 94/04693, WO 94/09144, WO 94/11520, WO 95/35026 o WO 97/20936 o enzimas involucradas en la producción de polifructosa, en especial del tipo de la inulina y levano, como se divulga en EP 0663956, WO 96/01904, WO 96/21023, WO 98/39460 y WO 99/24593, la producción de alfa-1,4-glucanos como se divulga en WO 95/31553, US 2002031826, la Patente de los EE.UU. N°: 6.284.479, la Patente de los EE.UU. N°: 5.712.107, WO 97/47806, WO 97/47807, WO 97/47808 y WO 00/14249, la producción de alfa-1,4-glucanos ramificados en alfa-1,6, como se divulga en WO 00/73422, la producción de alternano, como se divulga, por ejemplo, en WO 00/47727, WO 00/73422, EP 06077301.7, la Patente de los EE.UU. n.°: 5.908.975 y EP 0728213, la producción de hialurona, como se divulga, por ejemplo, en WO 2006/032538, WO 2007/039314, WO 2007/039315, WO 2007/039316, JP 2006304779 y WO 2005/012529.

- Genes que mejoran la resistencia a la sequía. Por ejemplo, en WO 2013122472 se divulga que la ausencia o un nivel reducido de la proteína funcional de la ubiquitina proteína ligasa (UPL), más específicamente, UPL3, conduce a una menor necesidad de agua o a una mayor resistencia a sequía en la planta. Otros ejemplos de plantas transgénicas con una mayor tolerancia a sequía se divulgan, por ejemplo, en US 2009/0144850, en US 2007/0266453 y en WO 2002/083911. En US2009/0144850 se describe una planta que presenta un fenotipo de tolerancia a la sequía debido a una expresión alterada de un ácido nucleico DR02. En US 2007/0266453 se describe una planta que tiene un fenotipo de tolerancia a sequía debido a una expresión alterada de un ácido nucleico DR03 y en WO 2002/08391 1 se describe una planta que tiene una mayor tolerancia al estrés por sequía debido a una actividad reducida de un transportador ABC que se expresa en células oclusivas. Otro ejemplo es el trabajo de Kasuga y coautores (1999), quienes describen que la sobreexpresión del ADNc que codifica DREB1 A en plantas transgénicas activó la expresión de muchos genes de tolerancia al estrés bajo condiciones de crecimiento normales y dio como resultado una tolerancia mejorada a la sequía, a la carga de sales y al congelamiento. Sin embargo, la expresión de DREB1A también dio como resultado un retardo severo del crecimiento bajo condiciones de crecimiento normales (Kasuga (1999) *Nat Biotechnol* 17(3) 287-291).

En otras formas de realización particulares, se pueden mejorar las plantas de cultivo afectando algunas características específicas de las plantas. Por ejemplo, mediante el desarrollo de plantas resistentes a plaguicidas, la mejora de la resistencia a enfermedades en plantas, la mejora de la resistencia a insectos y nematodos en plantas, la mejora de la resistencia de plantas contra malezas parásitas, la mejora de la tolerancia a la sequía en plantas, la mejora del valor nutricional de las plantas, la mejora de la tolerancia al estrés en plantas, evitando la autopolinización, biomasa digestible forrajera de plantas, rendimiento de granos, etc. Más adelante se proveen unos pocos ejemplos no taxativos específicos.

Además de la mutación dirigida de genes individuales, los complejos de Cpf1CRISPR se pueden diseñar para permitir una mutación dirigida de múltiples genes, la supresión de fragmentos cromosómicos, integración específica del sitio de transgenes, mutagénesis dirigida al sitio in vivo y reemplazo preciso de genes o intercambio de alelos en plantas. Por ello, los métodos que se describen en la presente tienen amplias aplicaciones en el descubrimiento y la validación de genes, la cría mutacional y cisgénica, así como la cría de híbridos. Estas aplicaciones facilitan la producción de una nueva generación de cultivos modificados genéticamente con varias características agronómicas mejoradas tal como la resistencia a herbicidas, la resistencia a enfermedades, la tolerancia al estrés abiótico, un rendimiento elevado y una calidad superior.

Uso del gen Cpf1 para crear plantas estériles masculinas

Las plantas híbridas típicamente ofrecen características agronómicas ventajosas en comparación con las plantas endogámicas. Sin embargo, en el caso de las plantas autopolinizantes, la generación de híbridos puede ser un desafío. Se han identificado genes en diferentes tipos de plantas que son importantes para la fertilidad de las plantas, más particularmente para la fertilidad masculina. Por ejemplo, en maíz, se han identificado por lo menos dos genes que son importantes para la fertilidad (Amitabh Mohanty International Conference on New Plant Breeding Molecular Technologies Technology Development And Regulation, 9-10 de octubre, 2014, Jaipur, India; Svitashev y col., Plant Physiol., octubre de 2015; 169(2): 931-45; Djukanovic y col., Plant J. diciembre de 2013;76(5): 888-99). Los métodos provistos en la presente se pueden usar para dirigir los genes diana necesarios para la fertilidad masculina para así generar plantas estériles masculinas que se puedan cruzar fácilmente para generar híbridos. En formas de realización particulares, el sistema Cpf1 CRISPR provisto en la presente se usa para una mutagénesis dirigida del gen tipo citocromo P450 (MS26) o el gen de meganucleasa (MS45) confiriendo de esa manera esterilidad masculina a la planta de maíz. Las plantas de maíz que están alteradas genéticamente de esta manera se pueden usar en los programas de cría de híbridos.

Incremento de la etapa de fertilidad en plantas

En formas de realización particulares, los métodos provistos en la presente se usan para prolongar la etapa de fertilidad de una planta, tal como de una planta de arroz. Por ejemplo, se puede usar como diana un gen de la etapa de fertilidad de arroz, tal como Ehd3, a fin de generar una mutación en el gen y se pueden seleccionar las plántulas para una etapa de fertilidad vegetal de regeneración prolongada (como se describe en CN 104004782).

Uso de Cpf1 para generar variación genética en un cultivo de interés

La disponibilidad de germoplasma salvaje y de variaciones genéticas en plantas de cultivo es la clave para los programas de mejoras de cultivos, pero la diversidad disponible en los germoplasmas de las plantas de cultivo es limitada. La presente invención contempla métodos para generar una diversidad de variaciones genéticas en un germoplasma de interés. En esta solicitud del sistema Cpf1 CRISPR, se provee una biblioteca de ARN guía dirigida a diferentes ubicaciones en el genoma vegetal y se introduce en las células vegetales junto con la proteína efectora Cpf1. De esta manera se puede generar una colección de mutaciones puntuales a escala de genoma y de noqueos de genes. En formas de realización particulares, los métodos comprenden generar una parte de planta o una planta a partir de las células así obtenidas y seleccionar las células según una característica de interés. Los genes diana pueden incluir regiones codificantes y no codificantes. En formas de realización particulares, la característica es la tolerancia al estrés y el método es un método para generar variedades de cultivo tolerantes al estrés.

Uso de Cpf1 para afectar la maduración de frutos

La maduración es una fase normal en el proceso de maduración de frutas y verduras. Solamente unos pocos días después que comienza las frutas o las verduras ya no son comestibles. Este proceso provoca pérdidas significativas tanto para los granjeros como para los consumidores. En formas de realización particulares, los métodos de la presente invención se usan para reducir la producción de etileno. Esto se puede asegurar si se logra garantizar una o más de los siguientes: a. supresión de la expresión del gen de la ACC sintasa. La ACC (ácido 1-aminociclopropano-1-carboxílico) sintasa es la enzima responsable de la conversión de S-adenosilmetionina (SAM) en ACC; el segundo al último paso en la biosíntesis del etileno. La expresión de la enzima es impedida cuando se inserta una copia antisentido ("imagen especular") o truncada del gen de la sintasa en el genoma de la planta; b. Inserción del gen de la ACC desaminasa. El gen que codifica la enzima se obtiene de Pseudomonas chlororaphis, una bacteria no patógena común en el suelo. Convierte al ACC en un compuesto diferente reduciendo de esa manera la cantidad de ACC disponible para la producción de etileno; c. Inserción del gen de la SAM hidrolasa. Este abordaje es similar al de la ACC desaminasa donde la producción de etileno es impedida cuando se reduce la cantidad de su metabolito precursor; en este caso, se convierte SAM en homoserina. El gen que codifica la enzima se obtiene del bacteriófago T3 de E. coli y d. Supresión de la expresión del gen de la ACC oxidasa. La ACC oxidasa es la enzima que cataliza la oxidación de ACC en etileno, el último paso en la vía de biosíntesis de etileno. Cuando se usan los métodos descritos en la presente, una expresión disminuida del gen de la ACC oxidasa da como resultado la supresión de la producción de etileno, demorando de esa manera la maduración del fruto. En formas de realización particulares, adicionalmente o como una alternativa de las modificaciones descritas precedentemente, los métodos que se describen en la presente se usan para modificar los receptores de etileno, para de esa manera interferir con las señales de etileno obtenidas por la fruta. En formas de realización particulares, se modifica, más particularmente se suprime, la expresión del gen ETR1, que codifica una proteína de unión a etileno. En formas de realización particulares, adicionalmente o como una alternativa de las modificaciones descritas precedentemente, los métodos que se describen en la presente se usan para modificar la expresión del gen que codifica la poligalacturonasa (PG), que es la enzima responsable de la degradación de pectina, la sustancia que mantiene la integridad de las paredes de las células vegetales. La degradación de la pectina se produce al comienzo del proceso de maduración que da como resultado el ablandamiento de la fruta. Por lo tanto, en formas de realización particulares, los métodos que se describen en la presente se usan para introducir una mutación en el gen PG o para suprimir la activación del gen PG a fin de reducir la cantidad de enzima PG producida y demorando de esa manera la degradación de pectina.

Por consiguiente, en formas de realización particulares, los métodos comprenden el uso del sistema Cpf1 CRISPR para asegurar una o más modificaciones del genoma de una célula vegetal tal como se describió con anterioridad y

regenerar una planta a partir de la misma. En formas de realización particulares, la planta es una planta de tomate.

Aumento de la vida útil de plantas

En formas de realización particulares, los métodos reivindicados de la presente invención se usan para modificar los genes involucrados en la producción de aquellos compuestos que afectan la vida útil de la planta o parte de planta. Más particularmente, la modificación tiene lugar en un gen que impide la acumulación de azúcares reductores en tubérculos de patata. Tras un procesamiento a temperatura elevada, estos azúcares reductores reaccionan con aminoácidos libres, lo que da como resultado productos marrones, de sabor amargo, y niveles elevados de acrilamida, que es un posible carcinógeno. En formas de realización particulares, los métodos provistos en la presente se usan para reducir o inhibir la expresión del gen de la invertasa vacuolar (VInv), que codifica una proteína que degrada sacarosa en glucosa y fructosa (Clasen y col., DOI: 10.1111/pbi.12370).

Uso del sistema Cpf1 CRISPR para asegurar una característica de valor agregado

En formas de realización particulares, el sistema Cpf1 CRISPR se usa para producir cultivos agronómicos nutricionalmente mejorados. En formas de realización particulares, los métodos provistos en la presente se adaptan para generar "alimentos funcionales", es decir un alimento o un ingrediente de un alimento modificado que puede proporcionar un beneficio saludable más allá de los nutrientes tradicionales que ya contiene y/o un "nutracéutico", es decir, sustancias que pueden considerarse como un alimento o como una parte de un alimento y que proveen beneficios para la salud, incluyendo la prevención y el tratamiento de enfermedades. En formas de realización particulares, el nutracéutico es de utilidad en la prevención y/o el tratamiento de uno o más entre cáncer, diabetes, enfermedades cardiovasculares e hipertensión.

Los ejemplos de cultivos nutricionalmente mejorados incluyen (Newell-McGloughlin, Plant Physiology, julio de 2008, volumen 147, páginas 939-953):

- calidad, contenido y/o composición de aminoácidos modificados de proteínas, tal como se ha descrito para el pasto de Bahía (Luciani y col., 2005, *Florida Genetics Conference Poster*), canola (Roesler y col., 1997, *Plant Physiol* 113 75-81), maíz (Cromwell y col., 1967, 1969 *J Anim Sci* 26 1325-1331, O'Quin y col., 2000 *J Anim Sci* 78 2144-2149, Yang y col., 2002, *Transgenic Res* 11 11-20, Young y col., 2004, *Plant J* 38 910-922), patata (Yu J y Ao, 1997 *Acta Bot Sin* 39 329-334; Chakraborty y col., 2000, *Proc Natl Acad Sci USA* 97 3724-3729; Li y col., 2001) *Chin Sci Bull* 46 482-484, arroz (Katsube y col., 1999, *Plant Physiol* 120 1063-1074), soja (Dinkins y col., 2001, Rapp 2002, *In Vitro Cell Dev Biol Plant* 37 742-747), batata (Egnin y Prakash 1997, *In Vitro Cell Dev Biol* 33-52A).

- contenido de aminoácidos esenciales tal como se ha descrito para canola (Falco y col., 1995, *Bio/Technology* 13 577-582), lupina (White y col., 2001, *J Sci Food Agric* 81 147-154), maíz (Lai y Messing, 2002, base de datos de cultivos GM *Agbios* 2008 (11 de marzo, 2008)), patata (Zeh y col., 2001, *Plant Physiol* 127 792-802), sorgo (Zhao y col., 2003, Kluwer Academic Publishers, Dordrecht, Los Países Bajos, páginas 413-416), soja (Falco y col., 1995 *Bio/Technology* 13: 577-582; Galili y col., 2002 *Crit Rev Plant Sci* 21 167-204).

- aceites y ácidos grasos tal como para canola (Dehesh y col., (1996) *Plant J* 9 167-172 [PubMed]; Del Vecchio (1996) *INFORM International News on Fats, Oils and Related Materials* 7 230-243; Roesler y col., (1997) *Plant Physiol* 113 75-81 [PMC, artículo libre] [PubMed]; Froman y Ursin (2002, 2003) resúmenes de publicaciones de *American Chemical Society* 223 U35; James y col., (2003) *Am J Clin Nutr* 77 1140-1145 [PubMed]; Agbios (2008, anterior); algodón (Chapman y col., (2001) *J Am Aceite Chem Soc* 78 941-947; Liu y col., (2002) *J Am Coll Nutr* 21 205S-211S [PubMed]; O'Neill (2007) *Australian Life Scientist.* http://www.biotechnews.com.au/index.php/id;866694817;fp;4;fpid;2 (17 de junio, 2008), linaza (Abbadi y col., 2004, *Plant Cell* 16: 2734-2748), maíz (Young y col., 2004, *Plant J* 38 910-922), palmera oleaginosa (Jalani y col., 1997, *J Am Aceite Chem Soc* 74 1451-1455; Parveez, 2003, AgBiotechNet 113 1-8), arroz (Anai y col., 2003, *Plant Cell Rep* 21 988-992), soja (Reddy y Thomas, 1996, *Nat Biotechnol* 14 639-642; Kinney y Kwolton, 1998, *Blackie Academic and Professional*, Londres, páginas 193-213), girasol (Arcadia, *Biosciences* 2008)

- carbohidratos, tales como fructanos, como se ha descrito para achicoria (Smeekens (1997) *Trends Plant Sci* 2 286-287, Sprenger y col., (1997) *FEBS Lett* 400 355-358, Sévenier y col., (1998) *Nat Biotechnol* 16 843-846), maíz (Caimi y col., (1996) *Plant Physiol* 110 355-363), patata (Hellwege y col., 1997 *Plant J* 12 1057-1065), remolacha (Smeekens y col., 1997, antes), inulina, tal como se ha descrito para patata (Hellewege y col., 2000, *Proc Natl Acad Sci USA* 97 8699-8704), almidón, tal como se ha descrito para arroz (Schwall y col., (2000) *Nat Biotechnol* 18 551-554, Chiang y col., (2005) *Mol Breed* 15 125-143),

- vitaminas y carotenoides, tal como se ha descrito para canola (Shintani y DellaPenna (1998) *Science* 282 2098-2100), maíz (Rocheford y col., (2002) *J Am Coll Nutr* 21 191S-198S, Cahoon y col., (2003) *Nat Biotechnol* 21 1082-1087, Chen y col., (2003) *Proc Natl Acad Sci USA* 100 3525-3530), semillas de mostaza (Shewmaker y col., (1999) *Plant J* 20 401-412, patata (Ducreux y col., 2005, *J Exp Bot* 56 81-89), arroz (Ye y col., (2000) *Science* 287 303-305, frutilla (Agius y col., (2003), *Nat Biotechnol* 21 177-181), tomate (Rosati y col., (2000) *Plant J* 24 413-419, Fraser y col., (2001) *J Sci Food Agric* 81 822-827, Mehta y col., (2002) *Nat Biotechnol* 20 613-618, Díaz de la Garza y col., (2004) *Proc Natl Acad Sci USA* 101 13720-13725, Enfissi y col., (2005) *Plant Biotechnol J* 3 17-27, DellaPenna (2007) *Proc Natl Acad Sci USA* 104 3675-3676.

- metabolitos secundarios funcionales, tal como se ha descrito para manzana (estilbenos, Szankowski y col., (2003) *Plant Cell Rep* 22:141-149), alfalfa (resveratrol, Hipskind y Paiva (2000) *Mol Plant Microbio Interact* 13 551-562), kiwi (resveratrol, Kobayashi y col., (2000) *Plant Cell Rep* 19 904-910), maíz y soja (flavonoides, Yu y col., (2000) *Plant Physiol* 124 781-794), patata (antocianina y glicósidos alcaloides, Lukaszewicz y col., (2004) *J Agric Food Chem* 52 1526-1533), arroz (flavonoides y resveratrol, Stark-Lorenzen y col., (1997) *Plant Cell Rep* 16 668-673, Shin y col., (2006) *Plant Biotechnol J* 4 303-315), tomate (+resveratrol, ácido clorogénico, flavonoides, estilbeno; Rosati y col., (2000) antes, Muir y col., (2001) *Nature* 19 470-474, Niggeweg y col., (2004) *Nat Biotechnol* 22 746-754, Giovinazzo y col., (2005) *Plant Biotechnol J* 3 57-69), sorgo (ácidos cafeico y ferúlico, resveratrol; *United Press International* (2002)); y

- disponibilidad de minerales tal como se ha descrito para alfalfa (fitasa, Austin-Phillips y col., (1999) http://www.molecularfarming.com/nonmedical.html), lechuga (hierro, Goto y col., (2000) *Theor Appl Genet* 100 658-664), arroz (hierro, Lucca y col., (2002) *J Am Coll Nutr* 21 184S-190S), maíz, soja y trigo (fitasa, Drakakaki y col., (2005) *Plant Mol Biol* 59 869-880, Denbow y col., (1998) *Poult Sci* 77 878-881, Brinch-Pedersen y col., (2000) *Mol Breed* 6 195-206).

En formas de realización particulares, la característica de valor agregado está relacionada con los beneficios para la salud contemplados de los compuestos presentes en la planta. Por ejemplo, en formas de realización particulares, el cultivo de valor agregado se obtiene mediante aplicación de los métodos reivindicados de la invención para asegurar la modificación o para inducir/aumentar la síntesis de uno o más de los siguientes compuestos:

- Carotenoides, tal como el α-caroteno presente en zanahorias que neutraliza los radicales libres que pueden causar daño en las células o el β-caroteno presente en varias frutas y verduras que neutraliza los radicales libres
- Luteína, presente en verduras verdes, que contribuye en el mantenimiento de una visión sana
- Licopeno, presente en tomate y productos de tomate, que se cree reduce el riesgo de cáncer de próstata
- Zeaxantina, presente en cítricos y maíz, que contribuye en el mantenimiento de una visión sana
- Fibras de la dieta, tal como la fibra insoluble presente en salvado de sorgo, que puede reducir el riesgo de cáncer de mama y/o de colon, y el β-glucano presente en avena, fibras solubles presentes en *Psylium* y granos enteros de cereales que pueden reducir el riesgo de enfermedades cardiovasculares (CVD)
- Ácidos grasos, tales como ácidos grasos ω-3, que pueden reducir el riesgo de CVD y mejorar las funciones mentales y visuales, ácido linoleico conjugado, que puede mejorar la composición corporal, puede disminuir riesgo de determinados cánceres y GLA que puede reducir el riesgo de inflamación, de cáncer y CVD, que puede mejorar la composición corporal
- Flavonoides, tales como hidroxicinamatos, presentes en sorgo que tienen actividades tipo antioxidante, pueden reducir el riesgo de enfermedades degenerativas, flavonoles, catequinas y taninos presentes en frutas y verduras, que neutralizan los radicales libres y pueden reducir el riesgo de cáncer
- Glucosinolatos, indoles, isotiocianatos, tal como el sulforafano, presentes en verduras Cruciferous (brócoli, col rizado), rábano picante, que neutralizan los radicales libres, pueden reducir el riesgo de cáncer
- Fenólicos, tales como estilbenos, presentes en vides, que pueden reducir el riesgo de enfermedades degenerativas, enfermedades cardíacas y cáncer, pueden tener un efecto sobre la longevidad, y el ácido cafeico y ácido ferúlico presentes en verduras y cítricos que tienen actividades tipo antioxidante, pueden reducir el riesgo de enfermedades degenerativas, enfermedades cardíacas y enfermedades oculares, y epicatequina presente en el cacao, que tiene actividades tipo antioxidante, puede reducir el riesgo de enfermedades degenerativas y enfermedades cardíacas
- Estanoles/esteroles vegetales, presentes en maíz, soja, sorgo y aceites de leñosas que pueden reducir el riesgo de enfermedades cardíacas coronarias por disminución de los niveles de colesterol en sangre
- Fructanos, inulinas, fructooligosacáridos, presentes en alcachofa de Jerusalén, echalotes, polvo de cebolla, que pueden mejorar la salud gastrointestinal
- Saponinas, presentes en soja, que pueden disminuir el colesterol LDL
- Proteína de soja, presente en soja, que puede reducir el riesgo de enfermedades cardíacas
- Fitoestrógenos, tales como isoflavonas, presentes en soja, que pueden reducir los síntomas de la menopausia, tales como sofocos, pueden reducir la osteoporosis y CVD y lignanos, presentes en lino, centeno y verduras, que pueden ofrecer protección contra enfermedades cardíacas y algunos cánceres, pueden disminuir el colesterol LDL, el colesterol total.
- Sulfuros y tioles, tal como el sulfuro de dialilo, presente en cebolla, ajo, aceitunas, puerro y cebollín; y trisulfuro de alilmetilo, ditioltionas, presentes en verduras de crucíferas, que pueden disminuir el colesterol LDL, ayudan a

mantener un sistema inmunológico sano

- Taninos, tales como proantocianidinas, presentes en arándanos, cacao, que pueden mejorar la salud del tracto urinario, pueden reducir riesgo de CVD y presión sanguínea alta

- Etc.

Además, los métodos reivindicados de la presente invención también contemplan modificar las características de funcionalidad, vida útil, sabor/estética, calidad de las fibras y alergenos, antinutrientes y reducción de toxinas de proteínas/almidón.

Por lo tanto, los métodos y composiciones reivindicados de la invención abarcan producir plantas con un valor nutricional agregado, donde dichos métodos comprenden introducir en una célula vegetal un gen que codifica una enzima involucrada en la producción de un componente de valor nutricional agregado usando el sistema Cpf1 CRISPR descrito en la presente y regenerar una planta a partir de dicha célula vegetal, donde dicha planta se caracteriza por una mayor expresión de dicho componente de valor nutricional agregado. En formas de realización particulares, el sistema Cpf1 CRISPR se usa para modificar la síntesis endógena de estos compuestos de manera indirecta, por ejemplo, mediante modificación de uno o más factores de la transcripción que controla el metabolismo de este compuesto. Los métodos para introducir un gen de interés en una célula vegetal y/o modificar un gen endógeno usando el sistema Cpf1 CRISPR se describieron precedentemente en la presente.

Algunos ejemplos específicos de modificaciones en plantas que se pueden modificar para conferir características de valor agregado son: plantas con un metabolismo de ácidos grasos modificado, por ejemplo, mediante transformación de una planta con un gen antisentido de la estearil-ACP desaturasa para aumentar el contenido de ácido esteárico de la planta. Véase Knultzon y col., Proc. Natl. Acad. Sci. U.S.A. 89: 2624 (1992). Otro ejemplo comprende disminuir el contenido de fitato, por ejemplo, mediante clonación y luego reintroducir el ADN asociado con el alelo individual que puede ser responsable de mutantes de maíz caracterizados por niveles bajos de ácido fítico. Véase Raboy y col., Maydica 35: 383 (1990).

De manera similar, la expresión de las Tfs C1 y R de maíz (Zea mays), que regula la producción de flavonoides en las capas de aleurona de maíz, bajo el control de un promotor fuerte, dio como resultado un índice de acumulación elevado de antocianinas en Arabidopsis (Arabidopsis thaliana), presumiblemente por activación de toda la vía (Bruce y col., 2000, Plant Cell 12: 65-80). DellaPenna (Welsch y col., 2007 Annu Rev Plant Biol 57: 711-738) encontraron que Tf RAP2.2 y su miembro de interacción, SINAT2, aumentó la carotenogénesis en hojas de Arabidopsis. La expresión de Tf Dof1 indujo una regulación aumentada de genes que codifican enzimas para la producción del esqueleto de carbono, un aumento marcado del contenido de aminoácidos y una reducción del nivel de Glc en Arabidopsis transgénico (Yanagisawa, 2004 Plant Cell Physiol 45: 386-391) y DOF Tf AtDof1.1 (OBP2) aumentó la expresión de todos los pasos en la vía de biosíntesis de glucosinolato en Arabidopsis (Skirycz y col., 2006 Plant J 47: 10-24).

Reducción de alergenos en plantas

En formas de realización particulares, los métodos provistos en la presente se usan para generar plantas con un nivel reducido de alergenos, volviéndolas más seguras para el consumidor. En formas de realización particulares, los métodos comprenden modificar la expresión de uno o más genes responsables de la producción de alergenos vegetales. Por ejemplo, en formas de realización particulares, los métodos comprenden disminuir la expresión de un gen Lol p5 en una célula vegetal, tal como una célula vegetal de pasto de centeno y regenerar una planta a partir de la misma para así reducir la alergenicidad del polen de dicha planta (Bhalla y col., 1999, Proc. Natl. Acad. Sci. USA volumen 96: 11676-11680).

Las alergias al maní y las alergias a las legumbres generalmente son un problema para la salud real y severa. El sistema de la proteína efectora Cpf1 usada en los métodos y composiciones reivindicados de la presente invención se puede usar para identificar y luego editar o silenciar los genes que codifican las proteínas alergénicas de dichas legumbres. En un sentido no taxativo, con respecto a dichos genes y proteínas, Nicolaou y col., identifican proteínas alergénicas en maníes, soja, lentejas, guisantes, lupina, habas verdes y frijol mungo. Véase, Nicolaou y col., Current Opinion in Allergy and Clinical Immunology 2011;11(3): 222).

Métodos de selección de los genes endógenos de interés

Los métodos provistos en la presente permiten además identificar genes de valor que codifican enzimas involucradas en la producción de un componente de valor nutricional agregado o genes que afectan en general características agronómicas de interés, entre especies, Filas y el reino vegetal. El direccionamiento selectivo, por ejemplo, de genes que codifican enzimas de las vías metabólicas de plantas usando el sistema Cpf1 CRISPR descrito en la presente, permite identificar los genes responsables de determinados aspectos nutricionales de una planta. De manera similar, el selectivamente direccionamiento genes que pueden afectar una característica agronómica deseable, permite identificar los genes relevantes. Por lo tanto, la presente invención abarca métodos de selección de genes que codifican las enzimas involucradas en la producción de compuestos con un valor nutricional y/o características agronómicas particulares.

Aplicaciones adicionales del sistema Cpf1 CRISPR en plantas y levaduras

Uso del sistema Cpf1 CRISPR en la producción de biocombustibles

El término "biocombustible" según se usa aquí, se refiere a un combustible alternativo que se hace a partir de plantas y recursos derivados de plantas. Los biocombustibles renovables se pueden extraer de la materia orgánica cuya energía se obtuvo mediante un proceso de fijación de carbono o se hacen mediante el uso o conversión de biomasa. Esta biomasa se puede utilizar directamente para biocombustibles o se puede convertir en sustancias convenientes que contienen energía por conversión térmica, conversión química, y conversión bioquímica. Esta conversión de biomasa puede dar como resultado un combustible en forma sólida, líquida, o gaseosa. Hay dos tipos de biocombustibles: bioetanol y biodiesel. El bioetanol se produce principalmente por el proceso de fermentación de azúcares de celulosa (o almidón), que principalmente se obtienen del maíz y la caña de azúcar. Por otro lado, el biodiesel se produce principalmente a partir de cultivos de oleaginosas tales como colza, palma, y soja. Los biocombustibles se utilizan principalmente para los transportes.

Mejora de las propiedades de las plantas para la producción de biocombustibles

En algunas formas de realización en particular, los métodos en los que se utiliza el sistema Cpf1 CRISPR según se describe aquí se utilizan para alterar las propiedades de la pared celular para facilitar el acceso de los agentes hidrolizantes clave para una liberación más eficiente de los azúcares para la fermentación. En algunas formas de realización en particular, se modifican la biosíntesis de celulosa y/o lignina. La celulosa es el componente principal de la pared celular. La biosíntesis de celulosa y lignina están co-reguladas. al reducir la proporción de lignina en una planta, se puede incrementar la proporción de celulosa. En algunas formas de realización en particular, los métodos que se describen aquí se utilizan para reducir la expresión de la biosíntesis de lignina en la planta de manera tal de incrementar la cantidad de carbohidratos fermentables. Más particularmente, los métodos que se describen aquí se utilizan para reducir la expresión por lo menos un primer gen de biosíntesis de lignina que se selecciona entre el grupo que consiste en 4-cumarato 3-hidroxilasa (C3H), fenilalanina amoníaco-liasa (PAL), cinamato 4-hidroxilasa (C4H), hidroxicinamoil transferasa (HCT), ácido cafeico O-metiltransferasa (COMT), cafeoil CoA 3-O-metiltransferasa (CCoAOMT), ferulato 5-hidroxilasa (F5H), alcohol cinamílico deshidrogenasa (CAD), cinamoil CoA-reductasa (CCR), 4-cumarato-CoA ligasa (4CL), glicosiltransferasa específica para monolignol-lignina, y aldehído deshidrogenasa (ALDH) según se divulga en WO 2008064289 A2.

En algunas formas de realización en particular, los métodos que se describen aquí se utilizan para producir una masa vegetal que produce menores niveles de ácido acético durante la fermentación (véase también WO 2010096488). Más particularmente, los métodos que se divulgan aquí se utilizan para generar mutaciones en homólogos a CasIL para reducir la acetilación de polisacáridos.

Modificación de levaduras para la producción de biocombustibles

En algunas formas de realización en particular, la enzima Cpf1 que se provee aquí se utiliza para la producción de bioetanol por microorganismos recombinantes. Por ejemplo, Cpf1 se puede utilizar para modificar microorganismos, por ejemplo, levaduras, para que generen biocombustibles o biopolímeros a partir de azúcares fermentables y opcionalmente para que sean capaces de degradar derivados de lignocelulosa de plantas que se obtienen de residuos agrícolas como fuente de azúcares fermentables. Más particularmente, la invención provee métodos mediante los cuales se utiliza el complejo Cpf1 CRISPR para introducir en microorganismos los genes extraños que son necesarios para la producción de biocombustibles y/o para modificar genes endógenos que pueden interferir con la síntesis de biocombustibles. Más particularmente los métodos incluyen introducir en un microorganismo tal como una levadura una o más secuencias de nucleótidos que codifican enzimas incluidas en la conversión de piruvato a etanol u otro producto de interés. En algunas formas de realización en particular los métodos aseguran la introducción de una o más enzimas que permiten que el microorganismo degrade celulosa, por ejemplo, una celulasa. En otras formas de realización aún adicionales, el complejo Cpf1 CRISPR se utiliza para modificar vías metabólicas endógenas que compiten con la vía de producción de biocombustibles.

Por lo tanto, en algunas formas de realización más particulares, los métodos que se describen aquí se utilizan para modificar un microorganismo de la siguiente manera:

introducir por lo menos un ácido nucleico heterólogo o aumentar la expresión de por lo menos un ácido nucleico endógeno que codifica una enzima que degrada la pared celular de la planta, de manera tal que dicho microorganismo es capaz de expresar dicho ácido nucleico y de producir y secretar dicha enzima que degrada la pared celular de la planta,

introducir por lo menos un ácido nucleico heterólogo o aumentar la expresión de por lo menos un ácido nucleico endógeno que codifica una enzima que convierte piruvato en acetaldehído opcionalmente combinada con por lo menos un ácido nucleico heterólogo que codifica una enzima que convierte acetaldehído en etanol de manera tal que dicha célula huésped sea capaz de expresar dicho ácido nucleico, y/o

modificar por lo menos un ácido nucleico que codifica una enzima en una vía metabólica en dicha célula huésped, donde dicha vía produce un metabolito diferente del acetaldehído a partir del piruvato o etanol a partir del acetaldehído, y donde dicha modificación da como resultado una menor producción de dicho metabolito, o introducir por lo menos un ácido nucleico que codifica un inhibidor de dicha enzima.

Modificación de algas y plantas para la producción de aceites vegetales o biocombustibles

Las algas u otras plantas transgénicas tales como la colza podrán ser particularmente útiles en la producción de aceites vegetales o biocombustibles tales como alcoholes (especialmente metanol y etanol), por ejemplo. Estas se podrán modificar para expresar o sobreexpresar niveles elevados de aceite o alcoholes para su uso en las industrias de los aceites o biocombustibles.

De acuerdo con algunas formas de realización en particular de la invención, el sistema Cpf1 CRISPR se utiliza para generar diatomeas ricas en lípidos que son útiles en la producción de biocombustibles.

En algunas formas de realización en particular se prevé la modificación específica de genes relacionados con la modificación de la cantidad de lípidos y/o el tipo de lípidos que producen las células de algas. Algunos ejemplos de genes que codifican enzimas incluidas en las vías de síntesis de ácidos grasos pueden codificar proteínas que por ejemplo tienen actividades de acetil-CoA carboxilasa, ácido graso sintasa, 3-cetoacil (proteína transportadora de acilo) sintasa III, glicerol-3-fosfato deshidrogenasa (G3PDH), enoil-(proteína transportadora de acilo) reductasa (Enoil-ACP-reductasa), glicerol-3-fosfato aciltransferasa, acil transferasa lisofosfatídica o diacilglicerol aciltransferasa, fosfolípido:diacilglicerol aciltransferasa, fosfatidato fosfatasa, ácido graso tioesterasa tal como palmitoil proteína tioesterasa, o enzima málica. En otras formas de realización adicionales se prevé la generación de diatomeas con una mayor acumulación de lípidos. Esto se puede realizar mediante el direccionamiento hacia genes que reducen la catabolización de lípidos. Para utilizar en los métodos de la presente invención son particularmente interesantes los genes relacionados con la activación tanto de triacilglicerol como de ácidos grasos libres, así como genes relacionados directamente con la β-oxidación de ácidos grasos, por ejemplo, con actividad acil-CoA sintetasa, 3-cetoacil-CoA tiolasa, acil-CoA oxidasa y fosfoglucomutasa. El sistema Cpf1 CRISPR y los métodos que se describen aquí se pueden utilizar para activar específicamente dichos genes en diatomeas para aumentar su contenido de lípidos.

Los organismos tales como las algas microscópicas se utilizan ampliamente para la biología de síntesis. Stovicek y col. (Metab. Eng. Comm., 2015; 2:13 describen la edición del genoma de la levadura industrial, por ejemplo, *Saccharomyces cerevisae*, para producir eficientemente cepas robustas para la producción industrial. Stovicek utiliza un sistema CRISPR-Cas9 optimizada por codones para la levadura para interrumpir simultáneamente ambos alelos de un gen endógeno y realizar el *knock in* de un gen heterólogo. Cas9 y ARNg se expresaron a partir de localizaciones genómicas o episómicas de vectores basados en 2μ. Los autores también mostraron que la eficiencia de la interrupción del gen se puede mejorar por optimización de los niveles de expresión de Cas9 y ARNg. Hlavová y col. (Biotechnol. Adv. 2015) exponen el desarrollo de especies o cepas de algas microscópicas usando técnicas tales como CRISPR para direccionar hacia genes nucleares y de cloroplastos para la mutagénesis insercional y cribado. Los métodos de Stovicek y Hlavová se pueden aplicar al sistema de proteína efectora Cpf1 de la presente invención.

US 8945839 describe un método para modificar algas microscópicas (células de la especie *Chlamydomonas reinhardtii*) usando Cas9. Usando herramientas similares, se pueden aplicar los métodos del sistema Cpf1 CRISPR que se describen aquí a especies de Chlamydomonas y otras algas. En algunas formas de realización en particular, se introducen Cpf1 y el ARN guía en algas y se expresan usando un vector que expresa Cpf1 bajo el control de un promotor constitutivo tal como Hsp70A-Rbc S2 o Beta2-tubulina. El ARN guía se suministrará usando un vector que contiene un promotor T7. Como alternativa, se puede suministrar ARNm de Cpf1 y ARN guía transcrito *in vitro* a células de algas. El protocolo de electroporación sigue el protocolo estándar recomendado por el conjunto de elementos de modificación de Chlamydomonas de GeneArt.

Uso de Cpf1 en la generación de microorganismos capaces de producir ácidos grasos

En algunas formas de realización en particular, los métodos de la invención se utilizan para la generación de microorganismos modificados genéticamente capaces de producir de ésteres grasos, por ejemplo ésteres metílicos de ácidos grasos ("FAME") y ésteres etílicos de ácidos grasos ("FAEE"),

Típicamente, las células huésped se pueden modificar para que produzcan ésteres grasos a partir de una fuente de carbono, por ejemplo, un alcohol, presente en el medio, por expresión o sobreexpresión de un gen que codifica una tioesterasa, un gen que codifica una acil-CoA sintasa, y un gen que codifica una éster sintasa. Por lo tanto, los métodos que se proveen aquí se utilizan para modificar un microorganismo de manera tal de sobreexpresar o introducir un gen de tioesterasa, un gen que codifica una acil-CoA sintasa, y un gen que codifica una éster sintasa. En algunas formas de realización en particular, el gen de tioesterasa se selecciona entre tesA, 'tesA, tesB, fatB, fatB2, fatB3, fatAl, o fatA. En algunas formas de realización en particular, el gen que codifica una acil-CoA sintasa se selecciona entre fadDJadK, BH3103, pfl-4354, EAV15023, fadDl, fadD2, RPC_4074, fadDD35, fadDD22, faa39, o un gen identificado que codifica una enzima con las mismas propiedades. En algunas formas de realización en particular, el gen que codifica una éster sintasa es un gen que codifica una sintasa/acil-CoA:diacilgliceril aciltransferasa de *Simmondsia chinensis, Acinetobacter* sp. ADP, *Alcanivorax borkumensis, Pseudomonas aeruginosa, Fundibacter jadensis, Arabidopsis thaliana*, o *Alcaligenes eutrophus*, o una variante de las mismas. Adicionalmente o como alternativa, los métodos que se proveen aquí se utilizan para reducir la expresión en dicho microorganismo de por lo menos un gen que codifica una acil-CoA deshidrogenasa, un gen que codifica una proteína receptora exterior de membrana, y un gen que codifica un regulador transcripcional de la biosíntesis de ácidos grasos. En algunas formas de realización en particular uno o más de dichos genes se inactiva, por ejemplo por introducción de una mutación. En algunas formas de realización en particular, el gen que codifica una acil-CoA deshidrogenasa es fadE. En algunas formas de realización en particular,

el gen que codifica un regulador transcripcional de biosíntesis de ácidos grasos codifica un represor de la transcripción de ADN, por ejemplo, fabR.

Adicionalmente o como alternativa, dicho microorganismo se modifica para reducir la expresión de por lo menos un gen que codifica una piruvato formiato liasa, un gen que codifica una lactato deshidrogenasa, o ambos. En algunas formas de realización en particular, el gen que codifica una piruvato formiato liasa es pflB. En algunas formas de realización en particular, el gen que codifica una lactato deshidrogenasa es ldhA. En algunas formas de realización en particular uno o más de dichos genes está inactivado, por ejemplo, por introducción de una mutación en el mismo.

En algunas formas de realización en particular, el microorganismo se selecciona entre los géneros *Escherichia, Bacillus, Lactobacillus, Rhodococcus, Synechococcus, Synechoystis, Pseudomonas, Aspergillus, Trichoderma, Neurospora, Fusarium, Humicola, Rhizomucor, Kluyveromyces, Pichia, Mucor, Myceliophtora, Penicillium, Phanerochaete, Pleurotus, Trametes, Chrysosporium, Saccharomyces, Stenotrophamonas, Schizosaccharomyces, Yarrowia* o *Streptomyces*.

Uso de Cpf1 en la generación de microorganismos capaces de producir ácidos orgánicos

Los métodos que se proveen aquí se utilizan también para modificar microorganismos capaces de producir ácidos orgánicos, más particularmente a partir de azúcares que son pentosas o hexosas. En algunas formas de realización en particular, los métodos comprenden introducir en un microorganismo un gen LDH exógeno. En algunas formas de realización en particular, adicionalmente o como alternativa, la producción de ácidos orgánicos en dichos microorganismos aumenta por inactivación de genes endógenos que codifican proteínas incluidas en una vía metabólica endógena que produce un metabolito diferente del ácido orgánico de interés y/o donde la vía metabólica endógena consume el ácido orgánico. En algunas formas de realización en particular, la modificación asegura que se reduzca la producción del metabolito diferente del ácido orgánico de interés. De acuerdo con algunas formas de realización en particular, los métodos que se utilizan para introducir por lo menos una supresión y/o inactivación de un gen modificado de una vía endógena en la cual se consume el ácido orgánico o un gen que codifica un producto incluido en una vía endógena que produce un metabolito diferente del ácido orgánico de interés. En algunas formas de realización en particular, la supresión o inactivación del por lo menos un gen modificado es en uno o más genes que codifican una enzima que se selecciona entre el grupo que consiste en piruvato descarboxilasa (pdc), fumarato reductasa, alcohol deshidrogenasa (adh), acetaldehído deshidrogenasa, fosfoenolpiruvato carboxilasa (ppc), D-lactato deshidrogenasa (d-ldh), L-lactato deshidrogenasa (l-ldh), lactato 2-monooxigenasa. En otras formas de realización adicionales la supresión y/o inactivación del por lo menos un gen modificado es en un gen endógeno que codifica la piruvato descarboxilasa (pdc).

En otras formas de realización adicionales, el microorganismo se modifica para producir ácido láctico y la supresión y/o inactivación del por lo menos un gen modificado es en un gen endógeno que codifica la lactato deshidrogenasa. Adicionalmente o como alternativa, el microorganismo comprende la supresión o inactivación del por lo menos un gen modificado de un gen endógeno que codifican una lactato deshidrogenasa dependiente de citocromo, por ejemplo una L-lactato deshidrogenasa dependiente del citocromo B2.

Uso de Cpf1 en la generación de xilosa o celobiosa mejorada utilizando cepas de levaduras

En algunas formas de realización en particular, el sistema Cpf1 CRISPR se puede aplicar para seleccionar para obtener xilosa o celobiosa mejorada utilizando cepas de levadura. Se puede utilizar la PCR propensa a errores para amplificar uno (o más) genes incluidos en las vías de utilización de xilosa o utilización de celobiosa. Algunos ejemplos de genes incluidos en las vías de utilización de xilosa y vías de utilización de celobiosa puede incluir, sin limitación, a aquellos que se han descrito en Ha, S.J., y col. (2011) Proc. Natl. Acad. Sci. USA 108(2):504-9 y Galazka, J.M., y col. (2010) Science 330(6000):84-6. Las bibliotecas de moléculas de ADN de doble hebra que se obtienen como resultado, cada una de los cuales comprende una mutación al azar en un gen seleccionado con dichas características se pueden co-transformar con los componentes del sistema Cpf1 CRISPR en una cepa de levadura (por ejemplo, S288C) y se pueden seleccionar las cepas con capacidad mejorada de utilización de la xilosa o celobiosa, según se describe en WO2015138855.

Uso de Cpf1 en la generación de cepas de levaduras mejoradas para utilizar en biosíntesis de isoprenoides

Tadas Jakočiūnas y col. han descrito la aplicación exitosa de un sistema CRISPR/Cas9 múltiple para la modificación del genoma de hasta loci genómicos 5 diferentes en un paso de transformación en levadura *Saccharomyces cerevisiae* para horno (*Metabolic Engineering* Volumen 28, marzo de 2015, Páginas 213–222) para obtener como resultado cepas con alta producción de mevalonato, un intermediario clave para la vía de biosíntesis de isoprenoides, que es industrialmente importante. En algunas formas de realización en particular, el sistema Cpf1 CRISPR se puede aplicar en un método multiplex de modificación del genoma según se describe aquí para identificar cepas de levadura de alta producción adicionales para utilizar en la síntesis de isoprenoides.

Uso de Cpf1 en la generación de cepas de levaduras productoras de ácido láctico

En otra forma de realización, se abarca la aplicación exitosa de un sistema multiplexado Cpf1 CRISPR. Operando de manera análoga a Vratislav Stovicek y col. (Metabolic Engineering Communications, Volumen 2, diciembre 2015,

Páginas 13–22), se pueden diseñar y obtener cepas productoras de ácido láctico mejoradas en un único evento de transformación. En una forma de realización en particular, el sistema Cpf1 CRISPR se utiliza para insertar simultáneamente el gen heterólogo de lactato deshidrogenasa y para la interrupción de dos genes endógenos PDC1 y genes PDC5.

Aplicaciones adicionales del sistema Cpf1 CRISPR en plantas

En algunas formas de realización en particular, el sistema CRISPR, y preferiblemente el sistema Cpf1 CRISPR que se describe aquí, se pueden utilizar para la visualización de elementos genéticos dinámicos. Por ejemplo, mediante la obtención de imágenes de CRISPR se pueden visualizar secuencias genómicas ya sea repetitivas o no repetitivas, informar cambios de longitud de telómeros y desplazamientos de telómeros y monitorear la dinámica de los loci de genes durante todo el ciclo celular (Chen y col., Cell, 2013). Dichos métodos también se pueden aplicar a las plantas.

Otras aplicaciones del sistema CRISPR, y preferiblemente el sistema Cpf1 CRISPR que se describe aquí, es la interrupción dirigida del gen cribado por selección positiva *in vitro* e *in vivo* (Malina y col., *Genes and Development*, 2013). Dichos métodos también se pueden aplicar a plantas.

En algunas formas de realización en particular, la fusión de endonucleasas Cpf1 inactivas con enzimas modificadoras de histona puede introducir cambios a medida en el complejo epigenoma (Rusk y col., *Nature Methods*, 2014). Dichos métodos también se pueden aplicar a plantas.

En algunas formas de realización en particular, el sistema CRISPR, y preferiblemente el sistema Cpf1 CRISPR que se describe aquí, se pueden utilizar para purificar una parte específica de la cromatina e identificar las proteínas asociadas, para dilucidar de esa manera sus papeles regulatorios en la transcripción (Waldrip y col., *Epigenetics*, 2014). Dichos métodos también se pueden aplicar a plantas.

En algunas formas de realización en particular, la presente invención se puede utilizar como terapia para la eliminación de virus en sistemas vegetales ya que es capaz de clivar tanto el ADN como el ARN viral. Ciertos estudios previos en sistemas humanos han demostrado el éxito del uso de CRISPR en la dirección del virus de ARN de hebra simple, hepatitis C (A. Price, y col., Proc. Natl. Acad. Sci, 2015) así como el virus de ADN de doble hebra, hepatitis B (V. Ramanan, y col., Sci. Rep, 2015). Dichos métodos también se pueden adaptar al uso del sistema Cpf1 CRISPR en plantas.

En algunas formas de realización en particular, la presente invención se puede utilizar para alterar la complejidad del genoma. En otra forma de realización particular adicional, el sistema CRISPR, y preferiblemente el sistema Cpf1 CRISPR que se describe aquí, se pueden utilizar para interrumpir o alterar número de cromosomas y generar plantas haploides, que solo contiene cromosomas de un progenitor. Dichas plantas se pueden inducir para someterlas a una duplicación de cromosomas y convertirlas en plantas diploides que solo contienen alelos homocigotas (Karimi-Ashtiyani y col., PNAS. 2015; Anton y col., *Nucleus*, 2014). Dichos métodos también se pueden aplicar a plantas.

En algunas formas de realización en particular, el sistema Cpf1 CRISPR que se describe aquí, se puede utilizar para el autoclivaje. En dichas formas de realización, el promotor de la enzima Cpf1 y ARNg puede ser un promotor constitutivo y se introduce un segundo ARNg en el mismo cassette de transformación, pero controlado por un promotor inducible. Este segundo ARNg se puede diseñar para inducir el clivaje específico del sitio en el gen Cpf1 para crear un Cpf1 no funcional. En otra forma de realización particular adicional, el segundo ARNg induce el clivaje en ambos extremos del cassette de transformación, para obtener como resultado la eliminación del cassette del genoma huésped. Este sistema ofrece una duración controlada de la exposición celular a la enzima Cas y minimiza adicionalmente la edición fuera del diana. Además, se puede utilizar el clivaje de ambos extremos de un cassette CRISPR/Cas para generar plantas T0 libres transgenes con mutaciones bi-alélicas (como describen para Cas9, por ejemplo, Moore y col., *Nucleic Acids Research*, 2014; Schaeffer y col., *Plant Science*, 2015). Los métodos de Moore y col. se pueden aplicar a sistemas Cpf1 CRISPR que se describen aquí. Sugano y col. (Plant Cell Physiol., marzo de 2014; 55(3):475-81. doi: 10.1093/pcp/pcu014. Epub 2014 enero de 18) informan la aplicación de CRISPR-Cas9 a la mutagénesis dirigida en la hepática *Marchantia polymorpha* L., que ha emergido como una especie modelo para el estudio de la evolución de plantas terrestres. Se identificó el promotor U6 de *M. Polymorpha* y se clonó para expresar el ARNg. La secuencia diana del ARNg se diseñó de manera de alterar el gen que codifica al factor 1 de respuesta a auxina (ARF1) en *M. Polymorpha*. Usando una transformación mediada por Agrobacterium, Sugano y col. aislaron mutantes estables en la generación de gametofitos de *M. Polymorpha*. La mutagénesis dirigida hacia el sitio basada en CRISPR-Cas9 *in vivo* se realizó usando ya sea el virus mosaico de la coliflor 35S o el promotor de EF1α de *M. Polymorpha* para expresar Cas9. Los individuos mutantes aislados que mostraron un fenotipo resistente a auxina no fueron quiméricos. Además, los mutantes estables produjeron plantas T1 de reproducción asexual. Se establecieron fácilmente alelos arf1 múltiples usando basados en mutagénesis dirigida hacia CRIPSR-Cas9. Los métodos de Sugano y col. se pueden aplicar al sistema de proteína efectora Cpf1 de la presente invención.

Kabadi y col. (*Nucleic Acids Res.* 2014 Oct 29; 42(19): e147. doi: 10.1093/nar/gku749. Epub 13 de agosto 2014) desarrollaron un sistema lentiviral simple para expresar una variante Cas9, un gen informante y hasta cuatro ARNg de promotores de ARN polimerasa III independientes que se incorporaron al vector mediante un método de clonación Golden Gate conveniente. Cada ARNg se expresó eficientemente y puede mediar la edición de múltiples genes y la activación transcripcional sostenida en células humanas inmortalizadas y primarias. Los métodos de Kabadi y col. se

pueden aplicar al sistema de proteína efectora Cpf1 de la presente invención.

Ling y col. (*BMC Plant Biology* 2014, 14:327) desarrollaron un conjunto de vector binario CRISPR-Cas9 basado en el esqueleto principal pGreen o pCAMBIA, así como ARNg. Este conjunto de herramientas no requiere enzimas de restricción además de BsaI para generar construcciones finales que alojan Cas9 optimizada por codones de maíz y uno o más ARNg de alta eficiencia en tan poco como un paso de clonación. El conjunto de herramientas se validó usando protoplastos de maíz, líneas de maíz transgénico, y líneas de Arabidopsis transgénico y se mostró que tenían una alta eficiencia y especificidad. Más importante, usando este conjunto de herramientas, se detectaron mutaciones dirigidas de tres genes de Arabidopsis en plantones transgénicos de la generación T1. Además, la próxima generación puede heredar las mutaciones en múltiples genes. Se puede usar el conjunto de vector (ARN guía)módulo, como conjunto de herramientas para la edición de genoma multiplexado en plantas. El conjunto de herramientas de Lin y col. se puede aplicar al sistema de proteína efectora Cpf1 de la presente invención.

También se pueden obtener protocolos para la edición dirigida hacia el genoma de la planta por CRISPR-Cpf1 en base a aquellos que se divulgan para el sistema CRISPR-Cas9 en el Volumen 1284 de la serie Methods in Molecular Biology pp 239-255, 10 de febrero de 2015. Se describe un procedimiento detallado para diseñar, construir, y evaluar ARNg duales para edición del genoma mediado por Cas9 optimizada por codones de plantas (pcoCas9) usando modelos celulares de sistemas de protoplastos de *Arabidopsis thaliana* y *Nicotiana benthamiana*. También se exponen estrategias para aplicar el sistema CRISPR-Cas9 para generar modificaciones dirigidas hacia el genoma en plantas completas. Los protocolos descritos en el capítulo se pueden aplicar al sistema de proteína efectora Cpf1 de la presente invención.

Petersen ("Towards precisely glycol engineered plants", Plant Biotech Denmark Annual meeting 2015, Copenhagen, Dinamarca) desarrollaron un método para usar CRISPR/Cas9 para realizar cambios en el genoma de Arabidopsis, por ejemplo, para modificar por glicosilación a Arabidopsis para la producción de proteínas y productos con las modificaciones posteriores a la traducción que se desean. Hebelstrup y col. (Front Plant Sci. 23 de abril de 2015; 6:247) resume la biomodificación del almidón *in planta*, para proveer cultivos que expresen enzimas modificadoras del almidón y producir directamente productos que normalmente se hacen por química industrial y/o tratamientos físicos de almidones. Los métodos de Petersen y Hebelstrup se pueden aplicar al sistema de proteína efectora Cpf1 de la presente invención.

Ma y col. (Mol Plant. 3 de agosto de 2015; 8(8):1274-84. doi: 10.1016/j.molp.2015.04.007) informan un robusto sistema de vector CRISPR-Cas9, utilizando a gen de Cas9 optimizado por codones de plantas, para una edición conveniente y con alta eficiencia de genomas multiplexados en plantas monocotiledóneas y dicotiledóneas. Ma y col. diseñaron procedimientos basados en PCR para generar rápidamente múltiples casetes de expresión de ARNg, que se pueden armar en los vectores CRISPR-Cas9 binarios en un solo ciclo de clonación usando ligación Golden Gate o ensamblaje Gibson. Con este sistema, Ma y col. editaron 46 sitios diana en arroz con una tasa de mutación promedio de 85,4%, principalmente en estado bialélico y homocigota. Ma y col. proveen ejemplos de mutaciones del gen con pérdida de la función en arroz T0 y Plantas T1 de Arabidopsis por direccionamiento simultáneo de múltiples miembros de una familia de genes (hasta ocho), múltiples genes en una vía de biosíntesis, o múltiples sitios en un único gen. Los métodos de Ma y col. se pueden aplicar al sistema de proteína efectora Cpf1 de la presente invención.

Lowder y col. (Plant Physiol. 21 de agosto de 2015. Pii: pp. 00636.2015) también desarrollaron un conjunto de herramientas CRISPR-Cas9 que permite la edición de genomas multiplexados y la regulación transcripcional de expresada, silenciado o genes no codificantes en plantas. Este conjunto de herramientas provee a los investigadores con un protocolo y reactivos para montar rápida y eficientemente construcciones CRISPR-Cas9 ADN-T funcional para monocotiledóneas y dicotiledóneas usando los métodos de clonación Golden Gate y Gateway. El mismo viene con un conjunto completo de capacidades, que incluyen edición de genes multiplexados y activación transcripcional o represión de genes endógenos de plantas. La tecnología de transformación basada en ADN-T es fundamental para la moderna biotecnología, genética, biología molecular y fisiología de las plantas. Por lo tanto, los Inventores desarrollaron un método para el conjunto de Cas9 (WT, nickasa o dCas9) y ARNg(s) en un vector ADN-T de destino de interés. El método de ensamblaje se basa tanto en ensamblaje de Golden Gate y recombinación MultiSite Gateway. Son necesarios tres módulos por conjunto. El primer módulo es un vector de entrada Cas9, que contiene Cas9 sin promotor o sus genes derivados flanqueados por sitios attL1 y attR5. El segundo módulo es un vector de entrada de ARNg que contiene cassettes de expresión de ARNg de entrada flanqueados por sitios attL5 y attL2. El tercer módulo incluye attR1-attR2-que contiene vectores de ADN-T de destino que provee promotores de elección para la expresión de Cas9. El conjunto de herramientas de Lowder y col. se puede aplicar al sistema de proteína efectora Cpf1 de la presente invención.

En una forma de realización ventajosa, la planta puede ser un árbol. En la presente invención también se puede utilizar el sistema CRISPR Cas que se divulga aquí para sistemas herbáceos (véase, por ejemplo, Belhaj y col., Plant Methods 9: 39 y Harrison y col., Genes & Development 28: 1859–1872). En una forma de realización particularmente ventajosa, el sistema CRISPR Cas de la presente invención puede direccionarse hacia polimorfismos de un único nucleótido (SNPs) en árboles (véase, por ejemplo, Zhou y col., Nuevo Phytologist, Volumen 208, Entrega 2, páginas 298–301, octubre de 2015). En el estudio de Zhou y col., los autores aplicaron un sistema CRISPR Cas en el Populus [álamo] leñoso perenne usando la familia de genes de 4-cumarato:CoA ligasa (4CL) como caso de estudio y obtuvieron un 100% de eficiencia mutacional para dos genes 4CL que se usaron como diana, donde cada transformante que se

examinó tenía modificaciones bialélicas. En el estudio de Zhou y col., el sistema CRISPR-Cas9 era altamente sensible a los polimorfismos de un único nucleótido (SNPs), cuando se abolió el clivaje para un tercer gen 4CL debido a SNPs en la secuencia diana. Dichos métodos se pueden aplicar al sistema de proteína efectora Cpf1 de la presente invención.

Los métodos de Zhou y col. (Nuevo Phytologist, Volumen 208, Entrega 2, páginas 298–301, octubre de 2015) se pueden aplicar a la presente invención de la siguiente manera. Dos genes 4CL, 4CL1 y 4CL2, asociados a la biosíntesis de lignina y flavonoides, se direccionan respectivamente para la edición de CRISPR-Cas9. El clon de *Populus tremula* × *alba* 717-1B4 que se utiliza rutinariamente para transformación es divergente del *Populus trichocarpa* cuyo genoma está secuenciado. Por lo tanto, los ARNg de 4CL1 y 4CL2 que se diseñaron a partir del genoma de referencia se interrogan con datos de la secuencia 717 de ARN propios de los investigadores para asegurar la ausencia de SNPs que puedan limitar la eficiencia de Cas. También se incluye un tercer ARNg diseñado para 4CL5, un duplicado del genoma de 4CL1. La correspondiente secuencia 717 aloja un SNP en cada alelo cercano/dentro del PAM, y se espera que ambos abolan el direccionamiento por 4CL5-ARNg. Los tres sitios diana de ARNg están situados dentro del primer exón. Para la transformación 717, el ARNg se expresa desde el promotor Medicago U6.6, junto con un Cas optimizado por codones humanos bajo control del promotor CaMV 35S en un vector binario. La transformación con el único vector Cas puede servir como control. Las líneas 4CL1 y 4CL2 seleccionadas al azar se someten a secuenciación por amplicón. Luego los datos se procesan y se confirman las mutaciones bialélicas en todos los casos. Dichos métodos se pueden aplicar al sistema de proteína efectora Cpf1 de la presente invención.

En las plantas, los patógenos frecuentemente son específicos del huésped. Por ejemplo, *Fusarium oxysporum* f. Sp. *lycopersici* causa la marchitez del tomate, pero solo ataca al tomate, y *F. oxysporum f. dianthii Puccinia graminis* f. Sp. *tritici* solo ataca al trigo. Las plantas poseen defensas existentes e inducidas para resistir a la mayoría de los patógenos. Los eventos de recombinación y mutaciones a lo largo de generaciones de plantas dan lugar a una variabilidad genética que ocasiona susceptibilidad, especialmente debido a que los patógenos se reproducen con más frecuencia que las plantas. En las plantas puede existir una resistencia de tipo no huésped, por ejemplo, el huésped y el patógeno son incompatibles. También puede existir una resistencia horizontal, por ejemplo, resistencia parcial contra todas las razas de un patógeno, normalmente controlada por muchos genes y la resistencia vertical, por ejemplo, resistencia completa a algunas razas de un patógeno pero no a otras razas, controlada normalmente por unos pocos genes. En un nivel gen-por-gen, las plantas y los patógenos evolucionan juntos y los cambios genéticos en uno equilibran los cambios en el otro. En consecuencia, utilizando la variabilidad natural, los obtentores combinan los genes más útiles respecto al rendimiento, calidad, uniformidad, robustez y resistencia. Las fuentes de genes de resistencia incluyen variedades nativas o foráneas, variedades reliquia, relacionadas con plantas naturales y mutaciones inducidas, por ejemplo, tratando un material vegetal con agentes mutagénicos. Utilizando la presente invención, se proporciona a los obtentores de plantas una nueva herramienta para inducir mutaciones. En consecuencia, el experto en la técnica puede analizar el genoma de las fuentes de genes de resistencia, y emplear la presente invención con las variedades que tengan las características o rasgos deseados para inducir la aparición de genes de resistencia, con más precisión que los agentes mutagénicos previos y, por lo tanto, acelerar y mejorar los programas de cultivo selectivo de plantas.

Células mejoradas de plantas y levaduras

También se describen, pero no se reivindican específicamente en la presente, células de plantas y levadura que se pueden obtener y se obtienen por los métodos que se proveen aquí. Las plantas mejoradas que se obtienen por los métodos que se describen aquí pueden ser útiles en la producción de alimentos o piensos mediante la expresión de genes que, por ejemplo, aseguren la tolerancia a las plagas de las plantas, herbicidas, inundaciones, bajas o altas temperaturas, exceso de agua, etc.

Las plantas mejoradas que se obtienen por los métodos que se describen aquí, especialmente los cultivos y algas pueden ser útiles en la producción de alimentos o piensos mediante la expresión, por ejemplo, de los niveles de proteínas superiores, carbohidratos, nutrientes o vitaminas que normalmente se pueden ver en el tipo salvaje. Al respecto, se prefieren las plantas mejoradas, especialmente legumbres y tubérculos.

Las algas u otras plantas transgénicas tales como la colza podrán ser particularmente útiles en la producción de aceites vegetales o biocombustibles tales como alcoholes (especialmente metanol y etanol), por ejemplo. Estas se podrán modificar para expresar o sobreexpresar niveles elevados de aceite o alcoholes para su uso en las industrias de los aceites o biocombustibles.

También se describen, pero no se reivindican específicamente en la presente, partes mejoradas de una planta. Las partes de las plantas incluyen, pero de manera no taxativa, hojas, tallos, raíces, tubérculos, semillas, endospermo, óvulos, y polen. Las partes de las plantas que se prevén aquí pueden ser viables, no viables, regenerable, y/o no regenerable.

La presente también abarca la provisión de células de plantas y plantas generadas de acuerdo con los métodos de la invención. Los gametos, semillas, embriones, ya sea cigóticos o somáticos, progenie o híbridos de plantas que comprenden la modificación genética, que se producen por métodos de cría tradicionales, también están incluidos dentro del alcance de la presente invención. Dichas plantas pueden contener una secuencia de ADN heteróloga o extraña insertada en o en vez de una secuencia diana. Como alternativa, dichas plantas pueden contener solo una

alteración (mutación, supresión, inserción, sustitución) en uno o más nucleótidos. Por lo tanto, dichas plantas solo serán diferentes de sus plantas progenitoras por la presencia de la modificación particular.

Por lo tanto, la invención proporciona una planta, animal o célula producidos por los métodos de la presente o una progenie de estos. La progenie podrá ser un clon de la planta o animal producidos o podrá ser el resultado de la reproducción sexual por cruzamiento con otros individuos de la misma especie para lograr la introgresión de rasgos deseables adicionales en su descendencia. La célula puede ser *in vivo* o *ex vivo* en los casos de organismos multicelulares, en particular animales o plantas.

Los complejos de proteína efectora Cpf1 se pueden utilizar en organismos / animales no humanos

También se describe, pero no se reivindica en la presente, un organismo eucariótico no humano, tal como un organismo eucariótico multicelular, que comprende una célula huésped eucariótica como se describe en la presente. También se describe, pero no se reivindica en la presente, un organismo eucariótico; tal como un organismo eucariótico multicelular, que comprende una célula huésped eucariótica como se describe en la presente. El organismo puede ser un animal; por ejemplo, un mamífero. También, el organismo puede ser un artrópodo tal como un insecto. El organismo también puede ser una planta. Además, el organismo puede ser un hongo.

También se describen, pero no se reivindican específicamente en la presente, otras aplicaciones agrícolas tales como, por ejemplo, animales de granja y producción. Por ejemplo, los cerdos tienen muchas características que los hacen atractivos como modelos biomédicos, especialmente en medicina regenerativa. En particular, los cerdos con inmunodeficiencia combinada grave (SCID) pueden proveer modelos útiles para la medicina regenerativa, xenotrasplantes (se exponen aquí también en otra parte), y desarrollo tumoral y contribuirán al desarrollo de terapias para pacientes humanos de SCID. Lee y col., (Proc Natl Acad Sci U S A. 20 de mayo de 2014; 111(20):7260-5) utilizaron un sistema de nucleasa efectora similar al activador de la transcripción guiado por informante (TALEN) para generar dirigida hacia gen activador de modificaciones de recombinación (RAG) 2 en células somáticas con alta eficiencia, incluyendo a algunos que afectaban a ambos alelos. La proteína efectora Cpf1 se puede aplicar a un sistema similar.

Los métodos de Lee y col., (Proc Natl Acad Sci USA. 20 de mayo de 2014; 111(20):7260-5) se pueden aplicar a la presente invención de manera análoga a lo siguiente. Los cerdos mutados se producen por modificación de RAG2 dirigida en células de fibroblasto fetal seguido de SCNT y transferencia de embriones. Las construcciones que codifican para CRISPR Cas y un informante se electroporan en células de fibroblasto derivadas de fetos. Luego de 48 h, células transfectadas que expresan la proteína fluorescente verde se clasifican en pocillos individuales de una placa de 96 pocillos con una dilución estimada de una única célula por pocillo. La modificación dirigida de RAG2 se criba por amplificación de un fragmento de ADN genómico que flanquea a cualquier sitio de corte CRISPR Cas seguido de secuenciación de los productos de PCR. Luego del cribado y de asegurar la ausencia de mutaciones fuera del sitio, para la SCNT se utilizan células que llevan modificación dirigida de RAG2. El cuerpo polar, junto con una parte del citoplasma de oocito adyacente, que presumiblemente contiene la placa de metafase II, se elimina, y se dispone una célula donante en el perivitelino. Luego se electroporan los embriones reconstruidos para fusionar la célula donante con el oocito y luego se activan químicamente. Los embriones activados se incuban en medio para cigotos porcinos 3 (PZM3) con Scriptaid 0,5 µM (S7817; Sigma-Aldrich) durante 14–16 h. Luego los embriones se lavan para eliminar el Scriptaid y se cultivan en PZM3 hasta que se transfieren a los oviductos de cerdos sustitutos.

La presente invención se puede aplicar también para modificar SNPs de otros animales, por ejemplo, vacas. Tan y col. (Proc Natl Acad Sci U S A. 2013 Oct 8; 110(41): 16526-16531) expandieron el conjunto de herramientas para edición de genes del ganado para que incluya reparación dirigida por homología (HDR) estimulada por nucleasas efectoras de tipo activador de la transcripción (TAL) (TALEN) y grupos de repeticiones palindrómicas cortas en intervalos regulares (CRISPR)/Cas9 usando plásmido, rAAV y moldes de oligonucleótidos. Se clonaron secuencias de ARNg específicas del gen en el vector de ARNg del laboratorio del Dr. Church (Addgene ID: 41824) de acuerdo con sus métodos (Mali P, y col. (2013) *RNA-Guided Human Genome Engineering via Cas9*. Science 339(6121):823-826). La nucleasa Cas9 se suministró ya sea por co-transfección del plásmido hCas9 (Addgene ID: 41815) o ARNm sintetizado a partir de RCIScript-hCas9. Este RCIScript-hCas9 se construyó por sub-clonación del fragmento XbaI-AgeI a partir del plásmido hCas9 (que abarca al ADNc de hCas9) en el plásmido RCIScript.

Heo y col. (Stem Cells Dev. 1 de febrero de 2015; 24(3):393-402. doi: 10.1089/scd.2014.0278. Epub 3 de noviembre de 2014) informaron un direccionamiento de genes altamente eficiente en el genoma bovino usando células pluripotenciales bovinas y grupos de repetición palindrómica corta agrupada en intervalos regulares (CRISPR)/ nucleasa Cas9. Primero, Heo y col. generaron células pluripotenciales indiferenciadas inducidas (iPSCs) a partir de fibroblastos somáticos bovino por la expresión ectópica de factores yamanaka y tratamiento con inhibidor de GSK3β y MEK (2i). Heo y col. observaron que dichos iPSCs bovinos son altamente similares a células pluripotenciales indiferenciadas no tratadas previamente con respecto a la expresión de un gen y potencial de desarrollo en teratomas. Además, CRISPR-nucleasa Cas9, que era específica para el locus de NANOG bovino, mostraron una edición altamente eficiente del genoma bovino en iPSCs bovinas y embriones.

Igenity® provee un perfil de análisis de animales, por ejemplo, vacas, para llevar a cabo y transmitir rasgos con importancia económica, por ejemplo composición de la res, calidad de la res, rasgos maternales y reproductivos y ganancia diaria promedio. El análisis de un perfil Igenity® amplio comienza con el descubrimiento de marcadores de

ADN (más frecuentemente polimorfismos de un único nucleótido o SNPs). Todos los marcadores detrás del perfil Igenity® fueron descubiertos por científicos independientes en instituciones de investigación, que incluyen universidades, organizaciones de investigación, y entidades gubernamentales tales como la USDA. Luego, los marcadores se analizaron en Igenity® en poblaciones de validación. Igenity® utiliza múltiples poblaciones de recurso que representan diversos ambientes de producción y tipos biológicos, frecuentemente trabajando con socios de la industria de los segmentos de: abastecimiento de semilla, cría de vacas-terneros, *feedlots* y/o envasado, de la industria de la carne, para recolectar fenotipos que no sean comúnmente asequibles. Las bases de datos del genoma del ganado se encuentran ampliamente disponibles, véase, por ejemplo, el NAGRP Cattle Genome Coordination Program (http://www.animalgenome.org/cattle/maps/db.html). Por lo tanto, la presente invención se puede aplicar para direccionar SNPs bovinos. Alguien con experiencia en el arte puede utilizar los anteriores protocolos para el direccionamiento SNPs y aplicarlos a los SNPs bovinos según han descrito, por ejemplo, Tan y col. o Heo y col.

Qingjian Zou y col. (Journal of Molecular Cell Biology Advance Access, publicado el 12 de octubre de 2015) demostraron un incremento de la masa muscular en perros por direccionamiento el primer exón del gen de miostatina (MSTN) del perro (un regulador negativo de la masa muscular esquelética). Primero, se validó la eficiencia de los ARNg, usando cotransfección del ARNg direccionado hacia MSTN con un vector Cas9 en fibroblastos embriónicos caninos (CEFs). Luego de eso, se generaron perros MSTN KO por micro-inyección de embriones con morfología normal con una mezcla de ARNm de Cas9 y ARNg de MSTN y auto-trasplante de los cigotos en el oviducto de la misma perra. Los cachorros noqueados mostraron un fenotipo muscular obvio en los muslos al compararlos con su hermana de tipo salvaje de la misma camada. Esto también se puede llevar a cabo usando los sistemas Cpf1 CRISPR que se proveen aquí.

Ganado – Cerdos

En algunas formas de realización, los blancos virales en el ganado pueden incluir CD163 porcino, por ejemplo, en macrófagos porcinos. CD163 se asocia a la infección (se cree que es por la entrada viral en la célula) por PRRSv (virus del síndrome reproductivo y respiratorio porcino, un arterivirus). La infección por PRRSv, especialmente de macrófagos alveolares porcinos (encontrados en el pulmón), da como resultado un síndrome porcino que anteriormente era incurable ("enfermedad misteriosa del cerdo" o "enfermedad de la oreja azul") que causa padecimientos que incluyen fracaso reproductivo, pérdida de peso y altas tasas de mortalidad en los cerdos domésticos. Frecuentemente se ven infecciones oportunistas, por ejemplo, neumonía enzoótica, meningitis y edema en las orejas, debido a deficiencias inmunitarias por pérdida de actividad de macrófagos. La misma también tiene significativas repercusiones económicas y ambientales debido al aumento del uso de antibióticos y pérdidas financieras (una estimación de $660m al año).

Como informaron Kristin M Whidosrth y Dr Randall Prather y col. (Nature Biotech 3434 publicado en línea el 07 de diciembre de 2015) en la University of Missouri y en colaboración con Genus Plc, CD163 se direccionó usando CRISPR-Cas9 y la progenie de los cerdos editados fue resistente cuando se expuso a PRRSv. Se cruzaron un fundador macho y una hembra fundadora, ambos con mutaciones en el exón 7 de CD163, para obtener una progenie. El fundador macho poseía una supresión de 11 pb en el exón 7 en un alelo, lo que da como resultado una mutación de desplazamiento del marco y traducción con sentido erróneo en el aminoácido 45 del domino 5 y un subsiguiente codón de parada prematura en el aminoácido 64. El otro alelo tenía una adición de 2 pb en el exón 7 y una supresión de 377 pb en el intrón precedente, que se predijeron que darían como resultado en la expresión de los primeros 49 aminoácidos del domino 5, seguido de un código de parada prematuro en el aminoácido 85. La cerda tenía una adición de 7 pb en un alelo que cuando se tradujo se predijo que expresaría los primeros 48 aminoácidos del domino 5, seguido de un codón de parada prematura en el aminoácido 70. Otro alelo se la cerda no era amplificable. Se predijo que la progenie seleccionada sería un animal deficiente (CD163-/-), es decir con noqueo para CD163.

Por lo tanto, en algunas formas de realización, los macrófagos alveolares porcinos se pueden direccionar por la proteína CRISPR. En algunas formas de realización, el CD163 porcino se puede direccionar por la proteína CRISPR. En algunas formas de realización, el CD163 porcino se puede noquear por inducción de un DSB o por inserciones o supresiones, por ejemplo, direccionamiento de la supresión o modificación del exón 7, incluyendo uno o más de aquellos que se han descrito anteriormente, o en otras regiones del gen, por ejemplo, supresión o modificación del exón 5.

También se prevén un cerdo editado y su progenie, por ejemplo, un cerdo con CD163 noqueado. Esto puede ser para propósitos de ganado, cría o modelado (es decir un modelo porcino). También se provee semen que comprende el noqueo del gen.

CD163 es miembro de la superfamilia de receptores secuestrantes ricos en cisteína (SRCR). En base a los estudios *in vitro* del domino 5 de SRCR de la proteína es el domino responsable del desempaquetamiento y liberación del genoma viral. Por lo tanto, otros miembros de la superfamilia SRCR también se pueden direccionar para evaluar la resistencia a otros virus. PRRSV es también un miembro del grupo de los arterivirus de mamíferos, que también incluye virus murino elevador de la lactato deshidrogenasa, virus de la fiebre hemorrágica de simios y virus de la arteritis en equinos. Los arterivirus comparten importantes propiedades de patogénesis, que incluyen el tropismo de macrófagos y la capacidad de causar tanto enfermedades graves como infección persistente. Por lo tanto, los arterivirus, y en particular el virus murino elevador de la lactato deshidrogenasa, el virus de la fiebre hemorrágica de simios y el virus

de la arteritis en equinos, se pueden direccionar, por ejemplo mediante el CD163 porcino u homólogos del mismo también se proveen en otras especies, y modelos murinos, de simio y equinos y noqueados.

Por cierto, este enfoque se puede extender a los virus o bacterias que causan otras enfermedades del ganado que pueden ser transmitidas a los seres humanos, por ejemplo, cepas de virus de gripe porcina (SIV) incluyendo a la influenza C y los subtipos de influenza A conocidos como H1N1, H1N2, H2N1, H3N1, H3N2 y H2N3, así como neumonía, meningitis y edema mencionados anteriormente.

Direccionamiento terapéutico con complejo de proteína efectora Cpf1 guiada por ARN

Como será evidente, se prevé que el presente sistema se puede utilizar para direccionar cualquier secuencia de polinucleótidos de interés en los métodos reivindicados. Los métodos de la invención comprenden el uso de una composición que no se encuentra naturalmente o modificada, o uno o más polinucleótidos que codifican componentes de dicha composición, o vectores o sistemas de administración que comprenden uno o más polinucleótidos que codifican componentes de dicha composición para utilizar en la modificación de una célula diana *ex vivo* o *in vitro* y, que se pueden conducir de manera que alteren la célula de manera tal que una vez modificada, la progenie o línea celular de la célula modificada con CRISPR retenga el fenotipo alterado. Las células y la progenie modificadas pueden estar en una muestra obtenida a partir de un organismo multicelular tal como una planta o animal con la aplicación *ex vivo* del sistema CRISPR a los tipos de células que se desean. Se describe, pero no se reivindica en la presente, un método de tratamiento terapéutico. El método de tratamiento terapéutico puede comprender un gen o la edición del genoma, o una terapia génica.

Tratamiento de patógenos, tales como patógenos bacterianos, fúngicos y parasitarios

Se describen, pero no se reivindican en la presente, métodos para tratar organismos patógenos bacterianos, fúngicos y parasitarios. La mayoría de los esfuerzos de investigación se ha enfocado en el desarrollo de nuevos antibióticos, que, una vez desarrollados, sin embargo pueden estar sujetos a los mismos problemas de resistencia a las fármacos. La solicitud provee novedosas alternativas basadas en CRISPR que solucionan dichas dificultades. Además, al contrario de los antibióticos existentes, los tratamientos basados en CRISPR se pueden hacer específicos para un patógeno, inducir muerte de una célula bacteriana de un patógeno diana a la vez que se evitan las bacterias beneficiosas.

Jiang y col. ("*RNA-guided editing of bacterial genomes using CRISPR-Cas systems*", Nature Biotechnology vol. 31, p. 233-9, marzo de 2013) utilizaron un sistema CRISPR-Cas9 para mutar o matar *S. pneumoniae* y *E. coli*. El trabajo, que introdujo mutaciones precisas en los genomas, se basó en clivaje dirigido a ARN doble:Cas9 en el sitio genómico diana para matar células no mutadas y evitar la necesidad de marcadores seleccionable o sistemas de contra-selección. Los sistemas CRISPR se deben utilizar para revertir la resistencia a antibióticos y eliminar la transferencia de resistencia entre cepas. Bickard y col. mostraron que Cas9, reprogramada para direccionarla hacia genes de virulencia, mata a las *S. aureus* virulentas, pero no a las no virulentas. La reprogramación de la nucleasa para dirigirla hacia genes de resistencia a antibióticos destruyó plásmidos estafilocócicos que alojaban genes de resistencia a antibióticos e inmunizados contra la dispersión de genes de resistencia transportados por plásmidos. (véase, Bikard y col., "*Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials*", Nature Biotechnology vol. 32, 1146-1150, doi: 10.1038/nbt.3043, publicado en línea el 05 octubre de 2014). Bikard y col. mostraron que los antimicrobianos CRISPR-Cas9 funcionan *in vivo* para matar *S. aureus* en un modelo de colonización de la piel del ratón. De manera similar, Yosef y col. utilizaron un sistema CRISPR para direccionar hacia genes que codifican enzimas que confieren resistencia a antibióticos β-lactámicos (véase Yousef y col., "*Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria*", Proc. Natl. Acad. Sci. USA, vol. 112, p. 7267–7272, doi: 10.1073/pnas.1500107112 publicado en línea 18 de mayo de 2015). Los sistemas CRISPR se pueden utilizar para editar genomas de parásitos que son resistentes a otros enfoques genéticos. Por ejemplo, se mostró que un sistema CRISPR-Cas9 introduce roturas en la de doble hebra en el genoma de *Plasmodium yoelii* (véase, Zhang y col., "*Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System*", mBio. vol. 5, e01414-14, jul-ago 2014). Ghorbal y col. ("*Genome editing in the human malaria parasite Plasmodium falciparumusing the CRISPR-Cas9 system*" Nature Biotechnology, vol. 32, p. 819-821, doi: 10.1038/nbt.2925, publicado en línea 1 de junio de 2014) modificaron las secuencias de dos genes, *orc1* y *kelch13*, que tenían papeles putativos en silenciamiento de genes y resistencia emergente a la artemisinina, respectivamente. Los parásitos que fueron alterados en los sitios apropiados se recuperaron con muy alta eficiencia, a pesar de que no hubo una selección directa para la modificación, indicando que usando este sistema se pueden generar mutaciones neutras o aún perjudiciales. CRISPR-Cas9 se utiliza también para modificar los genomas de otros parásitos patogénicos, que incluyen *Toxoplasma gondii* (véase Shen y col., "*Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9*" mBio vol. 5:e01114-14, 2014; y Sidik y col., "*Efficient Genome Engineering of* Toxoplasma gondii *Using CRISPR/Cas9*" PLoS One vol. 9, e100450, doi: 10.1371/journal.pone.0100450, publicado en línea, 27 de junio de 2014).

Vyas y col. ("*A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families*" Science Advances, vol. 1, e1500248, DOI: 10.1126/sciadv.1500248, 3 de abril de 2015) emplearon un sistema CRISPR para solucionar obstáculos persistentes a largo plazo para la modificación genética en *C. albicans* y mutar eficientemente en un único experimento ambas copias de varios genes diferentes. En un organismo donde varios mecanismos contribuyen a la resistencia a los fármacos, Vyas produjo mutantes dobles homocigotas que ya mostraron

la hiper-resistencia al fluconazol o cicloheximida que mostró el aislado clínico parental Can90. Vyas también obtuvo mutaciones homocigotas de pérdida de la función en genes esenciales de *C. albicans* por creación de alelos condicionales. Los alelos nulos de *DCR1*, que es necesario para el procesamiento ribosómico del ARN, son letales a baja temperatura, pero viables a alta temperatura. Vyas utilizó un molde de reparación que introdujo una mutación sin sentido y aisló mutantes *dcr1/dcr1* que no pudieron crecer a 16 °C.

El sistema CRISPR de la presente invención para utilizar en *P. falciparum* por interrupción de loci cromosómicos. Ghorbal y col. ("*Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system*", Nature Biotechnology, 32, 819-821 (2014), DOI: 10.1038/nbt.2925, 1 de junio de 2014) emplearon un sistema CRISPR para introducir genes noqueados específicos y sustituciones en un único nucleótido en el genoma de malaria. Para adaptar el sistema CRISPR-Cas9 a *P. falciparum*, Ghorbal y col. generaron vectores de expresión para controlar elementos regulatorios plasmodiales en el episoma de pUF1-Cas9 que también lleva el marcador seleccionable por fármaco y*dhodh*, que otorga resistencia a DSM1, un inhibidor de la dihidroorotato deshidrogenasa de *P. falciparum* (PfDHODH) y para la transcripción de los ARNg, se utilizaron elementos regulatorios nucleares pequeños de ARN(np) U6 de *P. falciparum* para disponer el ARN guía y el molde de ADN donante para la reparación de la recombinación homóloga en el mismo plásmido, pL7. Véase también, Zhang C. y col. ("*Efficient editing of malaria parasite genome using the CRISPR/Cas9 system*", MBio, 1 de julio de 2014; 5(4):E01414-14, doi: 10.1128/MbIO0.1414-14) y Wagner y col. ("*Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum*", *Nature Methods* 11, 915-918 (2014), DOI: 10.1038/nmet.3063).

Tratamiento de patógenos, patógenos similares a los virales tales como el VIH

La edición del genoma mediada por Cas se podría utilizar para introducir mutaciones protectoras en tejidos somáticos para combatir enfermedades no genéticas o complejas. Por ejemplo, inactivación del receptor CCR5 mediada por NHEJ en linfocitos (Lombardo y col., Nat Biotechnol. Noviembre de 2007; 25(11):1298-306) puede ser una estrategia viable para eludir la infección por VIH, mientras que la supresión de PCSK9 (Cohen y col., Nat Genet. febrero de 2005; 37(2):161-5) orangiopoyetina (Musunuru y col., N Engl J Med. 2 de diciembre de 2010, 363(23):2220-7) pueden proveer efectos terapéuticos contra la hipercolesterolemia o hiperlipidemia resistente a estatinas. Aunque dichos objetivos se pueden realizar también usando mediado por ARNip por *knockdown* de proteínas, una ventaja única de la inactivación de genes mediada por NHEJ es la capacidad de obtener un beneficio terapéutico permanente sin necesidad de tratamiento continuo. Al igual que con todas las terapias génicas, por supuesto será importante establecer que cada uso terapéutico propuesto tiene una proporción beneficio-riesgo favorable.

La administración hidrodinámica de ADN de plásmido que codifican ARN guía nd de Cas9 junto con un molde de reparación en el hígado de un modelo de tirosinemia en ratón adulto se mostró que es capaz de corregir el gen Fah mutante y rescatar la expresión de la proteína Fah de tipo salvaje en ~1 de cada 250 células (Tebas y col., N Engl J Med. 6 de marzo de 2014; 370(10):901-10). Ambos resultados demuestran que las nucleasas programables son prometedoras como una nueva plataforma terapéutica.

En otra forma de realización, se pueden usar los vectores lentivirales autoinactivantes con un ARNip direccionado hacia un exón en común compartido por tat/rev de VIH, un señuelo para la localización nucleolar de TAR, y una ribozima cabeza de martillo específica anti-CCR5 (véase, por ejemplo, DiGiusto y col. (2010) Sci Transl Med 2:36ra43) y/o adaptar al sistema CRISPR-Cas de la presente invención. Se podrán recoger un mínimo de 2,5 × 10^6 células CD34+ por kilogramo de peso del paciente y preestimular durante 16 a 20 horas en medio X-VIVO 15 (Lonza) que contiene 2 µmol/L-glutamina, factor citoblástico (100 ng/ml), ligando Flt-3 (Flt-3L) (100 ng/ml) y trombopoyetina (10 ng/ml) (CellGenix) con una densidad de 2 × 10^6 células/ml. Las células preestimuladas se podrán transducir con lentivirus con una multiplicidad de infección de 5 durante 16 a 24 horas en matraces de cultivo tisular de 75 cm^2 recubiertos con fibronectina (25 mg/cm^2) (RetroNectin, Takara Bio Inc.).

Con el conocimiento en la técnica y las enseñanzas en la presente divulgación, el experto en la materia puede corregir los HSC en una afección de inmunodeficiencia, tal como VIH / SIDA que comprende poner en contacto un HSC con un sistema CRISPR-Cas9 que se dirige y atenúa a CCR5. Un ARN guía (y de manera ventajosa a enfoque de guía doble, por ejemplo, se ponen en contacto con HSCs un par de ARN guía diferentes; por ejemplo, direccionamiento de ARN guía de dos genes clínicamente relevantes, B2M y CCR5, en células T CD4+ humanas primarias y Células CD34+ del tallo hematopoyético y progenitoras (HSPCs)) que direccionan y noquean partículas que contienen proteína CCR5-y-Cpf1. Las células que se ponen en contacto de esa manera se pueden administrar; y opcionalmente tratar / expandir; cf. Cartier. Véase también Kiem, "*Hematopoietic stem cell-based gene therapy for HIV disease*" Cell Stem Cell. 3 de febrero de 2012; 10(2): 137–147; Mandal et al, "*Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9*," Cell Stem Cell, Volumen 15, 5ª edición, p643-652, 6 de noviembre de 2014. También se hace mención a Ebina, "*CRISPR/Cas9 system to suppress HIV-1 expression by editing HIV-1 integrated proviral DNA*" SCIENTIFIC REPORTS | 3: 2510 | DOI: 10.1038/srep02510, como otro medio para combatir el VIH/SIDA usando un sistema CRISPR-Cas9.

El fundamento para la edición del genoma para el tratamiento de VIH es la observación de que los individuos homocigotas para las mutaciones de pérdida de función en CCR5, un co-receptor celular para el virus, son altamente resistentes a las infecciones y saludables en todos los demás aspectos, lo que sugiere que el mimetizar esta mutación con una edición del genoma puede ser una estrategia terapéutica segura y eficaz [Liu, R., y col., *Cell* 86, 367-377

(1996)]. Esto idea fue validada clínicamente cuando a un paciente infectado con VIH se le dio un trasplante alogénico de médula ósea de un donante homocigota para una mutación de pérdida de función de CCR5, dio como resultado unos niveles indetectables de VIH y restauración de un conteo normal de células T CD4 [Hutter, G., y col. *The New England journal of medicine* 360, 692-698 (2009)]. Aunque los trasplantes de médula ósea no son una estrategia de tratamiento realista para la mayoría de los pacientes de VIH, debido a su costo y al potencial de enfermedad de injerto contra huésped, son deseables unas terapias para VIH que conviertan las células T del mismo paciente en CCR5-.

Los primeros estudios usando ZFNs y NHEJ para noquear CCR5 en modelos de ratón humanizado de VIH mostraron que los trasplantes de células T CD4 con CCR5 editado mejoraron carga viral y el conteo de células T CD4 [Perez, E.E., y col. Nature Biotechnology 26, 808-816 (2008)]. De manera importante, estos modelos también demostraron que la infección por VIH dio como resultado la selección respecto de células nulas para CCR5, lo que sugiere que la edición confiere una ventaja de adaptación y permite potencialmente que un pequeño número de células editadas creen un efecto terapéutico.

Como resultado de este y otros prometedores estudios preclínicos, la terapia de edición del genoma que noquea a CCR5 en las células T del paciente tiene ahora se ha probado en seres humanos [Holt, N., y col. *Nature Biotechnology* 28, 839-847 (2010); Li, L., y col. *Molecular therapy: the journal of the American Society of Gene Therapy* 21, 1259-1269 (2013)]. En una reciente prueba clínica de fase I, se extrajeron células T CD4+ de pacientes con VIH, se editaron con ZFNs modificados para noquear el gen CCR5, y se trasplantaron de manera autóloga de vuelta a los pacientes [Tebas, P., y col. *The New England journal of medicine* 370, 901-910 (2014)].

En otro estudio (Mandal y col., *Cell Stem Cell*, Volumen 15, Entrega 5, p643–652, 6 de noviembre de 2014), se direccionó a CRISPR-Cas9 hacia dos genes con relevancia clínica, B2M y CCR5, en células T CD4+ humano y células CD34+ del tallo hematopoyético y progenitoras (HSPCs). El uso de ARN guías simples llevó a una mutagénesis altamente eficiente en HSPCs pero no en las células T. Un enfoque de guía doble mejoró la eficacia de supresión del gen en ambos tipos de células. Las HSPCs que se sometieron a edición del genoma con CRISPR-Cas9 retuvieron su potencial multilinaje. Se examinaron las mutaciones predichas en el diana y fuera del mismo por secuenciación con captura de blancos en HSPCs y se observaron bajos niveles de mutagénesis fuera del diana en un solo sitio. Dichos resultados demuestran que CRISPR-Cas9 puede realizar eficientemente una ablación de genes en HSPCs con una mínima mutagénesis fuera del diana, lo que tiene una amplia aplicabilidad para las terapias basadas en células hematopoyéticas.

Wang y col. (PLoS One. 26 de diciembre de 2014; 9(12):e115987. doi: 10.1371/journal.pone.0115987) silenciaron CCR5 mediante la proteína 9 asociada a CRISPR (Cas9) y ARN guía simples (ARN guía) con vectores lentivirales que expresan Cas9 y ARN guía de CCR5. Wang y col. mostraron que una única vuelta transducción de vectores lentivirales que expresaban Cas9 y ARN guía de CCR5 en células CD4+ humanas susceptibles de VIH-1 da altas frecuencias de interrupción del gen de CCR5. Las células con el gen CCR5 interrumpido no solo son resistentes al VIH-1 R5-trópico, incluyendo a los aislados de VIH-1 transmitido/fundador (T/F), sino que también tienen una ventaja selectiva sobre células con el gen CCR5 ni interrumpido durante la infección por VIH-1 R5-trópico. Mutaciones en el genoma en sitios potencialmente fuera del diana que son altamente homólogos con dichos ARN guía de CCR5 en células transducidas de manera estable aún a los 84 días después de la transducción no fueron detectados por un ensayo con endonucleasa I de T7.

Fine y col. (Sci Rep. julio de 2015 1; 5:10777. doi: 10.1038/srep10777) identificaron un sistema de dos cassettes que expresa partes de la proteína Cas9 de S. Pyogenes (SpCas9) que realizan un corte y se empalman en las células para formar una proteína funcional capaz de realizar un clivaje de ADN específico para el sitio. Con hebras guía específicas para CRISPR, Fine y sus colaboradores demostraron la eficacia de este sistema para clivar los genes HBB y CCR5 en células HEK-293T humanas como una Cas9 simple y como un par de nickasa Cas9 s. La SpCas9 de transempalme (tsSpCas9) mostró ~35% de la actividad nucleasa al compararlos con la SpCas9 de tipo salvaje (wtSpCas9) en dosis de transfección estándar, pero tuvo una actividad sustancialmente menor con menores niveles de dosificación. La longitud muy reducida del marco de lectura abierto de tsSpCas9 en relación con wtSpCas9 permite empaquetar elementos genéticos potencialmente más complejos y de mayor longitud en un vector AAV que incluye promotores específicos para un tejido, expresión de ARN guía multiplexado, y fusiones de dominios efectores con SpCas9.

Li y col. (J Gen Virol. agosto 2015; 96(8):2381-93. doi: 10.1099/vir.0.000139. Epub 8 de abril de 2015) demostraron que CRISPR-Cas9 puede mediar eficientemente la edición del locus de CCR5 en líneas celulares, para obtener como resultado el noqueo de la expresión de CCR5 sobre la superficie de la célula. La secuenciación de la próxima generación reveló que se introdujeron diversas mutaciones alrededor del sitio de clivaje de CCR5 predicho. Para cada una de los tres ARN guía más eficaces que se analizaron, no se detectaron efectos significativos fuera del diana en los 15 potenciales sitios con mayor puntaje. Al construir adenovirus Ad5F35 quiméricos que llevan componentes de CRISPR-Cas9, Li y col. transdujeron eficientemente linfocitos T CD4+ primarios e interrumpieron la expresión de CCR5, y a las células transducidos positivamente se les confirió resistencia al VIH-1.

Alguien con experiencia en el arte puede utilizar los anteriores estudios, por ejemplo, de Holt, N., y col. Nature Biotechnology 28, 839-847 (2010), Li, L., y col. *Molecular therapy: the journal of the American Society of Gene Therapy* 21, 1259-1269 (2013), Mandal y col., Cell Stem Cell, Volumen 15, Entrega 5, p643–652, 6 de noviembre de 2014, Wang y col. (PLoS One. 26 de diciembre de 2014; 9(12):e115987. doi: 10.1371/journal.pone.0115987), Fine y col. (Sci

Rep. 1 de julio de 2015; 5:10777. doi: 10.1038/srep10777) y Li y col. (J Gen Virol. agosto de 2015, 96(8):2381-93. doi: 10.1099/vir.0.000139. Epub 8 de abril de 2015) para el direccionamiento hacia CCR5 con el sistema CRISPR Cas de la presente invención.

Tratamiento de patógenos, patógenos similares a los virales, tales como VHB

Se describen, pero no se reivindican en la presente, métodos que se pueden aplicar para tratar el virus de la hepatitis B (VHB). Sin embargo, el sistema CRISPR Cas se debe adaptar para evitar las deficiencias de las iARN, tales como el riesgo de sobresaturar rutas de ARN pequeño endógeno mediante, por ejemplo, la optimización de la dosis y secuencia (véase, por ejemplo, Grimm y col., Nature vol. 441, 26 de mayo de 2006). Por ejemplo, se contemplan dosis bajas, por ejemplo, de aproximadamente 1-10 × 10^{14} partículas por ser humano. En otra forma de realización, el sistema CRISPR Cas dirigido contra el VHB se podrá administrar en liposomas, tales como una partícula de ácido nucleico-lípido estable (SNALP) (remítase a, por ejemplo, Morrissey y col., Nature Biotechnology, vol. 23, n.º 8, agosto de 2005). Se contemplan inyecciones intravenosas diarias de aproximadamente 1, 3 o 5 mg/kg/día de un CRISPR Cas específico dirigido a ARN de HBV en una SNALP. El tratamiento diario podrá prolongarse durante aproximadamente tres días y a continuación semanalmente durante aproximadamente cinco semanas. En otra forma de realización, el sistema de Chen y col. (Gen Therapy (2007) 14, 11-19) se puede usar y/o adaptar para los sistemas CRISPR Cas descritos en la presente. Chen et al. utilizan un vector de pseudotipo del virus adenoasociado 8 bicatenario (AAV2bc/8) para suministrar ARNhc. Una simple administración del vector dsAAV2/8 (1 × 10^{12} genomas de vector por ratón), portadores de ARNhc específico de VHB, suprimieron eficazmente el nivel estable de la proteína de VHB, ARNm y ADN replicativo en el hígado de ratones VHB transgénicos, que lleva a obtener una reducción de hasta 2–3 $\log_{10}$ de la carga de VHB en circulación. Se mantuvo una supresión de HBV significativa durante al menos 120 días tras la administración del vector. El efecto terapéutico del ARNhc no fue dependiente de la secuencia diana y no conllevó la activación de interferón. Como se describe en la presente, un sistema CRISPR Cas direccionado hacia VHB se puede clonar en un vector AAV, por ejemplo, un vector dsAAV2/8 y se puede administrar a un ser humano, por ejemplo, con una dosificación de entre aproximadamente 1 × 10^{15} genomas de vector y aproximadamente 1 × 10^{16} genomas de vector por ser humano. En otra forma de realización, el método de Wooddell y col. (Molecular Therapy vol. 21 no. 5, 973–985 mayo de 2013) se puede usar y/o adaptar a los sistemas CRISPR Cas descritos en la presente. Woodell y col. muestran que la simple coinyección de un péptido de tipo melitina conjugado con N-acetilgalactosamina, cuya diana son los hepatocitos (NAG-MLP) con un ARNip conjugado con colesterol hepatotrópico (col-ARNip) cuya diana es el factor de coagulación VII (F7) daba como resultado una inactivación génica de F7 eficaz en ratones y primates no humanos sin cambios en la química clínica ni la inducción de citocinas. Utilizando modelos en ratones transgénicos y transitorios de la infección de VHB, Wooddell et al. muestran que una única coinyección de NAG-MLP con col-ARNip potente cuya diana son las secuencias de VHB conservadas dio como resultado una represión multilog de ARN vírico, proteínas y ADN vírico con una duración prolongada del efecto. Para los métodos aquí descritos, se podrán contemplar coinyecciones intravenosas, por ejemplo, de aproximadamente 6 mg/kg de NAG-MLP y 6 mg/kg de CRISPR Cas específico de VHB. Como alternativa, se podrán suministrar en el día uno aproximadamente 3 mg/kg de NAG-MLP y 3 mg/kg de CRISPR Cas específico de VHB, seguidos por la administración de aproximadamente 2-3 mg/kg de NAG-MLP y 2-3 mg/kg de CRISPR Cas específico de HBV dos semanas más tarde.

Lin y col. (Mol Ther Nucleic Acids. 19 de agosto de 2014; 3:e186. doi: 10.1038/mtna.2014.38) diseñaron ocho ARNg contra VHB del genotipo A. Con los ARNg específicos para el VHB, el sistema CRISPR-Cas9 redujo significativamente la producción del núcleo del VHB y proteínas superficiales en células Huh-7 transfectadas con un VHB-vector de expresión. Entre ocho ARNg cribados, se identificaron dos eficaces. Un ARNg direccionado hacia la secuencia del VHB conservada actuó contra diferentes genotipos. Usando un modelo de persistencia-hidrodinámica de VHB en ratón, Lin y col. demostraron además que este sistema puede clivar el plásmido que contiene el genoma de VHB intrahepático y facilitar su eliminación *in vivo*, para obtener como resultado una reducción de niveles en suero de antígeno superficial. Dichos datos sugieren que el sistema CRISPR-Cas9 puede alterar los moldes que expresan el VHB tanto *in vitro* como *in vivo*, indicando su potencial para erradicar la infección persistente por VHB.

Dong y col. (Antiviral Res. junio de 2015; 118:110-7. doi: 10.1016/j.antiviral.2015.03.015. Epub 3 de abril de 2015) utilizaron el sistema CRISPR-Cas9 para direccionar el genoma de VHB e inhibir eficientemente la infección por VHB. Dong y col. sintetizaron cuatro ARN guía simples (ARN guía) que direccionan hacia las regiones conservadas de VHB. La expresión de dichos ARN guía con Cas9 redujo la producción viral en células Huh7 así como la replicación de VHB en células HepG2.2.15. Dong y col. demostraron además que CRISPR-Cas9 direcciona el clivaje y ocurrió una mutagénesis mediada por clivaje del ADNccc de VHB de células transfectadas. En el modelo de ratón portador de ADNccc de VHB, la inyección rápida de plásmidos de ARN guía-Cas9 en la vena de la cola dio como resultado un bajo nivel de ADNccc y proteína de VHB.

Liu y col. (J Gen Virol. agosto de 2015, 96(8):2252-61. doi: 10.1099/vir.0.000159. Epub abril de 2015 22) diseñaron ocho ARN guías (ARNg) que direccionaban hacia las regiones conservadas de diferentes genotipos de VHB, que pueden inhibir significativamente la replicación del VHB tanto *in vitro* como *in vivo* para investigar la posibilidad de usar el sistema CRISPR-Cas9 para alterar los moldes de ADN de VHB. El sistema específico para el VHB de ARNg/Cpf1 pudo inhibir la replicación del VHB de diferentes genotipos en células, y el ADN viral se redujo significativamente por un simple sistema ARNg/Cpf1 y se eliminó mediante una combinación de diferentes sistemas ARNg/Cpf1.

Wang y col. (World J Gastroenterol. 28 de agosto de 2015; 21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554) diseñaron

15 ARNg contra VHB de genotipos A-D. Se seleccionaron once combinaciones de los dos anteriores ARNg (ARN dobles) que abarcaban la región regulatoria del VHB. Se estudió la eficiencia de cada ARNg y 11 ARN dobles sobre la supresión de la replicación de VHB (genotipos A-D) por medición del antígeno superficial del VHB (HBsAg) o el antígeno (HBeAg) en el sobrenadante del cultivo. Se estudió la destrucción del vector que expresaba VHB en células Huh7 co-transfectadas con ARN dobles y vectores que expresaban VHB usando el método de la reacción en cadena de la polimerasa (PCR) y secuenciación, y se estudió la destrucción de ADNccc en células HepAD38 usando precipitación con KCl, digestión con DNasa dependiente de ATP segura para el plásmido (PSAD), método combinado amplificación de círculo rodante y PCR cuantitativa. La citotoxicidad de dichos ARNg se evaluó con un ensayo mitocondrial con tetrazolio. Todos los ARNg pudieron reducir significativamente la producción de HBsAg o HBeAg en el sobrenadante del cultivo, de manera dependiente de la región contra la cual actuaba el ARNg. Todos los ARNg duales pudieron suprimir eficientemente HBsAg y/o la producción de HBeAg para VHB de genotipos A-D, y la eficacia de ARNg duales en la supresión de HBsAg y/o la producción de HBeAg aumentó significativamente en comparación con el ARNg simple utilizado por sí solo. Además, mediante la secuenciación directa por PCR los inventores confirmaron que dichos ARNg duales pudieron destruir específicamente al molde que expresa el VHB por eliminación del fragmento entre los sitios de clivaje de los dos ARNg que se utilizan. Más importante, la combinación de ARNg-5 y ARNg-12 no solo pudo suprimir eficientemente a HBsAg y/o la producción de HBeAg, sino que también destruye los reservorios de ADNccc en células HepAD38.

Karimova y col. (Sci Rep. Septiembre de 2015 3; 5:13734. doi: 10.1038/srep13734) identificaron secuencias conservadas de VHB de genotipo cruzado en las regiones S y X del genoma de VHB que estaban direccionadas para un clivaje específico y eficaz por una nickasa Cas9. Con este enfoque no solo se alteró el ADNccc episómico y los sitios diana del VHB integrados cromosómicamente en líneas celulares informantes, sino también la replicación del VHB en líneas celulares de hepatoma infectadas crónicamente y de novo.

Alguien con experiencia en el arte puede utilizar los anteriores estudios, por ejemplo, de Lin y col. (Mol Ther Nucleic Acids. 19 de agosto de 2014; 3:e186. doi: 10.1038/mtna.2014.38), Dong y col. (Antiviral Res. junio de 2015; 118:110-7. doi: 10.1016/j.antiviral.2015.03.015. Epub 3 de abril de 2015), Liu y col. (J Gen Virol. agosto de 2015, 96(8):2252-61. doi: 10.1099/vir.0.000159. Epub 22 de abril de 2015), Wang y col. (World J Gastroenterol. 28 de agosto de 2015; 21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554) y Karimova y col. (Sci Rep. 3 de septiembre de 2015, 5:13734. doi: 10.1038/srep13734) para el direccionamiento contra VHB con el sistema CRISPR Cas como se describe en la presente.

La infección crónica con el virus de la hepatitis B (VHB) es predominante, mortal, y rara vez se cura, debido a la persistencia del ADN viral episómico (ADNccc) en las células infectadas. Ramanan y col. (Ramanan V, Shlomai A, Cox DB, Schwartz RE, Michailidis E, Bhatta A, Scott DA, Zhang F, Arroz CM, Bhatia SN. Sci Rep. 2 de junio de 2015; 5:10833. doi: 10.1038/srep10833, publicado en línea el 2 de junio de 2015) mostraron que el sistema CRISPR/Cas9 puede direccionarse específicamente y clivar regiones conservadas en el genoma de VHB, para obtener como resultado una robusta supresión de la expresión del gen viral y su replicación. Con una expresión sostenida de Cas9 y ARN guía seleccionados apropiadamente, ellos demostraron el clivaje de ADNccc por Cas9 y una impresionante reducción tanto del ADNccc como de otros parámetros de expresión del gen viral y su replicación. Por lo tanto, ellos mostraron que el direccionamiento directo al ADN viral episómico es un novedoso enfoque terapéutico para controlar el virus y posiblemente curar los pacientes. Esto se describe también en WO2015089465 A1, a nombre de The Broad Institute y col..

Por lo tanto, en algunas formas de realización se prefiere el direccionamiento al ADN viral episómico en VHB.

También se describen, pero no se reivindican en la presente, métodos para tratar organismos patógenos, por ejemplo, patógenos bacterianos, fúngicos y parasitarios. La mayoría de los esfuerzos de investigación se ha enfocado en el desarrollo de nuevos antibióticos, que, una vez desarrollados, puedan sin embargo estar sujetos a los mismos problemas de resistencia a los fármacos. La solicitud provee novedosas alternativas basadas en CRISPR que resuelven dichas dificultades. Además, al contrario de los antibióticos existentes, los tratamientos basados en CRISPR se pueden hacer específicos para un patógeno, induciendo la muerte de una célula bacteriana de un patógeno diana, pero sin afectar las bacterias beneficiosas.

También se describen, pero no se reivindican en la presente, métodos para tratar el virus de la hepatitis C (VHC). Los métodos de Roelvinki y col. (Molecular Therapy vol. 20 no. 9, 1737-1749 septiembre de 2012) se pueden aplicar al sistema CRISPR/Cas. Por ejemplo, un vector que se contempla puede ser un vector AAV tal como AAV8 y por ejemplo se puede contemplar una dosificación de aproximadamente $1,25 \times 10^{11}$ a $1,25 \times 10^{13}$ genomas de vector por kilogramo de peso corporal (vg/kg). La solicitud también describe métodos que se puede aplicar para tratar organismos patógenos, por ejemplo, patógenos bacterianos, fúngicos y parasitarios. La mayoría de los esfuerzos de investigación se ha enfocado en el desarrollo de nuevo antibióticos, que, una vez desarrollado, pueden estar sujetos a los mismos problemas de resistencia a los fármacos. La solicitud provee novedosas alternativas basadas en CRISPR que resuelven dichas dificultades. Además, al contrario de los antibióticos existentes, los tratamientos basados en CRISPR se pueden hacer específicos para un patógeno, induciendo la muerte de una célula bacteriana de un patógeno diana pero sin afectar las bacterias beneficiosas.

Jiang y col. ("*RNA-guided editing of bacterial genomes using CRISPR-Cas systems*", Nature Biotechnology vol. 31, p.

233-9, marzo de 2013) utilizaron un sistema CRISPR-Cas9 para mutar o matar *S. Pneumoniae* y *E. coli*. El trabajo, donde se introdujeron mutaciones precisas en los genomas, se basó en clivaje dirigido a ARN doble:Cas9 en el sitio genómico diana para matar células no mutadas y evitar la necesidad de marcadores seleccionables o sistemas de contra-selección. Los sistemas CRISPR se deben utilizar para revertir la resistencia a antibióticos y eliminar la transferencia de la resistencia entre cepas. Bickard y col. mostraron que la Cas9, reprogramada para direccionarla hacia genes de virulencia, mata a las *S. aureus* virulentas, pero no a las no virulentas. La reprogramación de la nucleasa para direccionar hacia genes de resistencia a antibióticos destruyó plásmidos estafilocócicos que alojan genes de resistencia a antibióticos e inmunizó contra la dispersión de genes de resistencia transportados por plásmidos. (véase, Bikard y col., "*Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials*", Nature Biotechnology vol. 32, 1146-1150, doi: 10.1038/nbt.3043, publicado en línea el 05 de octubre de 2014). Bikard mostró que los antimicrobianos CRISPR-Cas9 tienen *in vivo* la función de matar a *S. aureus* en un modelo de colonización de la piel del ratón. De manera similar, Yosef y col. utilizaron un sistema CRISPR para direccionar hacia genes que codifican enzimas que confieren resistencia a los antibióticos β-lactámicos (véase Yousef y col., "*Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria*", Proc. Natl. Acad. Sci. USA, vol. 112, p. 7267–7272, doi: 10.1073/pnas.1500107112 publicado en línea 18 de mayo de 2015).

Los sistemas CRISPR se pueden utilizar para editar genomas de parásitos que son resistentes a otros enfoques genéticos. Por ejemplo, se mostró que un sistema CRISPR-Cas9 introduce roturas en la doble hebra del genoma de *Plasmodium yoelii* (véase, Zhang y col., "*Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System*" mBio. vol. 5, e01414-14, julio-agosto de 2014). Ghorbal y col. ("*Genome editing in the human malaria parasite Plasmodium falciparumusing the CRISPR-Cas9 system*", Nature Biotechnology, vol. 32, p. 819-821, doi: 10.1038/nbt.2925, publicado en línea el 1 de junio de 2014) modificaron las secuencias de dos genes, *orc1* y *kelch13*, que tienen papeles putativos en el silenciamiento de genes y la resistencia emergente a artemisinina, respectivamente. Se recuperaron parásitos que fueron alterados en los sitios apropiados con muy alta eficiencia, a pesar de no haber una selección directa para la modificación, indicando que se pueden generar mutaciones neutras o aún perjudiciales usando este sistema. CRISPR-Cas9 se utiliza también para modificar los genomas de otros parásitos patogénicos, incluyendo a *Toxoplasma gondii* (véase Shen y col., "*Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9*" mBio vol. 5:e01114-14, 2014; y Sidik y col., "*Efficient Genome Engineering of Toxoplasma gondii Using CRISPR/Cas9*" PLoS One vol. 9, e100450, doi: 10.1371/journal.pone.0100450, publicado en línea, 27 de junio de 2014).

Vyas y col. ("A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families" Science Advances, vol. 1, e1500248, DOI: 10.1126/sciadv.1500248, 3 de abril de 2015) emplearon un sistema CRISPR para solucionar obstáculos persistentes a largo plazo para la modificación genética en *C. albicans* y mutar eficientemente en un único experimento ambas copias de varios genes diferentes. En un organismo donde varios mecanismos contribuyen a la resistencia a los fármacos, Vyas produjo mutantes dobles homocigotas que ya no mostraron la hiper-resistencia a fluconazol o cicloheximida que mostraba el aislado clínico parental Can90. Vyas también obtuvo mutaciones homocigotas de pérdida de la función en genes esenciales de *C. albicans* por creación de alelos condicionales. Los alelos nulos de *DCR1*, que son necesarios para el procesamiento ribosómico del ARN, son letales a baja temperatura, pero viables a altas temperaturas. Vyas utilizó un molde de reparación que introdujo una mutación sin sentido y aisló *dcr1/dcr1* mutantes que no pudieron crecer a 16 °C.

Tratamiento de enfermedades con aspectos genéticos o epigenéticos

Los sistemas CRISPR-Cas de la presente invención se pueden utilizar para corregir mutaciones genéticas que previamente se habían intentado con éxito limitado usando TALEN y ZFN y se identificaron como objetivos potenciales para sistemas Cas9, que incluyen a las aplicaciones publicadas de Editas Medicine donde se describen métodos para usar sistemas Cas9 para direccionar loci para tratar terapéuticamente enfermedades con terapia génica, incluyendo a WO 2015/048577 CRISPR-RELATED METHODS AND COMPOSITIONS de Gluckmann y col.; WO 2015/070083 CRISPR-RELATED METHODS AND COMPOSITIONS WITH GOVERNING gRNAS de Glucksmann y col.. También se describen, pero no se reivindican en la presente, métodos para el tratamiento, la profilaxis o el diagnóstico de glaucoma primario de ángulo abierto (POAG, por las siglas en inglés de *Primary Open Angle Glaucoma*). El diana es preferiblemente el gen MYOC. Esto se describe en WO2015153780.

Se menciona WO2015/134812 CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING USHER SYNDROME AND RETINITIS PIGMENTOSA de Maeder y col., mediante las descripciones de la presente, la invención abarca a los métodos y materiales de dichos documentos aplicados en conjunto con las descripciones de la presente dentro del alcance de los métodos y composiciones reivindicados en la presente. La solicitud prevé que un método de terapia génica ocular y auricular, los métodos y composiciones para tratar el síndrome de Usher y retinitis pigmentosa se pueden adaptar al sistema CRISPR-Cas descrito en la presente (véase, por ejemplo, WO 2015/134812). También se describe en la presente WO 2015/134812 incluye un tratamiento o el retardo del inicio o el progreso del síndrome de Usher de tipo IIA (USH2A, USH11A) y la retinitis pigmentosa 39 (RP39) por edición de genes, por ejemplo, usando métodos mediados por CRISPR-Cas9 para corregir la supresión de la guanina en la posición 2299 en el gen USH2A (por ejemplo, reemplazo del residuo de guanina suprimido en la posición 2299 en el gen USH2A). Se puede obtener un efecto similar con Cpf1. También se describe en la presente, se direcciona una mutación por clivado ya sea con una o más nucleasas, una o más nickasas, o con una combinación de las mismas, por ejemplo, para inducir HDR con un molde donante que corrige la mutación puntual (por ejemplo, la supresión de un único nucleótido, por ejemplo,

guanina). La alteración o corrección del gen mutante USH2A puede estar mediada por cualquier mecanismo. Algunos mecanismos indicativos que pueden estar asociados con la alteración (por ejemplo, la corrección) del gen mutante HSH2A incluyen, pero de manera no taxativa, unión de extremos no homólogos, unión de extremos mediada por microhomología (MMEJ), por reparación dirigida por homología (por ejemplo, mediada por molde donante endógeno), SDSA (templado de hebra dependiente de síntesis), templado de hebra simple o invasión de hebra simple. En una forma de realización, el método que se utiliza para tratar el síndrome de Usher y la retinitis pigmentosa puede incluir la adquisición de conocimientos sobre la mutación que tiene el sujeto, por ejemplo, por secuenciación la parte apropiada del gen USH2A.

También se menciona a WO 2015/138510 y mediante las descripciones en la presente la solicitud describe (usando un sistema CRISPR-Cas9) proveer un tratamiento o retrasar el inicio o el progreso de la amaurosis congénita de Leber 10 (LCA 10). La LCA 10 es causada por una mutación en el gen CEP290, por ejemplo, una mutación adenina por guanina c.2991+1655, en el gen CEP290 que origina un sitio de empalme críptico en el intron 26. Esto es una mutación en el nucleótido 1655 del intron 26 de CEP290, por ejemplo, una mutación A a G. CEP290 también se conoce como: CT87; MKS4; POC3; rd16; BBS14; JBTS5; LCAJO; NPHP6; SLSN6; y 3H11Ag (véase, por ejemplo, WO 2015/138510). En un aspecto de terapia génica, la invención incluye introducir una o más roturas cerca del sitio de la posición diana de LCA (por ejemplo, c.2991 + 1655; A a G) en por lo menos un alelo del gen CEP290. La alteración de la posición diana de LCA10 se refiere a (1) la introducción inducida por rotura de un indel (que aquí también se denomina una introducción de un indel mediada por NHEJ) estrechamente cercana a una posición diana de LCA10 o que la incluye (por ejemplo, c.2991+1655A a G), o (2) la supresión inducida por rotura (que aquí también se denomina una supresión mediada por NHEJ) de la secuencia genómica que incluye a la mutación en una posición diana de LCA10 (por ejemplo, c.2991+1655A a G). Ambos enfoques originan la pérdida o la destrucción del sitio de empalme críptico que es el resultado de la mutación en la posición diana de LCA 10. Por lo tanto, se prevé específicamente el uso de Cpf1 en el tratamiento de LCA.

Los investigadores contemplan si se pueden emplear terapias génicas para tratar un amplio rango de enfermedades. Se describen, pero no se reivindican específicamente en la presente, que los sistemas CRISPR de la presente invención basados en proteínas efectoras Cpf1, incluyendo, pero de manera no taxativa, a las áreas diana y métodos de administración adicionales que se ejemplifican, según se dicen más adelante. Aquí también se proveen algunos ejemplos de condiciones o enfermedades que se podrían tratar de manera útil usando el presente sistema, incluyendo a los ejemplos de genes y referencias que se incluyen aquí y que actualmente se asocian con aquellas condiciones. Los ejemplos de genes y condiciones no son exhaustivos.

Tratamiento de enfermedades del sistema circulatorio

La presente solicitud también contempla suministrar el sistema CRISPR-Cas, específicamente los novedosos sistemas de proteína efectora CRISPR que se describen aquí, a la sangre o células hematopoyética indiferenciadas s. Los exosomas en plasma de Wahlgren y col. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) se han descrito previamente y se pueden utilizar para suministrar el sistema CRISPR Cas a la sangre. El sistema de direccionamiento hacia ácidos nucleicos descrito en la presente se contempla también, pero no se reivindica, en la presente solicitud, para tratar hemoglobinopatías, por ejemplo talasemias y anemia falciforme. Véase, por ejemplo, la Publicación Internacional de Patente n.º WO 2013/126794 para determinar blancos potenciales a las que se puede dirigir el sistema CRISPR Cas.

Drakopoulou, "Artículo de revisión, The Ongoing Challenge of Hematopoietic Stem Cell-Based Gene Therapy for β-Thalassemia," Stem Cells International, Volumen 2011, ID del artículo 987980, 10 páginas, doi:10.4061/2011/987980, tal como si se hubiese expuesto en su totalidad, discute la modificación de HSC usando un lentivirus que administra un gen para β-globina o γ-globina. En contraste al uso de lentivirus, con el conocimiento en la técnica y las enseñanzas en la presente divulgación, el experto en la materia puede corregir a las HSC respecto de la β-talasemia usando un sistema CRISPR-Cas9 que se dirige a y corrige la mutación (por ejemplo, con un molde de HDR adecuado que administra una secuencia codificante para β-globina o γ-globina, ventajosamente β-globina o γ-globina no formadoras de células falciformes); específicamente, el ARNsg puede dirigirse a la mutación que da lugar a la β-talasemia, y el HDR puede proporcionar codificación para la expresión correcta de β-globina o γ-globina. Un ARN guía direccionado hacia la partícula que contiene la mutación-y-proteína Cas se pone en contacto con HSCs que llevan la mutación. La partícula también puede contener un molde de HDR apropiado para corregir la mutación para la correcta expresión de β-globina o γ-globina; o la HSC se puede poner en contacto con una segunda partícula o un vector que contiene o suministra el molde de HDR. Las células que se ponen en contacto de esa manera se pueden administrar; y opcionalmente tratar / expandir; cf. Cartier. Al respecto se hace mención de: Cavazzana, "Resultados of Gene Therapy for β-Talasemia Major via Transplantation of Autologous Hematopoietic Stem Cells Transduced *Ex vivo* with a Lentiviral β^{A-T87Q}-Globin Vector". tif2014.org/abstractFiles/Jean%20Antoine%20Ribeil_Abstract.pdf; Cavazzana-Calvo, "Transfusion independence and HMGA2 activation after gene therapy of human β-talasemi", Nature 467, 318–322 (16 de septiembre de 2010) doi: 10.1038/nature09328; Nienhuis, "Development of Gene Therapy for Talasemia, Cold Spring Harbor Perpsectives in Medicine, doi: 10.1101/cshperspect.a011833 (2012), LentiGlobin BB305, un vector lentiviral que contiene un gen de β-globina modificado (βA-T87Q); y Xie y col., "Seamless gene correction of β-talasemi mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyback" Genome Research gr.173427.114 (2014) http://www.genome.org/cgi/doi/10.1101/gr.173427.114 (Cold Spring Harbor Laboratory Press); es decir, en el tema del trabajo de Cavazzana que implica β-talasemia humana y el tema del trabajo de Xie. Como se describe en la presente,

el molde de HDR puede proporcionar la expresión del gen de β-globina en las HSC (por ejemplo, βA-T87Q), o de β-globina como en Xie.

Xu y col. (Sci Rep. julio de 2015 9; 5:12065. doi: 10.1038/srep12065) diseñaron TALENs y CRISPR-Cas9 para direccionar directamente al sitio de la mutación IVS2-654 del intron2 en el gen de globina. Xu y col. observaron diferentes frecuencias de roturas de doble-hebra (DSBs) en los loci IVS2-654 usando TALENs y CRISPR-Cas9, y los TALENs mediaron una mayor eficiencia del direccionamiento a genes homólogos en comparación con CRISPR-Cas9 cuando se combinan con el donante del transposón piggyBac. Además, se observaron eventos fuera del diana más obvios para CRISPR-Cas9 en comparación con los TALENs. Por último, se seleccionaron clones iPSC corregidos por TALENs para la diferenciación de eritroblastos usando el sistema de co-cultivo OP9 y se detectó una transcripción relativamente mayor de HBB que con las células no corregidas.

Song y col. (Stem Cells Dev. 1 de mayo de 2015; 24(9):1053-65. doi: 10.1089/scd.2014.0347. Epub 5 de febrero de 2015) utilizaron CRISPR/ Cas9 para corregir β-Thal de iPSCs; las células con los genes corregidos muestran cariotipos normales y pluripotencia completa como células indiferenciadas embriónicas humanas (hESCs) no mostraron efectos fuera del direccionamiento. Luego, Song y col. evaluaron la eficiencia de diferenciación de las iPSCs con los genes β-Thal corregidos. Song y col. descubrieron que, durante la diferenciación hematopoyética, las iPSCs con los genes β-Thal corregidos mostraron una mayor proporción de cuerpos embrioides y diversos porcentajes de células progenitoras hematopoyéticas. Más importante, las líneas de iPSC con los genes β-Thal corregidos restituyó la expresión de HBB y redujo la producción de especies de oxígeno reactivo al compararlas con las del grupo no corregido. El estudio de Song y col. sugirió que la eficiencia hematopoyética de la diferenciación de iPSCs con β-Thal mejoró mucho una vez corregidas por el sistema CRISPR-Cas9. Otros métodos similares se pueden llevar a cabo utilizando los sistemas CRISPR-Cas que se describen aquí, por ejemplo, sistemas que comprenden proteínas efectoras Cpf1.

La anemia de células falciformes es una enfermedad genética autosómica recesiva en la que los glóbulos rojos sanguíneos adoptan una forma falciforme. La misma es causada por una simple sustitución de bases en el gen de β-globina, que está situado en el brazo corto del cromosoma 11. Como resultado, se produce valina en vez de ácido glutámico, causando la producción de hemoglobina falciforme (HbS). Esto da como resultado la formación de una forma distorsionada de los eritrocitos. Debido a esta forma anormal, pueden bloquearse los vasos sanguíneos pequeños, causando un gran daño a los tejidos óseos, del bazo y de la piel. Esto puede causar episodios de dolor, infecciones frecuentes, síndrome del pie y la mano o incluso fallo multiorgánico. Los eritrocitos deformados también son más susceptibles a la hemólisis, lo que origina una anemia grave. Como en el caso de la â-talasemia, la anemia de células falciformes puede corregirse modificando los HSC con el sistema CRISPR/Cas9. El sistema permite la edición específica del genoma de la célula cortando su ADN y después dejando que se repare por sí mismo. La proteína Cas9 se inserta y dirige por un ARN guía al punto mutado y después corta el ADN en ese punto. Simultáneamente, se inserta una versión sana de la secuencia. Esta secuencia se usa por el propio sistema de reparación de la célula para reparar el corte inducido. De esta manera, el sistema CRISPR-Cas permite corregir la mutación en las células indiferenciadas que se obtuvieron previamente. Con el conocimiento en la técnica y las enseñanzas en la presente divulgación, el experto en la materia puede corregir a las HSC respecto de la anemia de células falciformes usando un sistema CRISPR-Cas9 que se dirige a y corrige la mutación (por ejemplo, con un molde de HDR adecuado que administra una secuencia codificante para β-globina, ventajosamente β-globina no formadora de células falciformes); específicamente, el ARNsg puede dirigirse a la mutación que da lugar a la anemia de células falciformes, y el HDR puede proporcionar codificación para la expresión correcta de β-globina. Una partícula que contiene ARN guía direccionado hacia la mutación-y-proteína Cas se pone en contacto con HSCs que llevan la mutación. La partícula también puede contener un molde de HDR apropiado para corregir la mutación para la correcta expresión de β-globina; o la HSC se puede poner en contacto con una segunda partícula o un vector que contiene o suministra el molde de HDR. Las células que se ponen en contacto de esa manera se pueden administrar; y opcionalmente tratar / expandir; cf. Cartier. El molde de HDR puede proveer para la HSC para expresar un gen de β-globina modificado (por ejemplo, βA-T87Q), o β-globina, como en Xie.

Williams, "*Broadening the Indications for Hematopoietic Stem Cell Genetic Therapies," Cell Stem Cell* 13:263-264 (2013), como si se expusiera por completo, comunica la transferencia génica mediada por lentivirus en células HSC/P de pacientes con la enfermedad del almacenamiento lisosómico, leucodistrofia metacromática (MLD), una enfermedad genética causada por una deficiencia de la arilsulfatasa A (ARSA), que da como resultado la desmielinización de los nervios; y la transferencia génica en HSC de pacientes con síndrome de Wiskott-Aldrich (WAS) (pacientes con la proteína WAS defectuosa, un efector de la GTPasa pequeña CDC42 que regula la función citoesquelética en linajes celulares sanguíneos y que por lo tanto padecen de inmunodeficiencia con infecciones recurrentes, síntomas autoinmunitarios y trombocitopenia con plaquetas anormalmente pequeñas y disfuncionales que dan lugar a un sangrado excesivo y a un riesgo aumentado de leucemia y linfoma. Al contrario del uso de lentivirus, con los conocimientos del arte y las descripciones de la presente invención, una persona con experiencia puede corregir las HSCs con una MLD (deficiencia de arilsulfatasa A (ARSA)) usando un sistema CRISPR-Cas direccionado hacia la mutación (deficiencia de arilsulfatasa A (ARSA)) (por ejemplo, con un molde de HDR apropiado que suministre una secuencia codificante para ARSA) y que la corrige; específicamente, el ARN guía puede direccionarse hacia una mutación que origina una MLD (ARSA deficiente), y la HDR puede proveer codificación para la correcta expresión de ARSA. Se pone en contacto una partícula que contiene ARN guía direccionado hacia la mutación-y-proteína Cas con HSCs que llevan la mutación. La partícula también puede contener un molde de HDR apropiado para corregir la mutación para la correcta expresión de ARSA; o la HSC se puede poner en contacto con una segunda partícula o un

vector que contiene o suministra el molde de HDR. Las células que se ponen en contacto de esa manera se pueden administrar; y opcionalmente tratar / expandir; cf. Cartier. Al contrario del uso de lentivirus, con los conocimientos en el arte y las descripciones en la presente invención, una persona con experiencia puede corregir HSCs como WAS usando un sistema CRISPR-Cas direccionado hacia la mutación (deficiencia de proteína WAS) (por ejemplo, con un molde de HDR apropiado que suministre una secuencia codificante para la proteína WAS) y que la corrige; específicamente, el ARN guía puede direccionarse hacia una mutación que origina una WAS (deficiencia de proteína WAS), y la HDR puede proveer codificación para la correcta expresión de la proteína WAS. Se pone en contacto una partícula que contiene ARN guía direccionado hacia la mutación-y-proteína Cpf1 con HSCs que llevan la mutación. La partícula también puede contener un molde de HDR apropiado para corregir la mutación para la correcta expresión de proteína WAS, o la HSC se puede poner en contacto con una segunda partícula o un vector que contiene o suministra el molde de HDR. Las células que se ponen en contacto de esa manera se pueden administrar; y opcionalmente tratar / expandir; cf. Cartier.

Watts, "Hematopoietic Stem Cell Expansion and Gene Therapy" Cytotherapy 13(10):1164-1171. doi:10.3109/14653249.2011.620748 (2011), como si se expusiera por completo, discute la terapia génica con células madre hematopoyéticas (HSC), por ejemplo, la terapia génica con células madre hematopoyéticas (HSC) mediada por virus, como una opción de tratamiento altamente atractiva para muchos trastornos, incluyendo afecciones hematológicas, inmunodeficiencias, incluyendo VIH/SIDA, y otros trastornos genéticos, tales como enfermedades de almacenamiento lisosómico, incluyendo SCID-X1, ADA-SCID, β-talasemia, CGD ligado a X, síndrome de Wiskott-Aldrich, anemia de Fanconi, adrenoleucodistrofia (ALD) y leucodistrofia metacromática (MLD).

Las publicaciones de patente de los EE.UU. n.º 20110225664, 20110091441, 20100229252, 20090271881 y 20090222937 asignadas a Cellectis, se refiere a variantes de CREI, donde por lo menos uno de los dos monómeros I-CreI tiene por lo menos dos sustituciones, uno en cada uno de los dos subdominos funcionales del domino núcleo LAGLIDADG (SEQ ID NO: 26) situados respectivamente a partir de las posiciones 26 a 40 y 44 a 77 de I-CreI, donde dicha variante es capaz de clivar una secuencia de ADN diana del gen de la cadena gamma del receptor de interleuquina-2 humano (IL2RG) que también se denomina gen de la cadena gamma común del receptor de citoquina o gen gamma C. Las secuencias diana identificadas en las publicaciones de patente de los EE.UU. n.º 20110225664, 20110091441, 20100229252, 20090271881 y 20090222937 se pueden utilizar para el sistema de direccionamiento hacia ácidos nucleicos descrito en la presente.

La inmunodeficiencia combinada grave (SCID, por las siglas en inglés de *Severe Combined Immune Deficiency*) es el resultado de un defecto en la maduración de los linfocitos T, siempre asociada con un defecto funcional en linfocitos B (Cavazzana-Calvo y col., Annu. Rev. Med., 2005, 56, 585-602; Fischer y col., Immunol. Rev., 2005, 203, 98-109). Se estima que la incidencia global es de 1 de cada 75 000 nacimientos. Los pacientes con SCID no tratada están sujetos a múltiples infecciones por microorganismos oportunistas y, en general, no viven más de un año. La SCID se puede tratar mediante transferencia de células madre hematopoyéticas alógenas, procedentes de un donante de la familia. La histocompatibilidad con el donante puede variar enormemente. En el caso de la deficiencia de la adenosina-desaminasa (ADA), una de las formas de SCID, los pacientes pueden ser tratados mediante inyección de la enzima adenosina-desaminasa recombinante

Como se ha mostrado que el gen ADA está mutado en pacientes de SCID (Giblett y col., Lancet, 1972, 2, 1067-1069), se han identificado otros varios genes incluidos en SCID (Cavazzana-Calvo y col., Annu. Rev. Med., 2005, 56, 585-602; Fischer y col., Immunol. Rev., 2005, 203, 98-109). Existen cuatro causas principales para la SCID: (i) la forma más frecuente de SCID, SCID-X1 (SCID ligada al cromosoma X o X-SCID), está causada por una mutación en el gen de IL2RG, y da como resultado la ausencia de linfocitos T maduros y linfocitos citolíticos naturales. IL2RG codifica la proteína C gamma (Noguchi, y col., Cell, 1993, 73, 147-157), un componente común de al menos cinco complejos receptores interleucínicos. Estos receptores activan varias blancos mediante la cinasa JAK3 (Macchi y col., Nature, 1995, 377, 65-68), cuya inactivación da como resultado el mismo síndrome que la inactivación de la C gamma; (ii) la mutación en el gen de ADA da como resultado un defecto en el metabolismo de las purinas que es letal para los precursores de linfocitos, que a su vez da como resultado la ausencia casi total de linfocitos B, T y linfocitos citolíticos naturales; (iii) la recombinación V(D)J es un paso esencial en la maduración de los receptores de linfocitos T (TCR) e inmunoglobulinas. Las mutaciones en el gen activador de la recombinación 1 y 2 (RAG1 y RAG2) y Artemis, tres genes incluidos en este proceso, dan como resultado la ausencia de linfocitos T y B maduros; y (iv) también se han informado mutaciones en otros genes tales como CD45, incluidas en la señalización específica de las células T, aunque las mismas representan una minoría de los casos (Cavazzana-Calvo y col., Annu. Rev. Med., 2005, 56, 585-602; Fischer y col., Immunol. Rev., 2005, 203, 98-109). Desde que se identificaron sus bases genéticas, las diferentes formas de SCID se han convertido en un paradigma para las estrategias de terapia génica (Fischer y col., Immunol. Rev., 2005, 203, 98-109) por dos razones principales. La primera, como en todas las enfermedades sanguíneas, se puede pensar en un tratamiento ex vivo. Se pueden extraer células madre hematopoyéticas (HSC, por sus siglas en inglés) de la médula ósea y mantener sus propiedades pluripotentes durante unas pocas divisiones celulares. Por lo tanto, se pueden tratar in vitro, y se pueden volver a inyectar al paciente, donde repueblan la médula ósea. La segunda es que ya que la maduración de los linfocitos está alterada en los pacientes con SCID, las células corregidas tienen una ventaja selectiva. Por lo tanto, un pequeño número de células corregidas puede restaurar la funcionalidad del sistema inmunitario. Esta hipótesis se validó varias veces por (i) la restauración parcial de las funciones inmunes asociadas a la reversión de mutaciones en pacientes de SCID (Hirschhorn y col., Nat. Genet., 1996, 13, 290-295; Stephan y col., N. Engl. J. Med., 1996, 335, 1563-1567; Bousso y col., Proc. Natl., Acad. Sci. USA, 2000, 97, 274-278; Wada y col.,

Proc. Natl. Acad. Sci. USA, 2001, 98, 8697-8702; Nishikomori y col., Blood, 2004, 103, 4565-4572), (ii) la corrección de deficiencias SCID-X1 *in vitro* en células hematopoyéticas (Candotti y col., Blood, 1996, 87, 3097-3102; Cavazzana-Calvo y col., Blood, 1996, Blood, 88, 3901-3909; Taylor y col., Blood, 1996, 87, 3103-3107; Hacein-Bey y col., Blood, 1998, 92, 4090-4097), (iii) la corrección de SCID-X1 (Soudais y col., Blood, 2000, 95, 3071-3077; Tsai y col., Blood, 2002, 100, 72-79), JAK-3 (Bunting y col., Nat. Med., 1998, 4, 58-64; Bunting y col., Hum. Gene Ther., 2000, 11, 2353-2364) y deficiencias de RAG2 (Yates y col., Blood, 2002, 100, 3942-3949) in vivo en modelos en animales y (iv) mediante el resultado de ensayos clínicos de terapia génica (Cavazzana-Calvo y col., Science, 2000, 288, 669-672; Aiuti y col., Nat. Med., 2002; 8, 423-425; Gaspar y col., Lancet, 2004, 364, 2181-2187).

Publicación de patente de los EE.UU. No. 20110182867 asignada al Children's Medical Center Corporation y el *President and Fellows* del Harvard College se refiere a métodos y usos para modular la expresión fetal de la hemoglobina (HbF) en células hematopoyéticas progenitoras mediante la expresión o actividad de inhibidores de BCL11A, por ejemplo, ARNi y anticuerpos. Los blancos divulgados en la Publicación de Patente de los EE.UU. n.º 20110182867, tal como BCL11A, podrán ser los blancos del sistema CRISPR Cas descrito en la presente para modular la expresión de hemoglobina fetal. Véase también Bauer y col. (Science 11 de octubre de 2013: Vol. 342 no. 6155 pp. 253-257) y Xu y col. (Science 18 noviembre de 2011: Vol. 334 no. 6058 pp. 993-996) para blancos BCL11A adicionales.

Con los conocimientos en el arte y las descripciones de la presente invención, una persona con experiencia puede corregir HSCs como un trastorno hematológico genético, por ejemplo, β-Talasemia, hemofilia, o a enfermedad genética de almacenamiento lisosómico.

HSC - Administración a células hematopoyética indiferenciadas y edición de las mismas, y condiciones particulares.

La expresión "célula hematopoyética indiferenciada" o "HSC" se utiliza con la intención de incluir ampliamente a aquellas células que se consideran HSC, por ejemplo, células sanguíneas que originan todas las otras células sanguíneas y se obtienen del mesodermo; situadas en la médula ósea roja, que está contenida en el núcleo de la mayoría de los huesos. Las HSC, como se describen en la presente, incluyen células con un fenotipo de células hematopoyéticas indiferenciadas, identificado por falta de marcadores de linaje (lin) de tamaño pequeño, y marcadores que pertenecen al agrupamiento de series de diferenciación, como: CD34, CD38, CD90, CD133, CD105, CD45, y también c-kit, - el receptor del factor de células indiferenciadas. Las células hematopoyéticas indiferenciadas son negativas para los marcadores que se utilizan para la detección de compromiso del linaje, y, por lo tanto, se denominan Lin-; y, durante su purificación por FACS, se han identificado por marcadores un número de hasta 14 diferentes marcadores de linaje de sangre madura, por ejemplo, CD13 & CD33 para mieloide, CD71 para eritroide, CD19 para células B, CD61 para megacariocítica, etc. Para humanos; y, B220 (CD45 murina) para células B, Mac-1 (CD11b/CD18) para monocitos, Gr-1 para Granulocitos, Ter119 para células eritroides, IL7Ra, CD3, CD4, CD5, CD8 para células T, etc. Marcadores de HSC en ratón: CD34lo/-, SCA-1+, Thy1.1+/lo, CD38+, C-kit+, lin-, y marcadores de HSC humanas: CD34+, CD59+, Thy1/CD90+, CD38lo/-, C-kit/CD117+, y lin-. Por lo tanto, en las formas de realización que se exponen aquí, las HSCs pueden ser células CD34+. Las HSC también pueden ser células hematopoyéticas indiferenciadas que son CD34-/CD38-. Las células indiferenciadas que pueden carecer de c-kit en la superficie de la célula que se consideran en el arte como HSCs se encuentran dentro del ámbito de los métodos descritos en la presente, así como las células CD133+ que de manera similar se consideran HSCs en el arte.

El sistema CRISPR-Cas (por ejemplo, Cpf1) se puede modificar para direccionar hacia un locus genético o loci de HSCs. Se puede preparar una proteína Cas (por ejemplo, Cpf1), optimizada de manera ventajosa por codones para una célula eucariótica y especialmente una célula de mamífero, por ejemplo, una célula humana, por ejemplo, HSC, y direccionar el ARNg hacia un locus o loci de HSC, por ejemplo, hacia el gen EMX1. Esta se puede suministrar mediante partículas. Las partículas se pueden formar mezclando la proteína Cas (por ejemplo, Cpf1) y el ARNg. La mezcla de ARNg y la proteína Cas (por ejemplo, Cpf1) se puede combinar por ejemplo con una mezcla que comprende o que consiste esencialmente en o consiste en: tensioactivo, fosfolípido, polímero biodegradable, lipoproteína y alcohol, mediante lo cual se pueden formar partículas que contienen el ARNg y la proteína Cas (por ejemplo, Cpf1). La solicitud también abarca la producción de partículas y las partículas que se obtienen de un método con dichas características, así como los usos de las mismas.

De manera más general, se pueden formar partículas usando un proceso eficiente. En primer lugar, la proteína Cas9 y el ARNsg que se dirige al gen EMX1 o al gen de control LacZ se mezclaron juntas a una proporción adecuada, por ejemplo, una relación molar de 3:1 a 1:3 o 2:1 a 1:2 o 1:1 a una temperatura adecuada, por ejemplo, de entre 15 y 30C, por ejemplo, de entre 20 y 25C, por ejemplo, a temperatura ambiente durante un tiempo adecuado, por ejemplo, de entre 15 y 45, por ejemplo 30 minutos, ventajosamente en solución reguladora de pH estéril libre de nucleasas, por ejemplo, PBS 1X. Por separado, los componentes de la partícula tales como: un tensioactivo, por ejemplo, un lípido catiónico, por ejemplo, 1,2-dioleoil-3-trimetilamonio-propano (DOTAP); un fosfolípido, por ejemplo, dimiristoilfosfatidilcolina (DMPC); un polímero biodegradable, por ejemplo un polímero de etilenglicol o PEG, y una lipoproteína, por ejemplo una lipoproteína de baja densidad, por ejemplo, colesterol; o que comprenden dichos materiales, se pueden disolver en un alcohol, de manera ventajosa un alcohol C1-6 alquílico, por ejemplo metanol, etanol, isopropanol, por ejemplo, etanol al 100%. Las dos soluciones se pueden mezclar entre sí para formar partículas que contienen los complejos Cas (por ejemplo, Cpf1)-ARNg. En ciertas formas de realización la partícula puede contener un molde de HDR. Que puede ser una partícula co-administrada con una partícula que contiene

ARNg+Proteína Cas (por ejemplo, Cpf1), o es decir, además de poner en contacto una HSC con una partícula que contiene ARNg+Proteína Cas (por ejemplo Cpf1), la HSC se pone en contacto con una partícula que contiene un molde de HDR, o la HSC se pone en contacto con una partícula que contiene a todo el ARNg, a la Cas (por ejemplo Cpf1) y a molde de HDR. El molde de HDR se puede suministrar mediante un vector separado, mediante lo cual en una primera instancia la partícula penetra una célula HSC y el vector separado también penetra la célula, donde el genoma de la HSC es modificado por el ARNg+Cas (por ejemplo, Cpf1) y el molde de HDR también está presente, mediante lo cual los loci genómicos son modificados por la HDR; por ejemplo, esto puede dar como resultado la corrección de una mutación.

Luego de las formar las partículas, las HSC se pueden transfectar en placas de 96 pocillos con 15 ug de proteína Cas (por ejemplo, Cpf1) por pocillo. Tres días después de la transfección, las HSCs se pueden cosechar, y se puede cuantificar el número de inserciones y supresiones (indels) en el locus de EMX1.

Esto ilustra cómo las HSC se pueden modificar usando el direccionamiento de CRISPR-Cas (por ejemplo, Cpf1) a un locus o loci genómico(s) de interés en la HSC. Mientras que este aspecto no se reivindica en la presente, las HSC que se deben modificar pueden estar *in vivo*, es decir, en un organismo no humano, por ejemplo, un ser humano o un eucariota no humano, por ejemplo, un animal, por ejemplo, un pez, por ejemplo, un pez cebra, un mamífero, por ejemplo, un primate, por ejemplo, un simio, un chimpancé, un macaco, un roedor, por ejemplo, un ratón, conejo, rata, un canino o un perro, ganado (vaca / bovino, oveja / ovino, cabra o cerdo), aves o aves de corral, por ejemplo, pollos. Las HSCs que se deben modificar pueden estar *in vitro*, es decir, afuera de un organismo con dichas características. Y, las HSCs modificadas se pueden utilizar *ex vivo*, es decir, se puede(n) obtener o aislar del organismo una o más HSCs de un organismo con dichas características, opcionalmente la(s) HSC(s) se puede(n) expandir, la(s) HSC(s) se modifican mediante una composición que comprende un sistema CRISPR-Cas (por ejemplo Cpf1) direccionado hacia locus o loci genético(s) en la HSC, por ejemplo, poniendo en contacto a la(s) HSC(s) con la composición, por ejemplo, donde la composición comprende una partícula que contiene a la enzima CRISPR y uno o más ARNg direccionado(s) hacia el locus genético o los loci en la HSC, por ejemplo una partícula que se obtiene o que se puede obtener por mezcla de una mezcla de ARNg y proteína Cas (por ejemplo Cpf1) con una mezcla que comprende o que consiste esencialmente en o consiste en: tensioactivo, fosfolípido, polímero biodegradable, lipoproteína y alcohol (donde uno o más ARNg objetivos el locus genético o los loci en la HSC), opcionalmente expandir las HSCs modificadas que se obtienen como resultado y suministrar al organismo las HSCs modificadas que se obtienen como resultado. En algunas instancias las HSCs que se aislaron u obtuvieron pueden provenir de un primer organismo, por ejemplo, un organismo de la misma especie que el segundo organismo, y el segundo organismo puede ser el organismo al cual se le suministran las HSCs modificadas que se obtienen como resultado, por ejemplo, el primer organismo puede ser un donante (tal como un pariente, tal como un progenitor o hermano) para el segundo organismo. Las HSCs modificadas pueden tener modificaciones genéticas para resolver o aliviar o reducir síntomas de un estado de enfermedad o condición de un individuo o sujeto o paciente. Las HSCs modificadas, por ejemplo, en la instancia de un primer organismo donante a un segundo organismo, puede tener modificaciones genéticas para hacer que las HSCs tengan una o más proteínas por ejemplo marcadores superficiales o proteínas más parecidas a las del segundo organismo. Las HSCs modificadas pueden tener modificaciones genéticas para simular un estado de enfermedad o condición de un individuo o sujeto o paciente y se pueden volver a suministrar a un organismo no humano de manera tal de preparar un modelo animal. La expansión de HSCs se encuentra dentro del ámbito de una persona con experiencia al ver la presente invención y los conocimientos del arte, véase, por ejemplo, Lee, "*Improved expansion ex vivo of adult hematopoietic stem cells by overcoming CUL4-mediated degradation of HOXB4.*" Blood. 16 de mayo de 2013; 121(20):4082-9. doi: 10.1182/blood-2012-09-455204. Epub 21 de marzo de 2013.

Como se indicó para mejorar la actividad, el ARNg se puede complejar de antemano con la proteína Cas (por ejemplo, Cpf1), antes de formular el complejo completo en una partícula. Las formulaciones pueden producirse con una relación molar de diferentes componentes conocidos por promover el suministro de ácidos nucleicos en células (por ejemplo 1,2-dioleoil-3-trimetilamonio-propano (DOTAP), 1,2-ditetradecanoil-sn-glicero-3-fosfocolina (DMPC), polietilenglicol (PEG), y colesterol). Por ejemplo, las relaciones molares de DOTAP: DMPC: PEG: Colesterol pueden ser DOTAP 100, DMPC 0, PEG 0, colesterol 0; o DOTAP 90, DMPC 0, PEG 10, colesterol 0; o DOTAP 90, DMPC 0, PEG 5, colesterol 5. DOTAP 100, DMPC 0, PEG 0, colesterol 0. Por lo tanto, la invención abarca la mezcla de ARNg, proteína Cas (por ejemplo, Cpf1) y componentes que forman una partícula; así como partículas que se obtienen al realizar dicha mezcla.

En una forma de realización preferida, se pueden formar partículas que contienen los complejos Cas (por ejemplo, Cpf1)-ARNg mezclando entre sí proteína Cas (por ejemplo, Cpf1) y uno o más ARNg, preferiblemente en una proporción molar 1:1 de enzima:ARN guía. Por separado, se disuelven los diferentes componentes que se sabe que promueven el suministro de ácidos nucleicos (por ejemplo, DOTAP, DMPC, PEG y colesterol), preferentemente en etanol. Las dos soluciones se mezclan entre sí para formar partículas que contienen los complejos Cas (por ejemplo, Cpf1)-ARNg. Luego de formar las partículas, los complejos Cas (por ejemplo, Cpf1)-ARNg se pueden transfectar en células (por ejemplo, HSCs). Puede aplicarse una codificación de códigos de barras. Pueden marcarse con un código de barras las partículas, la Cas-9 y/o el ARNsg.

Los métodos y composiciones reivindicados de la invención pueden comprender preparar una partícula que contiene ARNg-y-Proteína Cas (por ejemplo, Cpf1) que comprende mezclar una mezcla de ARNg y proteína Cas (por ejemplo, Cpf1) con una mezcla que comprende o que consiste esencialmente en o consiste en: tensioactivo, fosfolípido, polímero biodegradable, lipoproteína y alcohol. Una forma de realización abarca una partícula que contiene ARNg-y-

Proteína Cas (por ejemplo, Cpf1) que se obtiene por dicho método. La invención, en una forma de realización, comprende el uso de la partícula en un método reivindicado en la presente, que comprende poner en contacto una célula que contiene el locus genómico de interés con la partícula, donde el ARNsg se dirige al locus genómico de interés; o un método como se reivindica en la presente puede implicar modificar un locus genómico de interés, mediante manipulación de una secuencia diana en un locus genómico de interés, que comprende poner en contacto el locus genómico de interés con la partícula, donde el ARNsg se dirige al locus genómico de interés. En dichas formas de realización, el locus genómico de interés es de manera ventajosa un locus genómico en una HSC.

Consideraciones para aplicaciones terapéuticas descritas, pero no reivindicadas en la presente: Una consideración en la terapia de edición del genoma es la elección de una nucleasa específica para una secuencia, por ejemplo una variante de una nucleasa Cpf1. Cada variante de nucleasa puede poseer su propio conjunto único de puntos fuertes y débiles, muchos de los cuales se deben equilibrar en el contexto del tratamiento para maximizar el beneficio terapéutico. Hasta ahora, dos enfoques de edición terapéutica con nucleasas han mostrado ser significativamente prometedores: interrupción de genes y corrección de genes. La interrupción de genes implica la estimulación de la NHEJ para crear indel dirigidos en elementos genéticos, a menudo dando como resultado mutaciones de pérdida de función que son beneficiosas para pacientes (figura 13A). Al contrario, en la corrección de genes se utiliza HDR para revertir directamente una enfermedad causando una mutación, restableciendo la función a la vez que se protege la regulación fisiológica del elemento corregido. También se puede utilizar la HDR para insertar un transgen terapéutico en un determinado locus 'de alojamiento seguro' en el genoma para recuperar la pérdida de función del gen. Para que una terapia de edición específica sea eficaz, tiene que lograrse un nivel suficientemente alto de modificaciones en poblaciones de células diana para revertir los síntomas de la enfermedad. Este "umbral" de modificación terapéutica está determinado por la adecuación de las células editadas después del tratamiento y la cantidad de producto génico necesario para revertir los síntomas. Con respecto a la adecuación, la edición crea tres potenciales resultados en las células tratadas en relación con sus contrapartes sin editar: mayor, neutra, o menor adecuación. En el caso de la mayor adecuación, por ejemplo, en el tratamiento de SCID-X1, las células progenitoras hematopoyéticas modificadas se expanden de manera selectiva en relación a sus homólogos no editados. La SCID-X1 es una enfermedad causada por mutaciones en el gen IL2RG, cuya función es necesaria para corregir el desarrollo del linaje de los linfocitos hematopoyéticos [Leonard, W.J., y col. Immunological Reviews 138, 61-86 (1994); Kaushansky, K. & Williams, W.J. Williams Hematology, (McGraw-Hill Medical, New York, 2010)]. En pruebas clínicas con pacientes que recibieron terapia génica viral para SCID-X1, y un poco frecuente ejemplo de una corrección espontánea de mutación de SCID-X1, las células hematopoyéticas progenitoras corregidas pueden ser capaces de solucionar este bloque de desarrollo y expandirlo en relación con sus contrapartes enfermas para mediar la terapia [Bousso, P., y col. Proceedings of the National Academy of Sciences of the United States of America 97, 274-278 (2000); Hacein-Bey-Abina, S., y col. The New England Journal of Medicine 346, 1185-1193 (2002); Gaspar, H.B., y col. Lancet 364, 2181-2187 (2004)]. En este caso, en donde las células editadas poseen una ventaja selectiva, pueden amplificarse números incluso bajos de células editadas mediante expansión, proporcionando un beneficio terapéutico al paciente. Por el contrario, la edición para otras enfermedades hematopoyéticas, tales como trastornos granulomatosos crónicos (CGD), podrían no inducir un cambio en la adecuación para las células progenitoras hematopoyéticas, aumentando el umbral de modificación terapéutica. Los CGD son causados por mutaciones en genes que codifican proteínas oxidasa fagocíticas, que normalmente son utilizadas por los neutrófilos para generar especies de oxígeno reactivo que matan patógenos [Mukherjee, S. & Thrasher, A.J. Gen 525, 174-181 (2013)]. Ya que la disfunción de estos genes no influencia a la adecuación o al desarrollo de las células progenitoras hematopoyéticas, sino solo a la capacidad del tipo celular hematopoyético maduro para combatir las infecciones, es probable que no hubiese una expansión preferencial de las células editadas en esta enfermedad. Por cierto, en las pruebas de terapia génica no se observaron ventajas selectivas de las células con el gen corregido en los CGD, lo que lleva a obtener dificultades con el proceso de injerto a largo plazo de la célula [Malech, H.L., y col. Proceedings of the National Academy of Sciences of the United States of America 94, 12133-12138 (1997); Kang, H.J., y col., Molecular therapy: The journal of the American Society of Gene Therapy 19, 2092-2101 (2011)]. Como tales, se requerirían niveles significativamente mayores de edición para tratar enfermedades como la CGD, en donde la edición crea una ventaja de adecuación neutra, en relación a enfermedades donde la edición crea una adecuación aumentada para las células diana. Si la edición impone una desventaja de adecuación, tal como sería el caso para restaurar la función a un gen supresor tumoral en células cancerosas, las células modificadas podrían verse superadas por la competición con sus homólogos enfermos, haciendo que el beneficio del tratamiento sea bajo en relación a las tasas de edición. Esta última clase de enfermedades podría ser particularmente difícil de tratar con terapia de edición genómica.

Además de la adecuación celular, la cantidad de producto génico necesario para tratar la enfermedad también influencia al nivel mínimo de edición genómica terapéutica que tiene que lograrse para revertir los síntomas. La hemofilia B es una enfermedad en donde un pequeño cambio en los niveles de producto génico puede dar como resultado cambios significativos en los resultados clínicos. Esta enfermedad está causada por mutaciones en el gen que codifica el factor IX, una proteína secretada normalmente en el hígado a la sangre, donde funciona como componente de la cascada de coagulación. La gravedad clínica de la hemofilia B está relacionada con la cantidad de actividad del factor IX. Mientras que la enfermedad grave está asociada a menos del 1% de la actividad normal, las formas más suaves de las enfermedades están asociadas con más del 1% de actividad del factor IX [Kaushansky, K. & Williams, W.J. Williams, Hematology, (McGraw-Hill Medical, New York, 2010); Lofqvist, T., y col. Journal of Internal Medicine 241, 395-400 (1997)]. Esto sugiere que las terapias de edición que puedan restaurar la expresión del factor IX incluso en un pequeño porcentaje de células hepáticas podría tener un gran impacto en los resultados clínicos. Un

estudio usando ZFNs para corregir un modelo en ratón de hemofilia B poco después del nacimiento demostró que un 3-7% de corrección fue suficiente para revertir los síntomas de la enfermedad, proporcionando evidencia preclínica para dicha hipótesis [Li, H., y col. Nature 475, 217-221 (2011)].

Los trastornos en donde un pequeño cambio en los niveles de producto génico puede influenciar los resultados clínicos y las enfermedades en donde hay una ventaja de adecuación para las células editadas son blancos ideales para la terapia de edición genómica, ya que el umbral de modificación terapéutica es lo suficientemente bajo como para permitir grandes probabilidades de éxito dada la tecnología actual. Al apuntar hacia dichas enfermedades ahora ha dado como resultado éxitos en la terapia de edición en el nivel preclínico y en las pruebas clínicas de fase I. Son necesarias mejoras en la manipulación de la ruta de reparación de los DSB y la administración de nucleasas para extender estos resultados prometedores a enfermedades con una ventaja de adecuación neutra para las células editadas o donde se necesitan cantidades más grandes de producto génico para el tratamiento. La siguiente tabla muestra algunos ejemplos de aplicaciones de edición del genoma a modelos terapéuticos.

Tipo de enfermedad	**Plataforma de nucleasa empleada**	**Estrategia terapéutica**	**Referencias**
Hemofilia		B ZFN Inserción mediada por HDR de la secuencia genética correcta 217-221 (2011)	Li, H., y col. Nature 475,
SCID		ZFN Inserción mediada por HDR Genovese,de secuencia génica correcta Nature	P., y col. 510, 235-240 (2014)
Tirosinemia hereditaria	CRISPR mediada por HDR	Corrección de mutaciones en Biotechnology 32, 551-el hígado	Yin, H., y col. Nature 553 (2014)

La solución de cada una de las condiciones de la anterior tabla, usando el sistema CRISPR-Cas (por ejemplo Cpf1) para dirigirla ya sea hacia la corrección de la mutación mediada por HDR, o mediado por HDR por inserción de una secuencia correcta del gen, de manera ventajosa mediante un sistema de administración a la que se presenta aquí, por ejemplo, un sistema de administración basado en partículas, según se encuentra dentro del ámbito de una persona con experiencia al ver la presente invención y en base a los conocimientos del arte. Por lo tanto, una forma de realización abarca poner en contacto HSC que porta una mutación de hemofilia B, SCID (por ejemplo, SCID-X1, ADA-SCID) o Tirosinemia hereditaria con una partícula que contiene un sistema de ARNg-y-Proteína Cas (por ejemplo, Cpf1) direccionado hacia un locus genómico de interés como hemofilia B, SCID (por ejemplo, SCID-X1, ADA-SCID) o tirosinemia hereditaria (por ejemplo, como en Li, Genovese o Yin). La partícula también puede contener un molde de HDR apropiado para corregir la mutación; o la HSC se puede poner en contacto con una segunda partícula o un vector que contiene o suministra el molde de HDR. Al respecto, se menciona que la Hemofilia B es un trastorno ligado al cromosoma X recesivo causado por mutaciones de pérdida de la función en el gen que codifica al factor IX, un componente crucial de la cascada de coagulación. Una actividad de recuperación del factor IX mayor al 1% de sus niveles en individuos gravemente afectados puede transformar la enfermedad en una forma significativamente más leve, ya que la infusión de Factor IX recombinante en dichos pacientes profilácticamente desde una edad temprana para obtener dichos niveles mejora mucho las complicaciones clínicas. Con el conocimiento en la técnica y las enseñanzas en la presente divulgación, el experto en la materia puede corregir HSC para hemofilia B usando un sistema CRISPR-Cas9 que se dirige a y corrige la mutación (trastorno recesivo ligado a X causado por mutaciones de pérdida de función en el gen que codifica al factor IX) (por ejemplo, con un molde de HDR adecuado que administra una secuencia codificante para el factor IX); específicamente, el ARNsg puede dirigirse a la mutación que da lugar a la hemofilia B y el la HDR puede proporcionar la codificación para la expresión adecuada del factor IX. Una partícula que contiene al sistema de ARNg-y-Proteína Cas (por ejemplo, Cpf1) direccionado hacia la mutación se pone en contacto con HSCs que llevan la mutación. La partícula también puede contener un molde de HDR apropiado para corregir la mutación para la correcta expresión del Factor IX; o la HSC se puede poner en contacto con una segunda partícula o un vector que contiene o suministra el molde de HDR. Las células que se ponen en contacto de esa manera se pueden administrar; y opcionalmente tratar / expandir; cf. Cartier, según se expone aquí.

En Cartier, "MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-Linked Adrenoleukodystrophy," Brain Pathology 20 (2010) 857-862, como si se expusiera por completo, se reconoce que se utilizó el trasplante de células madre hematopoyéticas (HSCT) alogénicas para administrar enzima lisosómica normal al cerebro de un paciente con enfermedad de Hurler y una discusión acerca de la terapia génica de HSC para tratar la ALD. Se recogieron en dos pacientes células CD34+ periféricas después de su movilización con factor estimulante de colonias de granulocitos (G-CSF) y se transdujeron con el vector lentiviral de (MND)-ALD con un potenciador del virus del sarcoma mieloproliferativo, con la región de control negativo eliminada y sustitución en el sitio de unión a cebador dl587rev. Se transdujeron las células CD34+ de los pacientes con el vector MND-ALD durante 16 h en presencia de citocinas a bajas concentraciones. Las células CD34+ se congelaron tras la transducción para llevar a cabo pruebas de seguridad en el 5% de las células que incluían en particular tres ensayos de lentivirus competente para replicación (RCL). La eficacia de la transducción de las células

CD34+ osciló entre el 35% y el 50% con un número medio de copias lentivirales integradas de entre 0,65 y 0,70. Luego de descongelar las células CD34+ transducidas, los pacientes se reinfundieron con más de 4,106 células CD34+ transducidas/kg luego de una mieloablación completa con busulfan y ciclofosfamida. Se suprimieron las HSC de los pacientes para favorecer el injerto de las HSC corregidas génicamente. La recuperación hematológica de los pacientes tuvo lugar entre los días 13 y 15 para los dos pacientes. La recuperación inmunológica prácticamente completa tuvo lugar a los 12 meses para el primer paciente y a los 9 meses para el segundo paciente. Al contrario del uso de lentivirus, con los conocimientos en el arte y las descripciones en la presente invención, una persona con experiencia puede corregir HSCs como una ALD usando un sistema CRISPR-Cas (Cpf1) direccionado hacia la mutación y que la corrige (por ejemplo, con un molde de HDR apropiado); específicamente, el ARNg puede dirigirse hacia mutaciones en ABCD1, un gen situado en el cromosoma X que codifica a ALD, una proteína transportadora de membrana peroxisómica, y la HDR puede proveer la codificación para corregir la expresión de la proteína. Una partícula que contiene ARNg direccionado hacia la mutación-y-Proteína Cas (Cpf1) se pone en contacto con HSCs, por ejemplo, con células CD34+ que llevan la mutación como en Cartier. La partícula también puede contener un molde de HDR apropiado para corregir la mutación para la expresión de la proteína transportadora de membrana peroxisómica; o la HSC se puede poner en contacto con una segunda partícula o un vector que contiene o suministra el molde de HDR. Las células que se ponen en contacto de esa manera opcionalmente se pueden tratar como en Cartier. Las células que se ponen en contacto de esa manera se pueden administrar como en Cartier.

Se menciona WO 2015/148860, a través de las enseñanzas en la presente, la presente divulgación comprende métodos y materiales de dichos documentos aplicados en conjunto con las descripciones de la presente. En un aspecto de terapia génica para enfermedades relacionadas con la sangre, lo métodos y composiciones para tratar beta talasemia se pueden adaptar al sistema CRISPR-Cas descrito en la presente (véase, por ejemplo, WO 2015/148860). En una forma de realización, WO 2015/148860 incluye el tratamiento o prevención de la beta talasemia, o sus síntomas, por ejemplo, alterando el gen para CLL de células B/linfoma 11A (BCL11A). El gen BCL11A también se conoce como CLL de células B/linfoma 11A, BCL11A -L, BCL11A -S, BCL11AXL, CTIP 1, HBFQTL5 y ZNF. BCL11A codifica una proteína de dedo de cinc que está incluida en la regulación de la expresión del gen de globina. Mediante la alteración del gen BCL11A (por ejemplo, uno o ambos alelos del gen BCL11A), se pueden incrementar los niveles de gamma globina. La gamma globina puede reemplazar a la beta globina en el complejo hemoglobina y transporta eficazmente el oxígeno a los tejidos, mejorando de esa manera los fenotipos de la enfermedad denominada beta talasemia.

También se menciona a WO 2015/148863 y a través de las enseñazas en la presente, los métodos y composiciones reivindicados de la invención comprende métodos y materiales de dichos documentos que se pueden adaptar al sistema CRISPR-Cas usado en los métodos y reivindicaciones de la presente invención. En un aspecto del tratamiento y la prevención de la enfermedad de la anemia falciforme, que es una enfermedad hematológica hereditaria, WO 2015/148863 abarca la alteración del gen BCL11A. Mediante la alteración del gen BCL11A (por ejemplo, uno o ambos alelos del gen BCL11A), se pueden incrementar los niveles de gamma globina. La gamma globina puede reemplazar a la beta globina en el complejo hemoglobina y transporta eficazmente el oxígeno a los tejidos, mejorando de esa manera los fenotipos de enfermedad de la anemia falciforme.

En un aspecto de la invención, se abarca a los métodos y composiciones que incluyen la edición para el direccionamiento hacia una secuencia de ácido nucleico, o modular la expresión de una secuencia de ácido nucleico a la que se direcciona, y aplicaciones de esto en relación con la inmunoterapia para tratar el cáncer por adaptación del sistema CRISPR-Cas de la presente invención. Se hace referencia a la aplicación de la terapia génica en WO 2015/161276 que incluye métodos y composiciones que se pueden utilizar para afectar la proliferación, supervivencia y/o función de las células T alterando uno o más genes expresados en las células T, por ejemplo, uno o más de los genes FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC y/o TRBC. En un aspecto relacionado, la proliferación de células T se puede afectar por alteración de uno o más genes expresados en células T, por ejemplo, el(los) gen(es) CBLB y/o PTPN6, el(los) gen(es) FAS y/o BID, el(los) gen(es) CTLA4 y/o PDCDI y/o TRAC y/o TRBC.

Las células T con receptores de antígeno quimérico (CAR)19 muestran efectos antileucémicos sobre las malignidades en el paciente. Sin embargo, frecuentemente los pacientes de leucemia no tienen suficientes células T que recolectar, lo que significa que el tratamiento debe incluir células T de donantes modificadas. Por lo tanto, existe interés en establecer un banco de células T de donantes. Qasim y col. ("First Clinical Application of Talen Engineered Universal CAR19 T Cells in B-ALL" ASH 57th Annual Meeting and Exposition, 5-8 de diciembre de 2015, Resumen 2046 (https://ash.confex.com/ash/2015/webprogram/Paper81653.html publicado en línea, noviembre de 2015) exponen la modificación de células T CAR19 para eliminar el riesgo de enfermedad de injerto-contra-huésped por interrupción de expresión del receptor en células T y direccionamiento hacia CD52. Además, las células CD52 se direccionaron de manera tal que las mismas se volvieron insensibles al Alemtuzumab, y de esa manera permitieron que el Alemtuzumab impida el rechazo mediado por el huésped de células T CAR19 sin coincidencia para el antígeno leucocitario humano (HLA). Los investigadores utilizaron un vector lentiviral auto-inactivante de tercera generación que codifican a 4g7 CAR19 (CD19 scFv-4-1BB-CD3ζ) conectado a RQR8, luego trataron por electroporación células con dos pares de ARNm de TALEN para el direccionamiento múltiple tanto para el locus de la cadena constante alfa del receptor de células T (TCR) y el locus del gen CD52. Las células aún expresaban TCR luego de la expansión *ex vivo* se eliminaron usando eliminación de TCR α/β con CliniMacs, para obtener un producto de células T (UCART19) con <1% de expresión de TCR, 85% de la cual expresaba CAR19, y 64% se volvió CD52 negativa. Las células T CAR19 modificadas se administraron para tratar una recidiva de leucemia linfoblástica aguda del paciente. Las descripciones

que se dan aquí proveen métodos eficaces para proporcionar células hematopoyéticas indiferenciadas modificadas y la progenie de las mismas, incluyendo, pero de manera no taxativa, a células de los linajes de sangre mieloide y linfoide, incluyendo a las células T, células B, monocitos, macrófagos, neutrófilos, basófilos, eosinófilos, eritrocitos, células dendríticas, y megacariocitos o plaquetas, y células asesinas naturales y sus precursores y progenitores. Dichas células se pueden modificar por noqueo, knock in, o de otra manera modulando blancos, por ejemplo, para eliminar o modular CD52 como se describió antes, y otros blancos, tales como, sin limitación, CXCR4, y PD-1. Así, las composiciones, células, y métodos de la invención se pueden utilizar para modular respuestas inmunitarias y para tratar, sin limitación, malignidades, infecciones virales, y trastornos inmunitarios, en conjunto con la modificación de la administración de células T u otras células a los pacientes.

Se menciona WO 2015/148670 y mediante las descripciones de la presente la invención abarca la aplicación de métodos y materiales de este documento en conjunto con las descripciones de la presente. En un aspecto de terapia génica, se abarcan métodos y composiciones para la edición de una secuencia diana relacionada con o vinculada con el virus de inmunodeficiencia humana (VIH) y el síndrome de inmunodeficiencia adquirida (SIDA). En un aspecto relacionado, la invención que se describe aquí abarca la prevención y el tratamiento de la infección por VIH y SIDA, por introducción de una o más mutaciones en el gen del receptor de quimioquina C-C tipo 5 (CCR5). El gen CCR5 también se conoce como CKR5, CCR-5, CD195, CKR-5, CCCKR5, CMKBR5, IDDM22 y CC-CKR-5. En un aspecto adicional, la invención que se describe aquí abarca proveer la prevención o reducción de infección por VIH y/o prevención o reducción de la capacidad del VIH de entrar en las células huésped, por ejemplo, en sujetos que ya están infectados. Algunas células huésped indicativas para el VIH incluyen, pero de manera no taxativa, células CD4, células T, tejido linfoide asociado al intestino (GALT), macrófagos, células dendríticas, células precursoras mieloides, y microglía. La entrada viral en las células huésped requiere la interacción de las glicoproteínas virales gp41 y gp120 tanto con el receptor CD4 como con un co-receptor, por ejemplo, CCR5. Si no hay presente un co-receptor, por ejemplo, CCR5, sobre la superficie de las células huésped, el virus no se puede unir y entra en las células huésped. De esa manera, se impide el progreso de la enfermedad. Al noquear o realizar un *knock down* de CCR5 en las células huésped, por ejemplo, por introducción de una mutación protectora (tal como una mutación delta 32 en CCR5), se impide la entrada del virus VIH en las células huésped.

La enfermedad granulomatosa crónica ligada a X (CGD) es un trastorno hereditario de las defensas del huésped debido a una actividad ausente o reducida de la NADPH oxidasa de fagocitos. Usando un sistema CRISPR-Cas (Cpf1) direccionado hacia la mutación y que la corrige (actividad ausente o reducida de NADPH oxidasa de fagocitos) (por ejemplo, con un molde de HDR apropiado que suministre una secuencia codificante para NADPH oxidasa de fagocitos); específicamente, el ARNg puede direccionarse hacia una mutación que da origen a un CGD (deficiencia en NADPH oxidasa de fagocitos), y la HDR puede proveer la codificación necesaria para corregir la expresión de la NADPH oxidasa de fagocitos. Una partícula que contiene ARNg direccionado hacia la mutación-y-Proteína Cas (Cpf1) se pone en contacto con HSCs que llevan la mutación. La partícula también puede contener un molde de HDR apropiado para corregir la mutación para la correcta expresión de NADPH oxidasa de fagocitos; o la HSC se puede poner en contacto con una segunda partícula o un vector que contiene o suministra el molde de HDR. Las células que se ponen en contacto de esa manera se pueden administrar; y opcionalmente tratar / expandir; cf. Cartier.

Anemia de Fanconi: Las mutaciones en por lo menos 15 genes (FANCA, FANCB, FANCC, FANCD1/BRCA2, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ/BACH1/BRIP1, FANCL/PHF9/POG, FANCM, FANCN/PALB2, FANC0/Rad51C y FANCP/SLX4/BTBD12) pueden causar la Anemia de Fanconi. Las proteínas producidas por estos genes están implicadas en un proceso celular conocido como la ruta de FA. La ruta de FA se enciende (activa) cuando se bloquea el proceso de producción de nuevas copias de ADN, denominado replicación de ADN, debido al daño en el ADN. La ruta de FA envía determinadas proteínas al área de daño, lo que desencadena la reparación de ADN de tal forma que puede continuar la replicación del ADN. La ruta de FA es particularmente responsable de un tipo determinado de daño en el ADN conocido como sobrecruzamientos intercadena (ICL). Los ICL suceden cuando dos bloques de construcción de ADN (nucleótidos) en hebras de ADN enfrentadas se unen o enlazan entre sí de manera anómala, lo que detiene el proceso de replicación del ADN. Los ICL pueden estar causados por una acumulación de sustancias tóxicas producidas en el organismo o por el tratamiento con determinadas terapias para el cáncer. Ocho proteínas asociadas con la anemia de Fanconi se agrupan para formar un complejo conocido como complejo de núcleo de FA. El complejo de núcleo de FA activa a dos proteínas, denominadas FANCD2 y FANCI. La activación de estas dos proteínas lleva proteínas de reparación de ADN al área del ICL, de tal forma que puede retirarse el sobrecruzamiento y puede continuar la replicación. El complejo de núcleo de FA. Más en particular, el complejo de núcleo de FA es un complejo multiproteína nuclear que consiste en FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, y FANCM, funciona como una ubiquitina ligasa E3 y media la activación del complejo ID, que es un heterodímero compuesto de FANCD2 y FANCI. Una vez se ha monoubiquitinado, interactúa con supresores tumorales clásicos aguas abajo de la ruta de FA, incluyendo FANCD1BRCA2, FANCN/PALB2, FANCJBRIP1, y FANCO/Rad51C y de este modo contribuye a la reparación de ADN mediante recombinación de homólogos (HR). Entre el ochenta y el 90 por ciento de los casos FA se deben a mutaciones en uno de tres genes, FANCA, FANCC, y FANCG. Estos genes proporcionan instrucciones para producir componentes del complejo de núcleo de FA. Las mutaciones en dichos genes asociados con el complejo de núcleo de FA harán que el complejo no sea funcional e interrumpirá completamente la ruta de FA. Como resultado, el daño en el ADN no se repara de manera eficaz y se acumulan los ICL con el paso del tiempo. Geiselhart, "Artículo de revisión, Disrupted Signaling through the Fanconi Anemia Pathway Leads to Dysfunctional Hematopoietic Stem Cell Biology: Underlying Mechanisms and Potential Therapeutic Strategies",

Anemia Volume 2012 (2012), ID del artículo 265790, http://dx.doi.org/10.1155/2012/265790 discutió la FA y un experimento en animales que implicaba la inyección intrafemoral de un lentivirus que codificaba el gen FANCC, dando como resultado la corrección del HSC in vivo. Usando un sistema CRISPR-Cas (Cpf1) direccionado hacia una o más de las mutaciones asociadas con la FA, por ejemplo un sistema CRISPR-Cas (Cpf1) con uno o varios ARNg y molde(s) de HDR direccionados respectivamente hacia una o más de las mutaciones de FANCA, FANCC, o FANCG que originan una FA y proveen una expresión correctora de uno o más de FANCA, FANCC o FANCG; por ejemplo, el ARNg puede direccionarse hacia una mutación como FANCC, y la HDR puede proveer la codificación para corregir la expresión de FANCC. Una partícula que contiene ARNg direccionado hacia la(s) mutación(es) (por ejemplo, una o más incluidas en FA, por ejemplo mutación(es) tal como en uno o más de FANCA, FANCC o FANCG)-y-Proteína Cas (Cpf1) se pone en contacto con HSCs que llevan la(s) mutación(es). La partícula también puede contener un molde de HDR apropiado para corregir la mutación para la correcta expresión de una o más de las proteínas incluidas en FA, por ejemplo uno o más cualesquiera de FANCA, FANCC o FANCG; o la HSC se puede poner en contacto con una segunda partícula o un vector que contiene o suministra el molde de HDR. Las células que se ponen en contacto de esa manera se pueden administrar; y opcionalmente tratar / expandir; cf. Cartier.

La partícula en la presente exposición (por ejemplo, que contiene ARNg(s) y Cas (Cpf1), opcionalmente molde(s) de HDR, o molde(s) de HDR, por ejemplo de hemofilia B, SCID, SCID-X1, ADA-SCID, de tirosinemia hereditaria, β-talasemia, CGD conectada al cromosoma X, síndrome de Wiskott-Aldrich, anemia de Fanconi, adrenoleucodistrofia (ALD), leucodistrofia metacromática (MLD), VIH/SIDA, trastorno de inmunodeficiencia, condición hematológica, o enfermedad genética de almacenamiento lisosómico) se obtiene de manera ventajosa o se puede obtener por mezcla de un ARNg(s) y proteína Cas (Cpf1) (que opcionalmente contiene molde(s) de HDR o donde dicha mezcla solo contiene molde(s) de HDR cuando se desean partículas separadas como molde(s)) con una mezcla que comprende o que consiste esencialmente en o consiste en: tensioactivo, fosfolípido, polímero biodegradable, lipoproteína y alcohol (donde uno o más ARNg se han direccionado hacia el locus genético o los loci en la HSC).

Por cierto, la invención es especialmente apropiada para tratar trastornos genéticos hematopoyéticos por edición del genoma, y trastorno de inmunodeficiencias, por ejemplo, trastorno de inmunodeficiencias genéticos, especialmente mediante el uso de la tecnología de la partícula que se ha expuesto aquí. Las inmunodeficiencias de origen genético son enfermedades donde las intervenciones por edición del genoma de la presente invención pueden ser exitosas. Las razones incluyen: las células hematopoyéticas, de las cuales las células inmunitarias son un subconjunto, son terapéuticamente accesibles. Las mismas se pueden extraer del cuerpo y trasplantar de manera autóloga o alogénicamente. Además, ciertas inmunodeficiencias de origen genético, por ejemplo, la inmunodeficiencia combinada grave (SCID), crean una desventaja proliferativa para las células inmunitarias. La corrección de lesiones genéticas que causan SCID por las poco frecuentes mutaciones 'inversas' espontánea indican que la corrección de aún un linfocito progenitor puede ser suficiente para recuperar la función inmune en los pacientes. Véase Bousso, P., y col. "Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor", Proceedings of the National Academy of Sciences of the United States of America 97, 274-278 (2000). La ventaja selectiva de las células editadas permite que aún bajos niveles de edición permitan obtener un efecto terapéutico. Este efecto de la presente invención se puede ver en la SCID, en el síndrome de Wiskott-Aldrich, y las otras condiciones que se han mencionado aquí, incluyendo a otros trastornos hematopoyéticos genéticos tales como alfa- y beta-talasemia, donde las deficiencias de hemoglobina afectan de manera negativa la adecuación de los progenitores eritroides.

La actividad de reparación de DSB de NHEJ y HDR varía significativamente según el tipo de célula y el estado celular. La NHEJ no está altamente regulada por el ciclo celular y es eficaz entre tipos celulares, permitiendo altos niveles de disrupción génica en poblaciones de células diana accesibles. Al contrario, HDR actúa en principio durante la fase S/G2, y por lo tanto está restringido a células que se están dividiendo activamente, limitando los tratamientos que requieren precisas modificaciones del genoma a las células mitóticas [Ciccia, A. & Elledge, S.J. Molecular Cell 40, 179-204 (2010); Chapman, J.R., y col. Molecular Cell 47, 497-510 (2012)].

La eficiencia de la corrección por HDR se puede controlar por el estado epigenético o la secuencia del locus diana, o se puede utilizar la configuración específica del molde de reparación (de hebra simple contra doble hebra, homología de brazos largos contra cortos) [Hacein-Bey-Abina, S., y col. The New England journal of Medicine 346, 1185-1193 (2002); Gaspar, H.B., y col. Lancet 364, 2181-2187 (2004); Beumer, K.J., y col. G3 (2013)]. La actividad relativa de las maquinarias NHEJ y HDR en las células diana también puede afectar la eficiencia de la corrección génica, ya que dichas vías pueden competir para resolver las DSBs [Beumer, K.J., y col. Proceedings of the National Academy of Sciences of the United States of America 105, 19821-19826 (2008)]. La HDR también impone una dificultad de administración no observada con las estrategias de NHEJ, ya que requiere de la administración concurrente de nucleasas y moldes de reparación. En la práctica, estas restricciones han conducido hasta ahora a bajos niveles de HDR en tipos celulares terapéuticamente relevantes. Por lo tanto, la traducción clínica se ha enfocado principalmente en estrategias de NHEJ para tratar la enfermedad, aunque ahora se han tratamientos preclínicos por HDR de prueba del concepto descrito para modelos en ratón de hemofilia B y tirosinemia hereditaria [Li, H., y col. Nature 475, 217-221 (2011); Yin, H., y col. Nature biotechnology 32, 551-553 (2014)].

Cualquier aplicación de edición genómica dada puede comprender combinaciones de proteínas, moléculas de ARN pequeñas y/o moldes de reparación, haciendo que la administración de estas múltiples partes sea sustancialmente más complicada que los agentes terapéuticos de molécula pequeña. Se han desarrollado dos estrategias principales

de herramientas de edición genómica: *ex vivo* e *in vivo*. Los métodos de tratamiento no se reivindican en la presente y referencia a los mismos en la presente se proporcionan solo para información. En los tratamientos *ex vivo*, las células enfermas se extraen del cuerpo, se editan y luego se vuelven a trasplantar en el paciente. La edición *ex vivo* tiene la ventaja de permitir que la población de células diana esté bien definida y que se especifique la dosis específica de moléculas terapéuticas administrada a las células. Esta última consideración puede ser particularmente útil cuando son preocupantes las modificaciones fuera del diana, ya que la titulación de la cantidad de nucleasas puede reducir dichas mutaciones (Hsu y col., 2013). Otra ventaja de las estrategias *ex vivo* es las altas tasas de edición que pueden lograrse debido al desarrollo de sistemas de administración eficaces para proteínas y ácidos nucleicos en células en cultivo para investigación y aplicaciones de terapia génica.

Puede haber desventajas con enfoques *ex vivo* que limitan la aplicación a una baja cantidad de enfermedades. Por ejemplo, células diana debe ser capaz de sobrevivir a la manipulación fuera del cuerpo. Para muchos tejidos, tales como el cerebro, el cultivo de células fuera del organismo es un gran reto debido a que las células no logran sobrevivir, o pierden las propiedades necesarias para su función *in vivo*. Por lo tanto, en vista de esta invención y los conocimientos del arte, se posibilita la terapia *ex vivo* en lo que respecta a tejidos con poblaciones de células adultas indiferenciadas que se pueden someter a cultivos y manipulación *ex vivo*, como por ejemplo el sistema hematopoyético, mediante el sistema CRISPR-Cas (Cpf1). [Bunn, H.F. & Aster, J. *Pathophysiology of blood disorders*, (McGraw-Hill, New York, 2011)].

La edición del genoma *in vivo* incluye el suministro directo de sistemas de edición a tipos de células en sus tejidos nativos. La edición *in vivo* permite tratar enfermedades en las cuales la población de células afectadas no se puede someter a una manipulación *ex vivo*. Además, el suministro *in situ* de nucleasas a las células permite el tratamiento de múltiples tipos de tejidos y de células. Dichas propiedades probablemente permiten aplicar el tratamiento *in vivo* a un rango más amplio de enfermedades que las terapias *ex vivo*.

Hasta la fecha, la edición *in vivo* se ha logrado en gran medida mediante el uso de vectores víricos con tropismo específico de tejido definido. Dichos vectores están limitados en la actualidad en cuanto a su capacidad para portar una carga y tropismo, restringiendo este modo de terapia a sistemas de órganos en donde es eficaz la transducción con vectores clínicamente útiles, tales como el hígado, músculos y el ojo [Kotterman, M.A. y Schaffer, D.V. Genetics 15, 445-451 (2014); Nguyen, T.H. & Ferry, N. Gene therapy 11 Supl. 1, S76-84 (2004); Boye, S.E., y col. Molecular therapy: the journal of the American Society of Gene Therapy 21, 509-519 (2013)].

Una potencial barrera para el suministro *in vivo* es la respuesta inmunitaria que se puede crear en respuesta a las grandes cantidades de virus que son necesarias para el tratamiento, pero este fenómeno no es peculiar de la edición del genoma y se observa con otras terapias génicas basadas en virus [Bessis, N., y col. Gene therapy 11 Supl. 1, S10-17 (2004)]. También es posible que los péptidos de las nucleasas de edición se presenten en moléculas de clase I del CMH para estimular una respuesta inmunitaria, aunque hay pocas pruebas que apoyen que esto suceda a nivel preclínico. Otra dificultad importante con este modo de terapia es controlar la distribución, y por consiguiente la dosificación de las nucleasas editoras del genoma *in vivo*, dando lugar a perfiles de mutación fuera del diana que son difíciles de predecir. Sin embargo, en vista de esta invención y los conocimientos en el arte, incluyendo el uso de terapias basadas en virus y partículas que se utilizan en el tratamiento de tipos de cáncer, modificación *in vivo* de HSCs, por ejemplo, por suministro ya sea mediante partículas o virus, se encuentra dentro de las capacidades de las personas con experiencia.

Terapia de edición *ex vivo*: La gran experiencia clínica con la purificación, cultivo y trasplante de células madre hematopoyéticas han puesto el foco de atención sobre las enfermedades que afectan al sistema sanguíneo, tales como SCID, anemia de Fanconi, el síndrome de Wiskott-Aldrich y la anemia de células falciformes para la terapia de edición *ex vivo*. Otra razón para centrarse en las células hematopoyéticas es que, gracias a los esfuerzos previos para diseñar terapias génicas para trastornos sanguíneos, ya existen sistemas de administración con una eficacia relativamente alta. Con dichas ventajas, esta modalidad de terapia se puede aplicar a enfermedades donde las células editadas poseen una ventaja por su aptitud, de manera que una baja cantidad de células editadas injertadas puede expandirse y tratar la enfermedad. Una de estas enfermedades es el VIH, en donde la infección da como resultado una desventaja de adecuación para las células T CD4+.

La terapia de edición *ex vivo* se ha extendido recientemente para que incluya estrategias de corrección génica. Las barreras a la HDR *ex vivo* se superaron en una publicación reciente de Genovese y sus colegas, quienes consiguieron la corrección génica de un gen IL2RG mutado en células indiferenciadas hematopoyéticas (HSCs) que se obtuvieron de un paciente que padecía SCID-X1 [Genovese, P., y col. Nature 510, 235-240 (2014)]. Genovese y sus colaboradores lograron la corrección génica en HSCs usando una estrategia multimodal. En primer lugar, se transdujeron las HSC usando lentivirus deficiente para integración que contenía un molde de HDR que codificaba un ADNc terapéutico para IL2RG. Después de la transducción, se electroporó a las células con ARNm que codificaba a los ZFN que se dirigen al punto mutacional en IL2RG para estimular la corrección génica basada en HDR. Para aumentar las frecuencias de HDR, se optimizaron las condiciones de cultivo con moléculas pequeñas para estimular la división de las HSC. Con las condiciones de cultivo optimizadas, las nucleasas y los moldes de HDR, se obtuvieron HSC corregidas génicamente del paciente SCID-X1 en cultivo a tasas terapéuticamente relevantes. Las HSC de individuos no afectados que se sometieron al mismo procedimiento de corrección génica podían sostener hematopoyesis a largo plazo en ratones, el patrón de oro para la función de las HSC. Las HSCs son capaces de

originar todos los tipos de células hematopoyéticas y se pueden transplantar de manera autóloga, convirtiéndolas en una población de células extremadamente valiosa para todos los trastornos genéticos hematopoyéticos [Weissman, I.L. & Shizuru, J.A. Blood 112, 3543-3553 (2008)]. Las HSC corregidas génicamente podrían usarse, en principio, para tratar una gran variedad de trastornos genéticos hematológicos haciendo que este estudio sea un punto de partida excitante para la edición genómica terapéutica.

Terapia de edición *in vivo* (descrita, pero no reivindicada en la presente): La edición *in vivo* se puede utilizar de manera ventajosa basándose en esta invención y los conocimientos en el arte. Para los sistemas de órganos donde el suministro es eficiente, ya existen varios excitantes éxitos terapéuticos preclínicos. El primer ejemplo de terapia de edición *in vivo* exitosa se demostró en un modelo de hemofilia B en ratón [Li, H., y col. Nature 475, 217-221 (2011)]. Tal como se indicó anteriormente, la hemofilia B es un trastorno recesivo ligado a X causado por mutaciones de pérdida de función en el gen que codifica al factor IX, un componente crucial de la cascada de coagulación. La recuperación de la actividad del Factor IX hasta más del 1% de sus niveles en individuos gravemente afectados puede transformar la enfermedad en una forma significativamente más leve, ya que la infusión profiláctica del Factor IX recombinante en dichos pacientes desde una corta edad para obtener dichos niveles mejora mucho las complicaciones clínicas [Lofqvist, T., y col. Journal of Internal Medicine 241, 395-400 (1997)]. Por lo tanto, solo son necesarios bajos niveles de corrección génica por HDR para cambiar los resultados clínicos en los pacientes. Además, el factor IX se sintetiza y secreta por el hígado, un órgano que puede transducirse de manera eficaz por vectores víricos que codifican sistemas de edición.

Usando serotipos virales adeno-asociados (AAV) hepatotrópicos que codifican ZFNs y un molde de corrección por HDR, se consiguió hasta un 7% de corrección génica de un gen del Factor IX mutado, humanizado en el hígado murino [Li, H., y col. Nature 475, 217-221 (2011)]. Esto dio como resultado una mejora en la cinética de formación del coágulo, una medida de la función de la cascada de coagulación, que demuestra por primera vez que la terapia de edición *in vivo* no es solo factible, sino también eficaz. Como se dice aquí, en base a las descripciones de la presente y los conocimientos en el arte, las personas con experiencia pueden, por ejemplo, tratar la Hemofilia B con un molde de HDR que contiene partículas y un sistema CRISPR-Cas (Cpf1) direccionado hacia la mutación del trastorno recesivo relacionado con X para revertir la mutación que causa la pérdida de función (Li).

En base a este estudio, recientemente se han utilizado otros grupos en la edición *in vivo* del genoma del hígado con CRISPR-Cas para tratar de manera exitosa un modelo en ratón de tirosinemia hereditaria y para crear mutaciones que proveen protección contra la enfermedad cardiovascular. Dichas dos diferentes aplicaciones demuestran la versatilidad de este enfoque para los trastornos relacionados con la disfunción hepática [Yin, H., y col. Nature biotechnology 32, 551-553 (2014); Ding, Q., y col. Circulation research 115, 488-492 (2014)]. Es necesaria la aplicación de la edición *in vivo* a otros sistemas orgánicos para demostrar que esta estrategia es ampliamente aplicable. En la actualidad, se están realizando esfuerzos para mejorar los vectores tanto víricos como no víricos para expandir la gama de trastornos que pueden tratarse con este modo de terapia [Kotterman, M.A. y Schaffer, D.V. Genetics 15, 445-451 (2014); Yin, H., y col. Nature reviews. Genetics 15, 541-555 (2014)]. Como se dice aquí, en base a las descripciones de la presente y los conocimientos en el arte, las personas con experiencia pueden, por ejemplo, tratar la tirosinemia hereditaria con un molde de HDR que contiene partículas y un sistema CRISPR-Cas (Cpf1) direccionado hacia la mutación (Yin).

Supresión dirigida, aplicaciones terapéuticas: Puede preferirse la supresión de genes diana. Por lo tanto, se prefieren genes relacionados con el trastorno de inmunodeficiencia, condición hematológica, o enfermedad genética de almacenamiento lisosómico, por ejemplo, hemofilia B, SCID, SCID-X1, ADA-SCID, Tirosinemia hereditaria, β-talasemia, CGD relacionada con X, síndrome de Wiskott-Aldrich, anemia de Fanconi, adrenoleucodistrofia (ALD), leucodistrofia metacromática (MLD), VIH/SIDA, otros trastornos metabólicos, genes que codifican proteínas mal plegadas relacionados con enfermedades, genes que llevan a la pérdida de función relacionada con enfermedades; en general, mutaciones que pueden ser blancos en un HSC, usando cualquier sistema de suministro de los que se exponen aquí, donde se considera ventajoso el sistema de partículas.

En los métodos y composiciones reivindicados en la presente invención, la inmunogenicidad de la enzima CRISPR en particular se podrá reducir siguiendo la estrategia propuesta inicialmente por Tangri y col. con respecto a la eritropoyetina y desarrollada posteriormente. En consecuencia, se podrá utilizar la evolución dirigida o el diseño racional para reducir la inmunogenicidad de la enzima CRISPR (por ejemplo, una Cas9) en las especies hospedadoras (especie humana o de otro tipo).

Edición del genoma: Los sistemas de CRISPR/Cas (Cpf1) de la presente invención se pueden utilizar para corregir mutaciones genéticas que previamente se habían intentado con un éxito limitado usando TALEN y ZFN y lentivirus, incluyendo las que se exponen aquí, véase también WO2013163628.

Tratamiento de enfermedades del cerebro, sistema nervioso central e inmunitario

La presente solicitud también describe, pero no reivindica, suministro del sistema CRISPR-Cas al cerebro o a las neuronas. Por ejemplo, el ARN de interferencia (ARNi) ofrece potencial terapéutico para este trastorno por reducción de la expresión de *HTT*, el gen causante de enfermedad de la enfermedad de Huntington (véase, por ejemplo, McBride y col., Molecular Therapy vol. 19 no. 12 de diciembre de 2011, pp. 2152-2162), por lo tanto el Solicitante postula que se puede usar y/o adaptar al sistema CRISPR-Cas. Se podrá generar el sistema CRISPR-Cas utilizando un algoritmo

para reducir el potencial de actuar no específico de secuencias de sentido contrario. Las secuencias CRISPR-Cas podrán tener como diana una secuencia en el exón 52 de la huntingtina humana, de rhesus o de ratón y expresarse en un vector vírico, tal como AAV. A los animales, incluyendo a los seres humanos, se les puede inyectar aproximadamente tres microinyecciones por hemisferio (seis inyecciones en total): las primeras inyecciones en el rostro a 1 mm de la comisura anterior (12 µl) y las dos restantes (12 µl y 10 µl, respectivamente) separadas 3 y 6 mm en dirección caudal de la primera inyección con 1e12 vg/ml de AAV con una velocidad de aproximadamente 1 µl/minuto, y la aguja se dejó en su lugar durante 5 minutos adicional para permitir que lo inyectado difunda desde la punta de la aguja.

DiFiglia y col. (PNAS, 23 de octubre de 2007, vol. 104, n.º 43, 17204–17209) observaron que una única administración en el cuerpo estriado adulto de un ARNip cuyo diana es Htt puede silenciar un Htt mutante, atenuar la patología neuronal y retrasar el fenotipo conductual anómalo observado en un modelo en ratones transgénicos vírico con un inicio rápido de HD. DiFiglia inyectó ratones intraestriatalmente con 2 µl de marcado con Cy3 cc-ARNip-Htt o ARNip-Htt sin conjugar a 10 µM. Se podrá contemplar una dosificación similar de CRISPR Cas cuyo diana sea Htt para seres humanos, por ejemplo, se podrán inyectar aproximadamente 5-10 ml de CRISPR Cas 10 µM cuyo diana sea Htt en el interior del cuerpo estriado.

En otro ejemplo, Boudreau y col. (Molecular Therapy vol. 17 no. 6 de junio de 2009) inyectó 5 µl de serotipo de AAV recombinante 2/1 vectores que expresaban ARNi virus htt-específico (a 4×10^{12} genomas virales/ml) en el cuerpo estriado. Se puede contemplar una dosificación similar de CRISPR Cas direccionado hacia Htt para seres humanos, por ejemplo, se les puede inyectar intraestriatalmente aproximadamente 10-20 ml de 4×10^{12} genomas virales/ml) CRISPR Cas direccionado hacia Htt.

En otro ejemplo, se puede administrar de manera continua un CRISPR Cas diseccionado hacia HTT (véase, por ejemplo, Yu y col., Cell 150, 895-908, 31 de agosto de 2012). Yu et al. utilizan bombas osmóticas que suministran 0,25 mL/h (Modelo 2004) para suministrar 300 mg/día de ARNip-mc o solución salina de pH controlado al fosfato (Sigma Aldrich) durante 28 días y para suministrar 75 mg/día del control positivo MOE ASO durante 14 días se utilizaron bombas diseñadas para suministrar 0,5 µL/h (Modelo 2002). Se rellenaron las bombas (Durect Corporation) con ARNip-mc o MOE diluido en PBS estéril y a continuación se incubaron a 37 ºC durante 24 o 48 horas (Modelo 2004) antes de la implantación. Se anestesiaron los ratones con un 2,5% de isofluorano y se practicó una incisión en la línea media en la base del cráneo. Utilizando guías estereotácticas, se implantó una cánula en el ventrículo lateral derecho y se aseguró con adhesivo Loctite. A la cánula se unió un catéter incorporado a una minibomba osmótica Alzet y se colocó la bomba subcutáneamente en el área escapular media. Se cerró la incisión con 5,0 suturas de nailon. Se podrá contemplar una dosificación similar de CRISPR Cas cuyo diana sea Htt para seres humanos en la presente invención, por ejemplo, se podrán administrar de aproximadamente 500 a 1000 g/día de CRISPR Cas cuyo diana sea Htt.

En otro ejemplo de infusión continua, Stiles y col. (Experimental Neurology 233 (2012) 463–471) implantaron un catéter intraparenquimatoso con una punta de aguja de titanio en el putamen derecho. Se conectó el catéter a una bomba SynchroMed® II (Medtronic Neurological, Minneapolis, MN) implantada subcutáneamente en el abdomen. Después de una infusión de 7 días de solución salina de pH controlado al fosfato con 6 µl/día, se rellenaron las bombas con el artículo de prueba y se programaron para el suministro continuo durante 7 días. Se infundieron entre aproximadamente 2,3 y 11,52 mg/d de ARNip con velocidades de infusión variables de entre aproximadamente 0,1 y 0,5 µL/min. Se puede contemplar una dosificación similar de CRISPR Cas direccionado hacia Htt para seres humanos, por ejemplo, se puede administrar entre aproximadamente 20 y 200 mg/día de CRISPR Cas direccionado hacia Htt. En otro ejemplo, los métodos de Publicación de Patente de los EE.UU. n.º 20130253040 asignada a Sangamo también se puede adaptar a partir de TALES al sistema de direccionamiento hacia ácidos nucleicos para tratar la enfermedad de Huntington.

En otro ejemplo, también se podrán adaptar los métodos de la patente de EE.UU. con n.º de publicación 20130253040 adjudicada a Sangamo de TALES al sistema CRIPSR Cas para tratar la enfermedad de Huntington.

WO2015089354 A1 a nombre de The Broad Institute y col., describe una diana para la enfermedad de Huntington (HP). Los genes diana posibles del complejo CRISPR con respecto a la enfermedad de Huntington: PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; y TGM2. Por lo tanto, se pueden seleccionar uno o más de PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; y TGM2 como blancos para la enfermedad de Huntington.

Otros trastornos repetitivos trinucleótidos. Los mismos pueden incluir cualquiera de los siguientes: la Categoría I incluye a la Enfermedad de Huntington (HD) y las ataxias espinocerebelares; las expansiones de Categoría II son fenotípicamente diferentes a las expansiones heterogéneas que generalmente son de pequea magnitud, pero también se encuentran en los exones de genes; y la Categoría III incluye síndrome de X frágil, distrofia miotónica, dos de las ataxias espinocerebelares, epilepsia mioclónica juvenil, y ataxia de Friedreich.

También se describen, pero no se reivindican específicamente en la presente, métodos de uso del sistema CRISPR-Cas para corregir defectos en los genes EMP2A y EMP2B que se identificaron como asociados con la enfermedad de Lafora. La enfermedad de Lafora es una afección autosómica recesiva que se caracteriza por una epilepsia mioclónica progresiva que puede comenzar con convulsiones epilépticas en la adolescencia. Unos pocos casos de la enfermedad pueden estar provocados por mutaciones en genes que aún no se han identificado. La enfermedad provoca convulsiones, espasmos musculares, dificultad al caminar, demencia y en última instancia la muerte. En la actualidad

no se dispone de una terapia que haya mostrado ser eficaz contra la evolución de la enfermedad. El sistema CRISPR-Cas también podrá tener como diana otras anomalías genéticas asociadas con la epilepsia y los factores genéticos subyacentes se describen más detalladamente en Genetics of Epilepsy and Genetic Epilepsies, editado por Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009).

Los métodos de la publicación de patente de Estados Unidos n.º 20110158957 concedida a Sangamo BioSciences, Inc. Sobre la inactivación de genes del receptor de los linfocitos T (TCR) también se podrán modificar para adaptarlos al sistema CRISPR Cas. En otro ejemplo, los métodos de la Publicación de Patente de los EE.UU. n.º 20100311124 concedida a Sangamo BioSciences, Inc. y la Publicación de Patente de los EE.UU. n.º 20110225664 concedida a Cellectis, que tratan las dos sobre la inactivación de la expresión génica del gen de la glutamina sintetasa también se podrán modificar para adaptarlos al sistema CRISPR Cas.

Las opciones de suministro para el cerebro incluyen la encapsulación de enzima CRISPR y ARN guía en la forma ya sea de ADN o ARN en liposomas y conjugación con caballos de Troya moleculares para el suministro a través de la barrera sangre-cerebro (BBB). Los caballos de Troya moleculares han demostrado ser eficaces para el suministro de vectores de expresión B-gal el cerebro de primates no humanos. Se puede utilizar la misma estrategia para vectores de suministro que contengan la enzima CRISPR y el ARN guía. Por ejemplo, Xia CF y Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology." Mol Pharm. Mayo-junio de 2009; 6(3):747-51. doi: 10.1021/mp800194) describen cómo es posible el suministro de ARN de interferencia pequeño (ARNip) a las células en cultivo, e *in vivo*, con el uso combinado de un anticuerpo monoclonal (mAb) específico del receptor y la tecnología de avidina-biotina. Los autores también han publicado que debido a que el enlace entre el mAb de direccionamiento y el ARNip es estable con la tecnología de avidina-biotina, y los efectos de la iARN en sitios distantes tales como el cerebro se observan *in vivo* tras la administración intravenosa del ARNip dirigido.

Zhang y col. (Mol Ther. Enero de 2003; 7(1):11-8.)) describen cómo se encapsularon plásmidos de expresión que codifican indicadores, tales como la luciferasa, en el interior de un "virus artificial" que comprende un inmunoliposoma pegilado de 85 nm, que se direccionó al cerebro del mono rhesus *in vivo* con un anticuerpo monoclonal (MAb) contra el receptor de la insulina humana (HIR, por sus siglas en inglés). El HIRMAb hace posible que el liposoma que porta el gen exógeno experimente transcitosis a través de la barrera hematoencefálica y endocitosis a través de la membrana plasmática neuronal tras la inyección intravenosa. El nivel de la expresión del gen de la luciferasa en el cerebro fue 50 veces más elevado en el mono rhesus en comparación con la rata. La expresión neuronal extendida del gen de la beta-galactosidasa en el cerebro de primates se demostró tanto mediante microscopía confocal como técnicas histoquímicas. Los autores indican que esta estrategia hace viable llevar a cabo una modificación reversible de tipo transgénico en adultos en 24 horas. En consecuencia, se prefiere el uso de inmunoliposomas. Se podrán utilizar conjuntamente con anticuerpos que se dirijan a proteínas de la superficie celular o tejidos específicos.

Enfermedad de Alzheimer

La Publicación de Patente de los EE.UU. n.º 20110023153, describe el uso de nucleasas con dedos de cinc para modificar genéticamente células, animales y proteínas asociadas a la enfermedad de Alzheimer. Una vez modificados, las células y animales se podrán someter a más pruebas utilizando métodos conocidos para estudiar los efectos de las mutaciones objetivo en el desarrollo y/o evolución de la EA utilizando medidas que se utilizan habitualmente en el estudio de la EA tales como, sin limitación, el aprendizaje y memoria, ansiedad, depresión, adicción y funciones sensitivomotoras, así como también ensayos que midan la función bioquímica, metabólica, patológica, funcional y conductual.

La presente divulgación comprende la edición de cualquier secuencia cromosómicas que codifique proteínas asociadas con la EA. Las proteínas relacionadas con la EA se seleccionan normalmente en función de una asociación experimental entre el trastorno de tipo EA y la proteína relacionada con la EA. Por ejemplo, la velocidad de producción o la concentración de circulación de una proteína relacionada con la EA podrá estar elevada o disminuida en una población que padece EA respecto a una población que no padece EA. Se podrán evaluar las diferencias en los niveles de proteína utilizando técnicas proteómicas que incluyen, sin limitación, la inmunoelectrotransferencia, tinción inmunohistoquímica, ensayo de inmunoadsorción enzimática (ELISA) y espectrometría de masas. Como alternativa, se podrán identificar las proteínas relacionadas con la EA obteniendo los perfiles de expresión génica de los genes que codifican las proteínas utilizando técnicas genómicas que incluyen, sin limitación, análisis por microarreglo de ADN, análisis en serie de la expresión génica (SAGE) y reacción en cadena de la polimerasa en tiempo real cuantitativa (Q-PCR).

Algunos ejemplos de proteínas asociadas a la enfermedad de Alzheimer pueden incluir a la proteína receptor de lipoproteínas de muy baja densidad (VLDLR) codificada por el gen VLDLR, la enzima activadora del modificador similar a ubicuitina 1 (UBA1) codificada por el gen UBA1, o la proteína de la subunidad catalítica E1 de la enzima activadora de NEDD8 (UBE1C) codificada por ejemplo por el gen UBA3.

A manera de ejemplo no limitativo, las proteínas asociadas con la EA incluyen, pero de manera no taxativa, a las proteínas que se mencionan a continuación: Proteína codificada por la secuencia cromosómica ALAS2 Delta-aminolevulinato sintasa 2 (ALAS2) ABCA1 Transportador de casete de unión a ATP (ABCA1) ACE Enzima conversora de angiotensina I (ACE) APOE Precursor de la apolipoproteína E (APOE) APP proteína precursora del amiloide (APP)

AQP1 proteína acuaporina 1 (AQP1) BIN1 Proteína 1 de interacción dependiente de la secuencia myc o proteína 1 integradora de tipo puente (BIN1) BDNF factor neurotrófico derivado del cerebro (BDNF) BTNL8 Proteína 8 de tipo butirofilina (BTNL8) C1ORF49 Marco abierto de lectura 49 del cromosoma 1 CDH4 Cadherina-4 CHRNB2 Subunidad beta-2 del receptor de acetilcolina neuronal CKLFSF2 Proteína 2 que contiene el dominio transmembrana MARVEL de tipo CKLF (CKLFSF2) CLEC4E Familia 4 con dominio de lectina de tipo C, miembro e (CLEC4E) CLU proteína clusterina (también conocida como apoplipoproteína J) CR1 Receptor 1 del complemento de eritrocitos (CR1, también conocido como receptor de adherencia inmunitario y receptor CD35, C3b/C4b) CR1L Receptor 1 del complemento de eritrocitos (CR1L) CSF3R Receptor del factor 3 estimulante de colonias de granulocitos (CSF3R) CST3 Cistatina C o cistatina 3 CYP2C Citocromo P450 2C DAPK1 Proteína cinasa 1 asociada a la muerte (DAPK1) ESR1 Receptor estrogénico 1 FCAR Receptor del fragmento Fc de IgA (FCAR, también conocido como CD89) FCGR3B Receptor del fragmento Fc de IgG, IIIb de baja afinidad (FCGR3B o CD16b) FFA2 Receptor 2 de ácidos grasos libres (FFA2) FGA Fibrinógeno (Factor I) GAB2 Proteína 2 de unión asociada a GRB2 (GAB2) GAB2 Proteína 2 de unión asociada a GRB2 (GAB2) GALP Péptido de tipo galanina GAPDHS Gliceraldehído-3-fosfato-deshidrogenasa, espermatogénica (GAPDHS) GMPB GMBP HP Haptoglobina (HP) HTR7 Receptor 7 de 5-hidroxitriptamina (serotonina) (acoplado a la adenilato ciclasa) IDE Enzima que degrada insulina IF127 IFI6 Interferón, proteína 6 alfa inducible (IFI6) IFIT2 Proteína inducida por interferón con repeticiones de tetratricopéptidos 2 (IFIT2) IL1RN antagonista del receptor de interleuquina-1 (IL-1RA) IL8RA Receptor de interleuquina 8, alfa (IL8RA o CD181) IL8RB Receptor de interleuquina 8, beta (IL8RB) JAG1 Jagged 1 (JAG1) KCNJ15 Canal de potasio rectificador de entrada, subfamilia J, miembro 15 (KCNJ15) LRP6 Proteína 6 relacionada con el receptor de lipoproteína de baja densidad (LRP6) MAPT Proteína tau asociada a los microtúbulos (MAPT) MARK4 Cinasa 4 reguladora por afinidad a MAP/microtúbulos (MARK4) MPHOSPH1 Fosfoproteína 1 de la fase M MTHFR 5,10-metilenotetrahidrofolato-reductasa MX2 Proteína Mx2 de unión a GTP inducida por interferón NBN Nibrina, también conocida como NBN NCSTN Nicastrina NIACR2 Receptor 2 de niacina (NIACR2, también conocido como GPR109B) NMNAT3 nicotinamida nucleótido adenililtransferasa 3 NTM Neurotrimina (o HNT) ORM1 Orosmucoide 1 (ORM1) o glucoproteína ácida alfa-1 1 P2RY13 P2Y purinoceptor 13 (P2RY13) PBEF1 Nicotinamida fosforribosiltransferasa (NAmPRTasa o Nampt) también conocida como factor 1 potenciador de colonias de prelinfocitos B (PBEF1) o visfatina PCK1 Fosfoenolpiruvato carboxicinasa PICALM proteína de reclutamiento de clatrina de unión a fosfatidilinositol (PICALM) PLAU Activador del plasminógeno de tipo urocinasa (PLAU) PLXNC1 Plexina C1 (PLXNC1) PRNP Proteína priónica PSEN1 Proteína presenilina 1 (PSEN1) PSEN2 Proteína presenilina 2 (PSEN2) PTPRA Proteína receptor tipo A de la proteína tirosina fosfatasa (PTPRA) RALGPS2 Ral GEF con dominio PH y motivo 2 de unión a SH3 (RALGPS2) RGSL2 regulador de señalización de la proteína G tipo 2 (RGSL2) SELENBP1 Proteína 1 de unión a selenio (SELNBP1) SLC25A37 Mitoferrina-1 SORL1 Receptor L relacionado con sortilina (clase DLR) proteína que contiene repeticiones de A (SORL1) TF Transferrina TFAM Factor A de transcripción mitocondrial TNF Factor de necrosis tumoral TNFRSF10C Miembro 10C de la superfamilia del receptor del factor de necrosis tumoral (TNFRSF10C) TNFSF10 Superfamilia del receptor del factor de necrosis tumoral, miembro 10a (TRAIL) (TNFSF10) UBA1 Enzima 1 activadora con modificación de tipo ubiquitina (UBA1) UBA3 Proteína subunidad catalítica E1 de la enzima activante de NEDD8 (UBE1C) UBB Proteína ubiquitina B (UBB) UBQLN1 Ubiquilina-1 UCHL1 Proteína L1 ubiquitina carboxilo-terminal esterasa (UCHL1) UCHL3 Proteína isoenzima L3 ubiquitina carboxilo-terminal hidrolasa (UCHL3) VLDLR Proteína receptor de lipoproteínas de muy baja densidad (VLDLR).

En las realizaciones ilustrativas, las proteínas asociadas con la EA cuya secuencia cromosómica se edita podrán ser la proteína receptor de lipoproteínas de muy baja densidad (VLDLR) codificada por el gen VLDLR, la enzima 1 activadora con modificación de tipo ubiquitina (UBA1) codificada por el gen UBA1, la proteína subunidad catalítica E1 de la enzima activante de NEDD8 (UBE1C) codificada por el gen UBA3, la proteína acuaporina 1 (AQP1) codificada por el gen AQP1, la proteína L1 ubiquitina carboxilo-terminal esterasa (UCHL1) codificada por el gen UCHL1, la proteína isoenzima L3 ubiquitina carboxilo-terminal hidrolasa (UCHL3) codificada por el gen UCHL3, la proteína ubiquitina B (UBB) codificada por el gen UBB, la proteína tau asociada a microtúbulos (MAPT) codificada por el gen MAPT, la proteína receptor tipo A de la proteína tirosina fosfatasa (PTPRA) codificada por el gen PTPRA, la proteína de reclutamiento de clatrina de unión a fosfatidilinositol (PICALM) codificada por el gen PICALM, la proteína clusterina (también conocida como apolipoproteína J) codificada por el gen CLU, la proteína presenilina 1 codificada por el gen PSEN1, la proteína presenilina 2 codificada por el gen PSEN2, el receptor L relacionado con sortilina (clase DLR) proteína que contiene repeticiones de A (SORL1) proteína codificada por el gen SORL1, la proteína precursora del amiloide (APP) codificada por el gen APP, el precursor de la apolipoproteína E (APOE) codificada por el gen APOE o el factor neurotrófico derivado del cerebro (BDNF) codificado por el gen BDNF. En una forma de realización ilustrativa, el animal modificado genéticamente es una rata, y la secuencia cromosómica editada que codifica la proteína asociada con la EA es como las siguientes: APP proteína precursora del amiloide (APP) NM_019288 AQP1 proteína acuaporina 1 (AQP1) NM_012778 BDNF Factor neurotrófico derivado del cerebro NM_012513 CLU proteína clusterina (también conocida como NM_053021 apolipoproteína J) MAPT proteína tau asociada a microtúbulos NM_017212 (MAPT) PICALM proteína de reclutamiento de clatrina de unión a fosfatidilinositol NM_053554 (PICALM) PSEN1 proteína presenilina 1 (PSEN1) NM_019163 PSEN2 proteína presenilina 2 (PSEN2) NM_031087 PTPRA proteína receptor tipo A de la proteína tirosina fosfatasa NM_012763 (PTPRA) SORL1 receptor L relacionado con sortilina (clase DLR, NM_053519) proteína que contiene repeticiones de A XM_001065506 (SORL1) XM_217115 UBA1 enzima 1 activadora con modificación de tipo ubiquitina NM_001014080 (UBA1) UBA3 proteína subunidad catalítica E1 de la enzima activante de NEDD8 NM_057205 (UBE1C) UBB proteína ubiquitina B (UBB) NM_138895 UCHL1 proteína L1 ubiquitina carboxilo-terminal esterasa NM_017237 (UCHL1) UCHL3 proteína isoenzima L3 ubiquitina carboxilo-terminal hidrolasa NM_001110165 (UCHL3) VLDLR proteína receptor de lipoproteínas de muy baja densidad

NM_013155 (VLDLR).

El animal o célula podrá comprender 1, 2, 3, 4, 5, 6, 7, 8, 9,10, 11, 12, 13, 14, 15 o más secuencias cromosómicas alteradas que codifican una proteína asociada con la EA y cero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o más secuencias integradas cromosómicamente que codifican una proteína asociada con la EA.

La secuencia cromosómica integrada o editada se podrá modificar para que codifique una proteína alterada asociada con la MD. Se han asociado con la EA varias mutaciones en secuencias cromosómicas relacionadas con la EA. Por ejemplo, la mutación de aminoácido V7171 (es decir, se reemplaza valina de la posición 717 por isoleuquina) en APP causa EA familiar. Las mutaciones múltiples en la proteína presenilina-1, como por ejemplo H163R (es decir la histidina en la posición 163 se cambia por arginina), A246E (es decir, la alanina en la posición 246 se cambia por glutamato), L286V (es decir, la leucina en la posición 286 se cambia por valina) y C410Y (es decir la cisteína en la posición 410 se cambia por tirosina) causa la enfermedad de Alzheimer familiar de tipo 3. Las mutaciones en la proteína presenilina-2, como por ejemplo N141 I (es decir la asparagina en la posición 141 se cambia por isoleucina), M239V (es decir la metionina en la posición 239 se cambia por valina), y D439A (es decir el aspartato en la posición 439 se cambia por alanina) causa la enfermedad de Alzheimer familiar de tipo 4. Otras asociaciones de variantes genéticas en genes asociados a la EA y enfermedad son conocidas en el arte. Véase, por ejemplo, Waring y col. (2008) Arch. Neurol. 65:329-334.

Trastornos de la secretasa

La Publicación de Patente de los EE.UU. n.º 20110023146, describe el uso de nucleasas con dedos de cinc para modificar genéticamente células, animales y proteínas asociadas con trastornos asociados con la secretasa. Las secretasas son esenciales para procesar preproteínas y obtener sus formas biológicamente activas. Los defectos en diversos componentes de las rutas de la secretasa contribuyen a muchos trastornos, especialmente aquellos con placas amiloideas o amiloidogénesis distintivas, tal como la enfermedad de Alzheimer (EA).

Un trastorno asociado con la secretasa y las proteínas asociadas con dichos trastornos son conjunto variado de proteínas que tienen como efecto la susceptibilidad de numerosos trastornos, la presencia del trastorno, la gravedad del trastorno, o cualquier combinación de los mismos. La presente divulgación comprende editar cualquiera de las secuencias cromosómicas que codifican proteínas asociadas con un trastorno de la secretasa. Las proteínas asociadas con un trastorno de la secretasa se seleccionan normalmente en función de una asociación experimental entre las proteínas relacionadas con la secretasa y el desarrollo de un trastorno de la secretasa. Por ejemplo, la velocidad de producción o la concentración de circulación de una proteína asociada con un trastorno de la secretasa podrá estar elevada o disminuida en una población que padece un trastorno de la secretasa respecto a una población que no padece el trastorno de la secretasa. Se podrán evaluar las diferencias en los niveles de proteína utilizando técnicas proteómicas que incluyen, sin limitación, la inmunoelectrotransferencia, tinción inmunohistoquímica, ensayo de inmunoadsorción enzimática (ELISA) y espectrometría de masas. Como alternativa, se podrá identificar la proteína asociada con un trastorno de la secretasa obteniendo los perfiles de expresión génica de los genes que codifican las proteínas utilizando técnicas genómicas que incluyen, sin limitación, análisis por microarreglo de ADN, análisis en serie de la expresión génica (SAGE) y reacción en cadena de la polimerasa en tiempo real cuantitativa (Q-PCR).

A manera de ejemplo no limitativo, las proteínas asociadas con trastornos asociados con la secretasa incluyen a: PSENEN (homólogo del potenciador de presenilina 2 (*C. elegans*)), CTSB (catepsina B), PSEN1 (presenilina 1), APP (proteína precursora de beta amiloide (A4)), APH1B (homólogo B de faringe anterior defectuosa 1 (*C. elegans*)), PSEN2 (presenilina 2 (enfermedad de Alzheimer 4)), BACE1 (Enzima que escinde APP en el sitio beta 1), ITM2B (proteína integral de membrana 2B), CTSD (catepsina D), NOTCH1 (homólogo de Notch 1, asociado a la translocación (Drosophila)), TNF (factor de necrosis tumoral (superfamilia TNF, miembro 2)), INS (insulina), DYT10 (distonia 10), ADAM17 (dominio 17 de metalopeptidasa ADAM), APOE (apolipoproteína E), ACE (enzima convertidora de angiotensina I (peptidil-dipeptidasa A) 1), STN (estatina), TP53 (proteína tumoral p53), IL6 (interleuquina 6 (interferón, beta 2)), NGFR (receptor del factor de crecimiento nervioso (superfamilia TNFR, miembro 16)), IL1B (interleuquina 1, beta), ACHE (acetilcolinesterasa (grupo sanguíneo Yt)), CTNNB1 (catenina (proteína asociada a cadherina), beta 1, 88kDa), IGF1 (factor de crecimiento 1 similar a insulina (somatomedina C)), IFNG (interferón, gamma), NRG1 (neuregulina 1), CASP3 (caspasa 3, cisteína peptidasa relacionada con la apoptosis), MAPK1 (proteína quinasa 1 activada por mitógeno), CDH1 (cadherina 1, tipo 1, E-cadherina (epitelial)), APBB1 (proteína de unión a precursores de beta amiloide (A4), familia B, miembro 1 (Fe65)), HMGCR (3-hidroxi-3-metilglutaril-Coenzima A reductasa), CREB1 (proteína de unión al elemento de respuesta a cAMP 1), PTGS2 (prostaglandina-endoperóxido sintasa 2 (prostaglandina G/H sintasa y ciclooxigenasa)), HES1 (piloso y potenciador de split 1, (Drosophila)), CAT (catalasa), TGFB1 (factor de crecimiento transformante, beta 1), ENO2 (enolasa 2 (gamma, neuronal)), ERBB4 (v-erb-a leucemia viral eritroblástica oncogen homólogo 4 (aviar)), TRAPPC10 (complejo de transporte de partículas de proteína 10), MAOB (monoamino oxidasa B), NGF (factor de crecimiento nervioso (polipéptido beta)), MMP12 (metalopeptidasa de matriz 12 (elastasa de macrófagos)), JAG1 (Jagged 1 (síndrome de Alagille)), CD40LG (ligando CD40), PPARG (receptor gamma activado por proliferador de peroxisomas), FGF2 (factor de crecimiento de fibroblastos 2 (básico)), IL3 (interleuquina 3 (factor estimulador de colonias, multiple)), LRP1 (proteína relacionada con el receptor de lipoproteína de baja densidad 1), NOTCH4 (Homólogo 4 de Notch (Drosophila)), MAPK8 (proteína quinasa activada por mitógeno 8), PREP (prolil endopeptidasa), NOTCH3 (Homólogo de Notch 3 (Drosophila)), PRNP (proteína de prion), CTSG (catepsina G), EGF (factor de crecimiento epidérmico (beta-urogastrona)), REN (renina), CD44

(moléculas CD44 (Grupo sanguíneo hindú)), SELP (selectina P (proteína de membrana granular 140 kDa, antígeno CD62)), GHR (receptor de la hormona de crecimiento), ADCYAP1 (polipéptido activador 1 de la adenilato ciclasa (pituitaria)), INSR (receptor de insulina), GFAP (proteína ácida fibrilar glial), MMP3 (metalopeptidasa de matriz 3 (estromelisina 1, progelatinasa)), MAPK10 (proteína quinasa 10 activada por mitógeno), SP1 (factor de transcripción Sp1), MYC (homólogo del oncogen viral de mielocitomatosis v-myc (aviar)), CTSE (catepsina E), PPARA (receptor alfa activado por proliferador de peroxisomas), JUN (oncogen jun), TIMP1 (inhibidor de metalopeptidasa TIMP 1), IL5 (interleuquina 5 (factor estimulador de colonias, eosinófilos)), IL1A (interleuquina 1, alfa), MMP9 (metalopeptidasa de matriz 9 (gelatinasa B, gelatinasa de 92 kDa, colagenasa de tipo IV de 92 kDa)), HTR4 (receptor 4 de 5-hidroxitriptamina (serotonina)), HSPG2 (sulfato de heparan proteoglicano 2), KRAS (homólogo del oncogen viral del sarcoma de rata Kirsten, v-Ki-ras2), CYCS (citocromo c, somático), SMG1 (homólogo de SMG1, quinasa relacionada con la fosfatidilinositol 3-quinasa (*C. elegans*)), IL1R1 (receptor de interleuquina 1, tipo I), PROK1 (prokineticina 1), MAPK3 (proteína quinasa activada por mitógeno 3), NTRK1 (tirosina quinasa neurotrófica, receptor, tipo 1), IL13 (interleuquina 13), MME (metalo-endopeptidasa de membrana), TKT (transcetolasa), CXCR2 (receptor 2 de quimioquina (motivo C-X-C)), IGF1R (factor de crecimiento 1 similar al receptor de insulina), RARA (receptor del ácido retinoico, alfa), CREBBP (proteína de unión a CREB), PTGS1 (prostaglandina-endoperóxido sintasa 1 (prostaglandina G/H sintasa y ciclooxigenasa)), GALT (galactosa-1-fosfato uridililtransferasa), CHRM1 (receptor colinérgico, muscarínico 1), ATXN1 (ataxina 1), PAWR (PRKC, apoptosis, WT1, regulador), NOTCH2 (homólogo de Notch 2 (*Drosophila*)), M6PR (receptor de manosa-6-fosfato (dependiente de catión)), CYP46A1 (citocromo P450, familia 46, subfamilia A, polipéptido 1), CSNK1 D (caseína quinasa 1, delta), MAPK14 (proteína quinasa 1 activada por mitógeno 4), PRG2 (proteoglicano 2, médula ósea (activador de la célula asesina natural, proteína básica mayor de gránulos de eosinófilos)), PRKCA (proteína quinasa C, alfa), L1 CAM (moléculas de adhesión celular L1), CD40 (moléculas CD40, miembro de la superfamilia 5 de receptores de TNF), NR1I2 (subfamilia de receptores nucleares 1, grupo 1, miembro 2), JAG2 (*jagged* 2), CTNND1 (catenina (proteína asociada a cadherina), delta 1), CDH2 (cadherina 2, tipo 1, N-cadherina (neuronal)), CMA1 (quimasa 1, mastocito), SORT1 (sortilina 1), DLK1 (homólogo 1 tipo delta (*Drosophila*)), THEM4 (miembro 4 de la superfamilia de tioesterasas), JUP (placoglobina de la unión), CD46 (moléculas CD46, proteína reguladora del complemento), CCL11 (ligando 11 de la quimioquina (motivo C-C)), CAV3 (caveolina 3), RNASE3 (ribonucleasa, familia de RNasa A, 3 (proteína catiónica de eosinófilos)), HSPA8 (proteína 8 de 70kDa de shock de calor), CASP9 (caspasa 9, relacionada con la cisteína peptidasa de apoptosis), CYP3A4 (citocromo P450, familia 3, subfamilia A, polipéptido 4), CCR3 (receptor 3 de quimioquina (motivo C-C)), TFAP2A (factor de transcripción AP-2 alfa (proteína de unión 2 alfa activadora potenciadora)), SCP2 (proteína 2 transportadora de esterol), CDK4 (quinasa 4 dependiente de ciclina), HIF1A (factor 1 inducible por hipoxia, subunidad alfa (factor de transcripción básico de hélice-bucle-hélice)), TCF7L2 (2 tipo factor de transcripción 7 (específica de células T, caja HMG)), IL1R2 (receptor de interleuquina 1, tipo II), B3GALTL (tipo beta 1,3-galactosiltransferasa), MDM2 (homólogo de la proteína de unión a p53 Mdm2 (ratón)), RELA (homólogo A del oncogen viral de la reticuloendoteliosis v-rel (aviar)), CASP7 (caspasa 7, relacionada con la cisteína peptidasa de apoptosis), IDE (enzima degradadora de insulina), FABP4 (proteína de unión a ácidos grasos 4, adipocito), CASK (serina proteína quinasa dependiente de calcio/calmodulina (familia MAGUK)), ADCYAP1R1 (receptor tipo I del polipéptido activador 1 de la adenilato ciclasa (pituitaria)), ATF4 (factor activador de la transcripción 4 (elemento potenciador que responde a tax B67)), PDGFA (polipéptido alfa del factor de crecimiento derivado de plaquetas), C21 o f33 (marco de lectura abierto 33 del cromosoma 21), SCG5 (secretogranina V (proteína 7B2)), RNF123 (proteína 123 de dedo anular), NFKB1 (factor nuclear del potenciador génico del polipéptido 1 de la cadena liviana kappa en células B), ERBB2 (homólogo 2 del oncogen viral eritroblástico de leucemia v-erb-b2, homólogo del oncogen derivado de neuro/glioblastoma (aviar)), CAV1 (caveolina 1, proteína caveolae, 22 kDa), MMP7 (metalopeptidasa de matriz 7 (matrilisina, uterina)), TGFA (factor de crecimiento transformante, alfa), RXRA (receptor X retinoide, alfa), STX1A (sintaxina 1A (cerebro)), PSMC4 (proteasoma (prosoma, macropaína) subunidad 26S, ATPasa, 4), P2RY2 (receptor purinérgico P2Y, acoplado a proteína G, 2), TNFRSF21 (superfamilia de receptores del factor de necrosis tumoral, miembro 21), DLG1 (discos, homólogo grande 1 (*Drosophila*)), NUMBL (homólogo de *numb* (tipo *Drosophila*)), SPN (sialoforina), PLSCR1 (fosfolípido escramblasa 1), UBQLN2 (ubiquilina 2), UBQLN1 (ubiquilina 1), PCSK7 (proproteína convertasa subtilisina/quexina tipo 7), SPON1 (espondina 1, proteína de la matriz extracelular), SILV (homólogo de *silver* (ratón)), QPCT (glutaminil-péptido ciclotransferasa), HESS (piloso y potenciador de split 5 (*Drosophila*)), GCC1 (proteína 1 que contiene dominio GRIP y bucle enrollado), y cualquier combinación de los mismos.

El animal o célula modificados genéticamente podrán comprender 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más secuencias cromosómicas alteradas que codifican una proteína asociada con un trastorno de la secretasa y cero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más secuencias integradas cromosómicamente que codifican una proteína alterada asociada con un trastorno de la secretasa.

ELA

La Publicación de Patente de los EE.UU. n.º 20110023144, describe el uso de nucleasas con dedos de cinc para modificar genéticamente células, animales y proteínas asociadas con la enfermedad denominada esclerosis lateral amiotrófica (ELA). La ELA se caracteriza por la degeneración gradual e inexorable de ciertas células nerviosas en la corteza cerebral, el tronco cerebral y la médula espinal que participan en el movimiento voluntario.

Los trastornos de las neuronas motoras y las proteínas asociadas con dichos trastornos son un conjunto variado de proteínas que tienen como efecto la susceptibilidad de desarrollar un trastorno de las neuronas motoras, la presencia del trastorno de las neuronas motoras, la gravedad del trastorno de las neuronas motoras o cualquier combinación de

los mismos. La presente divulgación comprende editar cualquiera de las secuencias cromosómicas que codifican proteínas asociadas con la enfermedad ELA, un trastorno de las neuronas motoras específico. Las proteínas asociadas con la ELA se seleccionan normalmente en función de una asociación experimental entre las proteínas relacionadas con ELA y la ELA. Por ejemplo, la velocidad de producción o la concentración de circulación de una proteína asociada con la ELA podrá estar elevada o disminuida en una población que padece ELA respecto a una población que no padece ELA. Se podrán evaluar las diferencias en los niveles de proteína utilizando técnicas proteómicas que incluyen, sin limitación, la inmunoelectrotransferencia, tinción inmunohistoquímica, ensayo de inmunoadsorción enzimática (ELISA) y espectrometría de masas. Como alternativa, se podrán identificar las proteínas asociadas con la ELA obteniendo los perfiles de expresión génica de los genes que codifican las proteínas utilizando técnicas genómicas que incluyen, sin limitación, análisis por microarreglo de ADN, análisis en serie de la expresión génica (SAGE) y reacción en cadena de la polimerasa en tiempo real cuantitativa (Q-PCR).

A modo de ejemplo no limitante, las proteínas asociadas con los TEA incluyen, sin limitación, las siguientes proteínas: SOD1 superóxido dismutasa 1, ALS3 soluble esclerosis lateral amiotrófica 3, SETX senataxina, ALS5 esclerosis lateral amiotrófica 5, FUS fusionado en sarcoma, ALS7 esclerosis lateral amiotrófica 7, ALS2 esclerosis lateral amiotrófica 2, DPP6 dipeptidil-peptidasa 6, NEFH neurofilamento, PTGS1 pesada polipéptido de prostaglandina-endoperóxido sintasa 1, SLC1A2 familia de transportadores de solutos 1, TNFRSF10B factor de necrosis tumoral (superfamilia de receptores de gran afinidad de la glía, transportador de glutamato), miembro 10b miembro 2, PRPH periferina, HSP90AA1 proteína de shock de calor de alfa de 90 kDa (citosólica), miembro 1 de clase A, GRIA2 receptor de glutamato, IFNG interferon-gamma ionotrópico, AMPA 2 S100B S100 de unión a calcio, FGF2 proteína B de factores de crecimiento de fibroblastos 2, AOX1 aldehído oxidasa 1, CS citrate sintasa, TARDBP TAR proteína de unión a AND, TXN tiorredoxina, RAPH1 asociación a Ras, MAP3K5 proteína activada por mitógeno (RalGDS/AF-6) y dominios de homología con pleckstrina 1 de quinasa 5, NBEAL1 tipo neurobeachina 1, GPX1 glutatión peroxidasa 1, ICA1L autoantígeno de células de los islotes, RAC1 sustrato 1 de la toxina botulínica de tipo C3 de 1,69 kDa relacionada con ras, MAPT asociada a microtúbulos, ITPR2 proteína tau receptor de trifosfato 1,4,5-inositol, tipo 2, ALS2CR4 región cromosómica GLS de glutaminasa 2 (juvenil) de la esclerosis lateral amiotrófica, candidato 4; ALS2CR8 región cromosómica del receptor del factor neurotrófico ciliar CNTFR 2 (juvenil) de la esclerosis lateral amiotrófica, candidato 8; ALS2CR11 región cromosómica 2 (juvenil) de la FOLH1 folato hidrolasa 1 de la esclerosis amiotrófica lateral, candidato 11, FAM117B polipéptido beta del miembro B de la familia con similitud a la secuencia P4HB de la prolil 4-hidroxilasa 117, CNTF factor neurotrófico ciliar, SQSTM1 secuestosoma 1, STRADB proteína beta inhibidora adaptadora de apoptosis familia NLR de NAIP quinasa relacionada con STE20, YWHAQ transportadora de tirosina 3-SLC33A1 monooxigenasa/triptofano(acetil-CoA) del miembro 1 de la familia transportadora de solutos 33, polipéptido theta de proteína de activación de 5-monooxigenasa, TRAK2 proteína de tráfico, homólogo de la FIG. 4, SAC1 quinesina que se une al dominio lípido fosfatasa 2 que contiene NIF3L1 proteína de filamento intermedio 1 tipo factor neuronal 3 NIF3 de internexina INA que interactúa con NGG1, alfa PARD3B particionamiento par-3, COX8A subunidad VIIIA del homólogo B de la citocromo c oxidasa defectuosa 3, CDK15 quinasa dependiente de ciclina, HECW1 HECT, C2 y WW proteína ligasa 1 de E3 ubicuitina que contiene dominio 15, NOS1 óxido nítrico sintasa 1, MET proto-oncogen met, SOD2 superóxido dismutasa 2, HSPB1 proteína 1 mitocondrial de 27 kDa de shock de calor, NEFL neurofilamento, CTSB polipéptido liviano de catepsina B, ANG angiogenina, HSPA8 ribonucleasa de 70 kDa de shock de calor, RNasa A familia de proteína 8, 5 VAPB VAMP (vesícula-proteína de membrana asociada al receptor 1 de estrógeno ESR1)-proteína B asociada y C SNCA sinucleína, alfa HGF factor de crecimiento de hepatocitos, CAT catalasa, ACTB actina, beta NEFM neurofilamento, polipéptido mediano de tirosina hidroxilasa TH, BCL2 CLL de células B/linfoma 2, FAS Fas (superfamilia del receptor de TNF, miembro 6), CASP3 caspasa 3, apoptosis-CLU cisteína peptidasa relacionada con clusterina, SMN1 neuronas motoras de supervivencia, G6PD glucosa-6-fosfato 1, deshidrogenasa telomérica, BAX X asociado a BCL2, HSF1 factor 1 de la proteína de transcripción de shock de calor, RNF19A proteína 19A de dedo anular, JUN oncogen jun, ALS2CR12 región cromosómica de la proteína 5 de shock de calor de 70 kDa HSPA5 de la esclerosis lateral amiotrófica 2 (juvenil), candidato 12, MAPK14 proteína activada por mitógeno, IL10 interleuquina 10 quinasa, 14 APEX1 nucleasa APEX, TXNRD1 tiorredoxina reductasa 1 (enzima reparadora de ADN multifuncional), NOS2 óxido nítrico sintasa 2, TIMP1 inhibidor inducible de metalopeptidasa TIMP 1, CASP9 caspasa 9, apoptosis-XIAP inhibidor ligado a X de la cisteína peptidasa de apoptosis relacionada, GLG1 glicoproteína 1 de Golgi, EPO eritropoyetina, VEGFA ELN elastina del factor de crecimiento del endotelio vascular A, GDNF factor nuclear NFE2L2 (factor neurotrófico eritroide derivado 2) tipo 2 de células derivadas de la glía, SLC6A3 familia de transportadores de solutos 6, HSPA4 shock de calor de 70 kDa (transportador de la proteína neurotransmisora 4, dopamina), miembro 3 de APOE apolipoproteína E, PSMB8 subunidad de proteasoma 8 (prosoma, macropaína), tipo beta, DCTN1 dinactina 1, TIMP3 inhibidor de metalopeptidasa 3 TIMP, KIFAP3 asociado a quinesina, SLC1A1 proteína 3 de la familia transportadora de solutos 1 (transportador de glutamato de gran afinidad neuronal/epitelial, sistema Xag), miembro 1, SMN2 neuronas motoras de supervivencia, CCNC ciclina C 2, centromérica, MPP4 proteína de membrana, STUB1 proteína 1 que contiene caja 4 de homología de STIP1 y U-palmitoilado, ALS2 beta amiloide (A4), PRDX6 proteína precursora de peroxirredoxina 6, SYP sinaptofisina, CABIN1 proteína de unión a calcineurina 1, CASP1 caspasa 1, apoptosis-GART cisteína relacionada con fosforribosilglicinamida formiltransferasa, peptidasa de la fosforribosilglicinamida sintetasa, fosforribosilaminoimidazol sintetasa, CDK5 quinasa 5 dependiente de ciclina, ATXN3 ataxina 3, RTN4 reticulona 4, C1QB componente 1 del complemento, subcomponente q, cadena B del VEGFC factor de crecimiento nervioso, HTT receptor de Huntington, PARK7 enfermedad de Parkinson 7, XDH xantina deshidrogenasa, GFAP proteína fibrilar ácida de la glía, MAP2 proteína 2 asociada a microtúbulos, CYCS citocromo c, somático, FCGR3B fragment Fc de la IgG, IIIb de baja afinidad, CCS chaperona de cobre para UBL5 similar a la ubicuitina 5 superóxido dismutasa, MMP9 metalopeptidasa de la

matriz, SLC18A3 familia transportadora de solutos 18 9 ((acetilcolina vesicular), miembro 3, TRPM7 receptor transitorio del canal catiónico potencial de 27 kDa HSPB2 shock de calor, proteína 2 subfamilia M, miembro 7, AKT1 timoma v-akt de murino, DERL1 familia de dominios tipo Der1, homólogo 1 del oncogen viral, miembro 1, CCL2 quimioquina (motivo C--C), NGRN neugrina, asociada al ligando 2 de la extensión de neuronas, GSR glutatión reductasa, TPPP3 miembro 3 de la familia de proteínas promotoras de la polimerización de tubulina, APAF1 peptidasa apoptótica, BTBD10 BTB (POZ) factor activador 1 que contiene dominio 10, GLUD1 glutamato CXCR4 receptor 4 de la quimioquina (motivo C--X--C) deshidrogenasa 1, SLC1A3 familia transportadora de solutos 1, FLT1 miembro 3 de tirosina quinasa 1 relacionada con fms (transportador de glutamato de gran afinidad de la glía), PON1 paraoxonasa 1, AR receptor de andrógeno, LIF factor inhibidor de leucemia, ERBB3 homólogo 3 eritroblástico del oncogen viral de leucemia v-erb-b2, LGALS1 lectina,galactósido-CD44 moléculas de unión a CD44, soluble, 1, TP53 proteína tumoral p53, TLR3 receptor 3 tipo toll, GRIA1 receptor de glutamato, GAPDH gliceraldehído-3 ionotrópico, AMPA 1 fosfato deshidrogenasa, receptor de glutamato GRIK1, DES desmina ionotrópica, cainato 1, CHAT colina acetiltransferasa, FLT4 tirosina quinasa 4 relacionada con fms, CHMP2B proteína 2B modificadora de la cromatina, asociada a BAG1 atanógeno asociado a BCL2, MT3 metalotioneína 3, CHRNA4 receptor colinérgico, nicotínico, alfa 4 GSS glutatión sintetasa, BAK1 antagonista/asesina BCL2-1, KDR dominio del inserto quinasa, GSTP1 receptor de la glutatión-S-transferasa (una tirosina quinasa de receptor tipo III pi 1), OGG1 8-oxoguanina ADN, IL6 interleuquina 6 (interferón, glicosilasa beta 2).

El animal o la célula puede comprender 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más secuencias cromosómicas interrumpidas que codifican una proteína asociada a la ELA y cero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más secuencias integradas cromosómicamente que codifican la proteína interrumpida asociada a la ELA. Las proteínas asociadas con la ELA preferidas incluyen SOD1 (superóxido-dismutasa 1), ELA2 (esclerosis lateral amiotrófica 2), FUS (fusionado con el sarcoma), TARDBP (proteína de unión a ADN TAR), VAGFA (factor A de crecimiento endotelial vascular), VAGFB (factor B de crecimiento endotelial vascular) y VAGFC (factor C de crecimiento endotelial vascular) y cualquiera de sus combinaciones.

Autismo

La Publicación de Patente de los EE.UU. n.º 20110023145, describe el uso de nucleasas con dedos de cinc para modificar genéticamente células, animales y proteínas asociadas con trastornos del espectro autista (TEA). Los trastornos del espectro autista (TEA) son un grupo de trastornos caracterizados por un deterioro cualitativo en la interacción social y comunicación y patrones estereotipados de comportamiento, intereses y actividades repetitivos y restringidos. Los tres trastornos, autismo, síndrome de Asperger (SA) y trastornos generalizados del desarrollo no especificados de otra manera (PDD-NOS, por sus siglas en inglés) forman una serie continua del mismo trastorno con grados de gravedad variables, asociados con el funcionamiento intelectual y las afecciones médicas. Los TEA son trastornos determinados genéticamente de manera predominante con una heredabilidad de aproximadamente un 90%.

La Publicación de Patente de los EE.UU. n.º 20110023145 comprende editar cualquiera de las secuencias cromosómicas que codifican proteínas asociadas con los TEA que se podría aplicar al sistema CRISPR Cas. Las proteínas asociadas con los TEA se seleccionan normalmente en función de una asociación experimental de la proteína asociada con los TEA con una incidencia o señal de un TEA. Por ejemplo, la velocidad de producción o la concentración de circulación de una proteína asociada con los TEA podrá estar elevada o disminuida en una población que padece un trastorno de tipo TEA respecto a una población que no padece el trastorno de tipo TEA. Se podrán evaluar las diferencias en los niveles de proteína utilizando técnicas proteómicas que incluyen, sin limitación, la inmunoelectrotransferencia, tinción inmunohistoquímica, ensayo de inmunoadsorción enzimática (ELISA) y espectrometría de masas. Como alternativa, se podrán identificar las proteínas asociadas con los ASD obteniendo los perfiles de expresión génica de los genes que codifican las proteínas utilizando técnicas genómicas que incluyen, sin limitación, análisis por microarreglo de ADN, análisis en serie de la expresión génica (SAGE) y reacción en cadena de la polimerasa en tiempo real cuantitativa (Q-PCR).

Algunos ejemplos no limitantes de estados de enfermedad o trastornos que pueden estar asociados a las proteínas asociadas con TEA incluyen: autismo, síndrome de Asperger (AS), trastorno generalizado del desarrollo - no especificado de otra manera (PDD-NOS), síndrome de Rett, esclerosis tuberosa, fenilcetonuria, síndrome de Smith-Lemli-Opitz y síndrome de X frágil. A modo de ejemplo no limitante, las proteínas asociadas con los TEA incluyen, sin limitación, las siguientes proteínas: ATP10C aminofosfolípido- MET MET receptor transportador de ATPasa tirosina cinasa (ATP10C) BZRAP1 MGLUR5 (GRM5) Receptor metabotrópico de glutamato 5 (MGLUR5) CDH10 Caderina-10 MGLUR6 (GRM6) Receptor metabotrópico de glutamato 6 (MGLUR6) CDH9 Caderina-9 NLGN1 Neuroligina-1 CNTN4 Contactina-4 NLGN2 Neuroligina-2 CNTNAP2 Asociada a contactina SEMA5A Neuroligina-3 proteína-tipo 2 (CNTNAP2) DHCR7 7-deshidrocolesterol NLGN4X Neuroligina-4 X- reductasa (DHCR7) unida a DOC2A Dominio tipo C2 doble- NLGN4Y Neuroligina-4 proteína alfa que contiene y unida a DPP6 Dipeptidil NLGN5 Neuroligina-5 proteína tipo aminopeptidasa 6 EN2 engrailed 2 (EN2) NRCAM Molécula de adhesión celular neuronal (NRCAM) MDGA2 retraso mental ligado al cromosoma X frágil NRXN1 Neurexina-1 1 (MDGA2) FMR2 (AFF2) AF4/FMR2 miembro de la familia 2 OR4M2 Receptor olfatorio (AFF2) 4M2 FOXP2 Proteína de la secuencia cabeza de tenedor (forkhead) P2 OR4N4 Receptor olfatorio (FOXP2) 4N4 FXR1 Mental ligado al cromosoma X frágil OXTR receptor de oxitocina retraso, autosómico (OXTR) homólogo 1 (FXR1) FXR2 Frágil X mental PAH fenilalanina retraso, autosómico hidroxilasa (PAH) homólogo 2 (FXR2) GABRA1 Ácido gamma-aminobutírico PTEN Fosfatasa y subunidad del receptor

alfa-1 homólogo de tensina (GABRA1) (PTEN) GABRA5 GABAA (.gamma.-aminobutírico PTPRZ1 Receptor-tipo ácido) receptor alfa 5 tirosina-proteína subunidad (GABRA5) fosfatasa zeta (PTPRZ1) GABRB1 Ácido gamma-aminobutírico RELN Receptor de reelina subunidad beta-1 (GABRB1) GABRB3 GABAA (.gamma.-aminobutírico RPL10 60S ribosomal ácido) receptor.beta.3 subunidad proteína L10 (GABRB3) GABRG1 Ácido gamma-aminobutírico SEMA5A Semaforina-5A receptor subunidad gamma-1 (SEMA5A) (GABRG1) HIRIP3 proteína de interacción con HIRA 3 SEZ6L2 2 similar al homólogo 6 relacionado con las convulsiones (de ratón) HOXA1 Proteína con homeosecuencia Hox-A1 SHANK3 SH3 y múltiples (HOXA1) repeticiones de anquirina en dominios 3 (SHANK3) IL6 Interleuquina-6 SHBZRAP1 SH3 y dominios con múltiples repeticiones de anquirina 3 (SHBZRAP1) LAMB1 Subunidad beta-1 de la laminina SLC6A4 Serotonina (LAMB1) transportador (SERT) MAPK3 Proteína activada por mitógenos TAS2R1 Receptor cinasa del gusto 3 tipo 2 miembro 1 TAS2R1 MAZ Dedo de zinc asociado a Myc TSC1 Proteína de la esclerosis tuberosa proteína 1 MDGA2 MAM dominio que contiene TSC2 Esclerosis tuberosa ancla 2 de proteína 2 de glucosilfosfatidilinositol (MDGA2) MECP2 Unión a metil CpG UBE3A Proteína ubiquitina proteína 2 (MECP2) ligasa E3A (UBE3A) MECP2 unión a metil CpG WNT2 Proteína 2 de tipo sin alas (MECP2) MMTV familia del sitio de integración, miembro 2 (WNT2)

La identidad de la proteína asociada con la MD cuya secuencia cromosómica se edita puede variar y variará. En las realizaciones preferidas, las proteínas asociadas con los TEA cuya secuencia cromosómica se edita podrán ser la proteína 1 asociada con el receptor benzodiazepínico (periférico) (BZRAP1) codificada por el gen BZRAP1, la proteína 2 miembro de la familia AF4/FMR2 (AFF2) codificada por el gen AFF2 (también denominado MFR2), la proteína 1 homóloga autosómica del retraso mental ligado al cromosoma X frágil (FXR1) codificada por el gen FXR1, la proteína 2 homóloga autosómica del retraso mental ligado al cromosoma X frágil (FXR2) codificada por el gen FXR2, la proteína de anclaje 2 de glicosilfosfatidilinositol que contiene el dominio MAM (MDGA2) codificada por el gen MDGA2, la proteína 2 de unión a metil CpG (MECP2) codificada por el gen MECP2, el receptor de glutamato metabotrópico 5 (MGLUR5) codificado por el gen MGLUR5-1 (también denominado GRM5), la proteína neurexina 1 codificada por el gen NRXN1 o la proteína semaforina-5A (SEMA5A) codificada por el gen SEMA5A. En una forma de realización ilustrativa, el animal modificado genéticamente es una rata, y la secuencia cromosómica editada que codifica la proteína asociada con los TEA es como se enuncia a continuación: BZRAP1 proteína 1 asociada XM_213427 con el receptor benzodiazepínico XM_002727789 (periférico) (BZRAP1), XM_002724533, XM_001081125 AFF2 (FMR2) miembro 2 de la familia AF4/FMR2 XM_219832, (AFF2) XM_001054673 FXR1 Retraso mental ligado al cromosoma X frágil NM_001012179, homólogo autosómico 1 (FXR2) FXR2 retraso mental ligado al cromosoma X frágil NM_001100647, homólogo autosómico 2 (FXR2) MDGA2 anclaje 2 de glicosilfosfatidilinositol NM_199269 que contiene el dominio MAM (MDGA2) MECP2 proteína 2 NM_022673 de unión a metil CpG (MECP2) MGLUR5 receptor metabotrópico de glutamato 5 NM_017012 (GRM5) (MGLUR5) NRXN1 Neurexina-1 NM_021767 SEMA5A Semaforina-5A (SEMA5A) NM_001107659

Trastornos por repetición de la expansión de trinucleótidos

La Publicación de Patente de los EE.UU. n.º 20110016540, describe el uso de nucleasas con dedos de cinc para modificar genéticamente células, animales y proteínas asociadas con trastornos por repetición de la expansión de trinucleótidos. Los trastornos por repetición de la expansión de trinucleótidos son trastornos complejos y progresivos que comprometen el desarrollo neurobiológico y a menudo afectan a las funciones cognitivas, así como también a las sensitivomotoras.

Las proteínas debidas a la repetición de la expansión de trinucleótidos son un conjunto variado de proteínas asociadas con susceptibilidad por desarrollar un trastorno por repetición de la expansión de trinucleótidos, la presencia de un trastorno por repetición de la expansión de trinucleótidos, la gravedad de un trastorno por repetición de la expansión de trinucleótidos o cualquier combinación de los mismos. Los trastornos por repetición de la expansión de trinucleótidos se dividen en dos categorías determinadas por el tipo de repetición. La repetición más común es el triplete CAG el cual, cuando está presente en la región codificante de un gen, codifica el aminoácido glutamina (Q). Por lo tanto, estos trastornos se denominan trastornos poliglutamina (poliQ) y comprenden las siguientes enfermedades: enfermedad de Huntington (HD); atrofia muscular espinobulbar (SBMA); ataxias espinocerebelosas (SCA de los tipos 1, 2, 3, 6, 7 y 17); y atrofia dentato-rubro-pálido-luisiana (DRPLA). El resto de los trastornos de extensión por repeticiones de trinucleótidos no implican el triplete CAG o el triplete CAG no está en la región codificante del gen y, por lo tanto, se denominan trastornos no de poliglutamina. Los trastornos no poliglutamina comprenden el síndrome del cromosoma X frágil (FRAXA); retraso mental XE frágil (FRAXE); ataxia de Friedreich (FRDA), distrofia miotónica (DM); y ataxias espinocerebelosa (SCA de los tipos 8 y 12).

Las proteínas asociadas a los trastornos por repetición de la expansión de trinucleótidos típicamente se pueden seleccionar en base a una asociación experimental de la proteína asociada a un trastorno por repetición de la expansión de trinucleótidos. Por ejemplo, la velocidad de producción o la concentración de circulación de una proteína asociada con un trastorno por expansión de repetición de trinucleótidos podrá estar elevada o disminuida en una población que padece un trastorno por expansión de repeticiones de trinucleótidos respecto a una población que no padece el trastorno por expansión de repeticiones de trinucleótidos. Se podrán evaluar las diferencias en los niveles de proteína utilizando técnicas proteómicas que incluyen, sin limitación, la inmunoelectrotransferencia, tinción inmunohistoquímica, ensayo de inmunoadsorción enzimática (ELISA) y espectrometría de masas. Como alternativa, se podrán identificar las proteínas asociadas con trastornos por expansión de repeticiones de trinucleótidos obteniendo los perfiles de expresión génica de los genes que codifican las proteínas utilizando técnicas genómicas que incluyen,

sin carácter limitante, análisis por microarreglo de ADN, análisis en serie de la expresión génica (SAGE) y reacción en cadena de la polimerasa en tiempo real cuantitativa (Q-PCR).

Los ejemplos no limitantes de proteínas asociadas con trastornos por expansión de repeticiones de trinucleótidos incluyen el AR (receptor androgénico), FMR1 (retraso mental 1 asociado al cromosoma X frágil), HTT (huntingtina), DMPK (proteína cinasa de la distrofia miotónica), FXN (frataxina), ATXN2 (ataxina 2), ATN1 (atrofina 1), FEN1 (endonucleasa específica 1 de la estructura de tipo solapa), TNRC6A (6A que contiene repeticiones de trinucleótidos), PABPN1 (proteína nuclear 1 de unión a poli (A)), JPH3 (junctofilina 3), MED15 (subunidad 15 del complejo mediador), ATXN1 (ataxina 1), ATXN3 (ataxina 3), TBP (proteína de unión a la caja TATA), CACNA1A (subunidad alfa 1A, tipo P/Q del canal de calcio dependiente de voltaje), ATXN80S (hebra opuesta a ATXN8 (no codifica proteína)), PPP2R2B (subunidad reguladora B, beta, de la proteína fosfatasa 2), ATXN7 (ataxina 7), TNRC6B (6B que contiene repeticiones de trinucleótidos), TNRC6C (6C que contiene repeticiones de trinucleótidos), CELF3 (miembro 3 de la familia de tipo Elav, CUGBP), MAB21L1 (mab-21-tipo 1 (C. elegans)), MSH2 (homólogo 2 de mutS, cáncer de colon, tipo 1 no poliposis (E. coli)), TMEM185A (proteína transmembrana 185A), SIX5 (homeosecuencia 5 de SIX), CNPY3 (homólogo 3 de canopy (pez cebra)), FRAXE (sitio frágil, tipo ácido fólico, raro, fra(X)(q28) E), GNB2 (proteína (proteína G) de unión al nucleótido guanina, beta polipéptido 2), RPL14 (proteína ribosomal L14), ATXN8 (ataxina 8), INSR (receptor de la insulina), TTR (transtiretina), EP400 (proteína p400 de unión a E1A), GIGYF2 (proteína 2 GYF de interacción con GRB10), OGG1 (ADN-glicosilasa de la 8-oxoguanina), STC1 (estaniocalcina 1), CNDP1 (dipeptidasa 1 de la carnosina (familia M20 de metalopeptidasas)), C10orf2 (marco de lectura abierto 2 del cromosoma 10), MAML3 tipo mastermind 3 (Drosophila), DKC1 (disquerina, disqueratosis congénita 1), PAXIP1 (proteína 1 (con un dominio de activación de la transcripción) que interacciona con PAX), CASK (cinasa de la serínproteina dependiente de calcio/calmodulina (familia MAGUK)), MAPT (proteína tau asociada a microtúbulos), SP1 (factor de transcripción Sp1), POLG (polimerasa (dirigida por ADN), gamma), AFF2 (familia AF4/FMR2, miembro 2), THBS1 (trombospondina 1), TP53 (proteína tumoral p53), ESR1 (receptor de estrógeno 1), CGGBP1 (proteína 1 de unión a repeticiones del triplete CGG), ABT1 (activador de la transcripción basal 1), KLK3 (peptidasa 3 relacionada con la kalikreína), PRNP (proteína priónica), JUN (oncogen jun), KCNN3 (miembro 3, subfamilia N, canal pequeño/intermedio de potasio activado por conductancia de calcio), BAX (proteína X asociada a BCL2), FRAXA (sitio frágil, tipo ácido fólico, raro, fra(X)(q27.3) A (macroorquidismo, retraso mental)), KBTBD10 (10 que contiene el dominio BTB (POZ) y repetición kelch), MBNL1 (tipo muscleblind (Drosophila)), RAD51 (homólogo de RAD51 (homólogo de RecA, E. coli) (S. cerevisiae)), NCOA3 (coactivador 3 del receptor nuclear), ERDA1 (dominio de repetición expandida, CAG/CTG 1), TSC1 (esclerosis tuberosa 1), COMP (proteína matricial oligomérica del cartílago), GCLC (glutamato-cisteína-ligasa, subunidad catalítica), RRAD (relacionada con Ras asociada con la diabetes), MSH3 (homólogo 3 de mutS (E. coli)), DRD2 (receptor de dopamina D2), CD44 (molécula CD44 (grupo sanguíneo de la India)), CTCF (factor de unión a CCCTC (proteína con dedos de zinc)), CCND1 (ciclina D1), CLSPN (homólogo de claspina (Xenopus laevis)), MEF2A (factor 2A potenciador de miocitos), PTPRU (proteína tirosina-fosfatasa, tipo de receptor, U), GAPDH (gliceraldehído-3-fosfato-deshidrogenasa), TRIM22 (22 que contiene el motivo tripartito), WT1 (tumor de Wilms 1), AHR (receptor de arilhidrocarburo), GPX1 (glutatión-peroxidasa 1), TPMT (tiopurina S-metiltransferasa), NDP (enfermedad de Norrie (pseudoglioma)), ARX (homeosecuencia vinculada con un estado sin “arista”), MUS81 (homólogo de endonucleasas MUS81 (S. cerevisiae)), TYR (tirosinasa (albinismo oculocutáneo IA)), EGR1 (respuesta de crecimiento temprano 1), UNG (uracil-ADN-glicosilasa), NUMBL (de tipo homólogo de numb (Drosophila)), FABP2 (proteína de unión a ácidos grasos 2, intestinal), EN2 (homeosecuencia engrailed 2), CRYGC (cristalina, gamma C), SRP14 (partícula de 14 kDa de reconocimiento de la señal (proteína de unión a ARN homólogo Alu)), CRYGB (cristalina, gamma B), PDCD1 (muerte celular programada 1), HOXA1 (homeosecuencia A1), ATXN2L (tipo ataxina 2), PMS2 (de incremento 2 de la segregación posmeiótica PMS2 (S. cerevisiae)), GLA (galactosidasa, alfa), CBL (secuencia transformante retrovírica ecotrópica Cas-Br-M (murina)), FTH1 (ferritina, polipéptido pesado 1), IL12RB2 (receptor de interleuquina 12, beta 2), OTX2 (homeosecuencia 2 de orthodenticle),HOXA5 (homeosecuencia A5), POLG2 (polimerasa (dirigida por ADN), gamma 2, subunidad accesoria), DLX2 (homeosecuencia distal-less 2), SIRPA (proteína alfa de la regulación de señal), OTX1 (homeosecuencia orthodenticle 1), AHRR (represor del receptor de arilhidrocarburos), MANF (factor neurotrófico derivado de astrocitos mesencefálicos), TMEM158 (proteína transmembrana 158 (gen/pseudogen)) y ENSG00000078687.

Las proteínas preferidas asociadas con los trastornos por expansión de repeticiones de trinucleótidos incluyen las HTT (Huntingtina), AR (receptor androgénico), FXN (frataxina), Atxn3 (ataxina), Atxn1 (ataxina), Atxn2 (ataxina), Atxn7 (ataxina), Atxn10 (ataxina), DMPK (proteína cinasa de la distrofia miotónica), Atn1 (atrofina 1), CBP (proteina de unión a CREB), VLDLR (receptor de lipoproteínas de muy baja densidad) y combinaciones de estos.

Tratamiento de enfermedades de la audición

También se describen, pero no se reivindican en la presente, métodos de suministro del sistema CRISPR-Cas a uno o a ambos oídos.

Los investigadores están investigando qué terapia génica se puede utilizar para contribuir a los actuales tratamientos para la sordera - es decir, los implantes cocleares. La sordera está causada a menudo por la pérdida o el deterioro de las células ciliadas que no pueden transmitir señales a las neuronas auditivas. En tales casos, se podrán utilizar los implantes cocleares para que respondan al sonido y transmitan señales eléctricas a las células nerviosas. Pero estas neuronas a menudo se degeneran y se retraen de la cóclea ya que las células ciliadas dañadas liberan menos factores de crecimiento.

La solicitud de patente de los EE.UU. 20120328580 describe la inyección de una composición farmacéutica en el oído (por ejemplo, por administración auricular), como por ejemplo dentro del lumen de la cóclea (por ejemplo, la rampa media, la rampa vestibular, y la rampa timpánica), por ejemplo, usando una jeringa, por ejemplo, una jeringa de una única dosis. Por ejemplo, se pueden administrar uno o más de los compuestos descritos en la presente mediante inyección intratimpánica (por ejemplo, en el oído medio) y/o inyecciones en el oído externo, medio y/o interno. Dichos métodos se utilizan habitualmente en la técnica, por ejemplo, para la administración de esteroides y antibióticos a los oídos humanos. La inyección se puede realizar, por ejemplo, a través de la ventana redonda del oído o mediante la cápsula coclear. En la técnica se conocen otros métodos de administración al oído interno (véase, por ejemplo, Salt y Plontke, Drug Discovery Today, 10:1299-1306, 2005).

En otra modalidad de suministro, la composición farmacéutica se puede administrar *in situ*, mediante un catéter o bomba. Un catéter o una bomba pueden, por ejemplo, dirigir una composición farmacéutica al interior de la luz coclear o la ventana redonda del oído y/o la luz del colon. Los aparatos y métodos de administración de fármacos ejemplares adecuados para administrar uno o más de los compuestos descritos en el presente documento en el oído, por ejemplo, un oído humano, se describen por McKenna y col., (Publicación de Estados Unidos n.º 2006/0030837) y Jacobsen y col., (Patente de los EE.UU. En algunas realizaciones, se puede situar un catéter o una bomba, por ejemplo, en el oído (por ejemplo, el oído externo, medio y/o interno) de un paciente durante un procedimiento quirúrgico. En algunas realizaciones, se puede situar un catéter o una bomba, por ejemplo, en el oído (por ejemplo, el oído externo, medio y/o interno) de un paciente sin que sea necesario un procedimiento quirúrgico.

Como alternativa o además de eso, uno o más de los compuestos que se describen aquí se puede administrar en combinación con un dispositivo mecánico tal como un implante coclear o una ayuda para la audición, que se usa en el oído externo. Edge y col., (publicación de Estados Unidos n.º 2007/0093878) describen un implante coclear ilustrativo que es adecuado para su uso como se prevé en la presente.

En ciertas formas de realización, se pueden combinar las modalidades de administración descritas anteriormente en cualquier orden y se pueden realizar simultáneamente o intercaladas.

Como alternativa o además de eso, las composiciones previstas en la presente se pueden administrar de acuerdo con cualquiera de los métodos aprobados por la Food and Drug Administration, por ejemplo, según se describe en CDER Data Standards Manual, versión número 004 (que se puede obtener en fda.give/cder/dsm/DRG/drg00301.htm).

En general, los métodos de terapia celular descritos en la solicitud de patente de los EE.UU. 20120328580 se pueden utilizar para promover *in vitro* una diferenciación completa o parcial de una célula en o hacia un tipo celular maduro del oído interno (por ejemplo, una célula pilosa). Las células generadas mediante este tipo de métodos se pueden trasplantar o implantar a continuación en un paciente que necesite un tratamiento de este tipo. Los métodos de cultivo celular requeridos para llevar a la práctica estos métodos, incluidos los métodos para identificar y seleccionar tipos celulares adecuados, métodos para promover la diferenciación parcial o completa de células seleccionadas, métodos para identificar tipos celulares parcial o completamente diferenciados y métodos para implantar células parcial o completamente diferenciadas se describen posteriormente.

Las células adecuadas para su uso en los métodos reivindicados de la presente invención incluyen, sin carácter limitante, células que son capaces de diferenciarse parcial o completamente en una célula madura del oído interno, por ejemplo, una célula ciliada (por ejemplo, una célula ciliada interna y/o externa), cuando se pone en contacto, por ejemplo, *in vitro*, con uno o más de los compuestos descritos en la presente. Las células ilustrativas que son capaces de diferenciarse en una célula ciliada incluyen, sin carácter limitante, células madre (por ejemplo, células madre del oído interno, células madre adultas, células madre obtenidas a partir de médula ósea, células madre embrionarias no humanas, células madre mesenquimatosas, células madre de la piel, células iPS y células madre obtenidas a partir de la grasa), células progenitoras (por ejemplo, células progenitoras del oído interno), células de soporte (por ejemplo, células de Deiters, células pilares, células falángicas internas, células tectales y células de Hensen) y/o células germinales. El uso de células indiferenciadas para el reemplazo de células sensoriales del oído interno ha sido descrito en Li y col., (Publicación de los EE.UU. n.º 2005/0287127) y Li y col., (Patente de los EE.UU. n.º de Serie 11/953.797). El uso de células indiferenciadas derivadas de médula ósea para el reemplazo de células sensoriales del oído interno ha sido descrito en Edge y col., PCT/US2007/084654. Las células iPS se han descrito, por ejemplo, en Takahashi y col., Cell, Volumen 131, Número 5, Páginas 861-872 (2007); Takahashi y Yamanaka, Cell 126, 663-76 (2006); Okita y col., Nature 448, 260-262 (2007); Yu, J. y col., Science 318(5858):1917-1920 (2007); Nakagawa y col., Nat. Biotechnol. 26:101-106 (2008); y Zaehres y Scholer, Cell 131(5):834-835 (2007). Dichas células adecuadas pueden identificarse mediante análisis (por ejemplo, cualitativo o cuantitativo) de la presencia de uno o más genes específicos de tejido. Por ejemplo, se puede detectar la expresión génica detectando el producto proteico de uno o más genes específicos del tejido. Las técnicas de detección proteica conllevan la tinción de proteínas (por ejemplo, utilizando extractos celulares o células intactas) utilizando anticuerpos contra el antígeno apropiado. En este caso, el antígeno apropiado es el producto proteico de la expresión del gen específico del tejido. Aunque, en teoría, se puede marcar un primer anticuerpo (es decir, el anticuerpo que se une al antígeno) es más común (y mejora la visualización) utilizar un segundo anticuerpo dirigido contra el primero (por ejemplo, un anticuerpo anti-IgG). Este segundo anticuerpo está conjugado con fluorocromos, o enzimas apropiadas para reacciones colorimétricas, o perlitas de oro (para la microscopía electrónica) o con el sistema biotina-avidina, de modo que se pueda reconocer la ubicación del anticuerpo primario y, por lo tanto, del antígeno.

Las moléculas CRISPR Cas se pueden suministrar al oído por aplicación directa de la composición farmacéutica al oído externo, con composiciones modificadas, basadas en la Solicitud Publicada de los EE.UU. 20110142917. En ciertas formas de realización la composición farmacéutica se aplica al conducto auditivo. El suministro al oído también se puede denominar como suministro aural u ótico.

En ciertas formas de realización descritas en la presente, las moléculas de ARN se suministran en formulaciones de liposomas o lipofectina y otras similares y se pueden preparar por métodos bien conocidos por aquellos con experiencia en el arte. Dichos métodos se han descrito, por ejemplo, en las Patentes de los EE.UU. n.º 5.593.972, 5.589.466 y 5.580.859.

Se han desarrollado sistemas de suministro orientados específicamente al suministro potenciado y mejorado de ARNip en células de mamífero (remítase a, por ejemplo, Shen et al. FEBS Let. 2003, 539:111-114; Xia y col., Nat. Biotech. 2002, 20:1006-1010; Reich y col., Mol. Vision. 210-216; Sorensen y col., J. Mol. Biol. 2003, 327: 761-766; Lewis y col., Nat. Gen. 2002, 32: 107-108 y Simeoni y col., NAR 2003, 31, 11: 2717-2724) y se podrán aplicar. Recientemente se ha utilizado con éxito ARNip para la inhibición de la expresión génica en primates (remítase a, por ejemplo, Tolentino y col., Retina 24(4):660 que también se podrá aplicar.

Qi et al. divulgan métodos para una transfección eficaz de ARNip al oído interno a través de la ventana redonda intacta mediante una tecnología de suministro proteínico novedosa que se podrá aplicar al sistema CRISPR Cas descrito en la presente (véase, por ejemplo, Qi y col., Gene Therapy (2013), 1-9). En particular, una TAT de unión a dominios de ARN bicatenario (TAT-DRBD), con la que se puede transfectar ARNip marcado con Cy3 en células del oído interno, incluidas las células ciliadas internas y externas, cresta ampular, mácula utricular y mácula sacular, mediante permeación en la ventana redonda intacta tuvo éxito para suministrar ARNip bicatenarios *in vivo* para tratar varias dolencias del oído interno y preservar la función auditiva. Se podrán contemplar aproximadamente 40 µL de ARN 10 mM como la dosificación para la administración al oído.

De acuerdo con Rejali y col. (Hear Res. junio de 2007;228(1-2):180-7), se puede mejorar la función del implante coclear mediante una buena conservación de las neuronas del ganglio espiral que son el diana de la estimulación eléctrica del implante y se demostró previamente que el factor neurotrófico derivado del cerebro (BDNF, por sus siglas en inglés) potencia la supervivencia del ganglio espiral en oídos que se habían vuelto sordos experimentalmente. Rejali et al. probaron un diseño modificado del electrodo del implante coclear que incluye un recubrimiento de células de fibroblastos transducidos mediante un vector vírico con un inserto del gen de BDNF. Para lograr este tipo de transferencia génica *ex vivo*, Rejali et al. transdujeron fibroblastos de cobayas con un adenovirus con un inserto del casete del gen de BDNF y comprobaron que estas células secretaban BDNF, a continuación unieron las células secretoras de BDNF al electrodo del implante coclear mediante un gel de agarosa e implantaron el electrodo en la escala timpánica. Rejali et al. comprobaron que los electrodos que expresaban BDNF eran capaces de conservar significativamente más neuronas del ganglio espiral en la base de la cóclea 48 días después de la implantación cuando se comparó con electrodos de control y se demostró la viabilidad de combinar la terapia de implante coclear con transferencia génica *ex vivo* para potenciar la supervivencia de neuronas del ganglio espiral. Un sistema con dichas características se puede aplicar al sistema de direccionamiento hacia ácidos nucleicos descrito en la presente para suministrar al oído.

Mukherjea y col. (Antioxidants & Redox Signaling, Volumen 13, Número 5, 2010) dejan constancia de que la atenuación génica de NOX3 utilizando ARN de interferencia pequeño (ip) anuló la ototoxicidad del cisplatino, como lo evidencia la protección de OHC frente al daño y la reducción en las variaciones en el umbral en las respuestas auditorias del tronco encefálico (ABR, por sus siglas en inglés). Se administraron diferentes dosis de siNOX3 (0,3, 0,6, y 0,9 µg) a ratas y se evaluó la expresión de NOX3 por RT-PCR en tiempo real. La dosis más baja de ARNip de NOX3 que se utilizó (0,3 µg) no mostró ninguna inhibición del ARNm de NOX3 cuando se comparó con suministro transtimpánico de ARNip entremezclado o cócleas sin tratar. Sin embargo, el suministro de las mayores dosis de ARNip de NOX3 (0,6 y 0,9 µg) redujo la expresión de NOX3 en comparación con el control de ARNip entremezclado. Se podrá aplicar un sistema de este tipo al sistema CRISPR Cas descrito en la presente para la administración transtimpánica con una dosificación de aproximadamente 2 mg a aproximadamente 4 mg de CRISPR Cas para la administración a un ser humano.

Jung y col. (Molecular Therapy, vol. 21 n.º 4, 834-841 abril de 2013) demostraron que niveles de Hes5 en el utrículo disminuían después de la aplicación de ARNip y que el número de células ciliadas en estos utrículos era significativamente mayor que tras el tratamiento de control. Los datos sugieren que la tecnología de ARNpi podrá ser útil para inducir la reparación y regeneración en el oído interno y que la ruta de señalización de Notch es un diana potencialmente útil para una inhibición de la expresión génica específica. Jung et al. inyectaron 8 µg de ARNpi contra Hes5 en un volumen de 2 µL, preparado añadiendo solución salina normal estéril a ARNpi liofilizado, al epitelio vestibular del oído. Un sistema con dichas características se puede aplicar al sistema de direccionamiento hacia ácidos nucleicos descrito en la presente para suministrar al epitelio vestibular del oído con una dosificación de entre aproximadamente 1 y aproximadamente 30 mg de CRISPR Cas para suministrar a un ser humano.

Direccionamiento génico en células que no se dividen (neuronas y músculo)

Los tipos de células que no se dividen (especialmente las completamente diferenciadas que no se dividen) presentan problemas para el direccionamiento génico o la modificación del genoma, por ejemplo, porque generalmente la

recombinación homóloga (HR) se suprime en la fase G1 del ciclo celular. Sin embargo, aunque se estudian los mecanismos por los cuales las células controlan los sistemas de reparación del ADN normal, Durocher descubrió un interruptor previamente desconocido que mantiene a HR "apagado" en células que no se dividen y diseñaron una estrategia para volver a encender este interruptor. Orthwein y col. (Laboratorio de Daniel Durocher en el Hospital Monte Sinai en Ottawa, Canadá) recientemente informaron (Nature 16142, publicaron en línea 9 de diciembre de 2015) han mostrado que se puede hacer cesar la supresión de HR y el direccionamiento génico concluyó de manera exitosa en las células tanto de riñón (293T) como de osteosarcoma (U2OS). Se sabe que los supresores tumorales, BRCA1, PALB2 y BRAC2 promueven la reparación de DSB de ADN por HR. Ellos descubrieron que la formación de un complejo de BRCA1 con PALB2 - BRAC2 está regido por un sitio de ubicuitina en PALB2, de manera tal que la acción en el sitio por una E3 ubicuitina ligasa. Esta ubicuitina ligasa E3 está compuesta por KEAP1 (una proteína que interacciona con PALB2) formando un complejo con culina-3 (CUL3)-RBX1. La ubicuitinación de PALB2 suprime su interacción con BRCA1 y es contrarrestada por la desubicuitinasa USP11, que se encuentra en sí bajo el control del ciclo celular. El restablecimiento de la interacción BRCA1-PALB2 combinada con la activación de la resección de los extremos del ADN es suficiente para inducir la recombinación homóloga en G1, medida por varios métodos, incluyendo un ensayo de direccionamiento a genes basado en CRISPR-Cas9 dirigido a USP11 o KEAP1 (expresados a partir de un vector pX459). Sin embargo, cuando se restableció la interacción BRCA1-PALB2 en células G1 competentes para la resección usando ya sea eliminación de KEAP1 o expresión del mutante PALB2-KR, se detectó un robusto incremento de los eventos de direccionamiento hacia genes.

Por lo tanto, en ciertas formas de realización se prefiere la reactivación de HR en células, especialmente tipos de células que no se dividen, completamente diferenciadas. En ciertas formas de realización, en ciertas formas de realización se prefiere la promoción de la interacción BRCA1-PALB2. En ciertas formas de realización, no reivindicada en la presente, la célula diana es una célula que no se divide. En ciertas formas de realización no reivindicadas en la presente, la célula diana es una neurona o célula muscular. En ciertas formas de realización, la célula diana es un diana *in vivo*. En ciertas formas de realización, la célula está en G1 y HR se suprime. En ciertas formas de realización, el uso de la eliminación de KEAP1, por ejemplo, se prefiere la inhibición de la expresión de la actividad de KEAP1. La eliminación de KEAP1 se puede conseguir por ARNip, por ejemplo, como se muestra en Orthwein y col.. Como alternativa, la expresión del mutante PALB2-KR (se prefiere la carencia de los ocho residuos de Lys en el dominio de interacción BRCA1, ya sea en combinación con la eliminación de KEAP1 o solo. PALB2-KR interactúa con BRCA1 independientemente de la posición en el ciclo celular. Por lo tanto, en ciertas formas de realización se prefiere la promoción o el restablecimiento de la interacción BRCA1-PALB2, especialmente en células G1, especialmente donde las células diana no se dividen, o donde la eliminación y la restitución (direccionamiento génico *ex vivo*) es problemática, por ejemplo, neuronas o células musculares. El ARNip de KEAP1 se puede obtener de ThermoFischer. En ciertas formas de realización, se puede suministrar un complejo BRCA1-PALB2 a la célula G1. En ciertas formas de realización, la desubicuitinación de PALB2 puede ser promovida, por ejemplo, por el incremento de la expresión de la desubicuitinasa USP11, de manera que se prevé que se puede proveer una construcción para promover o incrementar la expresión o la actividad de la desubicuitinasa USP11.

Tratamiento de enfermedades del ojo

También se describe, pero no se reivindica en la presente, el suministro del sistema CRISPR-Cas a uno o ambos ojos.

En formas de realización particulares descritas en la presente, se podrá utilizar el sistema CRISPR-Cas para corregir defectos oculares que puedan surgir a partir de varias mutaciones genéticas descritas en más detalle en Genetic Diseases of the Eye, segunda edición, editado por Elias I. Traboulsi, Oxford University Press, 2012.

Para la administración al ojo, se prefieren especialmente los vectores lentivíricos, en particular los virus de la anemia infecciosa equina (EIAV, por sus siglas en inglés).

En otra forma de realización, también se contemplan vectores lentivíricos mínimos que no son de primates basados en el virus de la anemia infecciosa equina (EIAV, por sus siglas en inglés), especialmente para la terapia génica ocular (remítase a, por ejemplo, Balagaan, J Gene Med 2006; 8: 275 - 285, publicado electrónicamente el 21 de noviembre de 2005 en Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845). Se contempla que los vectores tengan un promotor de citomegalovirus (CMV) que impulse la expresión del gen diana. Se contemplan las inyecciones intracamerales, subretinales, intraoculares e intravítreas (véase, por ejemplo, Balagaan, J Gene Med 2006; 8: 275 - 285, publicado electrónicamente el 21 de noviembre de 2005 en Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845). Las inyecciones intraoculares se podrán realizar con la ayuda de un microscopio quirúrgico. Para las inyecciones subretinales e intravítreas, los ojos se podrán prolapsar mediante una compresión digital leve y visualizar el fondo del ojo utilizando un sistema de lente de contacto constituido por una gota de una solución de medio de acoplamiento en la córnea cubierta con una lámina de vidrio cubreobjetos para microscopio. Para las inyecciones subretinales, se podrá desplazar hacia adelante la punta de una aguja del calibre 34 de 10 mm, montada sobre una jeringa Hamilton de 5 µL, con visualización directa a través de la esclerótica ecuatorial superior tangencialmente hacia el polo posterior hasta que la apertura de la aguja sea visible en el espacio subretinal. A continuación, se podrán inyectar 2 µL de la suspensión con el vector para producir un desprendimiento de retina bulloso superior, y se confirma de esta manera la administración subretinal del vector. Esta estrategia crea una esclerotomía autosellante que permite retener la suspensión del vector en el espacio subretinal hasta que sea absorbida por el RPE, normalmente en las 48 h posteriores al procedimiento. Se podrá repetir este procedimiento en el hemisferio inferior para producir un

desprendimiento de retina inferior. Esta técnica conlleva la exposición de aproximadamente un 70% de la retina neurosensorial y RPE a la suspensión del vector. Para las inyecciones intravítreas, se podrá desplazar hacia adelante la punta de la aguja a través de la esclerótica 1 mm detrás del limbo corneoescleral e inyectar 2 µL de la suspensión con el vector al interior de la cavidad vítrea. Para las inyecciones intracamerales, se podrá desplazar hacia adelante la punta de la aguja mediante una paracentesis del limbo corneoescleral, dirigir hacia la córnea central y se podrán inyectar 2 µL de la suspensión con el vector. Para las inyecciones intracamerales, se podrá desplazar hacia adelante la punta de la aguja mediante una paracentesis del limbo corneoescleral, dirigir hacia la córnea central y se podrán inyectar 2 µL de la suspensión con el vector. Dichos vectores se pueden inyectar en títulos de ya sea 1,0-1,4 × 10^{10} o 1,0-1,4 × 10^{9} unidades de transducción (TU)/ml.

En otra forma de realización, también se contempla RetinoStat®, un vector de terapia génica lentivírica basado en el virus de la anemia infecciosa equina que expresa las proteínas angiostáticas endostatina y angiostatina que se administra mediante una inyección subretinal para el tratamiento de la forma húmeda de la degeneración macular asociada a la edad (véase, por ejemplo, Binley y col., HUMAN GENE THERAPY 23:980–991 (septiembre de 2012)). Se podrá modificar un vector de este tipo para el sistema CRISPR-Cas. Cada ojo se puede tratar ya sea con RetinoStat® en una dosis de 1,1 × 10^{5} unidades de transducción por ojo (TU/ojo) en un volumen total de 100 µl.

En otra forma de realización podría contemplarse un vector adenovírico con una eliminación de E1, parcial de E3 y eliminación de E4 para su suministro al ojo. Se administró una inyección intravítrea única de un vector adenovírico con una supresión de E1, supresión parcial de E3 y supresión de E4 que expresaba un factor derivado del epitelio pigmentario humano (AdPEDF.II) a veintiocho pacientes con degeneración macular asociada a la edad (AMD, por sus siglas en inglés) neovascular avanzada (véase, por ejemplo, Campochiaro y col., Human Gene Therapy 17:167-176 (febrero de 2006)). Se investigaron dosis dentro del rango entre 10^{6} y $10^{9,5}$ unidades de partículas (PU) y no hubo eventos adversos graves relacionados con AdPEDF.II ni tampoco toxicidades que limitasen la dosis (véase, por ejemplo, Campochiaro y col., Human Gene Therapy 17:167-176 (febrero 2006)). La transferencia ocular de genes mediada por vector adenoviral parece ser un enfoque viable para el tratamiento de trastornos oculares y se puede aplicar al sistema CRISPR Cas.

En otra forma de realización, se puede utilizar el sistema sd-rxRNA® de RXi Pharmaceuticals y/o adaptar para el suministro de CRISPR Cas al ojo. En este sistema, una única administración intravítrea de 3 µg de sd-rxRNA da como resultado reducción de los niveles de ARNm de PPIB específica para la secuencia durante 14 días. El sistema sd-rxRNA® se puede aplicar al sistema de direccionamiento hacia ácidos nucleicos, contemplándose la administración a un ser humano de una dosis de entre aproximadamente 3 y 20 mg de CRISPR.

Millington-Ward y col. (Molecular Therapy, vol. 19 no. 4, 642-649, abril 2011) describe vectores de virus adeno-asociado (AAV) para administrar un supresor de rodopsina basado en ARN de interferencia (ARNi) y un gen de reemplazo de rodopsina modificado por codones resistente a la supresión debida a alteraciones de nucleótidos en posiciones degeneradas sobre el sitio diana del ARNi. Millington-Ward y col. inyectaron subretinalmente una inyección de ya sea 6,0 × 10^{8} vp o 1,8 × 10^{10} vp AAV en los ojos. Los vectores AAV de Millington-Ward y col. se pueden aplicar al sistema CRISPR Cas, contemplando la administración a un ser humano de una dosis de entre aproximadamente 2 × 10^{11} y aproximadamente 6 × 10^{13} vp.

Dalkara y col. (Sci Transl Med 5, 189ra76 (2013)) también se refiere a una evolución dirigida *in vivo* para obtener un vector AAV que repare las versiones de tipo salvaje de genes defectuosos a través de toda la retina después de la inyección no perjudicial en el humor vítreo de los ojos. Dalkara describe a una biblioteca que expone péptidos de 7 aminoácidos y una biblioteca de AAV construida por entremezclado de ADN de genes cap de AAV1, 2, 4, 5, 6, 8, y 9. Las bibliotecas de vectores rcAAV y rAAV que expresaban GFP bajo el control de un promotor CAG o Rho se empaquetaron y se obtuvieron títulos genómicos resistentes a la desoxirribonucleasa por PCR cuantitativa. Se combinaron las colecciones y se realizaron dos ciclos de evolución, en los que cada uno consistía de una diversificación de la colección inicial seguida por tres pasos de selección *in vivo*. En cada uno de dichos pasos, se inyectaron por vía intravítrea ratones P30 rho-GFP con 2 ml de biblioteca purificada con iodixanol, dializada en solución salina amortiguadora de pH al fosfato (PBS) con a título genómico de aproximadamente 1 × 10^{12} vg/ml. Los vectores AAV de Dalkara y col. se pueden aplicar al sistema de direccionamiento hacia ácidos nucleicos, contemplando la administración a seres humanos de dosis de entre aproximadamente 1 × 10^{15} y aproximadamente 1 × 10^{16} vg/ml.

En otra forma de realización, el diana podrá ser el gen de la rodopsina para el tratamiento de la retinitis pigmentosa (RP, por sus siglas en inglés), donde el sistema de la Publicación de Patente de los EE.UU. n.º 20120204282 concedida a Sangamo BioSciences, Inc. Se podrá modificar para el sistema CRISPR Cas.

En otra forma de realización, los métodos de la Publicación de Patente de los EE.UU. n.º 20130183282 asignada a Cellectis, que se refiere a métodos para clivar una secuencia diana del gen de rodopsina humano, también se pueden modificar para usar con el sistema de direccionamiento hacia ácidos nucleicos.

La Publicación de Patente de los EE.UU. n.º 20130202678 concedida a Academia Sinica se refiere a métodos para tratar retinopatías y trastornos oftalmológicos que afectan a la visión que se refieren al suministro del gen Puf-A (que se expresa en el ganglio retinal y células pigmentadas de los tejidos oculares y presenta una actividad antiapoptósica única) en el espacio subretinal o intravitreal del ojo. Algunos blancos particularmente deseables son: zgc:193933, prdm1a, spata2, tex10, rbb4, ddx3, zp2.2, Blimp-1 y HtrA2, todos los cuales pueden ser blancos para el sistema de

direccionamiento hacia ácidos nucleicos.

Wu (Cell Stem Cell, 13:659–62, 2013) diseñó un ARN guía que condujo a Cas9 a una mutación de un único par de bases que provoca cataratas en ratones, donde induce el clivaje del ADN. A continuación, utilizando el otro alelo no modificado u oligos que se proporcionan a los mecanismos de reparación de zigotos, corrigieron la secuencia del alelo roto y corrigieron el defecto genético que causaba las cataratas en el ratón mutante.

La Publicación de Patente de los EE.UU. n.º 20120159653 describe la utilización de nucleasas con dedos de cinc para modificar genéticamente células, animales y proteínas asociados con la degeneración macular (MD). La degeneración macular (MD) es la causa principal de deficiencia visual en la población de edad avanzada, pero también es un síntoma distintivo de enfermedades de la infancia tales como la enfermedad de Stargardt, fondo de Sorsby y enfermedades neurodegenerativas de la infancia letales, con una edad de inicio muy temprana tal como el primer año de vida. La degeneración macular da como resultado la pérdida de visión del centro del campo visual (la mácula) debido al daño en la retina. En la actualidad, los modelos en animales existentes no manifiestan los distintivos principales de la enfermedad tal como se observa en seres humanos. Los modelos en animales disponibles que comprenden genes mutantes que codifican proteínas asociadas con la MD también producen fenotipos sumamente variables, lo que convierte en problemática su aplicación a la enfermedad humana y al desarrollo de una terapia.

Un aspecto de Publicación de Patente de los EE.UU. n.º 20120159653 se refiere a la edición de cualquier secuencia cromosómica que codifique proteínas asociadas con MD que se pueda aplicar al sistema de direccionamiento hacia ácidos nucleicos. Las proteínas asociadas con la MD se seleccionan normalmente en función de una asociación experimental de la proteína asociada con la MD con un trastorno de tipo MD. Por ejemplo, la velocidad de producción o la concentración de circulación de una proteína asociada con la MD podrá estar elevada o disminuida en una población que padece un trastorno de tipo MD respecto a una población que no padece el trastorno de tipo MD. Se podrán evaluar las diferencias en los niveles de proteína utilizando técnicas proteómicas que incluyen, sin limitación, la inmunoelectrotransferencia, tinción inmunohistoquímica, ensayo de inmunoadsorción enzimática (ELISA) y espectrometría de masas. Como alternativa, se podrán identificar las proteínas asociadas con la MD obteniendo los perfiles de expresión génica de los genes que codifican las proteínas utilizando técnicas genómicas que incluyen, sin limitación, análisis por microarreglo de ADN, análisis en serie de la expresión génica (SAGE) y reacción en cadena de la polimerasa en tiempo real cuantitativa (Q-PCR).

A modo de ejemplo no limitante, las proteínas asociadas con la MD incluyen, pero sin limitación, las siguientes proteínas: (ABCA4) casete de unión a ATP, subfamilia A (ABC1), miembro 4 de acromatopsia 1 ACHM1 (monocromatismo de los bastones) ApoE Apolipoproteína E (ApoE) C1QTNF5 (CTRP5) C1q y proteína 5 relacionada con el factor de necrosis tumoral (C1QTNF5) Componente 2 del complemento C2 (C2) Componentes del complemento C3 (C3) Ligando 2 de quimiocina CCL2 (motivo C-C) (CCL2) Receptor 2 de quimiocina (motivo C-C) CCR2 (CCR2) CD36 Agrupación de diferenciación 36 CFB Factor B del complemento CFH Factor H del complemento CFH CFHR1 1 relacionado con el factor H del complemento CFHR3 3 relacionado con el factor H del complemento CNGB3 subunidad beta 3 del canal operado por nucleótidos cíclicos CP ceruloplasmina (CP) CRP proteína reactiva C (CRP) CST3 cistatina C o cistatina 3 (CST3) CTSD Catepsina D (CTSD) CX3CR1 receptor 1 de la quimiocina (motivo C-X3-C) ELOVL4 Elongación de ácidos grasos de cadena muy larga 4 ERCC6 reparación de clivaje con complementación cruzada con la deficiencia de reparación en roedores, grupo de complementación 6 FBLN5 Fibulina-5 FBLN5 Fibulina 5 FBLN6 Fibulina 6 FSCN2 fascina (FSCN2) HMCN1 Hemicentrina 1 HMCN1 hemicentina 1 HTRA1 serínpeptidasa 1 HtrA (HTRA1) HTRA1 serínpeptidasa 1 HtrA IL-6 Interleuquina 6 IL-8 Interleuquina 8 LOC387715 Proteína hipotética PLEKHA1 miembro 1 de la familia A que contiene un dominio de homología con la pleckstrina (PLEKHA1) PROM1 Prominina 1(PROM1 o CD133) PRPH2 Periferina-2 RPGR regulador GTPasa de la retinitis pigmentosa SERPING1 inhibidor de la serpínpeptidasa, clado G, miembro 1 (inhibidor C1) TCOF1 Treacle TIMP3 Inhibidor 3 de metaloproteinasa (TIMP3) TLR3 Receptor 3 de tipo Toll.

La identidad de la proteína asociada con la MD cuya secuencia cromosómica se edita puede variar y variará. En realizaciones preferidas, las proteínas asociadas con la MD cuya secuencia cromosómica se edita podrán ser el casete de unión a ATP, proteína miembro 4 de la subfamilia A (ABC1)(ABCA4) codificada por el gen ABCR, la proteína apolipoproteína E (APOE) codificada por el gen APOE, la proteína ligando 2 de la quimiocina (motivo C-C) (CCL2) codificada por el gen CCL2, la proteína receptor 2 de la quimiocina (motivo C-C) (CCR2) codificada por el gen CCR2, la proteína ceruloplasmina (CP) codificada por el gen CP, la proteína catepsina D (CTSD) codificada por el gen CTSD o la proteína inhibidor 3 de metaloproteinasas (TIMP3) codificada por el gen TIMP3. En una forma de realización ilustrativa, el animal modificado genéticamente es una rata y la secuencia cromosómica editada que codifica la proteína asociada con la MD podrá ser: (ABCA4) casete de unión a ATP, NM_000350 subfamilia A (ABC1), miembro 4 APOE Apolipoproteína E NM_138828 (APOE) CCL2 Ligando 2 de quimiocina (motivo C-C NM_031530) (CCL2) CCR2 Receptor 2 de quimiocina (motivo C-C NM_021866) (CCR2) CP ceruloplasmina (CP) NM_012532 CTSD Catepsina D (CTSD) NM_134334 TIMP3 Inhibidor 3 de metaloproteinasas NM_012886 (TIMP3). El animal o célula podrá comprender 1, 2, 3, 4, 5, 6, 7 o más secuencias cromosómicas alteradas que codifican una proteína asociada con la MD y cero, 1, 2, 3, 4, 5, 6, 7 o más secuencias integradas cromosómicamente que codifican las proteínas alteradas asociadas con la MD.

La secuencia cromosómica editada o integrada se puede modificar para que codifique una proteína alterada asociada a MD. Varias mutaciones en secuencias cromosómicas relacionadas con la MD se han asociado con MD. Algunos

ejemplos no limitativos de mutaciones en secuencias cromosómicas asociadas a MD incluye a aquellos que pueden causar MD incluyendo a aquellas en la proteína ABCR, E471K (es decir que el glutamato en la posición 471 se cambia por lisina), R1129L (es decir que la arginina en la posición 1129 se cambia por leucina), T1428M (es decir que la treonina en la posición 1428 se cambia por metionina), R1517S (es decir que la arginina en la posición 1517 se cambia por serina), I1562T (es decir que la isoleucina en la posición 1562 se cambia por treonina), y G1578R (es decir que la glicina en la posición 1578 se cambia por arginina); en la proteína CCR2, V64I (es decir que la valina en la posición 192 se cambia por isoleucina); en la proteína CP, G969B (es decir que la glicina en la posición 969 se cambia por asparagina o aspartato); en proteína TIMP3, S156C (es decir que la serina en la posición 156 se cambia por cisteína), G166C (es decir que la glicina en la posición 166 se cambia por cisteína), G167C (es decir que la glicina en la posición 167 se cambia por cisteína), Y168C (es decir que la tirosina en la posición 168 se cambia por cisteína), S170C (es decir que la serina en la posición 170 se cambia por cisteína), Y172C (es decir que la tirosina en la posición 172 se cambia por cisteína) y S181C (es decir que la serina en la posición 181 se cambia por cisteína). En el arte existe constancia de otras asociaciones de variantes genéticas en genes asociados con la MD con enfermedades.

Los sistemas CRISPR son útiles para corregir enfermedades que son el resultado de genes dominantes autosómicos. Por ejemplo, CRISPR/Cas9 se utilizó para eliminar un gen dominante autosómico que causa la pérdida de receptores en el ojo. Bakondi, B. y col., *In vivo CRISPR/Cas9 Gene Editing Corrects Retinal Dystrophy in the S334ter-3 Rat Model of Autosomal Dominant Retinitis Pigmentosa. Molecular Therapy*, 2015; DOI: 10.1038/mt.2015.220.

Tratamiento de enfermedades circulatorias y musculares

También se describen, pero no se reivindican en la presente, métodos de suministro del sistema CRISPR-Cas descrito aquí, por ejemplo sistemas de proteína efectora Cpf1, al corazón. Para el corazón, se prefiere el virus adenoasociado miocardiotrópico (AAVM, por sus siglas en inglés), en particular AAVM41 que mostró una transferencia génica preferencial en el corazón (remítase a, por ejemplo, Lin-Yanga y col., PNAS, 10 de marzo de 2009, vol. 106, n.º 10). La administración podrá ser local o sistémica. Para el suministro sistémico se contempla una dosificación de aproximadamente 1-10 × 10^{14} genomas del vector. Véase también, por ejemplo, Eulalio y col. (2012) Nature 492: 376 y Somasuntharam y col. (2013) Biomaterials 34: 7790.

Por ejemplo, la Publicación de Patente de los EE.UU. n.º 20110023139 describe el uso de nucleasas de dedo de cinc para modificar genéticamente células, animales y proteínas asociadas con la enfermedad cardiovascular. Las enfermedades cardiovasculares incluyen por lo general presión sanguínea elevada, ataques cardíacos, insuficiencia cardíaca y accidentes cerebrovasculares y TIA. Se podrá utilizar cualquier secuencia cromosómica implicada en la enfermedad cardiovascular o la proteína codificada por cualquier secuencia cromosómica implicada en la enfermedad cardiovascular en los métodos descritos en esta divulgación. Las proteínas relacionadas con el sistema cardiovascular se seleccionan normalmente en función de una asociación experimental del desarrollo de la enfermedad cardiovascular con la proteína relacionada con el sistema cardiovascular. Por ejemplo, la velocidad de producción o la concentración de circulación de una proteína asociada con el sistema cardiovascular podrá estar elevada o disminuida en una población que padece un trastorno cardiovascular respecto a una población que no padece trastorno cardiovascular. Se podrán evaluar las diferencias en los niveles de proteína utilizando técnicas proteómicas que incluyen, sin limitación, la inmunoelectrotransferencia, tinción inmunohistoquímica, ensayo de inmunoadsorción enzimática (ELISA) y espectrometría de masas. Como alternativa, se podrán identificar las proteínas asociadas con el sistema cardiovascular obteniendo los perfiles de expresión génica de los genes que codifican las proteínas utilizando técnicas genómicas que incluyen, sin limitación, análisis por microarreglo de ADN, análisis en serie de la expresión génica (SAGE) y reacción en cadena de la polimerasa en tiempo real cuantitativa (Q-PCR).

A manera de ejemplo, la secuencia cromosómica puede comprender, pero de manera no taxativa, IL1B (interleuquina 1, beta), XDH (xantina deshidrogenasa), TP53 (proteína tumoral p53), PTGIS (prostaglandina 12 (prostaciclina) sintasa), MB (mioglobina), IL4 (interleuquina 4), ANGPT1 (angiopoyetina 1), ABCG8 (cassette de unión a ATP, sub-familia G (WHITE), miembro 8), CTSK (catepsina K), PTGIR (receptor de prostaglandina 12 (prostaciclina) (IP)), KCNJ11 (canal rectificador de entrada de potasio, subfamilia J, miembro 11), INS (insulina), CRP (proteína C reactiva, relacionada con pentraxina), PDGFRB (receptor del factor de crecimiento derivado de plaquetas, polipéptido beta), CCNA2 (ciclina A2), PDGFB (factor de crecimiento derivado de plaquetas, polipéptido beta (oncogen homólogo de sarcoma viral de simio (v-sis))), KCNJ5 (canal rectificador de entrada de potasio, subfamilia J, miembro 5), KCNN3 (canal de potasio de intermedia/baja activado por conductancia calcio, subfamilia N, miembro 3), CAPN10 (calpaína 10), PTGES (prostaglandina E sintasa), ADRA2B (receptor alfa-2B-adrenérgico), ABCG5 (cassette de unión a ATP, sub-familia G (WHITE), miembro 5), PRDX2 (peroxirredoxina 2), CAPN5 (calpaína 5), PARP14 (poli (ADP-ribosa) familia de polimerasa, miembro 14), MEX3C (homólogo C de mex-3 *C. elegans*)), ACE enzima convertidora de angiotensina I (peptidil-dipeptidasa A) 1), TNF (factor de necrosis tumoral (superfamilia TNF, miembro 2)), IL6 (interleuquina 6 (interferón, beta 2)), STN (estatina), SERPINE1 (inhibidor de serpina peptidasa, clado E (nexina, inhibidor del activador de plasminógeno tipo 1), miembro 1), ALB (albúmina), ADIPOQ (adiponectina, dominio que contiene C1Q y colágeno), APOB (apolipoproteína B (incluyendo al antígeno Ag(x))), APOE (apolipoproteína E), LEP (leptina), MTHFR (5,10-metilentetrahidrofolato reductasa (NADPH)), APOA1 (apolipoproteína A-I), EDN1 (endotelina 1), NPPB (precursor B del péptido natriurético), NOS3 (óxido nítrico sintasa 3 (célula endotelial)), PPARG (receptor gamma activado por proliferador de peroxisomas), PLAT (activador tisular del plasminógeno), PTGS2 (prostaglandina-endoperóxido sintasa 2 (prostaglandina G/H sintasa y ciclooxigenasa)), CETP (proteína de transferencia del éster de colesterilo, plasma), AGTR1 (receptor de angiotensina II, tipo 1), HMGCR (3-hidroxi-3-metilglutaril-Coenzima A

reductasa), IGF1 (factor de crecimiento 1 similar a insulina (somatomedina C)), SELE (selectina E), REN (renina), PPARA (receptor alfa activado por proliferador de peroxisomas), PON1 (paraoxonasa 1), KNG1 (quininógeno 1), CCL2 (ligando 2 de quimioquina (motivo C-C)), LPL (lipoproteína lipasa), VWF (factor de von Willebrand), F2 (factor de coagulación II (trombina)), ICAM1 (moléculas de adherencia intercelular 1), TGFB1 (factor de crecimiento transformante, beta 1), NPPA (precursor A del péptido natriurético), IL10 (interleuquina 10), EPO (eritropoyetina), SOD1 (superóxido dismutasa 1, soluble), VCAM1 (moléculas de adherencia a células vasculares 1), IFNG (interferón, gamma), LPA (lipoproteína, Lp(a)), MPO (mieloperoxidasa), ESR1 (receptor de estrógeno 1), MAPK1 (proteína quinasa 1 activada por mitógeno), HP (haptoglobina), F3 (factor de coagulación III (factor tisular de tromboplastina)), CST3 (cistatina C), COG2 (componente de oligomérico del aparato de golgi 2), MMP9 (metalopeptidasa de matriz 9 (gelatinasa B, 92 kDa gelatinasa, 92 kDa colagenasa de tipo IV)), SERPINC1 (inhibidor de serpina peptidasa, clado C (antitrombina), miembro 1), F8 (factor de coagulación VIII, componente procoagulante), HMOX1 (hemo oxigenasa (des-ciclante) 1), APOC3 (apolipoproteína C-III), IL8 (interleuquina 8), PROK1 (proquineticina 1), CBS (cistationina-beta-sintasa), NOS2 (óxido nítrico sintasa 2, inducible), TLR4 (receptor similar a toll 4), SELP (selectina P (proteína de membrana granular 140 kDa, antígeno CD62)), ABCA1 (cassette de unión a ATP, sub-familia A (ABC1), miembro 1), AGT (angiotensinógeno (inhibidor de serpina peptidasa, clado A, miembro 8)), LDLR (receptor de lipoproteína de baja densidad), GPT (glutamato-piruvato transaminasa (alanina aminotransferasa)), VEGFA (factor de crecimiento endotelial vascular A), NR3C2 (receptor nuclear, subfamilia 3, grupo C, miembro 2), IL18 (interleuquina 18 (factor inductor de interferón gamma)), NOS1 (óxido nítrico sintasa 1 (neuronal)), NR3C1 (receptor nuclear, subfamilia 3, grupo C, miembro 1 (receptor de glucocorticoides)), FGB (cadena beta de fibrinógeno), HGF (factor de crecimiento de hepatocitos (hepapoyetina A; factor de dispersión)), IL1A (interleuquina 1, alfa), RETN (resistina), AKT1 (v-akt timoma murino homólogo de oncogen viral 1), LIPC (lipasa, hepática), HSPD1 (proteína 1 de choque térmico de 60 kDa (chaperonina)), MAPK14 (proteína quinasa 1 activada por mitógeno 4), SPP1 (fosfoproteína 1 secretada), ITGB3 (integrina, beta 3 (glicoproteína de plaquetas 111a, antígeno CD61)), CAT (catalasa), UTS2 (urotensina 2), THBD (trombomodulina), F10 (factor de coagulación X), CP (ceruloplasmina (ferroxidasa)), TNFRSF11B (superfamilia de receptores del factor de necrosis tumoral, miembro 11b), EDNRA (receptor de endotelina de tipo A), EGFR (receptor del factor de crecimiento epidérmico (oncogen viral homólogo de leucemia eritroblástica (v-erb-b), aviar)), MMP2 (metalopeptidasa de matriz 2 (gelatinasa A, 72 kDa gelatinasa, 72 kDa colagenasa de tipo IV)), PLG (plasminógeno), NPY (neuropéptido Y), RHOD (familia de homólogos de gen ras, miembro D), MAPK8 (proteína quinasa activada por mitógeno 8), MYC (v-myc homólogo del oncogen viral de mielocitomatosis (aviar)), FN1 (fibronectina 1), CMA1 (quimasa 1, mastocito), PLAU (activador de plasminógeno, uroquinasa), GNB3 (proteína de unión a nucleótidos de guanina (proteína G), polipéptido beta 3), ADRB2 (receptor adrenérgico beta-2, superficial), APOA5 (apolipoproteína A-V), SOD2 (superóxido dismutasa 2, mitocondrial), F5 (factor de coagulación V (proacelerina, factor lábil)), VDR (receptor de vitamina D (1,25-dihidroxivitamina D3)), ALOX5 (araquidonato 5-lipooxigenasa), HLA-DRB1 (complejo mayor de histocompatibilidad, clase II, DR beta 1), PARP1 (poli (ADP-ribosa) polimerasa 1), CD40LG (ligando CD40), PON2 (paraoxonasa 2), AGER (receptor específico de producto final de glicosilación avanzada), IRS1 (sustrato del receptor de insulina 1), PTGS1 (prostaglandina-endoperóxido sintasa 1 (prostaglandina G/H sintasa y ciclooxigenasa)), ECE1 (enzima convertidora de endotelina 1), F7 (factor de coagulación VII (acelerador de conversión de protrombina en suero)), URN (antagonista del receptor de interleuquina 1), EPHX2 (epóxido hidrolasa 2, citoplasmática), IGFBP1 (proteína de unión del factor de crecimiento similar a insulina 1), MAPK10 (proteína quinasa 10 activada por mitógeno), FAS (Fas (Superfamilia del receptor del TNF, miembro 6)), ABCB1 (cassette de unión a ATP, sub-familia B (MDR/TAP), miembro 1), JUN (oncogen jun), IGFBP3 (proteína de unión 3 del factor de crecimiento similar a insulina), CD14 (moléculas CD14), PDE5A (fosfodiesterasa 5A, cGMP-específica), AGTR2 (receptor de angiotensina II, tipo 2), CD40 (moléculas CD40, superfamilia del receptor del TNF, miembro 5), LCAT (lecitina-colesterol aciltransferasa), CCR5 (receptor 5 de quimioquina (motivo C-C)), MMP1 (metalopeptidasa de matriz 1 (colagenasa intersticial)), TIMP1 (TIMP inhibidor de metalopeptidasa 1), ADM (adrenomedulina), DYT10 (distonia 10), STAT3 (transductor de señal y activador de transcripción 3 (factor de respuesta de fase aguda)), MMP3 (metalopeptidasa de matriz 3 (estromelisina 1, progelatinasa)), ELN (elastina), USF1 (factor de transcripción 1 hacia el extremo 3'), CFH (factor H del complemento), HSPA4 (proteína 4 de choque térmico de 70 kDa), MMP12 (metalopeptidasa de matriz 12 (elastasa de macrófagos)), MME (metalo-endopeptidasa de membrana), F2R (receptor del factor de coagulación II (trombina)), SELL (selectina L), CTSB (catepsina B), ANXA5 (anexina A5), ADRB1 (receptor beta-1-adrenérgico), CYBA (citocromo b-245, polipéptido alfa), FGA (cadena alfa de fibrinógeno), GGT1 (gamma-glutamiltransferasa 1), LIPG (lipasa, endotelial), HIF1A (factor 1 inducible por hipoxia, subunidad alfa (factor de transcripción básico hélice-bucle-hélice)), CXCR4 (receptor 4 de quimioquina (motivo C-X-C)), PROC (proteína C (desactivador de factores de coagulación Va y VIIIa)), SCARB1 (receptor secuestrante clase B, miembro 1), CD79A (moléculas CD79a, asociado a inmunoglobulina alfa), PLTP (proteína de transferencia de fosfolípido), ADD1 (aducina 1 (alfa)), FGG (cadena gamma de fibrinógeno), SAA1 (amiloide A1 del suero), KCNH2 (canal de potasio dependiente del potencial, subfamilia H (relacionado con eag), miembro 2), DPP4 (dipeptidil-peptidasa 4), G6PD (glucosa-6-fosfato deshidrogenasa), NPR1 (receptor A de péptido natriurético /guanilato ciclasa A (receptor A de atriopéptido natriurético)), VTN (vitronectina), KIAA0101 (KIAA0101), FOS (homólogo de oncogen viral de osteosarcoma murino FBJ), TLR2 (receptor 2 similar a toll), PPIG (peptidilprolil isomerasa G (ciclofilina G)), IL1R1 (receptor de interleuquina 1, tipo I), AR (receptor de andrógeno), CYP1A1 (citocromo P450, familia 1, subfamilia A, polipéptido 1), SERPINA1 (inhibidor de serpina peptidasa, clado A (alfa-1 antiproteinasa, antitripsina), miembro 1), MTR (5-metiltetrahidrofolato-homocisteína metiltransferasa), RBP4 (proteína de unión 4 a retinol, plasma), APOA4 (apolipoproteína A-IV), CDKN2A (inhibidor de quinasa 2A dependiente de ciclina (melanoma, p16, inhibe a CDK4)), FGF2 (factor de crecimiento de fibroblastos 2 (básico)), EDNRB (receptor de endotelina tipo B), ITGA2 (integrina, alfa 2 (CD49B, subunidad alfa 2 del receptor de VLA-2)), CABIN1 (proteína de unión a calcineurina 1), SHBG (globulina fijadora de hormonas sexuales),

HMGB1 (grupo de alta movilidad de caja 1), HSP90B2P (proteína de choque térmico 90 kDa beta (Grp94), miembro 2 (pseudogen)), CYP3A4 (citocromo P450, familia 3, subfamilia A, polipéptido 4), GJA1 (proteínas de uniones gap, alfa 1, 43 kDa), CAV1 (caveolina 1, proteína caveola, 22 kDa), ESR2 (receptor de estrógeno 2 (ER beta)), LTA (linfotoxina alfa (superfamilia TNF, miembro 1)), GDF15 (factor de diferenciación del crecimiento 15), BDNF (factor neurotrófico derivado de cerebro), CYP2D6 (citocromo P450, familia 2, subfamilia D, polipéptido 6), NGF (factor de crecimiento nervioso (polipéptido beta)), SP1 (factor de transcripción Sp1), TGIF1 (factor homeocaja 1 inducido por TGFB), SRC (homólogo de oncogen de sarcoma viral v-src (Schmidt-Ruppin A-2) (aviar)), EGF (factor de crecimiento epidérmico (beta-urogastrona)), PIK3CG (fosfoinositida-3-quinasa, catalítica, polipéptido gamma), HLA-A (complejo mayor de histocompatibilidad, clase I, A), KCNQ1 (canal de potasio dependiente del potencial, subfamilia similar a KQT, miembro 1), CNR1 (receptor 1 de canabinoides (cerebro)), FBN1 (fibrilina 1), CHKA (colina quinasa alfa), BEST1 (bestrofina 1), APP (proteína precursora de beta amiloide (A4)), CTNNB1 (catenina (proteína asociada a cadherina), beta 1, 88 kDa), IL2 (interleuquina 2), CD36 (CD36 moléculas (receptor de trombospondina)), PRKAB1 (proteína quinasa, activada por AMP, subunidad no catalítica beta 1), TPO (peroxidasa tiroidea), ALDH7A1 (familia de aldehído deshidrogenasa 7, miembro A1), CX3CR1 (receptor 1 de quimioquina (motivo C-X3-C)), TH (tirosina hidroxilasa), F9 (factor de coagulación IX), GH1 (hormona de crecimiento 1), TF (transferrina), HFE (hemocromatosis), IL17A (interleuquina 17A), PTEN (homólogo de fosfatasa y tensina), GSTM1 (glutationa S-transferasa mu 1), DMD (distrofina), GATA4 (proteína de unión GATA 4), F13A1 (factor de coagulación XIII, polipéptido A1), TTR (transtiretina), FABP4 (proteína de unión a ácidos grasos 4, de adipocitos), PON3 (paraoxonasa 3), APOC1 (apolipoproteína C-I), INSR (receptor de insulina), TNFRSF1B (superfamilia de receptores del factor de necrosis tumoral, miembro 1B), HTR2A (receptor 2A de 5-hidroxitriptamina (serotonina)), CSF3 (factor 3 estimulador de colonias (granulocito)), CYP2C9 (citocromo P450, familia 2, subfamilia C, polipéptido 9), TXN (tiorredoxina), CYP11B2 (citocromo P450, familia 11, subfamilia B, polipéptido 2), PTH (hormona paratiroidea), CSF2 (factor 2 estimulador de colonias (granulocito-macrófago)), KDR (dominio receptor de inserto quinasa (una receptor tirosina quinasa de tipo III)), PLA2G2A (fosfolipasa A2, grupo IIA (plaquetass, fluido sinovial)), B2M (beta-2-microglobulina), THBS1 (trombospondina 1), GCG (glucagón), RHOA (familia de homólogos del gen ras, miembro A), ALDH2 (familia de la aldehído deshidrogenasa 2 (mitocondrial)), TCF7L2 (factor de transcripción 7 similar a 2 (específico de células T, HMG-box)), BDKRB2 (receptor de bradiquinina B2), NFE2L2 (factor nuclear (derivado de eritroides 2) similar al 2), NOTCH1 (Homólogo de Notch 1, asociado a la translocación (Drosophila)), UGT1A1 (familia de UDP glucuroniltransferasa 1, polipéptido A1), IFNA1 (interferón, alfa 1), PPARD (receptor delta activado por proliferador de peroxisomas), SIRT1 (sirtuina (homólogo 2 de regulación de información de tipo de unión silente) 1 (S. cerevisiae)), GNRH1 (hormona liberadora de gonadotropina 1 (hormona liberadora de hormona luteinizante)), PAPPA (proteína A plasmática asociada con la gestación, papalisina 1), ARR3 (arrestina 3, retinal (arrestina X)), NPPC (precursor C de péptido natriurético), AHSP (proteína estabilizante de alfa hemoglobina), PTK2 (PTK2 proteína tirosina quinasa 2), IL13 (interleuquina 13), MTOR (diana de rapamicina (serina/treonina quinasa)), ITGB2 (integrina, beta 2 (componente 3 del complemento receptor 3 y subunidad 4)), GSTT1 (glutationa S-transferasa theta 1), IL6ST (interleuquina 6 transductor de señal (gp130, receptor de oncostatina M)), CPB2 (carboxipeptidasa B2 (plasma)), CYP1A2 (citocromo P450, familia 1, subfamilia A, polipéptido 2), HNF4A (factor nuclear 4 de hepatocitos, alfa), SLC6A4 miembro 4 de la familia de transportadores de solutos 6 (transportador de neurotransmisores, serotonina), PLA2G6 (fosfolipasa A2, grupo VI (citosólica, calcio-independiente)), TNFSF11 (factor de necrosis tumoral (ligand) superfamilia, miembro 11), SLC8A1 (familia vehículo de soluto 8 (sodio/calcio intercambiador), miembro 1), F2RL1 (factor de coagulación II (trombina) similar a receptor 1), AKR1A1 (familia 1 de aldo-ceto reductasas, miembro A1 (aldehído reductasa)), ALDH9A1 (familia de aldehído deshidrogenasa 9, miembro A1), BGLAP (proteína gamma-carboxiglutamato (gla) ósea), MTTP (proteína de transferencia microsómica de triglicéridos), MTRR (5-metiltetrahidrofolato-homocisteína metiltransferasa reductasa), SULT1A3 (familia de la sulfotransferasa, citosólica, 1A, con preferencia por fenol, miembro 3), RAGE (antígeno de tumor renal), C4B (componente del complemento 4B (grupo sanguíneo Chido), P2RY12 (receptor purinérgico P2Y, acoplado a proteína G, 12), RNLS (renalasa, amino oxidasa dependiente de FAD), CREB1 (proteína de unión al elemento de respuesta a cAMP 1), POMC (propiomelanocortina), RAC1 (sustrato 1 de toxina botulínica C3 relacionado con ras (familia rho, proteína pequeña de unión a GTP Rac1)), LMNA (lamina NC), CD59 (moléculas CD59, proteína reguladora del complemento), SCN5A (canal de sodio, dependiente del potencial, tipo V, subunidad alfa), CYP1B1 (citocromo P450, familia 1, subfamilia B, polipéptido 1), MIF (factor inhibidor de migración de macrófagos (factor de inhibidor de glicosilación)), MMP13 (metalopeptidasa de matriz 13 (colagenasa 3)), TIMP2 (TIMP inhibidor de metalopeptidasa 2), CYP19A1 (citocromo P450, familia 19, subfamilia A, polipéptido 1), CYP21A2 (citocromo P450, familia 21, subfamilia A, polipéptido 2), PTPN22 (proteína tirosina fosfatasa, de tipo no receptor 22 (linfoide)), MYH14 (miosina, cadena pesada 14, no muscular), MBL2 (lectina de unión a manosa (proteína C) 2, soluble (defecto opsónico)), SELPLG (ligando de selectina P), AOC3 (amino oxidasa que contiene cobre 3 (proteína de adherencia vascular 1)), CTSL1 (catepsina L1), PCNA (antígeno nuclear de proliferación celular), IGF2 (factor de crecimiento similar a insulina 2 (somatomedina A)), ITGB1 (integrina, beta 1 (receptor de fibronectina, polipéptido beta, antígeno CD29 incluyendo a MDF2, MSK12)), CAST (calpastatina), CXCL12 (quimioquina (motivo C-X-C) ligando 12 (factor 1 derivado de células del estroma)), IGHE (cadena epsilon constante pesada de inmunoglobulina), KCNE1 (canal de potasio dependiente del potencial, relacionado con familia Isk, miembro 1), TFRC (receptor de transferrina (p90, CD71)), COL1A1 (colágeno, tipo I, alfa 1), COL1A2 (colágeno, tipo I, alfa 2), IL2RB (receptor de interleuquina 2, beta), PLA2G10 (fosfolipasa A2, grupo X), ANGPT2 (angiopoyetina 2), PROCR (receptor de proteína C, endotelial (EPCR)), NOX4 (NADPH oxidasa 4), HAMP (péptido antimicrobiano de hepcidina), PTPN11 (proteína tirosina fosfatasa, de tipo no receptor 11), SLC2A1 (familia vehículo de soluto 2 (transportador de difusión facilitada de glucosa), miembro 1), IL2RA (receptor de interleuquina 2, alfa), CCL5 (quimioquina (motivo C-C) ligando 5), IRF1 (factor regulador de interferón 1), CFLAR (regulador de apoptosis similar a CASP8 y FADD), CALCA (polipéptido alfa relacionado con calcitonina), EIF4E

(factor de inicio de traducción eucariótico 4E), GSTP1 (glutationa S-transferasa pi 1), JAK2 (Jano quinasa 2), CYP3A5 (citocromo P450, familia 3, subfamilia A, polipéptido 5), HSPG2 (sulfato de heparán proteoglicano 2), CCL3 (quimioquina (motivo C-C) ligando 3), MYD88 (gen de respuesta primaria a la diferenciación mieloide (88)), VIP (péptido intestinal vasoactivo), SOAT1 (esterol O-aciltransferasa 1), ADRBK1 (receptor quinasa 1 beta adrenérgico), NR4A2 (subfamilia 4 de receptor nuclear, grupo A, miembro 2), MMP8 (metalopeptidasa de matriz 8 (colagenasa de neutrófilos)), NPR2 (receptor del péptido natriurético B/guanilato ciclasa B (atrio receptor del péptido natriurético B)), GCH1 (GTP ciclohidrolasa 1), EPRS (glutamil-prolil-ARNt sintetasa), PPARGC1A (receptor gamma activado por proliferador de peroxisomas, coactivador 1 alfa), F12 (factor de coagulación XII (factor de Hageman)), PECAM1 (moléculas de adherencia a plaquetas/células endoteliales), CCL4 (quimioquina (motivo C-C) ligando 4), SERPINA3 (inhibidor de serpina peptidasa, clado A (alfa-1 antiproteinasa, antitripsina), miembro 3), CASR (receptor sensor de calcio), GJA5 (proteínas de uniones gap, alfa 5, 40 kDa), FABP2 (proteína de unión a ácidos grasos 2, intestinal), TTF2 (factor de finalización de la transcripción, ARN polimerasa II), PROS1 (proteína S (alfa)), CTF1 (cardiotrofina 1), SGCB (sarcoglicano, beta (43 kDa glicoproteína asociada a distrofina)), YME1L1 (similar a YME1 1 (S. cerevisiae)), CAMP (péptido antimicrobiano catelicidina), ZC3H12A (dedos de cinc de tipo CCCH que contiene a 12A), AKR1B1 (familia 1 de aldo-ceto reductasa, miembro B1 (aldosa reductasa)), DES (desmina), MMP7 (metalopeptidasa de matriz 7 (matrilisina, uterina)), AHR (receptor de hidrocarburos arílicos), CSF1 (factor 1 estimulador de colonias (macrófago)), HDAC9 (histona desacetilasa 9), CTGF (factor tisular de crecimiento de tejido conectivo), KCNMA1 (potasio large canal de calcio activado por conductancia, subfamilia M, miembro alfa 1), UGT1A (familia UDP-glucuroniltransferasa 1, polipéptido A locus complejo), PRKCA (proteína quinasa C, alfa), COMT (catecol-.beta.-metiltransferasa), S100B (proteína S100 de unión a calcio B), EGR1 (respuesta de crecimiento temprano 1), PRL (prolactina), IL15 (interleuquina 15), DRD4 (receptor de dopamina D4), CAMK2G (proteína quinasa II gamma dependiente de calcio/calmodulina), SLC22A2 (familia vehículo de soluto 22 (transportador de catión orgánico), miembro 2), CCL11 (quimioquina (motivo C-C) ligando 11), PGF (B321 factor de crecimiento placentario), THPO (trombopoyetina), GP6 (glicoproteína VI (plaquetas)), TACR1 (receptor de taquiquinina 1), NTS (neurotensina), HNF1A (HNF1 homeocaja A), SST (somatostatina), KCND1 (canal de potasio dependiente del potencial, subfamilia relacionada con Shal, miembro 1), LOC646627 (inhibidor de fosfolipasa), TBXAS1 (tromboxano A sintasa 1 (plaquetas)), CYP2J2 (citocromo P450, familia 2, subfamilia J, polipéptido 2), TBXA2R (receptor de tromboxano A2), ADH1C (alcohol deshidrogenasa 1C (clase I), polipéptido gamma), ALOX12 (araquidonato 12-lipooxigenasa), AHSG (alfa-2-HS-glicoproteína), BHMT (betaína-homocisteína metiltransferasa), GJA4 (proteínas de uniones gap, alfa 4, 37 kDa), SLC25A4 (familia vehículo de soluto 25 (vehículo mitocondrial, translocador de nucleótido adenina), miembro 4), ACLY (ATP citrato liasa), ALOX5AP (proteína activadora de araquidonato 5-lipooxigenasa), NUMA1 (proteína 1 del aparato mitótico nuclear), CYP27B1 (citocromo P450, familia 27, subfamilia B, polipéptido 1), CYSLTR2 (receptor 2 de cisteinil leucotrieno), SOD3 (superóxido dismutasa 3, extracelular), LTC4S (leucotrieno C4 sintasa), UCN (urocortina), GHRL (prepropéptido grelina/obestatina), APOC2 (apolipoproteína C-II), CLEC4A (Dominio lectina de tipo C familia 4, miembro A), KBTBD10 (repetición de kelch y dominio que contiene BTB (POZ) 10), TNC (tenascin C), TYMS (timidilato sintetasa), SHCI (proteína transformante de SHC (dominio que contiene homología Src 2 a) 1), LRP1 (proteína relacionada con el receptor de lipoproteína de baja densidad 1), SOCS3 (supresor de señalización de citoquinas 3), ADH1B (alcohol deshidrogenasa 1B (clase I), polipéptido beta), KLK3 (peptidasa relacionada con calicreína 3), HSD11B1 (hidroxiesteroide (11-beta) deshidrogenasa 1), VKORC1 (complejo vitamina K epóxido reductasa, subunidad 1), SERPINB2 (inhibidor de serpina peptidasa, clado B (ovalbúmina), miembro 2), TNS1 (tensina 1), RNF19A (proteína dedo RING 19A), EPOR (receptor de eritropoyetina), ITGAM (integrina, alfa M (componente 3 de la subunidad del complemento receptor 3)), PITX2 (homeodominio similar al pareado 2), MAPK7 (proteína quinasa activada por mitógeno 7), FCGR3A (fragmento Fc de IgG, baja afinidad 111a, receptor (CD16a)), LEPR (leptina receptor), ENG (endoglina), GPX1 (glutationa peroxidasa 1), GOT2 (glutámico-oxaloacético transaminasa 2, mitocondrial (aspartato aminotransferasa 2)), HRH1 (receptor de histamina H1), NR112 (receptor nuclear subfamilia 1, grupo I, miembro 2), CRH (hormona liberadora de corticotropina), HTR1A (receptor 1A de 5-hidroxitriptamina (serotonina)), VDAC1 (canal de anión 1 dependiente del voltaje), HPSE (heparanasa), SFTPD (proteína tensioactiva D), TAP2 (transportador 2, cassette de unión a ATP, sub-familia B (MDR/TAP)), RNF123 (proteína dedo RING 123), PTK2B (PTK2B proteína tirosina quinasa 2 beta), NTRK2 (tirosina quinasa neurotrófica, receptor, tipo 2), IL6R (receptor de interleuquina 6), ACHE (acetilcolinesterasa (grupo sanguíneo Yt)), GLP1R (receptor del péptido similar a glucagón 1), GHR (receptor de la hormona de crecimiento), GSR (glutationa reductasa), NQO1 (NAD(P)H deshidrogenasa, quinona 1), NR5A1 (subfamilia 5 de receptores nucleares, grupo A, miembro 1), GJB2 (proteínas de uniones gap, beta 2, 26 kDa), SLC9A1 (familia vehículo de soluto 9 (intercambiador sodio/hidrógeno), miembro 1), MAOA (monoamino oxidasa A), PCSK9 (proproteína convertasa subtilisina/quexina de tipo 9), FCGR2A (fragmento Fc de IgG, de baja afinidad IIa, receptor (CD32)), SERPINF1 (inhibidor de serpina peptidasa, clado F (alfa-2 antiplasmina, factor derivado de pigmento de epitelio), miembro 1), EDN3 (endotelina 3), DHFR (dihidrofolato reductasa), GAS6 (factor específico del arresto del crecimiento 6), SMPD1 (esfingomielina fosfodiesterasa 1, ácido-lisosómica), UCP2 (proteína desacoplante 2 (mitocondrial, vehículo de protones)), TFAP2A (factor de transcripción AP-2 alfa (activador potenciador proteína de unión 2, alfa)), C4BPA (proteína de unión de componentes 4 del complemento, alfa), SERPINF2 (inhibidor de serpina peptidasa, clado F (alfa-2 antiplasmina, factor derivado de pigmento epitelial), miembro 2), TYMP (timidina fosforilasa), ALPP (fosfatasa alcalina, placentaria (isozima de Regan)), CXCR2 (receptor 2 de quimioquina (motivo C-X-C)), SLC39A3 (familia de vehículo de soluto 39 (transportador de cinc), miembro 3), ABCG2 (cassette de unión a ATP, sub-familia G (WHITE), miembro 2), ADA (adenosina desaminasa), JAK3 (Jano quinasa 3), HSPA1A (proteína 1A de choque térmico 70 kDa), FASN (ácido graso sintasa), FGF1 (factor de crecimiento de fibroblastos 1 (ácido)), F11 (factor de coagulación XI), ATP7A (ATPasa, transporte de Cu^{++}, polipéptido alfa), CR1 (componente del complemento (3b/4b) receptor 1 (grupo sanguíneo Knops)), GFAP (proteína ácida fibrilar glial), ROCK1 (asociado a Rho, hélice

superenrrollada que contiene proteína quinasa 1), MECP2 (proteína de unión 2 a metil CpG (síndrome de Rett)), MILK (quinasa de cadena liviana de la miosina), BCHE (butirilcolinesterasa), LIPE (lipasa, sensible a hormonas), PRDX5 (peroxirredoxina 5), ADORA1 (receptor de adenosina A1), WRN (síndrome de Werner, RecQ similar a helicasa), CXCR3 (receptor de quimioquina (motivo C-X-C) 3), CD81 (moléculas CD81), SMAD7 (familia SMAD, miembro 7), LAMC2 (laminina, gamma 2), MAP3K5 (proteína quinasa activada por mitógeno quinasa quinasa 5), CHGA (cromogranina A (proteína secretoria paratiroidea 1)), IAPP (polipéptido amiloide de los islotes), RHO (rodopsina), ENPP1 (ectonucleótido pirofosfatasa/fosfodiesterasa 1), PTHLH (hormona similar a hormona paratiroidea), NRG1 (neuregulina 1), VEGFC (factor de crecimiento endotelial vascular C), ENPEP (glutamil aminopeptidasa (aminopeptidasa A)), CEBPB (CCAAT/proteína de unión al potenciador (C/EBP), beta), NAGLU (N-acetilglucosaminidasa, alfa-), F2RL3 (factor de coagulación II (trombina) similar a receptor 3), CX3CL1 (quimioquina (motivo C-X3-C) ligando 1), BDKRB1 (receptor de bradiquinina B1), ADAMTS13 (metalopeptidasa ADAM con motivo de trombospondina de tipo 1, 13), ELANO (elastasa, expresada por neutrófilos), ENPP2 (ectonucleótido pirofosfatasa/fosfodiesterasa 2), CISH (proteína que contiene SH2 inducible por citoquina), GAST (gastrina), MYOC (miocilina, respuesta de glucocorticoides inducible de la red de células trabeculares), ATP1A2 (ATPasa, Na+/K+ transporte, alfa 2 polipéptido), NF1 (neurofibromina 1), GJB1 (proteínas de uniones gap, beta 1, 32 kDa), MEF2A (factor potenciador de miocitos 2A), VCL (vinculina), BMPR2 (receptor de proteína morfogenética ósea, tipo II (serina/treonina quinasa)), TUBB (tubulina, beta), CDC42 (proteína de ciclo de división celular 42 (proteína de unión a GTP, 25 kDa)), KRT18 (queratina 18), HSF1 (factor de transcripción de choque térmico 1), MYB (v- homólogo de oncogen de la mieloblastosis viral myb (aviar)), PRKAA2 (proteína quinasa, activada por AMP, subunidad catalítica alfa 2), ROCK2 (asociado a Rho, hélice superenrrollada que contiene proteína quinasa 2), TFPI (inhibidor de la vía del factor tisular (inhibidor de coagulación asociado a lipoproteínas)), PRKG1 (proteína quinasa, dependiente de cGMP, tipo I), BMP2 (proteína morfogenética del hueso 2), CTNND1 (catenina (proteína asociada a cadherina), delta 1), CTH (cistationasa (cistationina gamma-liasa)), CTSS (catepsina S), VAV2 (factor intercambiador de nucleótido guanina vav 2), NPY2R (receptor Y2 de neuropéptido Y), IGFBP2 (proteína de unión 2 al factor de crecimiento similar a insulina, 36 kDa), CD28 (moléculas CD28), GSTA1 (glutationa S-transferasa alfa 1), PPIA (peptidilprolil isomerasa A (ciclofilina A)), APOH (apolipoproteína H (beta-2-glicoproteína I)), S100A8 (S100 proteína de unión a calcio A8), IL11 (interleuquina 11), ALOX15 (araquidonato 15-lipooxigenasa), FBLN1 (fibulina 1), NR1H3 (receptor nuclear subfamilia 1, grupo H, miembro 3), SCD (estearoil-CoA desaturasa (delta-9-desaturasa)), GIP (polipéptido inhibidor gástrico), CHGB (cromogranina B (secretogranina 1)), PRKCB (proteína quinasa C, beta), SRD5A1 (esteroid-5-alfa-reductasa, polipéptido alfa 1 (3-oxo-5 alfa-esteroide delta 4-deshidrogenasa alfa 1)), HSD11B2 (hidroxiesteroide (11-beta) deshidrogenasa 2), CALCRL (calcitonina similar a receptor), GALNT2 (UDP-N-acetil-alfa-D-galactosamina:polipéptido N-acetilgalactosaminiltransferasa 2 (GalNAc-T2)), ANGPTL4 (proteína similar a angiopoyetina 4), KCNN4 (canal de potasio de conductancia intermedia/baja activado por calcio, subfamilia N, miembro 4), PIK3C2A (fosfoinositida-3-quinasa, clase 2, polipéptido alfa), HBEGF (factor de crecimiento similar al factor de crecimiento epidérmico de unión a la heparina), CYP7A1 (citocromo P450, familia 7, subfamilia A, polipéptido 1), HLA-DRB5 (complejo mayor de histocompatibilidad, clase II, DR beta 5), BNIP3 (proteína 3 de interacción con BCL2/adenovirus E1B 19 kDa), GCKR (regulador de glucoquinasa (hexoquinasa 4)), S100A12 (proteína S100 de unión a calcio A12), PADI4 (peptidil arginina desiminasa, tipo IV), HSPA14 (proteína 14 de choque térmico 70 kDa), CXCR1 (receptor 1 de quimioquina (motivo C-X-C)), H19 (H19, transcripción de impronta maternalmente expresada (que no codifica proteína)), KRTAP19-3 (queratina asociado proteína 19-3), IDDM2 (insulina-dependiente diabetes mellitus 2), RAC2 (sustrato de toxina botulínica C3 relacionado con ras 2 (familia rho, proteína pequeña de unión a GTP Rac2)), RYR1 (ryanodine receptor 1 (skeletal)), CLOCK (clock homólogo (ratón)), NGFR (factor de crecimiento nervioso receptor (TNFR superfamilia, miembro 16)), DBH (dopamina beta-hidroxilasa (dopamina beta-monooxigenasa)), CHRNA4 (colinérgico receptor, nicotinic, alfa 4), CACNA1C (canal de calcio, voltage-dependiente, L tipo, alfa 1C subunidad), PRKAG2 (proteína quinasa, activada por AMP, gamma 2 subunidad no catalítica), CHAT (colina acetiltransferasa), PTGDS (prostaglandina D2 sintasa 21 kDa (cerebro)), NR1H2 (receptor nuclear subfamilia 1, grupo H, miembro 2), TEK (TEK tirosina quinasa, endotelial), VEGFB (factor de crecimiento endotelial vascular B), MEF2C (factor potenciador de miocitos 2C), MAPKAPK2 (proteína quinasa activada por mitógeno -activado por proteína quinasa 2), TNFRSF11A (superfamilia de receptores del factor de necrosis tumoral, miembro 11a, NFKB activador), HSPA9 (choque térmico 70 kDa proteína 9 (mortalina)), CYSLTR1 (cisteinil leucotrieno receptor 1), MAT1A (metionina adenosiltransferasa I, alfa), OPRL1 (opiate similar a receptor 1), IMPA1 (inositol(myo)-1(o 4)-monofosfatasa 1), CLCN2 (cloruro de canal 2), DLD (dihidrolipoamida deshidrogenasa), PSMA6 (proteasome (prosome, macropain) subunidad, alfa tipo, 6), PSMB8 (proteasome (prosome, macropain) subunidad, beta tipo, 8 (large multifunctional peptidasa 7)), CHI3L1 (chitinase 3-similar a 1 (cartilage glicoproteína-39)), ALDH1B1 (aldehído deshidrogenasa 1 familia, miembro B1), PARP2 (poli (ADP-ribosa) polimerasa 2), STAR (steroidogenic acute proteína reguladoria), LBP (lipopolisaccharide proteína de unión), ABCC6 (cassette de unión a ATP, sub-familia C(CFTR/MRP), miembro 6), RGS2 (regulador de proteína G señalización 2, 24 kDa), EFNB2 (ephrin-B2), GJB6 (proteínas de uniones gap, beta 6, 30 kDa), APOA2 (apolipoproteína A-II), AMPD1 (adenosina monofosfato desaminasa 1), DYSF (dysferlin, limb girdle muscular distrofia 2B (autosómicos recesivo)), FDFT1 (farnesil-difosfato farnesiltransferasa 1), EDN2 (endotelina 2), CCR6 (quimioquina (motivo C-C) receptor 6), GJB3 (proteínas de uniones gap, beta 3, 31 kDa), IL1RL1 (interleuquina 1 similar a receptor 1), ENTPD1 (ectonucleoside trifosfato difosfohidrolasa 1), BBS4 (Bardet-Biedl síndrome 4), CELSR2 (cadherina, EGF LAG seven-pass G-tipo receptor 2 (flamingo homólogo, Drosophila)), F11R (F11 receptor), RAPGEF3 (Rap factor intercambiador de nucleótido guanina (GEF) 3), HYAL1 (hyaluronoglucosaminidase 1), ZNF259 (dedos de cinc proteína 259), ATOX1 (ATX1 antioxidante proteína 1 homólogo (yeast)), ATF6 (activador factor de transcripción 6), KHK (ketohexoquinasa (fructoquinasa)), SAT1 (spermidina/spermine N1-acetiltransferasa 1), GGH (gamma-glutamil hidrolasa (conjugase, folilpoligammaglutamil hidrolasa)), TIMP4 (TIMP inhibidor de metalopeptidasa 4), SLC4A4

(familia vehículo de soluto 4, sodio bicarbonato cotransportador, miembro 4), PDE2A (fosfodiesterasa 2A, cGMP-stimulated), PDE3B (fosfodiesterasa 3B, cGMP-inhibited), FADS1 (ácido graso desaturase 1), FADS2 (ácido graso desaturase 2), TMSB4X (thymosin beta 4, relacionado con X), TXNIP (tioredoxin proteína de interacción), LIMS1 (LIM y senescent célula antígeno-similar a dominios 1), RHOB (familia de homólogos de gen ras, miembro B), LY96 (lymfocyte antígeno 96), FOXO1 (forkhead box O1), PNPLA2 (patatin-similar a fosfolipasa dominio que contiene 2), TRH (thyrotropin-hormona liberadora de), GJC1 (proteínas de uniones gap, gamma 1, 45 kDa), SLC17A5 (familia vehículo de soluto 17 (anion/sugar transportador), miembro 5), FTO (fat mass y obesity asociado), GJD2 (proteínas de uniones gap, delta 2, 36 kDa), PSRC1 (proline/serina-rich hélice superenrrollada 1), CASP12 (caspasa 12 (gene/pseudogeno)), GPBAR1 (G proteína-acoplado bile ácido receptor 1), PXK (PX dominio que contiene serina/treonina quinasa), IL33 (interleuquina 33), TRIB1 (tribbles homólogo 1 (Drosophila)), PBX4 (pre-B-cell leucemia homeocaja 4), NUPR1 (nuclear proteína, transcripciónal regulador, 1), 15-Sep(15 kDa selenoprotein), CILP2 (cartilage intermedio capa proteína 2), TERC (telomerase ARN componente), GGT2 (gamma-glutamiltransferasa 2), MT-CO1 (mitocondrially codificada citocromo c oxidasa I), y UOX (urate oxidasa, pseudogeno). Cualquiera de estas secuencias puede ser una diana para el sistema CRISPR-Cas, por ejemplo, para abordar mutaciones.

En una forma de realización adicional, la secuencia cromosómica se podrá seleccionar adicionalmente entre Pon1 (paraoxonasa 1), LDLR (receptor de LDL), ApoE (Apolipoproteína E), Apo B-100 (Apolipoproteína B-100), ApoA (Apolipoproteína (a)), ApoA1 (Apolipoproteína A1), CBS (Cistatión B-sintasa), glucoproteína IIb/IIb, MTHRF (5,10-metilentetrahidrofolato-reductasa (NADPH) y combinaciones de estos. En una iteración, las secuencias cromosómicas y las proteínas codificadas por las secuencias cromosómicas implicadas en las enfermedades cardiovasculares se podrán escoger entre Cacna1C, Sod1, Pten, Ppar(alfa), ApoE, Leptina y combinaciones de estas como blancos para el sistema CRISPR-Cas.

Tratamiento de enfermedades del hígado y el riñón

También se describe, pero no se reivindica en la presente el suministro del sistema CRISPR-Cas que se describe aquí, por ejemplo, sistemas de proteína efectora Cpf1, al hígado y/o al riñón. Las estrategias de suministro para inducir la captación celular del ácido nucleico terapéutico incluyen la fuerza física o sistemas vectoriales tales como suministro con virus, lípidos o complejos o nanoportadores. Desde las aplicaciones iniciales con menores posibilidades de relevancia clínica, cuando se dirigieron ácidos nucleicos hacia células renales con inyección hidrodinámica a alta presión de manera sistémica, ya se ha aplicado un amplio rango de vehículos génicos terapéuticos, virales y no virales a eventos post-transcripcionales diana en diferentes modelos animales de enfermedad renal *in vivo* (Csaba Révész y Péter Hamar (2011). Delivery Methods to Target RNAs in the Kidney, Gene Therapy Applications, Prof. Chunsheng Kang (Ed.), ISBN: 978-953-307-541-9, InTech, que se puede obtener de: http://www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-inthe-kidney). Los métodos de suministro al riñón pueden incluir a aquellos que se mencionan en Yuan y col. (Am J Physiol Renal Physiol 295: F605–F617, 2008) ellos estudiaron si el suministro in vivo de ARN interferentes pequeños (ARNip) cuyo diana es la ruta 12/15-lipooxigenasa (12/15-LO) del metabolismo del ácido araquidónico puede mejorar la lesión renal y la nefropatía diabética (DN, por sus siglas en inglés) en modelo de diabetes de tipo 1 en ratones a los que se ha inyectado estreptozotocina. Para lograr un mayor acceso in vivo y expresión de ARNip en el riñón, Yuan y col. utilizaron oligonucleótidos ARNip contra 12/15-LO bicatenarios conjugados con colesterol. Se inyectaron aproximadamente 400 µg de ARNip subcutáneamente en ratones. El método de Yuang et al. se podrá aplicar al sistema CRISPR Cas, donde se contempla una inyección subcutánea de 1-2 g de CRISPR Cas conjugado con colesterol a un ser humano para el suministro a los riñones.

Molitoris y col. (J Am Soc Nephrol 20: 1754–1764, 2009) utilizan células tubulares proximales (PTC, por sus siglas en inglés), como el sitio de reabsorción de oligonucleótidos dentro del riñón para estudiar la eficacia de ARNip cuyo diana es p53, una proteína esencial en la ruta apoptótica, para prevenir la lesión en el riñón. El ARNip contra p53 sintético desnudo inyectado por vía intravenosa 4 h después de la lesión isquémica protegió de manera máxima tanto las PTC como la función renal. Los datos de Molitoris et al. indican que el suministro rápido de ARNip a las células tubulares proximales sigue a la administración intravenosa. Para los análisis de dosis-respuesta, se inyectaron las ratas con dosis de P53ip, se administraron 0,33; 1, 3 o 5 mg/kg en los mismos cuatro puntos temporales lo que dio como resultado dosis acumulativas de 1,32; 4, 12 y 20 mg/kg, respectivamente. Todas las dosis de ARNip estudiadas produjeron un efecto reductor de SCr el día uno donde las dosis más elevadas fueron eficaces a lo largo de aproximadamente cinco días en comparación con ratas de control isquémicas tratadas con PBS. Las dosis acumulativas de 12 y 20 mg/kg proporcionaron el mejor efecto protector. El método de Molitoris y col. se puede aplicar al sistema direccionado hacia ácidos nucleicos que contempla 12 y 20 mg/kg dosis acumulativas a un ser humano para suministrarlo a los riñones.

Thompson y col. (Nucleic Acid Therapeutics, Volumen 22, Número 4, 2012) informan las propiedades toxicológicas y farmacocinéticas del ARN pequeño de interferencia I5NP sintético luego de la administración intravenosa en roedores y primates no humanos. I5NP está diseñado para actuar mediante la ruta de interferencia por ARN (ARNi) para inhibir temporalmente la expresión de la proteína pro-apoptótica p53 y se está desarrollando para proteger las células de las lesiones por isquemia aguda/reperfusión tales como la lesión renal aguda que puede ocurrir durante la cirugía cardíaca mayor y función del injerto retrasada que puede ocurrir tras el trasplante renal. Se requirieron dosis de 800 mg/kg de I5NP en roedores y de 1000 mg/kg de I5NP en primates no humanos para suscitar efectos adversos, que en los monos se relacionaron de manera aislada con efectos directos en la sangre que incluyeron una activación subclínica del complemento y un ligero incremento de los tiempos de coagulación. En las ratas, no se observaron efectos adversos

adicionales con un análogo de I5NP para ratas, lo que indica que los efectos representan probablemente efectos de clase de ARN sintético bicatenario más que la toxicidad relacionada con la actividad farmacológica prevista de I5NP. En conjunto, estos datos respaldan la prueba clínica de la administración intravenosa de I5NP para preservar la función renal tras la lesión por isquemia aguda/reperfusión. El nivel sin efecto adverso observable (NOAEL, por las siglas en inglés de *No Observed Adverse Effect Level*) en monos fue de 500 mg/kg. En los monos no se observaron efectos sobre parámetros cardiovasculares, respiratorios, y neurológicos luego de la administración i.v. En niveles de dosis de hasta 25 mg/kg. Por lo tanto, se puede contemplar una dosificación similar para la administración intravenosa de CRISPR Cas a los riñones de un ser humano.

Shimizu y col. (J Am Soc Nephrol 21: 622–633, 2010) desarrollaron un sistema para el suministro dirigido de ARNip a los glomérulos mediante vehículos de poli(etilenglicol)-poli(L-lisina). El complejo ARNip/nanoportador tuvo un diámetro de aproximadamente 10 a 20 nm, un tamaño que le permitiría moverse a través del endotelio con microporos para acceder al mesangio. Tras la inyección intraperitoneal de complejos de ARNip marcado con fluorescencia/nanoportador, Shimizu et al. detectaron ARNip en la circulación sanguínea durante un tiempo prolongado. La administración intraperitoneal repetida de un complejo ARNip contra una proteína quinasa 1 activada por mitógenos (MAPK1)/nanoportador suprimió el ARNm de MAPK1 glomerular y la expresión de proteínas en un modelo de glomerulonefritis en ratones. Para el estudio de la acumulación de ARNip, se administraron a ratones BALB-c ARNip marcados con Cy5 complejados con nanoportadores de PIC (0,5 mL, 5 nmol de contenido de ARNip), ARNip marcados con Cy5 desnudos (0,5 mL, 5 nmol) o ARNip marcados con Cy5 encapsulados en HVJ-E (0,5 mL, 5 nmol de contenido de ARNip). El método de Shimizu y col. se puede aplicar al sistema direccionado hacia ácidos nucleicos que contempla una dosis de aproximadamente de 10-20 µmol de CRISPR Cas complejado con nanovehículos en aproximadamente 1-2 litros a un ser humano para la administración intraperitoneal y la suministro a los riñones.

Los métodos de suministro al riñón se resumen de la siguiente manera:

Método de suministro	Portador	ARN diana	Enfermedad	Modelo	Ensayos funcionales	Autor
Hidrodinámico/Lípido	Sistema de suministro génico in vivo TransIT-QOTAP	p85α	Lesión renal aguda	Isquemia-reperfusión	Captación, biodistribución	Larson y col., Surgery, (Ago 2007), Vol. 142, No. 2, pp. (262-268)
Hidrodinámico/Lípido	Lipofectamina 2000	Fas	Lesión renal aguda	Isquemia-reperfusión	Nitrógeno ureico en sangre, inmunohistoquímica Fas, apoptosis, calificación histológica	Hamar y col., Proc Natl Acad Sci, (Oct 2004), Vol. 101, No. 41, pp. (14883-14888)
Hidrodinámico	n.a.	Elementos de la cascada de apoptosis	Lesión renal aguda	Isquemia-reperfusión	n.a.	Zheng y col., Am J Pathol, (Oct 2008), Vol. 173, No. 4, pp. (973-980)
Hidrodinámico	n.a.	Factor nuclear kappa-b (NFkB)	Lesión renal aguda	Isquemia-reperfusión	n.a.	Feng y col., Transplantation, (May 2009), Vol. 87, No. 9, pp. (1283-1289)
Hidrodinámico/Vírico	Lipofectamina 2000	Factor de transcripción que actúa como antagonista respecto a la apoptosis (AATF)	Lesión renal aguda	Isquemia-reperfusión	Apoptosis, estrés oxidativo, activación de caspasas, peroxidación de lípidos de membrana	Xie & Guo, Am Soc Nephrol. (Dic 2006), Vol. 17, No. 12, pp. (3336-3346)
Hidrodinámico	Sistema de suministro hidrodinámico pBAsi mU6 Neo/ TransIT-EE	Gremlin	Nefropatía diabética	Diabetes inducida por estreptozotocina	Proteinuria, creatinina sérica, diámetro glomerular y tubular, expresión de colágeno de tipo IV/BMP7	Q. Zhang y col., PloS ONE, (Jul 2010), Vol. 5, No. 7, e11709, pp. (1-13)
Vírico/Lípido	vector pSUPER.Lipolectamina	Receptor de tipo II de TGF-β	Fibrosis renal intersticial	Obstrucción ureteral unilateral	Expresión de α-SMA, contenido de colágeno	Kushibiki y col., J Controlled Release, (Jul 2005), Vol. 105, No. 3, pp. (318-331)
Vírico	Virus adenoasociado 2	Receptor de corticoides minerales	Daño renal causado por la hipertensión	Hipertensión inducida por sal frío	Presión sanguínea, albúmina sérica, nitrógeno ureico sérico, creatinina sérica, peso del riñón, sodio urinario	Wang y col., Gene Therapy, (Jul 2006), Vol. 13, No. 14, pp. (1097-1103)

Método de suministro	Portador	ARN diana	Enfermedad	Modelo	Ensayos funcionales	Autor
Hidrodinámico/Vírico	Vector pU6	Luciferasa	n.a.	n.a.	Captación	Kobayashi y col., Journal of Pharmacology and Experimental Therapeutics, (Feb 2004), Vol. 308, No. 2, pp. (688-693)
Lípido	Lipoproteínas, albúmina	apoB1, apoM	n.a.	n.a.	Captación, afinidad de unión a lipoproteínas y albúmina	Wolfrum y col., Nature Biotechnology, (Sep 2007), Vol. 25, No. 10, pp. (1149-1157)
Lípido	Lipofectamina 2000	p53	Lesión renal aguda	Lesión aguda inducida por cisplatino e isquemia	Calificación histológica, apoptosis	Molitoris y col., J Am Soc Nephrol (agosto de 2009), Vol. 20, n.º 8, pp. (1754-1764)
Lípido	DOTAP/DOPE, DOTAP/DO PE/DOPE-PEG2000	COX-2	Adenocarcinoma de mama	Ratón que porta un xenoinjerto de cáncer de mama MDA-MB-231	Viabilidad celular, captación	Mikhaylova y col., Cancer Gene Therapy, (Mar 2011), Vol. 18, No. 3, pp. (217-226)
Lípido	Colesterol	12/15-lipoxigenasa	Nefropatía diabética	Diabetes inducida por estreptozotocina	Albuminuria, creatinina urinaria, histología, colágeno de tipo I y IV, TGF-β, fibronectina, inhibidor del activador del plasminógeno 1	Yuan y col., Am J Physiol Renal Physiol, (Jun 2008), Vol. 295, pp. (F605-F617)
Lípido	Lipofectamina 2000	Membrana mitocondrial 44 (TIM44)	Nefropatía diabética	Diabetes inducida por estreptozotocina	Proliferación y apoptosis celular, histología, ROS, importe mitocondrial de Mn-SOD y glutatión-peroxidasa, polarización de la membrana celular	Y. Zhang y col., J Am Soc Nephrol, (Abr 2006), Vol. 17, No. 4, pp. (1090--1101)
Hidrodinámico/Lípido	Proteoliposoma	RLIP76	Carcinoma renal	Ratón que porta un xenoinjerto de cáncer de riñón Caki-2	Captación	Singhal y col., Cancer Res, (May 2008), Vol. 68, No. 10, pp. (4244-4251)

Método de suministro	Portador	ARN diana	Enfermedad	Modelo	Ensayos funcionales	Autor
Polímero	PEI PEGilado	Luciferasa pGL3	n.a.	n.a.	Captación, biodistribución, agregación de eritrocitos	Malek y col., Toxicology and Applied Pharmacology, (Abr 2009), Vol. 236, No. 1, pp. (97-108)
Polímero	PEGilado poli-L-lisina	MAPK1	Glomerulonefritis del lupus	Glomerulonefritis	Proteinuria, glomerulosclerosis, TGF- β, fibronectina, inhibidor del activador del plasminógeno 1	Shimizu y col., J Am Soc Nephrology, (abr 2010), Vol. 21, No. 4, pp. (622-633)
Polímero/Nanopartícula	Ácido hialurónico/Punto cuántico/PEI	VEGF	Cáncer de riñón/melanoma	Ratón que porta un tumor de tipo melanoma B16F1	Biodistribución, citotoxicidad, volumen tumoral, endocitosis	Jiang y col., Molecular Pharmaceutics, (May-Jun 2009), Vol. 6, No. 3, pp. (727-737)
Polímero/Nanopartícula	Nanofibra de policaprolactona PEGilada	GAPDH	n.a.	n.a.	Viabilidad celular, captación	Cao et al. J Controlled Release, (Jun 2010), Vol. 144, No. 2, pp. (203-212)
Aptámero	Spiegelmer mNOX-E36	Ligando quimioquina CC 2	Esclerosis glomerular	Ratón sometido a uninefrectomía	Albúmina urinaria, creatinina urinaria, histopatología, tasa de filtración glomerular, recuento de macrófagos, Cd12 en suero, Mac 2+, Ki-67+	Ninichuk y col., Am J Pathol, (Mar 2008), Vol. 172, No. 3, pp. (628-637)
Aptámero	Aptámero NOX-F37	Vasopresina (AVP)	Insuficiencia cardíaca congestiva	n.a.	Afinidad de unión a D-AVP, inhibición de la señalización de AVP, Osmolalidad urinaria y Concentración de sodio	Purschke y col., Proc Natl Acad Sci, (Mar 2006), Vol. 103, No. 13, pp. (5173-5178)

Direccionamiento hacia el hígado o células hepáticas

Se provee el direccionamiento hacia células hepáticas. Esto puede ser *in vitro* o *in vivo (no reivindicado en la presente).* Se prefieren los hepatocitos. El suministro de la proteína CRISPR, por ejemplo, Cpf1 aquí se puede realizar mediante vectores virales, especialmente vectores AAV (y en particular AAV2/6). Estos se pueden administrar por inyección intravenosa.

Un diana preferido para el hígado, ya sea *in vitro* o *in vivo (no reivindicado en la presente)* es el gen de la albúmina. Este es un denominado "puerto seguro" ya que la albúmina se expresa a niveles muy altos y por lo tanto se puede tolerar cierta reducción de la producción de albúmina luego de una exitosa edición del gen. También se prefiere que los altos niveles de expresión que se ven con el promotor/potenciador de albúmina permite obtener niveles útiles de producción correcta o del transgen (del molde del donante insertado) aún si solo se edita una pequeña fracción de los hepatocitos.

Wechsler y col. (informado en la 57th Annual Meeting and Exposition of the American Society of Hematology - resumen disponible en línea en https://ash.confex.com/ash/2015/webprogram/Paper86495.html y presentado el 6 de diciembre de 2015) han mostrado que el intrón 1 de albúmina es un sitio apropiado como diana. En su trabajo utilizaron dedos de Zn para cortar el ADN en este sitio diana, y se pueden generar secuencias guía apropiadas para guiar el clivaje en el mismo sitio por una proteína CRISPR.

El uso de dianas dentro de genes altamente expresados (genes con potenciadores/promotores altamente activos), tales como la albúmina, también puede permitir el uso de un molde del donante sin promotor, como han informado Wechsler y col. y esto también se puede aplicar ampliamente aparte del direccionamiento hacia el hígado. Se conocen otros ejemplos de genes altamente expresados.

Otras enfermedades del hígado

En los métodos descritos, pero no reivindicados en la presente, las proteínas CRISPR se utilizan en el tratamiento de trastornos del hígado tales como amiloidosis por transtiretina (ATTR), deficiencia de alfa-1 antitripsina y otros errores congénitos de base hepática del metabolismo. La FAP es causad por una mutación en el gen que codifica la transtiretina (TTR). Aunque es una enfermedad autosómica dominante, no todos los portadores desarrollan la enfermedad. Hay más de 100 mutaciones en el gen de TTR que se sabe que están asociadas con la enfermedad. Algunos ejemplos de mutaciones comunes incluyen a V30M. El principio del tratamiento de TTR basado en el silenciamiento de genes se ha demostrado mediante estudios con ARNi (Ueda y col., 2014 Transl Neurogener. 03:19). La enfermedad de Wilson (WD) está causada por mutaciones en el gen que codifica a ATP7B, que se encuentra exclusivamente en el hepatocito. Hay más de 500 mutaciones asociadas con la WD, con mayor predominio en regiones específicas, tales como el Este de Asia. Otros ejemplos son A1ATD (una enfermedad autosómica recesiva causada por mutaciones en el gen SERPINA1) y PKU (una enfermedad autosómica recesiva causada por mutaciones en el gen de la fenilalanina hidroxilasa (PAH)).

Trastornos de la sangre asociados al hígado, especialmente hemofilia y en particular hemofilia B

Se ha conseguido una exitosa edición del gen de hepatocitos en ratones (tanto *in vitro* como *in vivo*) y en primates no humanos (*in vivo*), lo que muestra que el tratamiento de trastornos de la sangre por edición del gen /modificación del genoma en hepatocitos es factible. En particular, se ha mostrado que la expresión del gen F9 (hF9) humano en hepatocitos en primates no humanos es indicativa de un tratamiento para la hemofilia B en seres humanos.

Wechsler y col. informaron en la 57th Annual Meeting and Exposition of the American Society of Hematology (resumen presentado el 6 de diciembre de 2015 y disponible en línea en https://ash.confex.com/ash/2015/webprogram/Paper86495.html) que pudieron expresar con éxito el F9 humano (hF9) de hepatocitos en primates no humanos por edición *in vivo* del gen. Esto se realizó usando 1) dos nucleasas con dedos de zinc (ZFNs) direccionadas hacia el intrón 1 del locus de la albúmina, y 2) un molde de F9 humano de la construcción del donante. Los ZFNs y el molde del donante se codificaron en vectores de virus adeno-asociados hepatotrópicos del serotipo 2/6 (AAV2/6) separados inyectados por vía intravenosa, para obtener una inserción direccionada de una copia corregida del gen hF9 dentro del locus de la albúmina en una proporción de hepatocitos del hígado.

El locus de la albúmina se seleccionó como un "puerto seguro" ya que la producción de esta proteína más abundante en plasma supera los 10 g/día, y unas reducciones moderadas de dichos niveles se toleran bien. Los hepatocitos con el genoma editado produjeron hFIX (hF9) normales en cantidades terapéuticas, en vez de albúmina, impulsado por el potenciador/promotor de albúmina altamente activo. La integración direccionada del transgen hF9 en el locus de la albúmina y se mostró el empalme de dicho gen en el transcripto de la albúmina.

Estudios en ratones: A unos ratones C57BL/6 se les administró vehículo (n= 20) o vectores AAV2/6 (n= 25) que codifican reactivos sustituto de ratón a $1,0 \times 10^{13}$ de genoma del vector (vg)/kg por inyección en la vena de la cola. El análisis por ELISA de hFIX en plasma en los ratones tratados mostró niveles pico de 50-1053 ng/ml que se sostuvieron durante la duración del estudio de 6 meses. El análisis de actividad de FIX en plasma de ratón confirmó una bioactividad proporcional a los niveles de expresión.

Estudios en primates no humanos (NHP): una única co-infusión intravenosa de vectores AAV2/6 que codifican a los ZFNs específicos para albúmina direccionados hacia NHP y un donante de F9 humano a $1,2 \times 10^{13}$ vg/kg (n= 5/grupo) dio como resultado >50 ng/ml (>1% de lo normal) en este modelo animal grande. El uso de mayores dosis de AAV2/6 (de hasta 1,5 × 10^14 vg/kg) dio niveles de hFIX en plasma de hasta 1000 ng/ml (o 20% de lo normal) en varios animales y de hasta 2000 ng/ml (o 50% de lo normal) en un único animal, durante la duración del estudio (3 meses).

El tratamiento fue bien tolerado en ratones y NHPs, sin hallazgos toxicológicos significativos relacionados con el tratamiento con AAV2/6 ZFN + donante en todas las especies en dosis terapéuticas. Sangamo (CA, EE.UU.) desde entonces ha enviado una solicitud a la FDA, y se le ha otorgado, permiso de conducir la primera prueba clínica en humanos del mundo de una aplicación en edición de un genoma *in vivo*. Esto sigue a la aprobación de EMEA del tratamiento de terapia génica Glibera de deficiencia de lipoproteína lipasa.

Por lo tanto, en algunas formas de realización es preferible usar cualquiera de los siguientes, o todos ellos:

- Vectores AAV (especialmente AAV2/6), preferiblemente administrados por inyección intravenosa;
- Albúmina como diana para la edición del gen /inserción de transgen/molde - especialmente en el intrón 1 de albúmina;
- Molde de F9 humano del donante; y/o
- Un molde del donante sin promotor.

Hemofilia B

Por lo tanto, en algunas formas de realización de los métodos descritos, pero no reivindicados en la presente, es preferible que los métodos descritos se utilicen para tratar la hemofilia B. Por lo tanto, es preferible proveer un molde y el mismo es el gen F9 humano. Se apreciará que el molde de hF9 comprende la versión wt o 'correcta' de hF9 de manera tal que el tratamiento es eficaz.

En una forma de realización alternativa de los métodos descritos, pero no reivindicados en la presente, la versión de F9 de la hemofilia B se puede suministrar de manera tal de crear un organismo modelo, célula o línea celular (por ejemplo un organismo modelo, célula o línea celular murino o primate no humano), el organismo modelo, célula o línea celular con el fenotipo de hemofilia B o que lo porta, es decir una incapacidad de producir F9 wt.

Hemofilia A

En algunas formas de realización, el gen F9 (factor IX) puede ser reemplazado por el gen F8 (factor VIII) que se describió anteriormente, que permite realizar el tratamiento de la hemofilia A (proporcionando un gen F8 correcto) y/o la creación de un organismo, célula o línea celular modelo de hemofilia (proporcionando una versión incorrecta con hemofilia A del gen F8).

Hemofilia C

En algunas formas de realización, el gen F9 (factor IX) puede ser reemplazado por el gen F11 (factor XI) que se describió anteriormente, permitiendo realizar un tratamiento de la hemofilia C (proporcionando un gen F11 correcto) y/o la creación de un organismo, célula o línea celular modelo de hemofilia C (proporcionando una versión incorrecta, con hemofilia C del gen F11).

Tratamiento de enfermedades epiteliales y pulmonares

También se describen, pero no se reivindican en la presente, métodos de suministro del sistema CRISPR-Cas que se describe aquí, por ejemplo, sistemas de proteína efectora Cpf1, a uno o ambos pulmones.

Aunque los vectores derivados de AAV-2 se propusieron en un primer lugar para el suministro de CFTR a las vías respiratorias con CF, otros serotipos tales como AAV-1, AAV-5, AAV-6 y AAV-9 mostraron una eficacia de transferencia génica mejorada en varios modelos del epitelio pulmonar (véase, por ejemplo, Li y col., Molecular Therapy, vol. 17 n.º 12, 2067-2077 diciembre de 2009). Se demostró que AAV-1 es ~100 veces más eficiente que AAV-2 y AAV-5 en la transducción *in vitro* de células epiteliales de las vías respiratorias humanas, aunque AAV-1 transdujo *in vivo* epitelios de las vías respiratorias traqueales murinas con una eficiencia igual a la de AAV-5. Otros estudios mostraron que AAV-5 es 50 veces más eficiente que AAV-2 para la administración *in vitro* de genes al epitelio de las vías respiratorias humanas (HAE) y significativamente más eficiente en el epitelio de las vías respiratorias pulmonares del ratón *in vivo*. También se ha demostrado que AAV-6 es más eficaz que AAV-2 en las células epiteliales de las vías respiratorias humanas *in vitro* y en las vías respiratorias murinas *in vivo*.8 Se ha demostrado que el aislado más reciente, AAV-9, muestra una eficacia de transferencia génica mayor que AAV-5 en epitelio alveolar y nasal murino in vivo donde se detectó expresión génica durante más de 9 meses, lo que sugiere que AAV podrá posibilitar la expresión génica a largo plazo in vivo, una propiedad deseable para un vector de suministro génico de CFTR. Además, se ha demostrado que AAV-9 se podría volver a administrar al pulmón murino sin que haya pérdida de la expresión de CFTR y con consecuencias inmunitarias mínimas. Los cultivos de HAE de CF y no CF se podrán inocular en la superficie apical

con 100 µL de vectores de AAV durante horas (véase, por ejemplo, Li y col., Molecular Therapy, vol. 17 no. 12, 2067-2077 diciembre de 2009). La MOI puede variar entre 1×10^3 y 4×10^5 genomas de vector/célula, dependiendo de la concentración del virus y los propósitos de los experimentos. Se contemplan los vectores citados anteriormente para el suministro y/o administración de la invención.

Zamora y col. (Am J Respir Crit Care Med Vol 183. Pp 531-538, 2011) informaron un ejemplo de la aplicación de un ARN de interferencia terapéutico para el tratamiento de enfermedades infecciosas humanas y también una prueba aleatorizada de una droga antiviral en infectado con virus sincitial respiratorio (RSV) en pulmón, en receptores de trasplantes. Zamora et al. realizaron un ensayo controlado con placebo, con enmascaramiento doble, aleatorizado en receptores de LTX con infección en el aparato respiratorio por RSV. Se permitió que los pacientes recibieran el tratamiento asistencial habitual para RSV. Se administró placebo o ALN-RSV01 aerosolizado (0,6 mg/kg) a diario durante 3 días. Este estudio demuestra que una ARNi terapéutica cuyo diana sea RSV se puede administrar de manera segura a receptores de LTX con infección por RSV. Tres dosis diarias de ALN-RSV01 no dieron como resultado ninguna exacerbación de los síntomas del aparato respiratorio ni deterioro de la función pulmonar y no exhibieron ningún efecto proinflamatorio sistémico, tal como inducción de citoquinas o CRP. Los estudios farmacocinéticos muestran únicamente una exposición sistémica transitoria, baja tras la inhalación, coherente con los datos preclínicos en animales que mostraron que ALN-RSV01, administrado por vía intravenosa o por inhalación, se elimina rápidamente de la circulación mediante digestión mediada por exonucleasas y excreción renal. El método de Zamora y col. se puede aplicar al sistema direccionado hacia ácidos nucleicos descrito en la presente y se puede contemplar un CRISPR Cas aerosolizado, por ejemplo, con una dosificación de 0,6 mg/kg.

Los sujetos tratados de una enfermedad pulmonar pueden recibir por ejemplo una cantidad farmacéuticamente eficaz de sistema de vector AAV aerosolizado por pulmón suministrado endobronquialmente durante la respiración espontánea. En este sentido, se prefiere el suministro aerosolizado para el suministro de AAV en general. Se podrá utilizar un adenovirus o una partícula de AAV para el suministro. Las construcciones génicas adecuadas, cada una ligada operablemente a una o más secuencias reguladoras, se podrán clonar e introducir en el vector de suministro. En este caso, se proporcionan las siguientes construcciones como ejemplos: promotor Cbh o EF1a de Cas (Cpf1), promotor U6 o H1 de ARN guía): una disposición preferida es usar un CFTRdelta508 direccionado hacia guía, un molde de reparación para la mutación deltaF508 y una enzima Cpf1 optimizada por codones, que opcionalmente tiene uno o más señal o secuencia(s) de localización nuclear (NLS(s)), por ejemplo, dos (2) NLSs. También se conciben construcciones sin NLS.

<u>Tratamiento de enfermedades del sistema muscular</u>

También se describen, pero no se reivindican en la presente, métodos que comprenden el suministro del sistema CRISPR-Cas que se describe aquí, por ejemplo sistemas de proteína efectora Cpf1, al(a los) músculo(s).

Bortolanza y col. (Molecular Therapy vol. 19 no. 11, 2055-2064 Nov. 2011) muestra que la administración sistémica de cassettes de expresión de ARN de interferencia al ratón FRG1, luego del inicio de distrofia muscular facioescapulohumeral (FSHD), lleva a un *knockdown* dependiente de la dosis a largo plazo de FRG1 sin señales de toxicidad. Bortolanza y col. descubrieron que una única inyección intravenosa de 5×10^{12} vg de rAAV6-sh1FRG1 rescata la histopatología muscular y función muscular de los ratones FRG1. En detalle, se inyectaron 200 µl que contienen 2×10^{12} o 5×10^{12} vg de vector en solución fisiológica en la vena de la cola usando una jeringa Terumo número 25. El método de Bortolanza y col. se puede aplicar a un AAV que expresa CRISPR Cas y se inyectaron en seres humanos con una dosificación de aproximadamente 2×10^{15} o 2×10^{16} vg de vector.

Dumonceaux y col. (Molecular Therapy vol. 18 no. 5, 881-887 mayo de 2010) inhibieron la vía de miostatina usando la técnica del ARN de interferencia dirigido contra el ARNm del receptor de miostatina AcvRIIb (sh-AcvRIIb). La restauración de una cuasidistrofina estuvo mediada por la técnica para omitir exones U7 vectorizada (U7-DYS). Los vectores adenoasociados que portan la construcción sh-AcvrIIb solo, la construcción U7-DYS solo o una combinación de ambas construcciones se inyectaron en el músculo tibial posterior (TA, por sus siglas en inglés) de ratones mdx distróficos. Las inyecciones se llevaron a cabo con 10^{11} Genomas virales de AAV. El método de Dumonceaux y col. se puede aplicar a un AAV que expresa CRISPR Cas y se inyecta en seres humanos, por ejemplo, con una dosificación de entre aproximadamente 10^{14} y aproximadamente 10^{15} vg de vector.

Kinouchi y col. (Gen Therapy (2008) 15, 1126-1130) informan la eficacia de la administración *in vivo* de ARNip en músculos esqueléticos de ratones normales o enfermos por formación de nanopartículas de ARNip sin modificaciones químicas con atelocolágeno (ATCOL). La aplicación localizada mediada por ATCOL de ARNip cuyo diana es la miostatina, un regulador negativo del crecimiento del músculo esquelético, en músculos esqueléticos de ratón o por vía intravenosa, provocó un incremento notable en la masa muscular pocas semanas después de la aplicación. Estos resultados dan a entender que la aplicación mediada por ATCOL de ARNip es una herramienta poderosa para un futuro uso terapéutico en enfermedades incluida la atrofia muscular. Se mezclaron Mst-ARNip (concentración final, 10 mM) con ATCOL (concentración final para la administración localizada, 0,5%) (AteloGene, Kohken, Tokio, Japón) de acuerdo con las instrucciones del fabricante. Después de anestesiar ratones (machos C57BL/6 de 20 semanas de edad) mediante Nembutal (25 mg/kg, i.p.), el complejo Mst-ARNip/ATCOL se inyectó en los músculos masetero y bíceps femoral. Se podrá aplicar el método de Kinouchi y col. a un CRISPR Cas e inyectarlo en un ser humano con, por ejemplo, una dosificación de aproximadamente 500 a 1000 mL de una solución 40 µM en el músculo. Hagstrom y

col. (Molecular Therapy Vol. 10, No. 2, agosto de 2004) describe una metodología intravascular no viral que permite una administración eficiente y repetible de ácidos nucleicos a células de músculo (miofibras) a todos los músculos de los miembros de mamíferos. El procedimiento conlleva la inyección de ARNip o ADN plasmídico desnudo en una vena distal de una extremidad que se aísla temporalmente mediante un torniquete o un manguito de esfingomanómetro. El suministro del ácido nucleico a las miofibras está facilitado por su inyección rápida en un volumen suficiente para permitir la extravasación de la solución de ácido nucleico en el tejido muscular. Se lograron niveles elevados de la expresión del transgén en el músculo esquelético tanto en animales pequeños como grandes con una toxicidad mínima. También se obtuvo evidencia del suministro de ARNip a los músculos de las extremidades. Para la inyección intravenosa de ADN plasmídico en un mono rhesus, se conectó una llave de paso de tres vías a dos bombas de jeringa (Modelo PHD 2000; Harvard Instruments), cada una equipada con una jeringa única. Cinco minutos después de una inyección de papaverina, se inyectó ADNp (entre 15,5 y 25,7 mg en 40 -100 ml de solución salina) con una velocidad de 1,7 o 2,0 ml/s. Esto se puede aumentar de escala para ADN plasmídico que expresa el CRISPR Cas con una inyección de entre aproximadamente 300 y 500 mg en entre 800 y 2000 ml de solución salina para un ser humano. Para las inyecciones del vector adenoviral en rata, se inyectaron 2×10^9 partículas infecciosas en 3 ml de solución salina normal (NSS). Esto se puede aumentar de escala para un vector adenoviral que expresa el CRISPR Cas con una inyección de aproximadamente 1×10^{13} partículas infecciosas inyectada en 10 litros de NSS para un ser humano. Para el ARNip, se inyectó una rata en la vena safena magna con 12,5 µg de un ARNip y se inyectó un primate en la vena safena magna con 750 µg de ARNip. Esta escala se pudo aumentar para un CRISPR Cas, por ejemplo, con una inyección de aproximadamente 15 a aproximadamente 50 mg en la vena safena interna de un ser humano.

Por ejemplo, WO2013163628 A2, Genetic Correction of Mutated Genes, solicitud publicada de la Universidad de Duke describe los esfuerzos por corregir, por ejemplo, una mutación con desplazamiento del marco de lectura que provoca un codón de parada prematura y un producto génico truncado que se puede corregir mediante la unión de extremos no homólogos mediada por nucleasas tal como las responsables de la distrofia muscular de Duchenne ("DMD") un trastorno ligado al cromosoma X recesivo y letal que da como resultado degeneración muscular debida a mutaciones en el gen de la distrofina. La mayoría de las mutaciones en la distrofina que causan DMD son supresiones de exones que alteran el marco de lectura y causan una terminación de la traducción prematura en el gen de la distrofina. La distrofina es una proteína citoplasmática que proporciona estabilidad estructural al complejo distroglicano de la membrana celular que es responsable de regular la función e integridad de las células musculares. El gen de la distrofina o "gen DMD", utilizado indistintamente en la presente, tiene 2,2 megabases en el locus Xp21. La transcripción primaria mide aproximadamente 2400 kb teniendo el ARNm maduro aproximadamente 14 kb. 79 exones codifican la proteína que tiene más de 3500 aminoácidos. El exón 51 está frecuentemente adyacente a la supresión que altera el marco de lectura en pacientes con DMD y ha sido el diana de los ensayos clínicos para omitir exones basada en los oligonucleótidos. Un ensayo clínico con el compuesto eteplirsen que omite el exón 51 presentó recientemente un beneficio funcional significativo a lo largo de 48 semanas, con un promedio de un 47% de fibras positivas para la distrofina en comparación con los valores iniciales. Las mutaciones en el exón 51 son especialmente adecuadas para la corrección permanente por edición genómica mediante NHEJ.

Los métodos de la Publicación de Patente de los EE.UU. n.º 20130145487 asignada a Cellectis, que se refiere a una variante de meganucleasa para clivar una secuencia diana del gen de distrofina humano (DMD), también se pueden modificar para el sistema de direccionamiento hacia ácidos nucleicos descrito en la presente.

Tratamiento de enfermedades de la piel

La solicitud también describe, pero no reivindica, el suministro del sistema CRISPR-Cas que se describe aquí, por ejemplo, sistemas de proteína efectora Cpf1, a la piel.

Hickerson y col. (Molecular Therapy – Nucleic Acids (2013) 2, e129) se refieren a un dispositivo para administración a la piel con un conjunto de microagujas motorizado para la autoadministración de (sd)-ARNip a la piel de un ser humano o ratón. El desafío principal para adaptar los tratamientos de la piel con ARNip al ámbito de la atención sanitaria es el desarrollo de sistemas de suministro eficaces. Se han invertido esfuerzos sustanciales en varias tecnologías de suministro a la piel con un éxito limitado. En un estudio clínico en el que se trató la piel con ARNip, el dolor insufrible asociado con la inyección con una aguja hipodérmica imposibilitó la participación de pacientes adicionales en el ensayo, y resaltó cuán necesarias son estrategias de suministro mejoradas, más "cómodas para el paciente" (es decir, sin dolor o con poco dolor). Las microagujas representan una vía eficaz de suministrar cargamentos cargados grandes que incluyen ARNip a través de la barrera primaria, la capa córnea, y se consideran en general como menos dolorosas que las agujas hipodérmicas convencionales. Se ha demostrado que los dispositivos de microagujas de "tipo sello" motorizados, incluido el dispositivo de matriz de microagujas motorizado (MMNA, por sus siglas en inglés) utilizado por Hickerson y col., son seguros en estudios en ratones sin pelo y no causaron dolor, o muy poco, como lo evidencia (i) el uso extendido en la industria cosmética y (ii) la prueba limitada en la que casi todos los voluntarios encontraron el uso del dispositivo mucho menos doloroso que una inyección de la gripe, lo que sugiere que el suministro de ARNip utilizando este dispositivo conllevará mucho menos dolor del que se experimentó en los ensayos clínicos previos utilizando inyecciones con una agua hipodérmica. El dispositivo MMNA (comercializado como Triple-M o Tri-M por Bomtech Electronic Co, Seúl, Corea del Sur) se adaptó para el suministro de ARNip a piel de un ser humano y de ratón. Se introdujo una solución de ARNip-as (hasta 300 µL de 0,1 mg/mL de ARN) en la cámara de un cartucho de aguja Tri-M desechable (Bomtech), que se ajustó a una profundidad de 0,1 mm. Para tratar la piel humana, se estiró manualmente piel desidentificada (obtenida justo después de procedimientos quirúrgicos) y se clavó a una plataforma

de corcho antes del tratamiento. Se realizaron inyecciones intradérmicas utilizando una jeringa de insulina con una aguja de 0,5 pulgadas de calibre 28. El dispositivo MMNA y el método de Hickerson et al. se podrían utilizar y/o adaptar para suministrar el CRISPR Cas, por ejemplo, con una dosificación de hasta 300 µL de 0,1 mg/mL de CRISPR Cas a la piel.

Leachman y col. (Molecular Therapy, vol. 18 no. 2, 442-446 febrero de 2010) se refiere a una prueba clínica de fase Ib para el tratamiento de un trastorno de la piel poco frecuente, la paquioniquia congénita (PC), un síndrome autosómico dominante que incluye a la queratodermia plantar incapacitante, utilizando el primer terapéutico basado en ARN de interferencia pequeño (ARNip) para la piel. Este ARNip, denominado TD101, que tiene como diana potente y específicamente el ARNm mutante N171K de queratina 6a (K6a) sin afectar al ARNm de K6a no modificado.

Zheng y col. (PNAS, 24 de julio de 2012, vol. 109, no. 30, 11975-11980) muestra que nanopartículas esféricas de conjugados con ácido nucleico (SNA-NCs), núcleos de oro rodeados de un revestimiento denso de ARNip altamente orientado, inmovilizado covalentemente, penetran libremente casi el 100% de los queratinocitos *in vitro*, piel de ratón, y epidermis humana dentro de un período de horas luego de la aplicación. Zheng et al. demostraron que una única aplicación de SNA-NC del receptor del factor de crecimiento epidérmico 25 nM durante 60 h demostraba una inactivación génica eficaz en la piel humana. Se podrá contemplar una dosificación similar para CRISPR Cas inmovilizado en SNA-NC para la administración a la piel.

Cáncer

En algunas formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico del cáncer. El diana es preferiblemente uno o más de los genes FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC o TRBC. El cáncer puede ser uno o más de: linfoma, leucemia linfocítica crónica (CLL), leucemia linfocítica aguda asociada a células B (B-ALL), leucemia linfoblástica aguda, leucemia mieloide aguda, linfoma no de Hodgkin (NHL), linfoma difuso de células grandes (DLCL), mieloma múltiple, carcinoma de células renales (RCC), neuroblastoma, cáncer colorrectal, cáncer de mama, cáncer de ovario, melanoma, sarcoma, cáncer de próstata, cáncer de pulmón, cáncer de esófago, carcinoma hepatocelular, cáncer pancreático, astrocitoma, mesotelioma, cáncer de cabeza y cuello, y meduloblastoma. Esto se puede implementar con células T con receptores antigénicos quiméricos modificados (CAR). Esto se describe en WO2015161276, cuya divulgación se describe aquí más adelante.

Los genes diana apropiados para el tratamiento o la profilaxis del cáncer pueden incluir, en algunas formas de realización, a aquellos que se describen en WO2015048577.

Síndrome de Usher o retinitis pigmentosa-39

En algunas formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico del síndrome de Usher o retinitis pigmentosa-39. El diana es preferiblemente el gen USH2A. En algunas formas de realización, se provee la corrección de una supresión G en la posición 2299 (2299delG). Esto se describe en WO2015134812A1.

Fibrosis quística (CF)

En algunas formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico de la fibrosis quística. El diana es preferiblemente el gen SCNN1A o el gen CFTR. Esto se describe en WO2015157070.

Schwank y col. (*Cell Stem Cell*, 13:653-58, 2013) utilizaron CRISPR-Cas9 para corregir un defecto asociado a la fibrosis quística en células indiferenciadas humanas. El objetivo del equipo fue el gen para un canal iónico, receptor conductor transmembrana de la fibrosis quística (CFTR, por sus siglas en inglés). Una supresión en CFTR provoca que la proteína se pliegue de manera errónea en pacientes con fibrosis quística. Utilizando células madre intestinales cultivadas desarrolladas a partir de muestras celulares procedentes de dos niños con fibrosis quística, Schwank y col., fueron capaces de corregir el defecto utilizando CRISPR junto con un plásmido donante que contenía la secuencia reparadora que se iba a insertar. Los investigadores cultivaron posteriormente las células para obtener "organoides" intestinales, o intestinos en miniatura, y mostraron que funcionaban normalmente. En este caso, aproximadamente la mitad de los organoides clonales experimentaron la corrección genética oportuna.

VIH y SIDA

En algunas formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico de VIH y SIDA. El diana es preferiblemente el gen CCR5 en el VIH. Esto se describe en WO2015148670A1.

Beta talasemia

En algunas formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico de la Beta talasemia. El diana es preferiblemente el gen BCL11A. Esto se describe en WO2015148860.

Anemia de células falciformes (SCD)

En algunas formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico de la anemia de células falciformes (SCD, por las siglas en inglés de *Sickle Cell Disease*). El diana es preferiblemente el gen HBB o el gen

BCL11A. Esto se describe en WO2015148863.

Virus de herpes simple 1 y 2

En algunas formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico del HSV-1 (por las siglas en inglés de *Herpes Simplex Virus 1*, es decir virus de herpes simple 1). El diana son preferiblemente los genes UL19, UL30, UL48 o UL50 en el HSV-1. Esto se describe en WO2015153789.

En otras formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico de HSV-2 (Virus de herpes simple 2). El diana son preferiblemente los genes UL19, UL30, UL48 o UL50 en el HSV-2. Esto se describe en WO2015153791.

En algunas formas de realización, se provee el tratamiento, la profilaxis o el diagnóstico del glaucoma primario de ángulo abierto (POAG, por las siglas en inglés de *Primary Open Angle Glaucoma*). El diana es preferiblemente el gen MYOC. Esto se describe en WO2015153780.

Terapias celulares adoptivas

También se describe, pero no se reivindica en la presente, el uso del sistema CRISPR-Cas que se describe aquí, por ejemplo, sistemas de la proteína efectora Cpf1, para modificar células para terapias adoptivas. Por lo tanto, algunos aspectos de la invención se relacionan con la transferencia adoptiva de células del sistema inmunitario, por ejemplo, células T específicas para determinados antígenos, por ejemplo, antígenos asociados a tumores (véase Maus y col., 2014, *Adoptive Immunotherapy for Cancer or Viruses, Annual Review of Immunology*, Vol. 32: 189-225; Rosenberg y Restifo, 2015, *Adoptive cell transfer as personalized immunotherapy for human cancer*, Science Vol. 348 no. 6230 pp. 62-68; y, Restifo y col., 2015, *Adoptive immunotherapy for cancer: harnessing the T cell response*. Nat. Rev. Immunol. 12(4): 269-281); y Jenson and Riddell, 2014, *Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells*. Immunol Rev. 257(1): 127–144). Diversas estrategias se pueden emplear por ejemplo para modificar genéticamente las células T por alteración de la especificidad de los receptores de células T (TCR) por ejemplo por introducción de nuevas cadenas α y β de los TCR con especificidad por determinados péptidos (véase la Patente de los EE.UU. n.º 8.697.854; Publicaciones de Patente PCT: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; Patente de los EE.UU. n.º 8.088.379).

Como una alternativa a las modificaciones de TCR, o además de las mismas,, se pueden utilizar receptores de antígenos quiméricos (CARs) para generar células inmuno respondentes, por ejemplo células T, específicas para determinados blancos, por ejemplo células T malignas, con una amplia variedad de construcciones de receptores quiméricos que ya han sido descritos (véanse las Patentes de los EE.UU. n.º 5.843.728; 5.851.828; 5.912.170; 6.004.811; 6.284.240; 6.392.013; 6.410.014; 6.753.162; 8.211.422; y, Publicación PCT WO9215322). Otras construcciones de CAR alternativas se pueden caracterizar como pertenecientes a sucesivas generaciones. Los CARs de primera generación típicamente consiste en un fragmento variable de cadena simple de un anticuerpo específico para un antígeno, por ejemplo que comprende un VL conectado a un VH de un anticuerpo específico, conectado por un conector flexible, por ejemplo por un dominio bisagra CD8α y un dominio CD8α transmembrana, a los dominios de señalización transmembrana e intracelulares ya sea de CD3ζ o FcRγ (scFv-CD3ζ o scFv-FcRγ; véase la Patente de los EE.UU. n.º 7.741.465; Patente de los EE.UU. n.º 5.912.172; Patente de los EE.UU. n.º 5.906.936). Los CARs de segunda generación incorporan los dominios intracelulares de una o más moléculas coestimulantes, por ejemplo, CD28, OX40 (CD134), o 4-1BB (CD137) dentro del endodominio (por ejemplo, scFv-CD28/OX40/4-1BB-CD3ζ; véanse las Patentes de los EE.UU. n.º 8.911.993; 8.916.381; 8.975.071; 9.101.584; 9.102.760; 9.102.761). Los CARs de tercera generación incluye una combinación de endodominios coestimulantes, tales como una cadena CD3ζ, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, o CD28 dominios de señalización (por ejemplo, scFv-CD28-4-1BB-CD3ζ o scFv-CD28-OX40-CD3ζ; véase la Patente de los EE.UU. n.º 8.906.682; Patente de los EE.UU. n.º 8.399.645; U.S. Pat. n.º 5.686.281; Publicación PCT n.º WO2014134165; Publicación PCT n.º WO2012079000). Como alternativa, la coestimulación se puede orquestar por expresión de CARs en células T específicas para un antígeno, seleccionados de manera tal de que se activen y expandan luego de acoplarse con su αβTCR nativo, por ejemplo por células antigénicas en la presentación profesional de antígenos, con la concomitante coestimulación. Además, en las células inmuno respondentes se pueden proveer receptores modificados adicionales, por ejemplo para mejorar el direccionamiento del ataque hacia una célula T y/o minimizar los efectos secundarios.

Se pueden utilizar técnicas alternativas para transformar células diana inmuno respondentes, por ejemplo, protoplastos de fusión, lipofección, transfección o electroporación. Se puede usar una amplia variedad de vectores, por ejemplo, vectores retrovirales, vectores lentivirales, vectores adenovirales, vectores virales adeno-asociados, plásmidos o transposones, por ejemplo, se puede utilizar un transposón Sleeping Beauty (véanse las Patentes de los EE.UU. n.º 6.489.458; 7.148.203; 7.160.682; 7.985.739; 8.227.432), para introducir CARs, por ejemplo, usando señalización por CARs antígeno-específicos de segunda generación mediante CD3ζ y ya sea CD28 o CD137. Los vectores virales pueden incluir por ejemplo vectores basados en VIH, SV40, EBV, HSV o BPV.

Las células a las que se direcciona la transformación pueden incluir por ejemplo células T, células asesinas naturales (NK), linfocitos T citotóxicos (CTL), células T regulatorias, células indiferenciadas embriónicas no humanas, linfocitos

que se infiltran en tumor (TIL) o células T indiferenciadas pluripotentes entre las cuales las células linfoides pueden estar diferenciadas. Las células T que expresan un CAR que se desea se pueden seleccionar por ejemplo a través de co-cultivo con células activantes y propagantes (AaPC) irradiadas con rayos γ, que co-expresan el antígeno de cáncer y moléculas co-estimulatorias. Las células T CAR modificadas se pueden expandir, por ejemplo, por co-cultivo en AaPC en presencia de factores solubles, por ejemplo, IL-2 y IL-21. Esta expansión se puede realizar por ejemplo de manera tal de proveer células T CAR+ de memoria (que por ejemplo se pueden evaluar mediante un arreglo digital no enzimático y/o citometría de flujo multi-panel). De esta manera, se pueden proveer células T CAR que tengan una actividad citotóxica específica contra tumores que presentan antígeno (opcionalmente en conjunto con la producción de las quimioquinas que se desean tales como el interferón-γ). Las células T CAR de esta clase se pueden utilizar por ejemplo en modelos animales, por ejemplo, para tratar xenoinjertos tumorales.

Algunos enfoques tales como el anterior se pueden adaptar para proveer métodos para tratar y/o incrementar la supervivencia de un sujeto que padece una enfermedad, por ejemplo una neoplasia, por ejemplo por administración de una cantidad eficaz de una célula inmunorrespondente que comprende un receptor que reconoce un antígeno que se une a un determinado antígeno, donde la unión activa a la célula inmuno respondente, tratando o previniendo de esa manera la enfermedad (tal como una neoplasia, una infección por un patógeno, un trastorno autoinmune, o una reacción a un trasplante alogénico). La dosificación en las terapias de células T CAR puede incluir por ejemplo la administración de entre 106 y 109 células/kg, con un curso de supresión de linfocitos o sin la misma, por ejemplo, con ciclofosfamida.

En una forma de realización, el tratamiento se puede administrar a pacientes sometidos a un tratamiento inmunosupresor. Las células o la población de células, se pueden hacer resistentes a al menos un agente inmunosupresor debido a la desactivación de un gen que codifica un receptor para dicho agente inmunosupresor. Sin quedar atado a una teoría, el tratamiento inmunosupresor debería contribuir a la selección y expansión de las células T o inmunorrespondentes dentro del paciente.

La administración de las células o la población de células se puede llevar a cabo en cualquier forma conveniente, inclusive por inhalación de aerosol, inyección, ingestión, transfusión, implante o transplante. Las células o la población de células se pueden administrar a un paciente por vía subcutánea, por vía intradérmica, intratumoral, intranodal, intramedular, intramuscular, intravenosa o mediante inyección intralinfática, o por vía intraperitoneal. En una forma de realización, las composiciones de célula se administran preferiblemente por inyección intravenosa.

La administración de las células o la población de células puede consistir en la administración de entre 104 y 109 células por kg de peso corporal, preferiblemente entre 105 y 106 células / kg del peso corporal incluyendo a todos los números enteros de células dentro de esos rangos. La dosificación en las terapias de células T CAR puede incluir por ejemplo la administración de entre 10^6 y 10^9 células / kg, con un curso de linfoeliminacion o sin el mismo, por ejemplo, con ciclofosfamida. Las células o la población de células se pueden administrar en una o más dosis. En otra forma de realización, la cantidad eficaz de células se administra en una sola dosis. En otra forma de realización, la cantidad eficaz de células se administra como más de una dosis durante un período de tiempo. La temporización de la administración se encuentra dentro de la experiencia de los médicos y depende de la condición clínica del paciente. Las células o la población de células se pueden obtener de cualquier fuente, por ejemplo, de un banco de sangre o un donante. Aunque las necesidades individuales varían, la determinación de los rangos óptimos de cantidades eficaces de un tipo dado de células que se dan para una enfermedad o condiciones en particular se encuentran dentro de la experiencia de los expertos en el arte. Una cantidad eficaz significa una cantidad que provee un beneficio terapéutico o profiláctico. La dosificación administrada dependerá de la edad, la salud y el peso de receptor, del tipo de tratamiento concurrente, si lo hay, de la frecuencia de tratamiento y de la naturaleza del efecto que se desee.

En otra forma de realización, la cantidad eficaz de células o de la composición que comprende a dichas células se administra parenteralmente. La administración puede ser una administración intravenosa. La administración se puede hacer directamente por inyección dentro de un tumor

Como protección contra las posibles reacciones adversas, las células inmunorrespondentes modificadas pueden estar equipadas con un interruptor de seguridad transgénico, en la forma de un transgén que hace que las células sean vulnerables a la exposición a una señal específica. Por ejemplo, se puede utilizar el gen viral timidina quinasa (TK) del herpes simple, por ejemplo, mediante la introducción en los linfocitos T alogénicas utilizados como infusiones de linfocitos del donante después del trasplante de células madre (Greco, y col., Improving the safety of cell therapy with the TK-suicide gene. Front. Pharmacol. 2015; 6: 95). En dichas células, la administración de una profármaco nucleósido tal como el ganciclovir o el aciclovir causa la muerte celular. Las construcciones de interruptores de seguridad alternativas incluyen caspasa 9 inducible, por ejemplo, disparada por administración de un dimerizador de moléculas pequeñas que une dos moléculas de icasp9 no funcionales para formar la enzima activa. Se ha descrito una amplia variedad de enfoques alternativos para la implementación de controles de la proliferación celular (véase la Publicación de Patente de los EE.UU. No. 20130071414; Publicación de Patente PCT WO2011146862; Publicación de Patente PCT WO2014011987; Publicación de Patente PCT WO2013040371; Zhou y col. BLOOD, 2014, 123/25:3895 - 3905; Di Stasi y col., The New England Journal of Medicine 2011; 365:1673-1683; Sadelain M, The New England Journal of Medicine 2011; 365:1735-173; Ramos y col., Stem Cells 28(6):1107-15 (2010)).

En un refinamiento adicional de las terapias adoptivas, se puede utilizar la edición del genoma con un sistema CRISPR-

Cas según se describe aquí para adaptar las células inmuno respondentes a implementaciones alternativas, por ejemplo, proporcionando células T CAR editadas (véase Poirot y col., 2015, Multiplex genome edited T-cell manufacturing platform for "off-the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853). Por ejemplo, las células inmuno respondentes se pueden editar para suprimir la expresión de algunas o todas de las moléculas de la clase HLA de tipo II y/o tipo I, o el noqueo de determinados genes que pueden inhibir la respuesta inmunitaria que se desea, por ejemplo, el gen PD1.

Las células se pueden editar usando cualquier sistema CRISPR y forma de uso del mismo como se describe en el presente documento. Los sistemas CRISPR se pueden suministrar a una célula inmune por cualquier método descrito en este documento. En algunas formas de realización preferidas, las células se editan ex vivo y se transfieren a un sujeto que lo necesita. Las células inmunorrespondentes, las células T CAR o cualquier célula utilizadas para la transferencia adoptiva de células se pueden editar. La edición se puede realizar para eliminar receptores de células T potencialmente alorreactivas (TCR), interrumpir el objetivo de un agente quimioterapéutico, bloquear un punto de control inmune, activar una célula T, y / o aumentar la diferenciación y / o proliferación de células T CD8+ funcionalmente agotadas o disfuncionales (ver las Publicaciones de patente PCT: WO2013176915, WO2014059173, WO2014172606, WO2014184744 y WO2014191128). La edición puede dar como resultado la desactivación de un gen.

Con la desactivación de un gen se pretende que el gen de interés no se exprese en forma de una proteína funcional. En una forma de realización en particular, el sistema CRISPR cataliza específicamente el clivaje en un gen diana, desactivando de esta manera a dicho gen diana. Las rotura causadas en la hebra de ácido nucleico se reparan comúnmente a través de los distintos mecanismos de recombinación homóloga o de extremos no homólogos (NHEJ). Sin embargo, la NHEJ es un proceso de reparación imperfecto que frecuentemente resulta en cambios en la secuencia de ADN en el sitio de clivaje. La reparación a través de extremos no homólogos (NHEJ) frecuentemente resulta en pequeñas inserciones o supresiones (Indel) y se puede utilizar para la creación de noqueos de genes específicos. Las células en las que se ha producido un evento de mutagénesis inducida por clivaje se pueden identificar y / o seleccionar por métodos bien conocidos en el arte

Los receptores de células T (TCR) son receptores de la superficie celular que participan en la activación de células T en respuesta a la presentación del antígeno. Los TCR generalmente están formados por dos cadenas, α y β, que se ensamblan para formar un heterodímero y se asocian con las subunidades CD3 de transducción para formar el complejo receptor de las células T presente en la superficie celular. Cada cadena α y β del TCR consta de una variable de inmunoglobulina N-terminal (V) y la región constante (C), un dominio transmembrana hidrofóbico, y una región citoplasmática corta. En cuanto a las moléculas de inmunoglobulina, la región variable de las cadenas alfa y beta se genera por recombinación V(D)J, creando una gran diversidad de especificidades de antígeno dentro de la población de células T. Sin embargo, al contrario de las inmunoglobulinas que reconocen al antígeno intacto, las células T son activadas por fragmentos de péptidos procesados en asociación con una molécula de MHC, introduciendo una dimensión adicional en el reconocimiento del antígeno por las células T, conocido como restricción por MHC. El reconocimiento de las diferencias de MHC entre el donante y el receptor a través del receptor de células T conduce a la proliferación de células T y el desarrollo potencial de una enfermedad de injerto contra huésped (EICH). La desactivación de TCRα o TCR puede resultar en la eliminación del TCR de la superficie de las células T, impidiendo el reconocimiento de aloantígenos y por lo tanto originando una GVHD. Sin embargo, la interrupción por TCR generalmente resulta en la eliminación del componente de señalización CD3 y altera los medios de una mayor expansión de las células T

Las células alogénicas son rechazadas rápidamente por el sistema inmunitario del huésped. Se ha demostrado que los leucocitos alogénicos presentes en los productos sanguíneos no irradiados persistirán durante no más de entre 5 y 6 días (Boni, Muranski et al 2008 Blood 1; 112 (12):4746-54). Por lo tanto, para prevenir el rechazo de células alogénicas, el sistema inmunológico del huésped por lo general debe ser suprimido en cierta medida. Sin embargo, en el caso de la transferencia de células adoptivas el uso de fármacos inmunosupresores también tiene un efecto perjudicial sobre las células T terapéuticas introducidas. Por lo tanto, para utilizar con eficacia un enfoque de inmunoterapia adoptiva en estas condiciones, las células introducidas tendrían que ser resistentes al tratamiento inmunosupresor. Por lo tanto, en una forma de realización en particular, el método comprende además una etapa de modificación de las células T para que sean resistentes a un agente inmunosupresor, preferiblemente mediante la desactivación de al menos un gen que codifica un diana para un agente inmunosupresor. Un agente inmunosupresor es un agente que suprime la función inmune por uno de varios mecanismos de acción. Un agente inmunosupresor puede ser, pero de manera no taxativa, un inhibidor de calcineurina, un objetivo de la rapamicina, un receptor de bloqueador de la cadena α de interleuquina-2, un inhibidor de la inosina monofosfato deshidrogenasa, un inhibidor de la dihidrofolato reductasa, un corticosteroide o un antimetabolito inmunosupresor. El método descrito en la presente permite conferir inmunosupresión resistente a las células T para la inmunoterapia mediante la desactivación del diana del agente inmunosupresor en las células T. Como ejemplos no limitantes, los blancos de un agente inmunosupresor pueden ser un receptor de un agente inmunosupresor, tal como: CD52, receptores de glucocorticoides (GR), un miembro de la familia de genes FKBP y un miembro de la familia de genes de ciclofilina.

Los puntos de control inmunitario son vías inhibitorias que hacen más lentas o detienen las reacciones inmunes y previenen el daño tisular excesivo de la actividad descontrolada de las células inmunes. En ciertas realizaciones, el punto de control inmunitario diana es el gen (PDCD1) de muerte programada-1 (PD-1 o CD279). En otras formas de

realización, el punto de control inmunitario diana es el antígeno asociado al linfocito T citotóxico (CTLA-4). En otras formas de realización adicionales, el punto de control inmunitario específico es otro miembro de la superfamilia de Ig, CD28 y CTLA4 tales como BTLA, LAG3, ICOS, PDL1 o KIR. En otras formas de realización adicionales, el punto de control inmunológico específico es un miembro de la superfamilia TNFR tal como CD40, OX40, CD137, GITR, CD27 o TIM-3.

Otros puntos de control inmunitario adicionales incluyen a la proteína tirosina fosfatasa 1 con homología al Src 2 (SHP-1) (Watson HA, y col. SHP-1: the next checkpoint target for cancer immunotherapy? Biochem Soc Trans. 15 de abril de 2016; 44 (2): 356-62). SHP-1 es una proteína inhibidora de la tirosina fosfatasa (PTP) ampliamente expresada. En las células T, es un regulador negativo de la activación y la proliferación dependiente de antígeno. Es una proteína citosólica, y por lo tanto no es susceptibles a las terapias mediadas por anticuerpos, pero su papel en la activación y proliferación hace que sea un objetivo atractivo para la manipulación genética en las estrategias de transferencia adoptiva, como en las células T con receptores de antígenos quiméricos (CAR). Los puntos de control inmunitario también pueden incluir inmunoreceptores de células T con dominios Ig e ITIM (TIGIT/Vstm3/ WUCAM/ VSIG9) y VISTA (Le Mercier I, y col., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. Front. Immunol. 6:418).

WO2014172606 se refiere al uso de MT1 y / o inhibidores de MT1 para aumentar la proliferación y / o actividad de las células T CD8 + agotadas y para disminuir el agotamiento de las células T CD8 + (por ejemplo, disminuir las células inmunitarias CD8+ funcionalmente agotadas o no respondentes). En ciertas formas de realización, las metalotioneínas son el diana de la edición de genes en las células T transferidas de forma adoptiva.

En ciertas formas de realización, los objetivos de la edición del gen pueden ser al menos un locus específico relacionado con la expresión de una proteína de punto de control inmunológico. Tales objetivos pueden incluir, pero de manera no taxativa: CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBRII, TGFRBRI, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1, FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1 o TIM-3. En algunas formas de realización preferidas, el diana es el locus del gen involucrado en la expresión de los genes PD-1 o CTLA-4. En otras formas de realización preferidas, las combinaciones de genes se dirigen, por ejemplo, pero de manera no taxativa, a PD-1 y TIGIT.

En otras formas de realización, se editan al menos dos genes. Los pares de genes pueden incluir, pero de manera no taxativa, a PD1 y TCRα, PD1 y TCR, CTLA-4 y TCRα, CTLA-4 y TCR, LAG3 y TCRα, LAG3 y TCR, TIM3 y TCRα, TIM3 y TCR, BTLA y TCRα, BTLA y TCR, BY55 y TCRα, BY55 y TCR, TIGIT y TCRα, TIGIT y TCR, B7H5 y TCRα, B7H5 y TCR, LAIR1 y TCRα, LAIR1 y TCR, SIGLEC10 y TCRα, SIGLEC10 y TCR, 2B4 y TCRα, 2B4 y TCR.

Ya sea antes o después de la modificación genética de las células T, las células T pueden ser activadas y, en general se pueden ampliar usando los métodos que se describen, por ejemplo, en las patentes US 6.352.694; 6.534.055; 6.905.680; 5.858.358; 6.887.466; 6.905.681; 7.144.575; 7.232.566; 7.175.843; 5.883.223; 6.905.874; 6.797.514; 6.867.041; y 7.572.631. Las células T pueden expandirse *in vitro* o *in vivo*.

A menos que se indique lo contrario, al llevar a la práctica la presente invención se emplean técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica y ADN recombinante, que están dentro de la experiencia de la técnica. Véase MOLECULAR CLONING: A LABORATORY MANUAL, 2a edición (1989) (Sambrook, Fritsch y Maniatis); MOLECULAR CLONING: A LABORATORY MANUAL, 4a edición (2012) (Green y Sambrook); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1987) (F. M. Ausubel, y col. eds.); la serie METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (1995) (M.J. MacFerson, B.D. Hames y G.R. Tailor eds.); ANTIBODIES, A LABORATORY MANUAL (1988) (Harlow y Lane, eds.); ANTIBODIES A LABORATORY MANUAL, 2a edición (2013) (E.A. Greenfield ed.); y ANIMAL CELL CULTURE (1987) (R.I. Freshney, ed.).

A menos que se indique lo contrario, al llevar a la práctica la presente invención se emplean las técnicas convencionales para la generación de ratones modificados genéticamente. Véase Marten H. Hofker y Jan van Deursen, TRANSGENIC MOUSE METHODS AND PROTOCOLS, 2a edición (2011).

Impulso genético

La presente invención también describe, pero no reivindica, el uso del sistema CRISPR-Cas que se describe aquí, por ejemplo, sistemas de proteína efectora Cpf1, para proveer impulso genético dirigido por ARN, por ejemplo, en sistemas análogos al impulso genético que se describe en la Publicación de Patente PCT WO 2015/105928. Los sistemas de esta clase pueden proveer por ejemplo métodos para células eucarióticas de líneas germinales no humanas, por introducción en las células de líneas germinales a la secuencia de ácido nucleico que codifica una ADN nucleasa dirigida por ARN y uno o más ARN guía. los ARN guía se puede diseñar de manera que sean complementarios a una o más localizaciones diana en el ADN genómico de las células de líneas germinales. La secuencia de ácido nucleico que codifica la ADN nucleasa guiada por ARN y la secuencia de ácido nucleico que codifican los ARN guía se puede proveer en construcciones entre secuencias flanqueantes, con promotores dispuestos de manera tal que las células

de líneas germinales puedan expresar la ADN nucleasa guiada por ARN y los ARN guía, junto con cualquier carga de secuencias codificantes que se desee que también están situadas entre las secuencias flanqueantes. Las secuencias flanqueantes típicamente incluirán una secuencia que es idéntica a una secuencia correspondiente en un determinado cromosoma diana, de manera tal que las secuencias flanqueantes trabajen con los componentes codificados por la construcción para facilitar la inserción de las secuencias extrañas de ácido nucleico de la construcción en el ADN genómico en un sitio de corte diana por mecanismos tales como recombinación homóloga, para hacer que las células de líneas germinales sean homocigotas para la secuencia extraña de ácido nucleico. De esta manera, los sistemas impulsados por genes son capaces de introgresar que se desea genes de carga a través de toda una población de cría (Gantz y col., 2015, *Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi*, PNAS 2015, publicado antes de su impresión el 23 de noviembre de 2015, doi:10.1073/pnas.1521077112; Esvelt y col., 2014, *Concerning RNA-guided gene drives for the alteration of wild populations* eLife 2014;3:e03401). En determinadas formas de realización, se pueden seleccionar secuencias diana que tengan pocos sitios fuera del diana potenciales en un genoma. El direccionamiento hacia múltiples sitios dentro de un locus diana, usando múltiples ARN guía, puede incrementar la frecuencia de corte y dificultar la evolución de impulsar alelos resistentes. Los ARN guía truncados pueden reducir el corte fuera del diana. Se pueden utilizar nickasas pareadas en vez de una única nucleasa, para incrementar adicionalmente la especificidad. Las construcciones de impulso genético pueden incluir secuencias de carga que codifican reguladores transcripcionales, por ejemplo, para activar la recombinación homóloga de genes y/o reprimir la unión de extremos no homólogos. Se pueden seleccionar sitios diana dentro de un gen esencial, de manera tal que los eventos de unión de extremos no homólogos pueden causar letalidad en vez de crear un alelo resistente al impulso. Las construcciones de impulso genético se pueden modificar para que funcionen en un rango de huéspedes en un rango de temperaturas (Cho y col. 2013, *Rapid and Tunable Control of Protein Stability in Caenorhabditis elegans Using a Small Molecule*, PLoS ONE 8(8): e72393. doi:10.1371/journal.pone.0072393).

Xenotrasplante

La presente solicitud también describe, pero no reivindica, el uso del sistema CRISPR-Cas que se describe aquí, por ejemplo, sistemas de proteína efectora Cpf1, para proveer ADN nucleasas dirigidas por ARN adaptadas para utilizarlas para proveer tejidos modificados para transplante. Por ejemplo, las ADN nucleasas dirigidas por ARN se pueden utilizar para el noqueo, *knockdown* o interrupción de determinados genes en un animal, por ejemplo, un cerdo transgénico (tal como la línea de cerdo transgénico con hemo oxigenasa-1 humana), por ejemplo, por interrupción de la expresión de genes que codifican epítopos reconocidos por el sistema inmunitario humano, es decir genes de xenoantígenos. Los genes porcinos candidatos para la interrupción pueden incluir por ejemplo a los genes de α(l,3)-galactosiltransferasa y citidina monofosfato-ácido N-acetilneuramínico hidroxilasa (véase la Publicación de Patente PCT WO 2014/066505). Además, genes que codifican retrovirus encógenos se pueden interrumpir, por ejemplo, los genes que codifican a todos los retrovirus endógenos porcinos (véase Yang y col., 2015, *Genome-wide inactivation of porcine endogenous retroviruses (PERVs)*, Science 27 de noviembre de 2015: Vol. 350 no. 6264 pp. 1101-1104). Además, las ADN nucleasas dirigidas por ARN se pueden utilizar para dirigirlas hacia un sitio para la integración de genes adicionales en el xenotrasplante de animales donantes, por ejemplo, un gen CD55 humano para mejorar la protección contra el rechazo hiperagudo.

Consideraciones generales sobre la terapia génica

Algunos ejemplos de genes asociados a una enfermedad y polinucleótidos e información sobre enfermedades específicas se puede obtener de McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) y National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), que se puede encontrar en la World Wide Web.

Las mutaciones en estos genes y rutas pueden conllevar la producción de proteínas inadecuadas o proteínas en cantidades inadecuadas que afectan a la función. Otros ejemplos adicionales de genes, enfermedades y proteínas son de la Solicitud Provisional de los EE.UU. 61/736.527 presentada el 12 de diciembre de 2012. Dichos genes, proteínas y vías pueden ser el polinucleótido diana de un complejo CRISPR. En las tablas A y B se dan algunos ejemplos de genes asociados a una enfermedad y polinucleótidos. En la tabla C se mencionan algunos ejemplos de genes y polinucleótidos asociados a las vías bioquímicas de señalización.

Tabla A

ENFERMEDAD/TRASTORNOS	GEN(ES)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alfa; PPAR
	gamma; WT1 (Tumor de Wilms); miembros de la familia del receptor de FGF
	(5 miembros: 1, 2, 3, 4, 5); CDKN2a; APC; RB
	(retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR
	(Receptor androgénico); TSG101; IGF; Receptor de IGF; Igf1 (4
	variantes); Igf2 (3 variantes); Receptor de Igf 1; Receptor de Igf 2;
	Bax; Bcl2; familia de caspasas (9 miembros:
	1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Degeneración macular	Abcr; Ccl2; Cc2; cp (ceruloplasmina); Timp3; catepsina D;
senil	Vldlr; Ccr2
Esquizofrenia	Neurregulina1 (Nrg1); Erb4 (receptor de Neurregulina);
	Complexina1 (Cplx1); Tph1 Triptófano-hidroxilasa; Tph2
	Triptófano-hidroxilasa 2; Neurexina 1; GSK3; GSK3a;
	GSK3b
Trastornos	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA;
	DTNBP1; Dao (Dao1)
Repetición de trinucleótidos	HTT (diagnóstico de Huntington); SBMA/SMAX1/AR (diagnóstico de Kennedy);
Trastornos	); FXN/X25 (Ataxia de Friedrich); ATX3 (Dx de Machado-
	Joseph); ATXN1 y ATXN2 (ataxias espinocerebelosas);
	DMPK (distrofia miotónica); Atrofina-1 y Atn1
	(Diagnóstico de DRPLA); CBP (Creb-BP - inestabilidad global); VLDLR
	(Alzheimer); Atxn7; Atxn10
Síndrome del cromosoma X frágil	FMR2; FXR1; FXR2; mGLURS
Trastornos relacionados	APH-1 (alfa y beta); Presenilina (Psen1); nicastrina
Trastornos	(Ncstn); PEN-2
Otros	Nos1; Parp1; Nat1; Nat2
Trastornos relacionados con priones	Prp
ELA	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a;
	VEGF-b; VEGF-c)
Toxicomanía	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2;
	Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)

ENFERMEDAD/TRASTORNOS	GEN(ES)
Autismo	Mecp2; BZRAP1; MDGA2; SemaSA; Neurexina 1; Cromosoma X frágil
	(FMR2 (AFF2); FXR1; FXR2; Mglur5)
Enfermedad de Alzheimer	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterina; PS1;
	SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1,
	Acuaporina 1); Uchl1; Uchl3; APP
Inflamación	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-
	17b; IL-17c; IL-17d; IL-17f); Il-23; Cx3cr1; ptpn22; TNFa;
	NOD2/CARD15 para IBD; IL-6; IL-12 (IL-12a; IL-12b);
	CTLA4; Cx3cl1
Enfermedad de Parkinson	x-Sinucleína; DJ-1; LRRK2; Parkina; PINK1

Tabla B:

Enfermedades y trastornos de la sangre y la coagulación	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NTSC3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); Síndrome del linfocito desnudo (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFXS), Trastornos con pérdida de sangre (TBXA2R, P2RX1, P2X1); Factor H y 1 similar al factor H (HF1, CFH, HUS); Factor V y factor VIII (MCFD2); Deficiencia del factor VII (F7); Deficiencia del factor X (F10); Deficiencia del factor XI (F11); Deficiencia del factor XII (F12, HAF); Deficiencia del factor XIIIA (F13A1, F13A); Deficiencia del factor XIIIB (F13B); Anemia de Fanconi (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); Trastornos de tipo linfohistiocitosis hemofagocítica (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemofilia A (F8, F8C, HEMA); Hemofilia B (F9, HEMB), Trastornos hemorrágicos (PI, ATT, FS); Trastornos y deficiencias leucocitarias (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Anemia drepanocítica (HBB); Talasemia (HBA2, HBB, HBD, LCRB, HBA1).
Enfermedades y trastornos por regulación celular errónea y oncológicos	Linfoma no Hodgkin de linfocitos B (BCL7A, BCL7); Leucemia (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).

Enfermedades y trastornos relacionados con el sistema inmunitario y la inflamación	SIDA (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Síndrome linfoproliferativo autoinmunitario (TNFRSF6, APT1, FAS, CD95, ALPS1A); Inmunodeficiencia combinada, (IL2RG, SCIDX1, SCIDX, IMD4); VIH-1 (CCL5, SCYA5, D17S136E, TCP228), Infección por VIH o susceptibilidad a este (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKRS (CCRS)); Inmunodeficiencias (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSFS, CD40, UNG, DGU, HIGM4, TNFSFS, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflamación (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), Il-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 para IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Inmunodeficiencias combinadas graves (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Enfermedades y trastornos metabólicos, hepáticos, renales y proteicos	Neuropatía amiloide (TTR, PALB); Amiloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Fibrosis quística (CFTR, ABCC7, CF, MRP7); Enfermedades del almacenamiento de glucógeno (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Adenoma hepático, 142330 (TCF1, HNF1A, MODY3), Insuficiencia hepática de, inicio temprano, y trastorno neurológico (SCOD1, SCO1), deficiencia de Lipasa hepática (LIPC), Hepatoblastoma, cáncer y carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; enfermedad renal quística medular (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Fnilcetonuria (PAH, PKU1, QDPR, DHPR, PTS); Riñón poliquístico y enfermedad hepática (FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Enfermedades y trastornos musculares/esqueléticos	Distrofia muscular de Becker (DMD, BMD, MYF6), Distrofia muscular de Duchenne (DMD, BMD); Distrofia muscular de Emery-Dreifuss (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Distrofia muscular facioscapulohumeral (FSHMD1A, FSHD1A); Distrofia muscular (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Atrofia muscular (VAPB, VAPC, ELA8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).

Enfermedades y trastornos neurológicos y neuronales	ELA (SOD1, ELA2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); Enfermedad de Alzheimer (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autismo (Mecp2, BZRAP1, MDGA2, Sema5A, Neurexina 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Síndrome asociado al cromosoma X frágil (FMR2, FXR1, FXR2, mGLUR5); Enfermedad de Huntington y trastornos similares a la enfermedad (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Enfermedad de Parkinson (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARKS, PINK1, PARK6, UCHL1, PARKS, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); síndrome de Rett (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Sinucleína, DJ-1); Esquizofrenia (Neurregulina1 (Nrg1), Erb4 (receptor para la Neurregulina), Complexina1 (Cplx1), Tph1 Triptófano-hidroxilasa, Tph2, Triptófano-hidroxilasa 2, Neurexina 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drdia), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Trastornos relacionados con la secretasa (APH-1 (alfa y beta), Presenilina (Psen1), nicastrina, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trastornos por repeticiones de trinucleótidos (HTT (Diagnóstico de Huntington), SBMA/SMAX1/AR (Diagnóstico de Kennedy), FXN/X25 (Ataxia de Friedrich), ATX3 (Diagnóstico de Machado- Joseph), ATXN1 y ATXN2 (ataxias espinocerebelosas), DMPK (distrofia miotónica), Atrofina-1 y Atn1 (Diagnóstico DRPLA), CBP (Creb-BP - inestabilidad global), VLDLR (Alzheimer), Atxn7, Atxn10).
Enfermedades y trastornos oculares	Degeneración macular senil (Abcr, Ccl2, Cc2, cp (ceruloplasmina), Timp3, catepsinaD, Vldlr, Ccr2); Cataratas (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Opacidad y distrofia corneal (APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Córnea plana congénita (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Amaurosis congénita de Leber (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Distrofia macular (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

Tabla C:

FUNCIÓN CELULAR	GENES
Señalización de PI3K/AKT	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1;
	AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2;
	ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3;
	PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7;
	YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;
	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;

FUNCIÓN CELULAR	GENES
	HSP90AA1; RPS6KB1
Señalización de ERK/MAPK	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2;
	EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6;
	MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1;
	PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN;
	EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ;
	PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1;
	PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1;
	CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Señalización del receptor	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;
glucocorticoide	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13;
	RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1;
	PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3;
	MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP;
	CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;
	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1
Señalización de guía axonal	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12;
	IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2;
	ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2;
	CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11;
	PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;
	FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1;
	GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3;
	CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA

FUNCIÓN CELULAR	GENES
Señalización del receptor de efrina	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2;
	PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2;
	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;
	AKT3; SGK
Señalización del citoesqueleto de	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
actina	PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8;
	PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7;
	PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3;
	ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
Señalización de la enfermedad de	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;
Huntington	MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1;
	GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2;
	HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A;
	HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1;
	PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Señalización de la apoptosis	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;
	BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF;

FUNCIÓN CELULAR	GENES
	RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2;
	CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2;
	BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK;
	CASP3; BIRC3; PARP1
Señalización de receptores de células B	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11;
	AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1;
	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Señalización de la extravasación de	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
leucocitos	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Señalización mediante integrinas	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;
	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
Señalización de la respuesta	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
Fase aguda	AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14;
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;

FUNCIÓN CELULAR	GENES
	AKT3; IL1R1; IL6
Señalización mediante PTEN	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;
	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
Señalización de p53	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
Señalización del receptor	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;
de hidrocarburos de arilo	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
Señalización del metabolismo	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
xenobiótico	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP90AA1
Señalización de SAPK/JNK	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;

FUNCIÓN CELULAR	GENES
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;
	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1;
	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;
	CRKL; BRAF; SGK
Señalización de PPAr/RXR	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	ADIPOQ
Señalización de NK-KB	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6;
	TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2;
	KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF;
	INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL1R1
Señalización mediante neurregulina	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1;
	MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI;
	CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS;
	PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2;
	ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3;
	EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL;
	AKT3; PRKCA; HSP90AA1; RPS6KB1
Señalización mediante	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
Wnt y Beta catenina	AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A;
	WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK;
	LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1;
	GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B;

FUNCIÓN CELULAR	GENES
	AKT3; SOX2
Señalización del receptor de insulina	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1;
	PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3;
	MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1;
	SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN;
	MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1;
	GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK;
	RPS6KB1
Señalización de IL-6	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3;
	MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1;
	MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG;
	RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3;
	MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Colestasis hepática	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA;
	RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8;
	PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1;
	TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8;
	CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4;
	JUN; IL1R1; PRKCA; IL6
Señalización de IGF-1	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2;
	PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;
	IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;
	FOXO1; SRF; CTGF; RPS6KB1
Respuesta al estrés Oxidativo mediada por NRF2	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
	NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP;
	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;
	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Fibrosis hepática/Activación	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
de células estrelladas hepáticas	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;

FUNCIÓN CELULAR	GENES
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
Señalización de PPAR	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Señalización de Fc épsilon RI	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
Señalización del receptor	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
acoplado a la proteína G	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;
	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
Metabolismo de fosfato de inositol	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
	MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;
	PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
Señalización de PDGF	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB;
	PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC;
	PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2;
	PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
Señalización de VEGF	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;
	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3;
	BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN;
	RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN;

FUNCIÓN CELULAR	GENES
	VEGFA; AKT3; FOXO1; PRKCA
Señalización de linfocitos citolíticos naturales	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11;
	KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB;
	PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6;
	PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1;
	PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Ciclo celular: Punto de control de Regulación G1/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	GSK3B; RBL1; HDAC6
Señalización del receptor de linfocitos T	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS;
	NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN;
	MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10;
	JUN; VAV3
Señalización del receptor asociado a la muerte	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD;
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3
Señalización de FGF	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
Señalización de GM-CSF	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;
	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
Señalización de la esclerosis	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
lateral amiotrófica	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;

FUNCIÓN CELULAR	GENES
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
Señalización de JAK/Stat	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1
Metabolismo de nicotinato y	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1;
nicotinamida	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Señalización de quimioquinas	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
Señalización de IL-2	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;
	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Depresión sináptica	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
a largo plazo	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Señalización del receptor	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
de estrógeno	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Ruta de ubiquitinación	TRAF6; SMLTRF1; BIRC4; BRCA1; UCHL1; NEDD4;
de proteínas	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
Señalización de IL-10	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;
	IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6

FUNCIÓN CELULAR	GENES
Activación de VDR/RXR	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;
	LRP5; CEBPB; FOXO1; PRKCA
Señalización de TGF-beta	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGFBR1; SMAD4; JUN; SMADS
Señalización del receptor de tipo toll	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;
	NFKB1; TLR2; JUN
Señalización de p38 MAPK	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1;
	SRF; STAT1
Señalización de neurotrofina/TRK	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A;
	RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	CDC42; JUN; ATF4
Activación de FXR/RXR	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8;
	APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Depresión sináptica	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
a largo plazo	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
Señalización de calcio	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;
	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6
Señalización de EGF	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Señalización de hipoxia en el	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;

FUNCIÓN CELULAR	GENES
sistema cardiovascular	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
Inhibición mediada por LPS/IL-1	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1;
de la función RXR	MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2;
	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
Activación de LXR/RXR	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;
	SREBF1; IL1R1; CCL2; IL6; MMP9
Procesamiento del amiloide	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
Señalización de IL-4	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
Ciclo celular: G2/M y	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
Punto de control de daños	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
Regulación	PRKDC; ATM; SFN; CDKN2A
Señalización de óxido nítrico en el	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;
sistema cardiovascular	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
Metabolismo de purinas	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
Señalización mediada por AMPc	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Disfunción mitocondrial	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Señalización de Notch	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4
Ruta del estrés del	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
retículo endoplasmático	EIF2AK3; CASP3
Metabolismo de pirimidinas	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;
	NT5E; POLD1; NME1
Señalización del Parkinson	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7;
	PARK2; CASP3

FUNCIÓN CELULAR	GENES
Señalización	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC;
cardíaca y beta adrenérgica	PPP2R5C
Glucolisis/Gluconeogénesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Señalización de interferón	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Señalización de sonic Hedgehog	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Glicerofosfolípido	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Metabolismo de	
degradación de fosfolípidos	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Metabolismo de triptófano	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Degradación de lisina	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Ruta de reparación mediante	ERCC5; ERCC4; XPA; XPC; ERCC1
escisión de nucleótidos	
Metabolismo de	UCHL1; HK2; GCK; GPI; HK1
almidón y sacarosa	
Metabolismo de aminoazúcares	NQO1; HK2; GCK; HK1
Metabolismo del	PRDX6; GRN; YWHAZ; CYP1B1
Ácido araquidónico	
Señalización del ritmo circadiano	CSNK1E; CREB1; ATF4; NR1D1
Sistema de coagulación	BDKRB1; F2R; SERPINE1; F3
Señalización del receptor	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
de dopamina	
Metabolismo de glutatión	IDH2; GSTP1; ANPEP; IDH1
Metabolismo de glicerolípidos	ALDH1A1; GPAM; SPHK1; SPHK2
Metabolismo del ácido linoleico	PRDX6; GRN; YWHAZ; CYP1B1
Metabolismo de metionina	DNMT1; DNMT3B; AHCY; DNMT3A
Metabolismo de piruvato	GLO1; ALDH1A1; PKM2; LDHA
Metabolismo de arginina	ALDH1A1; NOS3; NOS2A
y prolina	
Señalización de eicosanoides	PRDX6; GRN; YWHAZ
Metabolismo de fructosa	HK2; GCK; HK1
y manosa	
Metabolismo de galactosa	HK2; GCK; HK1
Biosíntesis de estilbeno,	PRDX6; PRDX1; TYR
cumarina y lignina	
Ruta de presentación de	CALR; B2M

FUNCIÓN CELULAR	GENES
antígenos	
Biosíntesis de esteroides	NQO1; DHCR7
Metabolismo de butanoato	ALDH1A1; NLGN1
Ciclo de citrato	IDH2; IDH1
Metabolismo de ácidos grasos	ALDH1A1; CYP1B1
Metabolismo de	PRDX6; CHKA
glicerofosfolípido	
Metabolismo de histidina	PRMT5; ALDH1A1
Metabolismo de inositol	ERO1L; APEX1
Metabolismo de xenobióticos	GSTP1; CYP1B1
mediante el citocromo p450	
Metabolismo de metano	PRDX6; PRDX1
Metabolismo de fenilalanina	PRDX6; PRDX1
Metabolismo de propanoato	ALDH1A1; LDHA
Metabolismo de	PRMT5; AHCY
selenoaminoácido	
Metabolismo de esfingolípidos	SPHK1; SPHK2
Metabolismo de	PRMT5
aminofosfonato	
Metabolismo de andrógeno	PRMT5
y estrógeno	
Metabolismo de	ALDH1A1
ascorbato y aldarato	
Biosíntesis del ácido biliar	ALDH1A1
Metabolismo de cisteína	LDHA
Biosíntesis de ácidos grasos	FASN
Señalización del receptor de	GNB2L1
glutamato	
Respuesta al estrés oxidativo mediada por NRF2	PRDX1
Vía de pentosa fosfato	GPI
Interconversiones de	UCHL1
pentosa y glucuronato	
Metabolismo de retinol	ALDH1A1
Metabolismo de riboflavina	TYR

FUNCIÓN CELULAR	GENES
Metabolismo de tirosina	PRMT5, TYR
Biosíntesis de ubiquinona	PRMT5
Degradación de valina,	ALDH1A1
leucina e isoleucina	
Metabolismo de glicina,	CHKA
serina y treonina	
Degradación de lisina	ALDH1A1
Dolor/sabor	TRPM5; TRPA1
Dolor	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a
Función mitocondrial	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Neurología del desarrollo	BMP-4; Cordina (Chrd); Nogina (Nog); WNT (Wnt2;
	Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b;
	Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenina;
	Dkk-1; proteínas relacionadas con Frizzled; Otx-2; Gbx2; FGF-8;
	Reelin; Dab1; unc-86 (Pou4f1 o Brn3a); Numb; Reln

Las formas de realización de los métodos y composiciones reivindicados de la invención también se refieren a métodos y composiciones relacionadas con la inactivación de genes, amplificación de genes y reparación de mutaciones particulares asociadas con la inestabilidad por repeticiones del ADN y trastornos neurológicos (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Segunda edición, Academic Press, 13 de octubre de 2011 - Medical). se ha descubierto que algunos aspectos específicos de las secuencias repetitivas en tándem son responsables de más de veinte enfermedades humanas (New insights into repeat instability: role of RNA•DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. Sep-Oct 2010; 7(5):551-8). Los presentes sistemas de proteína efectora se pueden ser emplear para corregir dichos defectos de la inestabilidad genómica.

También se describen en la presente, pero no se reivindican, métodos que se refieren a la corrección de defectos asociados con una amplia gama de enfermedades genéticas que se describen más detalladamente en la página web de los National Institutes of Health en la subsección de tema de Trastornos genéticos (página web en health.nih.gov/topic/GeneticDisorders). Las enfermedades cerebrales genéticas podrán incluir, sin carácter limitante, Adrenoleucodistrofia, Agénesis del cuerpo calloso, síndrome de Aicardi, enfermedad de Alpers, enfermedad de Alzheimer, síndrome de Barth, enfermedad de Batten, CADASIL, degeneración cerebelosa, enfermedad de Fabry, enfermedad de Gerstmann-Straussler-Scheinker, enfermedad de Huntington y otros trastornos por repeticiones de tripletes, enfermedad de Leigh, síndrome de Lesch-Nyhan, enfermedad de Menkes, miopatías mitocondriales y colpocefalia de NINDS. Estas enfermedades se describen adicionalmente en la página web de los Institutos Nacionales de Salud en la subsección sobre Trastornos cerebrales genéticos.

Desarrollo y uso de Cas9

La presente invención se puede ilustrar adicionalmente y ampliar en base a aspectos del desarrollo y uso de CRISPR-Cas9 según se establece en los siguientes artículos y en particular en lo que se refiere a la administración de un complejo de proteína CRISPR y a los usos de una endonucleasa guiada por ARN en células y organismos:

> Multiplex genome engineering using CRISPR/Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. Science Feb 15;339(6121):819-23 (2013);

> RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013);

> One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome

Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013);

> Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Nature. Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23 (2013);

> Doble Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. Cell Aug 28. pii: S0092-8674(13)01015-5 (2013-A);

> DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013);

> Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. Nature Protocols Nov;8(11):2281-308 (2013-B);

> Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013). [publicación electrónica en proceso de impresión];

> Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Cell Feb 27, 156(5):935-49 (2014);

> Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. Nat Biotechnol. abr 20. doi: 10.1038/nbt.2889 (2014);

> CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F. Cell 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014);

> Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu PD, Lander ES, Zhang F., Cell. Jun 5;157(6):1262-78 (2014).

> Genetic screens in human cells using the CRISPR/Cas9 system, Wang T, Wei JJ, Sabatini DM, Lander ES., Science. January 3; 343(6166): 80–84. doi:10.1126/science.1246981 (2014);

> Rational design of highly active ARNgss for CRISPR-Cas9-mediated gene inactivation, Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE., (publicado en línea el 3 de septiembre de 2014) Nat Biotechnol. Dec;32(12):1262-7 (2014);

> *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F., (publicado en línea el 19 de octubre de 2014) Nat Biotechnol. Jan;33(1):102-6 (2015);

> Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015).

> A split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (publicado en línea el 02 de febrero de 2015) Nat Biotechnol. Feb;33(2):139-42 (2015);

> Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metástasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246–1260, March 12, 2015 (multiplex screen in mouse), y

> *In vivo* genome editing using *Staphylococcus aureus* Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (publicado en línea el 01 de abril de 2015), Nature. abr 9;520(7546):186-91 (2015).

> Shalem y col., "High-throughput functional genomics using CRISPR-Cas9," Nature Reviews Genetics 16, 299-311 (Mayo de 2015).

> Xu y col., "Sequence determinants of improved CRISPR sgRNA design," Genome Research 25, 1147-1157 (Agosto de 2015).

> Parnas y col., "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks,"

Cell 162, 675-686 (30 de julio de 2015).

> Ramanan y col., CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus," Scientific Reports 5:10833. doi: 10.1038/srep10833 (2 de junio de 2015)

> Nishimasu y col., Crystal Structure of Staphylococcus aureus Cas9," *Cell* 162, 1113-1126 (27 de agosto de 2015)

> *BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis*, Canver y col., *Nature* 527(7577):192-7 (12 de nov. de 2015) doi: 10.1038/nature15521. Epub 16 de sep 2015.

> *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System,* Zetsche y col., *Cell* 163, 759-71 (25 de sept. de 2015).

> *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems,* Shmakov y col., *Molecular Cell*, 60(3), 385-397 doi: 10.1016/j.molcel.2015.10.008 Epub 22 de octubre de 2015.

> *Rationally engineered Cas9 nucleases with improved specificity,* Slaymaker y col., *Science* 1 de enero de 2016 351(6268): 84-88 doi: 10.1126/science.aad5227. Epub 1 de dic. de 2015. [Epub previo a la impression]

cada una de los cuales se puede considerar al llevar a la práctica la presente invención y se expone brevemente a continuación:

> Cong et al. diseñaron sistemas CRISPR/Cas de tipo II para su uso en células eucariotas basándose tanto en Cas9 de Streptococcus thermophilus como también en Cas9 de Streptococcus pyogenes y demostraron que las nucleasas Cas9 pueden dirigirse por ARN cortos para inducir la escisión precisa del ADN en células humanas y de ratón. Su estudio mostró además que Cas9 se ha convertido en una enzima que realiza un corte monocatenario que se puede usar para facilitar la reparación dirigida a la homología en células eucariotas con mínima actividad mutagénica. Adicionalmente, su estudio ha demostrado que se podrán codificar múltiples secuencias guía en una única matriz CRISPR para permitir la edición simultánea de varias de estas en sitios de loci genómicos endógenos en el genoma de un mamífero, demostrando una fácil programabilidad y una amplia aplicabilidad de la tecnología de nucleasas guiadas por ARN. Esta capacidad de usar el ARN para programar la escisión del ADN específica de secuencias en células ha definido una nueva clase de herramientas de manipulación del genoma. Estos estudios mostraron además que es probable que otros loci CRISPR sean trasplantables a células de mamífero y puedan mediar también en la escisión del genoma de mamífero. De forma importante se podrá concebir que algunos aspectos del sistema CRISPR-Cas se puedan mejorar adicionalmente para aumentar su eficacia y versatilidad.

> Jiang et al. Usaron repeticiones palindrómicas cortas, agrupadas, regularmente separadas (CRISPR) asociadas con la endonucleasa Cas9 que forma complejo con ARN dobles para introducir mutaciones precisas en los genomas de Streptococcus pneumoniae y Escherichia coli. El enfoque se basa en la escisión dirigida a ARN doble:Cas9 en el sitio genómico dirigido para destruir células no mutadas y evitar la necesidad de marcadores seleccionables o sistemas contra selección. El estudio notificó la reprogramación de la especificidad del ARN doble:cas9 cambiando la secuencia del ARN CRISPR corto (ARNcr) para preparar cambios en uno o varios nucleótidos llevados a cabo en moldes de edición. El estudio mostró que el uso simultáneo de dos ARNcr permitió la mutagénesis multiplexada. Además, cuando se usó el enfoque combinado con ingeniería recombinógena, en S. pneumoniae, casi el 100% de células que se recuperaron usando el enfoque descrito contenían la mutación deseada, y en E. coli, el 65% que se recuperaron contenían la mutación.

> Wang y col. (2013) utilizaron el sistema CRISPR-Cas para la generación en un solo paso de ratones que portan mutaciones en múltiples genes que tradicionalmente se generaron en múltiples pasos por recombinación en secuencia en células indiferenciadas embriónicas y/o la cruza de ratones con una única mutación que demanda tiempo. El sistema CRISPR-Cas acelerará mucho el estudio *in vivo* de genes funcionalmente redundantes y de interacciones epistáticas de genes.

> Konermann y col. (2013) se enfocaron en la necesidad en el arte de disponer tecnologías versátiles y robustas que permitan la modulación óptica y química de los dominios de unión a ADN de CRISPR enzima Cas9 y también efectores similares al activador transcripcional

> Ran y col. (2013-A) describieron un enfoque que combina a mutante de Cas9 nickasa con ARN guía pareados para introducir roturas direccionadas de doble hebra. Esto resuelve el problema de la nucleasa Cas9 del sistema CRISPR-Cas microbiano direccionada a loci genómicos específicos por una secuencia guía, que puede tolerar ciertas faltas de coincidencia con el ADN diana y de esa manera promueve la mutagénesis fuera del diana no deseada. Como los cortes monocatenarios individuales del genoma se reparan con mucha fidelidad, se requiere la realización simultánea de cortes monocatenarios mediante ARN guías adecuadamente no específicos para roturas bicatenarias y ampliar el número de bases específicamente reconocidas para la escisión diana. Los autores han demostrado que el uso de cortes monocatenarios emparejados puede reducir la actividad no específica de 50 a 1,500 veces en líneas celulares y facilitar la inactivación génica en cigotos de ratón sin sacrificar la eficacia de escisión específica. Esta versátil estrategia permite una amplia variedad de aplicaciones de edición genómica que requieren una elevada especificidad.

> Hsu y col. (2013) caracterizaron la especificidad del direccionamiento de SpCas9 en células humanas para informar la selección de sitios diana y evitar los efectos fuera del diana. El estudio ha evaluado >700 variantes de ARN guías y los niveles de la mutación indel inducida por SpCas9 a >100 loci no específicos genómicos previstos en células 293T y 293FT. Los autores sostienen que SpCas9 tolera los emparejamientos incorrectos entre el ARN guía y el ADN diana en diferentes posiciones de una manera dependiente de secuencia, sensible al número posición y distribución de los emparejamientos incorrectos. Los autores mostraron además que la escisión mediada por SpCas9 no se ve afectada por la metilación del ADN y que la dosificación de SpCas9 y el ARNsg se puede valorar para minimizar la modificación no específica. Adicionalmente, para facilitar las aplicaciones de manipulación del genoma de mamífero, los autores notificaron proporcionar una herramienta de software basada en web para guiar la selección y validación de las secuencias diana así como el análisis no específico.

> Ran y col. (2013-B) describieron un conjunto de herramientas para la edición del genoma mediada por Cas9 *via* unión de extremos no homólogos (NHEJ) o reparación dirigida por homología (HDR) en células de mamífero, así como la generación de líneas celulares modificadas para estudios funcionales corriente abajo. Para minimizar la escisión no específica, los autores han descrito además una estrategia de doble corte utilizando la nickasa Cas9 mutante con ARN guías emparejados. El protocolo proporcionado por los autores derivó experimentalmente directrices para la selección de sitios diana, evaluación de la eficacia de escisión y análisis de la actividad no específica. Los estudios mostraron que, al comenzar con el diseño diana, las modificaciones génicas pueden conseguirse incluso en 1-2 semanas, y las líneas celulares clonales modificadas pueden derivarse en 2-3 semanas.

> Shalem et al. han descrito una nueva manera de interrogar la función génica en una amplia escala genómica. Sus estudios mostraron que el suministro de una biblioteca de CRISPR-Cas9 inactivado génicamente a escala genómica (GeCKO) dirigida a 18,080 genes con 64,751 secuencias guía únicas permitió el cribado de la selección negativa y positiva en células humanas. En primer lugar, los autores han mostrado el uso de la biblioteca GeCKO para identificar genes esenciales para la viabilidad celular en células cancerosas y citoblastos pluripotentes. En primer lugar, en un modelo de melanoma, los autores seleccionaron los genes cuya pérdida está implicada en la resistencia a vemurafenib, un agente terapéutico que inhibe la proteína quinasa mutante BRAF. Sus estudios han mostrado que los candidatos mejor clasificados incluyeron los genes anteriormente validados NF1 y MED12 así como las novedosos blancos NF2, CUL3, TADA2B, y TADA1. Los autores observaron un alto nivel de consistencia entre los ARN guías independientes que dirigían el mismo gen y una elevada tasa de confirmación de blancos, y de esta manera demostraron lo prometedor de la selección con Cas9 a escala genómica.

> Nishimasu et al. han notificado la estructura cristalina de Cas9 de Streptococcus pyogenes en el complejo con ARNsg y su ADN diana a una resolución de 2,5 Å. La estructura reveló una arquitectura bilobulada compuesta por un reconocimiento de la diana y de los lóbulos de nucleasa, que acomodaba el heteroduplete ARNsg:ADN en una ranura positivamente cargada en su interfase. Mientras que el lóbulo de reconocimiento es esencial para la unión del ARNsg y el ADN, el lóbulo de nucleasa contiene los dominios de la nucleasa HNH y RuvC, que se sitúan adecuadamente para la escisión de las hebras complementarias y no complementarias del ADN diana, respectivamente. El lóbulo de nucleasa contiene también un dominio carboxilo terminal responsable de la interacción del motivo adyacente al protoseparador (PAM). Esta estructura de alta resolución y los análisis funcionales que la acompañan han revelado el mecanismo molecular del direccionamiento del ADN guiado por el ARN mediante Cas9, facilitando de esta manera el modo del diseño racional de nuevas tecnologías editoras de genomas versátiles.

➢ Wu et al. cartografiaron los sitios de unión de todo el genoma de una Cas9 catalíticamente inactiva (dCas9) de Streptococcus pyogenes cargada con ARN guías individuales (ARNsg) en embriocitoblastos de ratón (mESC). Los autores han mostrado que cada uno de los cuatro ARNsg ensayados dirige la dCas9 hacia una cantidad entre decenas y miles de sitios genómicos, caracterizados frecuentemente por una región semilla de 5 nucleótidos en el ARNsg y un motivo adyacente con un protoseparador NGG (PAM). La inaccesibilidad de la cromatina disminuye la unión de dCas9 a otros sitios con secuencias semilla emparejadas; de esta manera, el 70% de los sitios no específicos se asocian con genes. Los autores han mostrado que la secuenciación dirigida de 295 sitios de unión a dCas9 en los mESC transfectados con Cas9 catalíticamente activa identificó solo un sitio mutado por encima de los niveles de fondo. Los autores han propuesto un modelo de dos estados para la unión y la escisión de Cas9, en el que un emparejamiento semilla estimula la unión, pero se requiere un emparejamiento extenso con el ADN diana para la escisión.

➢ Platt y col. establecieron un ratón con Cas9 dependiente de Cre noqueado. Los autores demostraron *in vivo* así como *ex vivo* la edición del genoma usando la administración de ARN guía mediada por virus adeno-asociado (AAV), lentivirus, o partículas, en neuronas, células del sistema inmunitario, y células endoteliales.

➢ Hsu y col. (2014) es un artículo de revisión donde se expone en general la historia de CRISPR-Cas9 desde el yogurt a la edición del genoma, incluyendo el cribado genético de células.

➢ Wang y col. (2014) se refieren a un enfoque de cribado genético combinado, con pérdida de función, apropiado para la selección tanto positiva como negativa, en el que se utiliza una biblioteca a escala de genoma de ARN guía simple lentivirales (ARNgs).

➢ Doench y col. crearon un conjunto de ARNgs combinados, abarcando a todos los sitios diana posibles de un

panel de seis genes endógenos de ratón y tres genes endógenos humanos y evaluaron cuantitativamente su capacidad de producir alelos nulos de su gen diana por tinción con anticuerpo y citometría de flujo. Los autores mostraron que la optimización de los PAM mejoró la actividad y también proporcionó una herramienta en línea para el diseño de ARNgs.

- Swiech y col. demostraron que edición del genoma con SpCas9 mediada por AAV puede permitir la realización de estudios de función génica en el cerebro por genética inversa.

- Konermann y col. (2015) realizaron una exposición sobre la capacidad de unir múltiples dominios efectores, por ejemplo, activadores transcripcionales, reguladores funcionales y epigenómicos en posiciones apropiada de la guía tal como el tallo o tetrabucle con conectores y sin ellos.

- Zetsche y col. demostraron que la enzima Cas9 se puede dividir en dos y de esa manera se puede controlar el ensamble de Cas9 para la activación.

- Chen y col. se refieren a un cribado múltiplex demostrando que un cribado *in vivo* de todo el genoma CRISPR-Cas9 en ratones revela a los genes que regulan las metástasis en el pulmón.

- Ran y col. (2015) se refieren a SaCas9 y a su capacidad de editar genomas y demostraron que no se puede extrapolar a partir de ensayos

- Shalem y col. (2015) describieron las formas en las que se utilizan fusiones catalíticamente inactivas de Cas9 (dCas9) para reprimir sintéticamente (CRISPRi) o activar la expresión (CRISPRa), lo que muestra avances en el uso de Cas9 para cribados a escala de genoma, incluyendo cribados en matrices y agrupados, enfoques de noqueo que inactivan los loci genómicos y estrategias que modulan la actividad transcripcional.

- Xu y col. (2015) evaluaron las características de las secuencias de ADN que contribuyen a la eficiencia del ARN guía simple (ARNgs) en cribados basados en CRISPR. Los autores exploraron la eficiencia del noqueo por CRISPR / Cas9 y la preferencia por los nucleótidos en el sitio de clivaje. Los autores también encontraron que la secuencia de preferencia para CRISPRi /a es sustancialmente diferente de la del noqueo CRISPR / Cas9.

- Parnas y col. (2015) introdujeron bibliotecas de CRISPR-Cas9 combinadas que abarcaban todo el genoma en células dendríticas (DC) para identificar los genes que controlan la inducción de factor de necrosis tumoral (TNF) por lipopolisacárido bacteriano (LPS). Se identificaron los reguladores conocidos del TLR4 de señalización y los candidatos previamente desconocidos y se clasificaron en tres módulos funcionales con efectos distintos sobre las respuestas canónicas a LPS.

- Ramanan y col. (2015) demostraron el clivaje de ADN episómico viral (ADNccc) en las células infectadas. El genoma del VHB existe en los núcleos de hepatocitos infectados como una especie de ADN episómico de doble hebra de 3,2 kb llamado ADN circular cerrado covalentemente (ADNccc), que es un componente clave en el ciclo de vida del HBV cuya replicación no es inhibida por las terapias actuales. Los autores mostraron que los ARNgs dirigidos específicamente hacia regiones altamente conservadas del HBV suprimen robustamente la replicación viral y el empobrecimiento del ADNccc.

- Nishimasu y col. (2015) informaron las estructuras cristalinas de SaCas9 formando un complejo con un ARN guía simple (ARNgs) y sus blancos de ADN de doble hebra, que contienen el extremo 5 'TTGAAT-3' PAM y el extremo 5 'TTGGGT-3' PAM. Una comparación estructural de SaCas9 con SpCas9 hizo notar tanto la conservación estructural como la divergencia, explicando sus especificidades por PAM distintas y el reconocimiento de ARNgs ortólogos.

- Canver y col. (2015) mostraron una investigación funcional basada en CRISPR-Cas9 de elementos genómicos no codificantes. Los autores desarrollaron bibliotecas ARN guía de CRISPR-Cas9 combinados para realizar *in situ* la saturación de mutagénesis de los potenciadores BCL11A humano y ratón, que revelando características críticas de los potenciadores.

- Zetsche y col. (2015) informaron la caracterización de CPF1, una nucleasa CRISPR clase 2 de *Francisella novicida* U112 tiene características distintas de Cas9. CPF1 es una endonucleasa guiada por ARN simple que carece ARNtracr, y utiliza un motivo protoespaciador adyacente rico en T, y el ADN se cliva por una rotura de doble cadena del ADN escalonado.

- Shmakov y col. (2015) informaron tres sistemas distintos de CRISPR-Cas Clase 2. Dos sistemas de enzimas CRISPR (C2c1 y C2c3) contienen dominios de endonucleasa de similar a RuvC lejanamente relacionados con Cpf1. A diferencia de Cpf1, C2c1 depende tanto de ARNcr como de ARNtracr para la escisión del ADN. La tercera enzima (C2C2) contiene dos dominios predichos HEPN RNasa y es independiente de ARNtracr.

- Slaymaker y col. (2016) informaron el uso de la ingeniería de proteínas orientada a su estructura para mejorar la especificidad de Cas9 de *Streptococcus pyogenes* (SpCas9). Los autores desarrollaron variantes de SpCas9 de " mayor especificidad" (eSpCas9) que mantenían un robusto clivaje en el diana con efectos reducidos fuera del mismo.

También, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014), se refiere a nucleasas FokI diméricas guiadas por ARN que reconocen secuencias extendidas y pueden editar genes endógenos con alta eficiencia en células humanas.

Las Patentes de los EE.UU. n.º 8.697.359, 8.771.945, 8.795.965, 8.865.406, 8.871.445, 8.889.356, 8.889.418, 8.895.308, 8.906.616, 8.932.814, 8.945.839, 8.993.233 y 8.999.641; Publicaciones de Patente de los EE.UU. US 2014-0310830 (Solicitud de EE.UU. n.º de Serie 14/105,031), US 2014-0287938 A1 (Solicitud de EE.UU. n.º de Serie 14/213.991), US 2014-0273234 A1 (Solicitud de EE.UU. n.º de Serie 14/293.674), US2014-0273232 A1 (Solicitud de EE.UU. n.º de Serie 14/290.575), US 2014-0273231 (Solicitud de EE.UU. n.º de Serie 14/259.420), US 2014-0256046 A1 (Solicitud de EE.UU. n.º de Serie 14/226.274), US 2014-0248702 A1 (Solicitud de EE.UU. n.º de Serie 14/258.458), US 2014-0242700 A1 (Solicitud de EE.UU. n.º de Serie 14/222.930), US 2014-0242699 A1 (Solicitud de EE.UU. n.º de Serie 14/183.512), US 2014-0242664 A1 (Solicitud de EE.UU. n.º de Serie 14/104.990), US 2014-0234972 A1 (Solicitud de EE.UU. n.º de Serie 14/183.471), US 2014-0227787 A1 (Solicitud de EE.UU. n.º de Serie 14/256.912), US 2014-0189896 A1 (Solicitud de EE.UU. n.º de Serie 14/105.035), US 2014-0186958 (Solicitud de EE.UU. n.º de Serie 14/105.017), US 2014-0186919 A1 (Solicitud de EE.UU. n.º de Serie 14/104.977), US 2014-0186843 A1 (Solicitud de EE.UU. n.º de Serie 14/104.900), US 2014-0179770 A1 (Solicitud de EE.UU. n.º de Serie 14/104.837) y US 2014-0179006 A1 (Solicitud de EE.UU. n.º de Serie 14/183.486), US 2014-0170753 (Solicitud de EE.UU. n.º de Serie 14/183.429); US 2015-0184139 (Solicitud de EE.UU. n.º de Serie 14/324.960); 14/054.414 Solicitudes de Patentes Europeas EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), y EP 2 784 162 (EP14170383.5); y WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO 2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO 2014/093701 (PCT/US2013/074800), WO 2014/018423 (PCT/US2013/051418), WO 2014/204723 (PCT/US2014/041790), WO 2014/204724 (PCT/US2014/041800), WO 2014/204725 (PCT/US2014/041803), WO 2014/204726 (PCT/US2014/041804), WO 2014/204727 (PCT/US2014/041806), WO 2014/204728 (PCT/US2014/041808), WO 2014/204729 (PCT/US2014/041809) WO 2015/089351 (PCT/US2014/069897), WO 2015/089354 (PCT/US2014/069902), WO 2015/089364 (PCT/US2014/069925), WO 2015/089427 (PCT/US2014/070068), WO 2015/089462 (PCT/US2014/070127), WO 2015/089419 (PCT/US2014/070057), WO 2015/089465 (PCT/US2014/070135), WO 2015/089486 (PCT/US2014/070175), PCT/US2015/051691, PCT/US2015/051830. También se hace referencia a las Solicitudes de Patente Provisionales de los EE.UU. 61/758.468; 61/802.174; 61/806.375; 61/814.263; 61/819.803 y 61/828.130, presentada el 30 de enero de 2013; 15 de marzo de 2013; 28 de marzo de 2013; 20 de abril de 2013; 6 de mayo de 2013 y 28 de mayo de 2013 respectivamente. También se hace referencia a la Solicitud de Patente Provisional de los EE.UU. 61/836.123, presentada el 17 de junio de 2013. Adicionalmente se hace referencia a las Solicitudes de Patente Provisionales de los EE.UU. 61/835.931, 61/835.936, 61/835.973, 61/836.080, 61/836.101, y 61/836.127, cada una de las cuales se presentó el 17 de junio de 2013. Además se hace referencia a las Solicitudes de Patente Provisionales de los EE.UU. 61/862.468 y 61/862.355 presentadas el 5 de agosto de 2013; 61/871.301 presentada el 28 de agosto de 2013; 61/960.777 presentada el 25 de septiembre de 2013 y 61/961.980 presentada el 28 de octubre de 2013. Aún adicionalmente, se hace referencia a: PCT/US2014/62558 presentada el 28 de octubre de 2014, y las Solicitudes de Patente Provisionales de los EE.UU. con los números de serie: 61/915.148, 61/915.150, 61/915.153, 61/915.203, 61/915.251, 61/915.301, 61/915.267, 61/915.260, y 61/915.397, cada una de las cuales se presentó el 12 de diciembre de 2013; 61/757.972 y 61/768.959, presentadas el 29 de enero de 2013 y 25 de febrero de 2013; 62/010.888 y 62/010.879, ambas presentadas el 11 de junio de 2014; 62/010.329, 62/010.439 y 62/010.441, cada una de las cuales se presentó el 10 de junio de 2014; 61/939.228 y 61/939.242, cada una de las cuales se presentó el 12 de febrero de 2014; 61/980.012, presentada el 15 de abril de 2014; 62/038.358, presentada el 17 de agosto de 2014; 62/055.484, 62/055.460 y 62/055.487, cada una de las cuales se presentó el 25 de septiembre de 2014; y 62/069.243, presentada el 27 de octubre de 2014. Se hace referencia a la Solicitud PCT designando, entre otras, solicitud de los EE.UU. n.º PCT/US14/41806, presentada el 10 de junio de 2014. Se hace referencia a la Solicitud de Patente provisional de los EE.UU. 61/930.214 presentada el 22 de enero de 2014. Se hace referencia a la Solicitud PCT designando, entre otras, a la solicitud de los EE.UU. n.º PCT/US14/41806, presentada el 10 de junio de 2014.

También se menciona a la Solicitud de los EE.UU. 62/180.709, 17-jun-15, PROTECTED GUIDE RNAS (PGRNAS); Solicitud de los EE.UU. 62/091.455, presentada, 12-dic-14, PROTECTED GUIDE RNAS (PGRNAS); Solicitud de los EE.UU. 62/096.708, 24-dic-14, PROTECTED GUIDE RNAS (PGRNAS); Solicitudes de los EE.UU. 62/091.462, 12-dic-14, 62/096.324, 23-dic-14, 62/180.681, 17-jun-2015, y 62/237,496, 5-oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; Solicitud de los EE.UU. 62/091.456, 12-dic-14 y 62/180.692, 17-jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; Solicitud de los EE.UU. 62/091.461, 12-dic-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); Solicitud de los EE.UU. 62/094.903, 19-dic-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; Solicitud de los EE.UU. 62/096.761, 24-dic-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; Solicitud de los EE.UU. 62/098.059, 30-dic-14, 62/181.641, 18-jun-2015, y 62/181.667, 18-jun-2015, RNA-TARGETING SYSTEM; Solicitud de los EE.UU. 62/096.656, 24-dic-14 y 62/181.151, 17-jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; Solicitud de los EE.UU. 62/096.697, 24-dic-14,

CRISPR HAVING OR ASSOCIATED WITH AAV; Solicitud de los EE.UU. 62/098.158, 30-dic-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; Solicitud de los EE.UU. 62/151.052, 22-apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; Solicitud de los EE.UU. 62/054.490, 24-sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; Solicitud de los EE.UU. 61/939.154, 12-FEB-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; Solicitud de los EE.UU. 62/055.484, 25-sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; Solicitud de los EE.UU. 62/087.537, 4-dic-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; Solicitud de los EE.UU. 62/054.651, 24-sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; Solicitud de los EE.UU. 62/067.886, 23-oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; Solicitudes de los EE.UU. 62/054.675, 24-sep-14 y 62/181.002, 17-jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; Solicitud de los EE.UU. 62/054.528, 24-sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; Solicitud de los EE.UU. 62/055.454, 25-sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); Solicitud de los EE.UU. 62/055.460, 25-sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; Solicitud de los EE.UU. 62/087.475, 4-dic-14 y 62/181.690, 18-jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; Solicitud de los EE.UU. 62/055.487, 25-sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; Solicitud de los EE.UU. 62/087.546, 4-dic-14 y 62/181.687, 18-jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; y Solicitud de los EE.UU. 62/098.285, 30-dic-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

Se mencionan las Solicitudes de los EE.UU. 62/181.659, 18-jun-2015 y 62/207,318, 19-ago-2015, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION. Se mencionan las Solicitudes de los EE.UU. 62/181.663, 18-jun-2015 y 62/245.264, 22-oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, Solicitudes de los EE.UU. 62/181.675, 18-jun-2015, 62/285.349, 22-oct-2015, 62/296.522, 17-Feb-2016, y 62/320.231, 8-apr-2016, NOVEL CRISPR ENZYMES AND SYSTEMS, Solicitud de los EE.UU. 62/232.067, 24-sep-2015, Solicitud de los EE.UU. 14/975.085, 18-dic-2015, Solicitud europea No. 16150428.7, Solicitud de los EE.UU. 62/205.733, 16-ago-2015, Solicitud de los EE.UU. 62/201.542, 5-ago-2015, Solicitud de los EE.UU. 62/193.507, 16-Jul-2015, y Solicitud de los EE.UU. 62/181.739, 18-jun-2015, cada una de las cuales se titula NOVEL CRISPR ENZYMES AND SYSTEMS y of Solicitud de los EE.UU. 62/245.270, 22-oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS. También se menciona la Solicitud de los EE.UU. 61/939.256, 12-Feb-2014, y WO 2015/089473 (PCT/US2014/070152), 12-dic-2014, cada una de las cuales se titula ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION. También se menciona PCT/US2015/045504, 15-ago-2015, Solicitud de los EE.UU. 62/180.699, 17-jun-2015, y Solicitud de los EE.UU. 62/038.358, 17-ago-2014, cada una de las cuales se titula GENOME EDITING USING CAS9 NICKASES.

Cada una de estas patentes, publicaciones de patente y solicitudes, y todos los documentos citados en la presente o durante su procesamiento ("documentos citados en la solicitud") y todos los documentos citados o referenciados en los documentos citados en la solicitud, junto con cualquier instrucción, descripción, especificaciones de producto y hojas de producto para cualquier producto mencionado en la presente o en cualquier documento en la presente e incorporado por referencia en la presente, puede ser empleado en la práctica de la invención.

La eficacia de la presente invención, se ha demostrado. Los complejos de CRISPR-CPF1 recombinantes prearmados que comprenden CPF1 y ARNcr se pueden transfectar, por ejemplo, mediante electroporación, dando como resultado altas tasas de mutación y la ausencia de mutaciones detectables fuera de objetivo. Hur, J.K. y col. Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596. [Epub previo a la impresión]. El análisis de todo el genoma muestra que Cpf1 es altamente específica. Par una medición, los sitios de clivaje *in vitro* determinados para SpCas9 en las células humanas HEK293T fueron significativamente menos que para SpCas9. Kim, D. y col., Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells, Nat Biotechnol. 6 de junio de 2016. doi: 10.1038/nbt.3609. [Epub previo a la impresión]. Se ha demostrado en Drosophila un sistema multiplexado eficiente que utiliza Cpf1 que emplea genes producidos a partir de una matriz que contiene los ARNt de la invención. Port, F. y col. Expansion of the CRISPR toolbox in an animal with tRNA-flanked Cas9 and Cpf1 gRNAs. doi: http://dx.doi.org/10.1101/046417.

La presente invención se ilustrará adicionalmente los siguientes Ejemplos que solo se dan con propósitos ilustrativos y sn intención de limitar la invención de ninguna manera.

EJEMPLOS

Ejemplo 1. Origen y evolución de los sistemas de inmunidad adaptativa

Clasificación y anotación de los sistemas CRISPR-Cas en genomas de arqueas y bacterias. Los loci en los sistemas CRISPR-Cas abarcan más de 50 familias de genes, y no hay genes estrictamente universales, ya que la evolución es rápida y la arquitectura de los loci es extremadamente diversa. Por consiguiente, no puede concebirse un único árbol y es necesario un abordaje con múltiples partes. Hasta el momento, se han identificado de manera exhaustiva 395 perfiles de genes *cas* que codifican 93 proteínas Cas. La clasificación abarca el perfil de los genes característicos más la arquitectura de los loci característicos.

En la figura 1 se propone una nueva clasificación de los sistemas CRISPR-Cas. La clase 2 abarca los complejos efectores relacionados con el ARNcr con subunidades múltiples (las cascadas), y la clase 2 abarca los complejos relacionados con el ARNcr con subunidades individuales (similares a Cas9). En la figura 2 se representa la organización molecular de los sistemas CRISPR-Cas. En la figura 3 se representa la estructura de los complejos efectores de los tipos I y III: hay una arquitectura y un linaje en común a pesar de la divergencia extensa a nivel de las secuencias. En la figura 4 se representa el motivo de reconocimiento del ARN (RRM) de un sistema centrado en CRISPR-Cas. En la figura 5 se representa la filogenia de Cas1, donde la recombinación de los módulos relacionados con la adaptación y los módulos efectores asociados al ARNcr es un aspecto importante de la evolución de los sistemas CRISPR-Cas. En la figura F se representa un censo de los sistemas CRISPR-Cas que tuvo por objeto determinar de manera específica la distribución de los tipos y los subtipos de los sistemas CRISPR-Cas en las arqueas y en las bacterias.

Cas1 no siempre está asociado a los sistemas CRISPR-Cas, por lo que es posible que haya dos ramas "individuales" de Cas1, a partir de lo cual puede inferirse que podría haber diferencias en la función y en el origen y que podría haber elementos móviles novedosos (véase Makarova, Krupovic, Koonin, Frontiers Genet, 2014). Sobre la base de la organización del genoma de tres familias de casposones, podría obtenerse información más detallada. Además de Cas1 y PolB, los casposones abarcan una variedad de genes, entre los cuales pueden mencionarse diversos genes que codifican diversas nucleasas (Krupovic y col., BMC Biology, 2014). Una familia abarca una polimerasa que es cebada por las proteínas, mientras que otra familia abarca una polimerasa que es cebada por el ARN. Además de diversos miembros de los filos Euryarqueasota y Thaumarqueasota, los casposones pueden hallarse en una variedad de bacterias, a partir de lo cual puede inferirse que hay una movilidad horizontal. Sobre la base de la naturaleza del casposón Cas1 (transposasa/integrasa), puede inferirse que hay un clado basal en la filogenia de Cas1.

En las bacterias y en las arqueas, los arreglos CRISPR son útiles para evadir la inmunidad adaptativa que puede hallarse en los procariotas o en los eucariotas, a través de la manipulación del genoma. Cas 1 podría ser una herramienta útil para manipular el genoma. Hay mecanismos similares a través de los cuales puede tener la integración en los casposones y en los arreglos CRISPR, específicamente la adquisición dependiente de la replicación sobre la base de un proceso de copia y pegado, y no a través de un proceso de corte y pegado (Krupovic y col., BMC Biology, 2014). Cas1 es una integrasa bona fide (Nuñez, J. K., Lee, A. S., Engelman, A., Doudna, J. A., Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity, Nature, 18 de Febrero de 2015). Hay una similitud entre las repeticiones invertidas en los extremos de los casposones y los arreglos CRISPR (Krupovic y col., BMC Biology, 2014). Los sistemas CRISPR-Cas pueden haberse originado en un casposón y comprenden un locus asociado a la inmunidad innata (Koonin, Krupovic, Nature Rev. Genet., 2015). La evolución de los sistemas relacionados con la inmunidad adaptativa en los procariotas y en los animales puede haber ocurrido a lo largo de cursos paralelos con la integración de los transposones en los loci asociados a la inmunidad innata (Koonin, Krupovic, Nature Rev. Genet., 2015). La transposasa RAG1 (la enzima fundamental en el contexto de la recombinación de V(D)J en los vertebrados) puede haberse originado en los transposones Transib (Kapitonov, V. V., Jurka, J., RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons, PLoS Biol., junio de 2005; 3(6): e181), pero ningún transposón Transib codifica RAG2. Los transposones que codifican RAG1 y RAG2 fueron descriptos en la publicación de Kapitonov, Koonin, Biol. Direct., 2015, al igual que la filogenia de la transposasa Transib. La eliminación defensiva del ADN en los ciliados evolucionó a partir de un transposón PiggyMAc y el ARNi, que constituyen un sistema asociado a la inmunidad innata (Swart, E. C., Nowacki, M., The eukaryotic way to defend and edit genomes by sRNA-targeted DNA deletion, Ann. NY Acad. Sci., 2015).

La estabilidad relativa de la clasificación implica que la mayoría de las variantes predominantes de los sistemas CRISPR-Cas ya son conocidas. Sin embargo, la existencia de variantes raras que todavía no pueden ser clasificadas en la actualidad implica que todavía podría ser necesario caracterizar otros tipos o subtipos (Makarova y col., 2015, Evolutionary classification of CRISPR-Cas systems and cas genes).

Los transposones proveen una contribución fundamental para la evolución de la inmunidad adaptativa y de los otros sistemas que participan en la manipulación del ADN. Los sistemas CRISPR-Cas de la clase 1 se han originado en los transposones, pero esto solamente se aplica a los módulos relacionados con la adaptación. Los sistemas CRISPR-Cas de la clase 2 presentan funciones adaptativas y efectoras, y los módulos pueden haber evolucionado a partir de una variedad de transposones.

Ejemplo 2. Nuevos sistemas CRISPR-Cas de la clase 2 predichos y evidencias de sus orígenes independientes a partir de elementos del tipo de los transposones

Los sistemas CRISPR-Cas que están relacionados con la inmunidad adaptativa de las bacterias y de las arqueas

presentan una diversidad extrema a nivel de la composición de sus proteínas y la arquitectura de sus loci genómicos. En general, estos sistemas se dividen en dos clases, la clase 1, que abarca los complejos efectores que comprenden múltiples subunidades, y la clase 2, que abarca los complejos efectores que comprenden subunidades individuales, tales como la proteína Cas9. Los solicitantes desarrollaron un abordaje informático sencillo para poder predecir otros sistemas CRISPR-Cas de la clase 2 potenciales. Mediante el análisis de una base de datos de genomas de bacterias que se llevó a cabo de acuerdo con este abordaje, fue posible identificar dos nuevas variantes, cada una de las cuales estuvo presente en diversas bacterias y comprendió los genes Cas1 y *cas2*, junto con un tercer gen que codificaba una proteína grande, de la cual se predijo que operaría como el módulo efector. En el primero de estos loci, la proteína efectora potencial (C2c1p) comprendió un dominio del tipo de las nucleasas que fue similar al de RuvC y fue semejante a la proteína Cpf1, que había sido descripta con anterioridad y que es el efector predicho para los sistemas CRISPR-Cas del tipo IV; por consiguiente, este nuevo sistema potencial fue clasificado como el subtipo V-B. Cuando se realizó una comparación más profunda entre las secuencias de las proteínas, fue posible inferir que las proteínas efectoras que contienen los motivos propios de RuvC, Cas9, Cpf1 y C2C1p, evolucionaron de manera independiente para generar una variedad de grupos de proteínas TnpB que están codificadas por transposones. El segundo grupo de loci novedosos potencialmente asociados a los sistemas CRISPR-Cas abarcó una proteína grande que contenía dos dominios HEPN altamente divergentes, de la que se predijo que podría presentar una actividad propia de una RNAsa. Debido al carácter novedoso de la proteína efectora predicha, estos loci fueron clasificados como nuevos sistemas CRISPR-Cas del tipo IV que probablemente tendrán como diana el ARNm. Sobre la base del resultado de este análisis, fue posible concluir que los sistemas CRISPR-Cas de la clase 2 evolucionaron en múltiples ocasiones independientes, en combinación con diversos módulos relacionados con la adaptación que codificaban Cas1 o Cas2 y con proteínas efectoras derivadas de una variedad de elementos móviles. Es probable que este camino evolutivo haya dado como resultado la generación de múltiples variantes de los sistemas de la clase 2 que todavía no hayan sido descubiertas.

Los sistemas CRISPR-Cas relacionados con la inmunidad adaptativa están presentes en el genoma de ~45% de las bacterias y en el genoma de ~90% de las arqueas y presentan una diversidad extrema a nivel de la composición y la secuencia de las proteínas y de la arquitectura de los loci genómicos. Sobre la base de la organización estructural de los complejos efectores asociados al ARNcr, estos sistemas son divididos en dos clases, es decir, la clase 1, que abarca los complejos efectores con subunidades múltiples, y la clase 2, que abarca los complejos con subunidades individuales (Makarova, 2015). Los sistemas de la clase 1 son mucho más comunes y diversos que los sistemas de la clase 2. En la actualidad, la clase 1 está representada por 12 subtipos diferentes, los cuales están codificados en el genoma de numerosas bacterias y en el genoma de numerosas arqueas, mientras que los sistemas de la clase 2 abarcan tres subtipos de sistemas del tipo II y los sistemas potenciales del tipo V, que en conjunto pueden hallarse en aproximadamente 10% de los genomas de las bacterianos que se han secuenciado (los sistemas de este tipo solamente han sido hallados en el genoma de una única arquea). Los sistemas de la clase 2 típicamente contienen tres o cuatro genes en el operón *cas*, es decir, el par de genes Cas1-*cas2*, que participan en la adaptación pero no en la interferencia, así como una única proteína efectora con múltiples dominios, la cual es responsable de la interferencia pero también contribuye al procesamiento y la adaptación preliminares del ARNcr, y frecuentemente también contienen un cuatro gen, cuya función todavía no ha sido establecida, que podría estar relacionado con al menos algunos de los sistemas del tipo II. En la mayoría de los casos, hay un arreglo CRISPR y un gen que codifica un tipo de ARN en particular, que puede ser conocido como ARNcrtrac (el ARN asociado a CRISPR que está codificado en trans) en una posición adyacente a un operón *cas* de la clase 2 (Chylinski, 2014). El ARNcrtrac presenta una homología parcial con las repeticiones que pueden hallarse en algunos arreglos CRISPR, y es esencial para el procesamiento del ARNcr preliminar, una acción que es catalizada por la ARNasa III, la cual a su vez es una enzima que puede hallarse de manera ubicua en las bacterias y que no está asociada a los loci relacionados con CRISPR-Cas (Deltcheva, 2011) (Chylinski, 2014; Chylinski, 2013).

La caracterización funcional y estructural de Cas9, una proteína efectora del tipo II que comprende múltiples dominios, ha sido muy detallada. En las diversas bacterias, las proteínas Cas9 comprenden entre aproximadamente 950 y aproximadamente 1400 aminoácidos, y contienen dos dominios propios de las nucleasas, que más precisamente son dominios similares a los que pueden hallarse en las nucleasas RuvC (una ARNasa con un pliegue en forma de H) y HNH (una nucleasa similar a McrA) (Makarova, 2011). Sobre la base de la estructura cristalina de Cas9, puede establecerse que la proteína comprende dos lóbulos diferentes: uno que participa en el reconocimiento del diana y otro que opera como nucleasa, donde este último es el que abarca los dominios propios de las nucleasas RuvC y HNH (Nishimasu, 2014) (Jinek, 2014). Cada uno de los dominios propios del tipo de las nucleasas de Cas9 es necesario para que sea posible clivar una de las cadenas del ADN diana (Jinek, 2012; Sapranauskas, 2011). Recientemente, se ha demostrado que Cas9 participa en las tres etapas de la respuesta asociada a los arreglos CRISPR, es decir, que no solamente participa en el clivaje del ADN (la interferencia), sino que también participa en la adaptación y en el procesamiento del ARNcr preliminar (Jinek, 2012). Más específicamente, se ha demostrado que un dominio distintivo que se encuentra en el lóbulo del tipo de las nucleasas de Cas9 puede reconocer el motivo asociado a los protoseparadores (PAM) que se encuentran en el ADN de origen viral durante la etapa de adaptación, para luego unirse a ellos (Nishimasu, 2014) (Jinek, 2014) (Heler, 2015; Wei, 2015).

En esta etapa de la respuesta relacionada con los arreglos CRISPR, Cas9 forma un complejo con Cas1 y Cas2, las dos proteínas que participan en la adquisición de los separadores en todos los sistemas CRISPR-Cas (Heler, 2015; Wei, 2015). La proteína Cas9, en combinación con el ARNcrtrac, recientemente se ha convertido en una herramienta fundamental para el desarrollo de métodos útiles en el contexto de la modificación y la alteración del genoma

(Gasiunas, 2013; Mali, 2013; Sampson, 2014; Cong, 2015). La utilidad de Cas9 en la modificación del genoma se basa en el hecho de que, en los sistemas CRISPR-Cas del tipo II, a diferencia de los sistemas CRISPR-Cas de otros tipos, todas las actividades que son necesarias para que tenga lugar el reconocimiento y el clivaje del ARN se basan en una misma proteína grande que comprende múltiples dominios. Esta característica de los sistemas del tipo II facilita en gran medida el desarrollo de herramientas eficientes para manipular el genoma. Vale destacar que no todas las variantes de Cas9 son iguales. Hasta el momento, la mayoría del trabajo ha sido realizado con la proteína Cas9 de *Streptococcus pyogenes*, pero podrían obtenerse resultados más ventajosos sobre la base del análisis de las proteínas Cas9 otras especies. A modo de ejemplo en este contexto, por medio de experimentos recientes que se llevaron a cabo con la proteína Cas9 de *Staphylococcus aureus*, que comprende 300 aminoácidos menos que la proteína de *S. pyogenes*, ha sido posible introducir la proteína Cas9 en un vector del tipo de los virus adenoasociados, con lo cual fue posible obtener una ventaja notable a nivel de la utilidad de los sistemas CRISPR-Cas en la modificación del genoma in vivo (Ran, 2015).

En la actualidad, los sistemas CRISPR-Cas del tipo II son clasificados en 3 subtipos (II-A, II-B y II-C) (Makarova, 2011) (Fonfara, 2014; Chylinski, 2013; Chylinski, 2014). Además de los genes Cas1, *cas2* y *cas9*, que están presentes en la totalidad de los loci propios del tipo II, el subtipo II-A se caracteriza por comprender un gen adicional, *csn2*, que codifica una ATPasa inactivada (Nam, 2011; Koo, 2012; Lee, 2012) cuya función en la adquisición de los separadores todavía no ha podido ser caracterizada en una medida apropiada (Barrangou, 2007; Arslan, 2013) (Heler, 2015). Los sistemas del subtipo II-B carecen de un gen *csn2*, pero en su lugar contienen un gen *cas4*, el cual es un gen propio de los sistemas del tipo I, que codifica una exonucleasa 5'-3' de la familia recB que participa en la adquisición de los separadores al generar extremos recombinogénicos en el ADN (Zhang, 2012) (Lemak, 2013; Lemak, 2014). Los genes Cas1 y *cas2* que se encuentran en el subtipo II-B están relacionados de una manera más estrecha con las proteínas respectivas de los sistemas CRISPR-Cas del tipo I, a partir de lo cual puede inferirse que el subtipo II podría haberse originado como resultado de un proceso de recombinación (Chylinski, 2014).

Los sistemas CRISPR-Cas del subtipo II-C constituyen una variante poco frecuente y tan solo comprenden los genes Cas1, *cas2* y *cas9* (Chylinski, 2013; Koonin, 2013; Chylinski, 2014). Sin embargo, en este contexto vale destacar que se ha demostrado que la adquisición de los separadores en *Campylobacter jejuni*, una acción que es ejercida por los sistemas del tipo II-C, depende de la participación de la proteína Cas4, que es codificada por un bacteriófago (Hooton, 2014). Otra característica propia de los sistemas del subtipo II-C es el hecho de que el ARNcr se forma como resultado de un proceso de transcripción complejo, en el cual hay una transcripción que tiene lugar a partir de promotores internos alternativos, en contraste con los procesos que pueden observarse en los otros sistemas CRISPR-Cas que han sido caracterizados en el ámbito experimental (Zhang, 2013).

Recientemente, ha sido posible predecir la existencia de los sistemas CRISPR-Cas del tipo V como resultado de un análisis comparativo de los genomas de diversas bacterias. Estos sistemas CRISPR-Cas potenciales novedosos han podido ser hallados en los genomas de diversas bacterias, particularmente en los genomas de las bacterias del género *Francisella*, así como en el genoma de una arquea, *Methanomethylophilus alvus* (Vestergaard, 2014). Otro sistema del tipo V potencial abarca los genes Cas1 y *cas2*, un gen específico que es conocido como *cpf1* y un arreglo CRISPR (Schunder, 2013) (Makarova, 2015). Cpf1 es una proteína grande (de aproximadamente 1300 aminoácidos) que contiene un dominio que es similar al de una nucleasa RuvC y que presenta homología con el domino correspondiente de Cas9, así como la contraparte del conjunto rico en arginina que es característico de Cas9. Sin embargo, Cpf1 carece del dominio propio de la nucleasa HNH que está presente en la totalidad de las proteínas Cas9, mientras que el dominio del tipo de las nucleasas similar al de RuvC se encuentra en una posición contigua en la secuencia que codifica Cpf1, en contraste con lo que puede observarse en Cas9, donde comprende porciones insertadas largas que abarcan el dominio HNH (Chylinski, 2014; Makarova, 2015). Sobre la base de estas diferencias importantes que hay a nivel de la arquitectura de los dominios de Cas9 y de Cpf1, puede concluirse que los sistemas que contienen Cpf1 podrían ser clasificados como un tipo novedoso. En función de la composición que presentan los sistemas potenciales del tipo V, podría inferirse que Cpf1 es un complejo efector que comprende una única subunidad, por lo cual los sistemas de este tipo podrían ser incluidos entre los sistemas CRISPR-Cas de la clase 2. Algunos de los loci asociados a los sistemas potenciales del tipo V codifican Cas4, por lo cual son similares a los loci propios del subtipo II-B, mientras que otros carecen de Cas4, por lo cual son análogos a los del subtipo II-C.

Se ha comprobado que los homólogos más cercanos de las proteínas Cas9 y Cpf1 son las proteínas TnpB, las cuales están codificadas por los transposones de la familia IS605 y contienen un dominio del tipo de las nucleasas similar al de RuvC y un dominio con dedos de Zn que es similar al que puede hallarse en Cpf1. Por otro lado, ha sido posible identificar homólogos de TnpB que contienen un dominio HNH insertado en el dominio del tipo de las nucleasas similar al de RuvC, cuya secuencia presenta una similitud importante con la de Cas9. La participación de TnpB en los transposones todavía no ha podido ser determinada, ya que se ha comprobado que esta proteína no es necesaria en el contexto del proceso de transposición.

Debido a la homología que hay entre Cas9, Cpf1 y las proteínas que están codificadas por los transposones, los solicitantes han planteado la hipótesis de que los sistemas CRISPR-Cas de la clase 2 podrían haber evolucionado en múltiples ocasiones como resultado de la recombinación entre un transposón y un locus Cas1-*cas2*. Por consiguiente, los solicitantes desarrollaron una estrategia informática simple para identificar los loci genómicos que pudieran ser variantes novedosas de los sistemas de la clase 2. En la presente, los solicitantes describen la primera aplicación de este abordaje, como resultado de la cual fue posible identificar dos grupos de variantes novedosas potenciales, una

de las cuales pareció ser un subtipo diferente del tipo V, mientras que la otra podría pertenecer al tipo VI. Las variantes novedosas de los sistemas CRISPR-Cas de la clase2 evidentemente podrían resultar de interés como herramientas potenciales en el contexto de la modificación del genoma y de la regulación de la expresión.

Estrategia basada en una búsqueda en una base de datos para detectar los loci novedosos potenciales relacionados con los sistemas CRISPR-Cas de la clase 2. Los solicitantes desarrollaron un abordaje informático sencillo con el propósito de identificar los loci novedosos potenciales relacionados con los sistemas CRISPR-Cas de la clase 2 (figura 7: abordaje). Debido a que la amplia mayoría de los loci relacionados con los sistemas CRISPR-Cas abarcan un gen Cas1 (Makarova, 2011; Makarova, 2015) y a que la secuencia de Cas1 se encuentra altamente conservada, en particular entre las proteínas Cas (Takeuchi, 2012), los solicitantes plantearon el postulado de que Cas1 podría ser la base más apropiada para poder identificar los loci novedosos potenciales, a través de un análisis del perfil de Cas1 mediante el uso de la herramienta de búsqueda PSI-BLAST. Después de detectar la totalidad de los contigs que codificaban Cas1, fue posible predecir los genes que codificaban proteínas mediante el uso del programa GenemarkS, donde estos genes se localizaron a una distancia de 20 kb tanto hacia el extremo 5' como hacia el extremo 3' del gen Cas1. Los genes que se predijeron fueron registrados en la base de datos CDD del NCBI, sobre la base de los perfiles específicos de las proteínas Cas, y los arreglos CRISPR fueron predichos mediante el uso del programa PILER-CR. Merced a este procedimiento, fue posible asignar los loci asociados a los sistemas CRISPR-Cas que se detectaron a los subtipos ya conocidos. Los sistemas novedosos potenciales de la clase 2 abarcaron loci asociados a CRISPR-Cas que todavía no habían sido clasificados, que codificaban proteínas grandes (de más de 500 aminoácidos), debido a la observación de características como la presencia de proteínas efectoras que comprendían subunidades individuales como las que pueden observarse en los sistemas de los tipos II y V (Cas9 y Cpf1, respectivamente). La totalidad de los 63 loci novedosos potenciales que se detectaron sobre la base de los criterios que se han descripto fueron analizados de manera individual con las herramientas PSI-BLAST y HHpred. A su vez, las secuencias de las proteínas que estuvieron codificadas por los loci novedosos potenciales fueron usadas para llevar a cabo búsquedas en las bases de datos metagenómicas, con el propósito de descubrir otros homólogos, y los contigs más largos que se detectaron como consecuencia de estas búsquedas fueron analizados de acuerdo con el protocolo que se describió con anterioridad. Merced al análisis que se llevó a cabo de acuerdo con el abordaje que se ha descripto, fue posible obtener dos grupos de loci con una conexión estrecha con los sistemas CRISPR-Cas.

Sistema potencial del tipo V-B. El primer grupo de loci novedosos potenciales, al cual se la había dado la designación provisoria de C2c1 (los loci novedosos potenciales 1 de la clase 2), está presente en el genoma de las bacterias de cuatro filos importantes, a saber, los bacilos, los verrucomicrobios, las alfa-proteobacterias y las delta-proteobacterias (figura 8, "Organización completa de los loci de los sistemas de la clase 2"). Todos los loci C2c1 codifican una fusión de Cas1 y Cas4, Cas2 y la proteína grande a la cual los solicitantes le han dado la designación C2c1p, y típicamente se encuentran en una posición adyacente a un arreglo CRISPR (figura 9, cercanías de C2c1). En el árbol filogenético de Cas1, las proteínas Cas1 respectivas están agrupadas como las de un sistema del tipo I-U (figura 10, árbol de Cas1), aunque también se ha hallado una fusión de Cas1 y Cas4. La longitud de las proteínas C2c1p que se identificaron en el contexto de la presente invención fue de aproximadamente 1200 aminoácidos, y por medio de una búsqueda con el algoritmo HHpred fue posible determinar que hay una similitud significativa entre la porción correspondiente al extremo C de las proteínas C2c1p y un subconjunto de las proteínas TnpB que están codificadas por los transposones de la familia IS605. En contraste, no fue posible detectar una similitud significativa entre C2c1p y Cas9 o Cpf1, que son similares a otras proteínas del grupo TnpB (Chylinski, 2014) (Makarova, 2015; Makarova, 2015). Por consiguiente, puede concluirse que la arquitectura de los dominios de C2c1p es similar a la de los de Cpf1 y es diferente de la de los Cas9, aunque las tres proteínas Cas parecen haber evolucionado a partir de la familia TnpB (figura 11, "Organización de los dominios de las familias de la clase 2"). La porción correspondiente al extremo N de C2c1p no presentó una similitud significativa con la de otras proteínas. Sobre la base de la predicción de la estructura secundaria, podría concluirse que esta región ha de presentar una conformación con forma de hélice alfa. Los dos segmentos en los que se observó una cierta similitud con las proteínas TnpB abarcaron los tres motivos catalíticos de la porción del tipo de las nucleasas similar a la de RuvC y presentaron el motivo de residuos de aminoácidos DED propio de su porción catalítica (Aravind y col., 2000, Nucleic Acids Res, vol. 28, 3417-3432) (figura 12, "Regiones en las que TnpB presenta homología con las proteínas de la clase 2"), la región que corresponde al puente de la hélice (que también es conocida como la agrupación rica en arginina), que en el caso de la proteína Cas9 participa en la unión al ARNcr, y una región que parecería corresponder a la porción que contiene los dedos de Zn de la proteína TnpB, aunque los residuos de cisteína propios de la porción que contiene los dedos de Zn están ausentes en la mayoría de las proteínas C2c1, por lo cual podría inferirse que estas proteínas no pueden unirse al cinc.

En función de la similitud que se observó a nivel de la arquitectura de los dominios de C2c1p y de Cpf1, puede inferirse que los loci de C2c1 podrían clasificarse como loci propios del subtipo V-B, en cuyo caso los loci que codificaran Cpf1 pertenecerían al subtipo V-A. A pesar de la similitud que se observó entre los genes Cas1 que estuvieron asociados a este sistema, las repeticiones en los arreglos CRISPR respectivos parecieron ser altamente heterogéneas, aunque todas ellas presentaron una longitud de 36-37 pares de bases y fueron clasificadas como repeticiones carentes de estructura (su energía para el plegamiento, ΔG, fue de -0,5-4,5 kcal/mol, mientras que el valor de ΔG para las proteínas en los arreglos CRISPR altamente palindrómicos fue inferior a -7). Sobre la base del esquema que se ha propuesto para clasificar los arreglos CRISPR (Lange, 2013), podría concluirse que varias de las repeticiones que pertenecen al subtipo V-B presentan una cierta similitud en su secuencia o en su estructura con las repeticiones del tipo II.

Si se tiene en cuenta la posibilidad de que los sistemas CRISPR-Cas del subtipo potencial V-B sean mecánicamente

análogos a los sistemas del tipo II, los solicitantes intentaron identificar el ARNcrtrac en los respectivos loci del genoma

A través de la comparación de los separadores de los arreglos CRISPR del tipo V-B con una base de datos de secuencias de nucleótidos no redundantes, fue posible hallar varias coincidencias con diversos genomas bacterianos. La relevancia de estas coincidencias es difícil de evaluar si se tiene en cuenta que no se conocen fagos asociados a las bacterias en las que se hallan los sistemas CRISPR-Cas del tipo V-B potenciales que se han descripto.

Sistemas potenciales del tipo VI. El segundo grupo de loci CRISPR-Cas potenciales, que fue conocido como C2c2, fue identificado en los genomas de 5 filos de bacterias, las alfa-proteobacterias, los bacilos, los clostridios, las fusobacterias y los bacteroidetes (figura 8, "Organización completa de los loci de los sistemas de la clase 2"). Al igual que en el caso de c2c1, los loci de C2d2 abarcaron genes Cas1 y *cas2* a lo largo de una proteína grande (C2c2p) y un arreglo CRISPR. Sin embargo, a diferencia de C2c1, C2c2p suele codificar una gran variedad de arreglos CRISPR pero ninguna fusión de Cas1 y Cas2 (figura 13, véase C2c2). En el árbol filogenético de Cas1, las proteínas de Cas1 que se originan en los loci relacionados con C2c2 se distribuyen entre dos clados. El primer clado abarca las proteínas Cas1 de los clostridios y se encuentra dentro del subárbol del tipo II, junto con una pequeña rama del tipo III (figura 10, árbol de Cas1). El segundo clado está compuesto por proteínas las Cas1 y C2c2 de *Leptotrichia* y comprende una rama mixta que contiene la mayoría de las proteínas Cas1 de los sistemas CRISPR-Cas del tipo III-A. Por medio de las búsquedas de bases de datos que se realizaron con las herramientas HHpred y PSI-BLAST, no fue posible detectar ninguna similitud entre la secuencia de C2c2p y las secuencias de las otras proteínas. Sin embargo, a través de la inspección de múltiples alineamientos de secuencias de proteínas C2c2p, fue posible identificar dos motivos conservados estrictamente del tipo RxxxxH que fueron característicos de los dominios HEPN (Anantharaman, 2013). Sobre la base de la estructura secundaria que se ha predicho, puede concluirse que estos motivos se encontrarían dentro de contextos estructurales compatibles con la estructura de un dominio HEPN, una conclusión similar a la que puede arribarse sobre la base de la estructura secundaria global que se ha predicho para las partes correspondientes de C2c2p. Los dominios HEPN son pequeños (comprenden ~150 bases), comprenden hélices alfa de las que se ha predicho que presentan actividad de RNasas y con frecuencia están asociados a diversos sistemas de defensa (Anantharaman, 2013) (figura 14, motivos HEPN RxxxxH de familia C2c2). Las secuencias de los dominios HEPN presentan una conservación escasa, excepto por el motivo catalítico RxxxxH. Por lo tanto, parece probable que C2c2p contenga dos dominios HEPN activos. El dominio HEPN no es nuevo en los sistemas CRISPR-Cas, ya que tiende a estar asociado al dominio CARF (los arreglos CRISPR que están asociados a pliegues de Rossmann) en las proteínas Csm6 y Csx1 que están presentes en diversos sistemas CRISPR-Cas del tipo III (Makarova, 2014). Estas proteínas no pertenecen a ninguno de los módulos relacionados con la adaptación ni a los complejos efectores, sino más bien parecen ser componentes de los módulos asociados a la inmunidad que están presentes en la mayoría de los sistemas CRISPR-Cas y participan en la muerte celular programada, así como en las funciones de regulación durante la respuesta mediada por los arreglos CRISPR (Koonin, 2013; Makarova, 2012; Makarova, 2013). Sin embargo, C2c2p difiere de Csm6 y Csx1 debido que se trata de una proteína mucho más grande y a que es la única que está codificada en los loci de C2c2, a excepción de Cas1 y Cas2. Por lo tanto, parece probable que C2c2p sea el efector de estos nuevos sistemas CRISPR-Cas potenciales, y que los dominios HEPN sean sus porciones catalíticas. Aparte de los dominios HEPN predichos, la secuencia de C2c1p no presentó una similitud detectable con otras proteínas, y se prevé que pueda adoptar una estructura secundaria mixta de hélices alfas y láminas beta.

Los arreglos CRISPR en el locus C2c2 son muy heterogéneas, presentan una longitud de entre 35 y 39 pares de bases y no están estructuradas (su energía de plegamiento es de entre -0,9 y 4,7 kcal/mol). De acuerdo con un análisis previo de los arreglos CRISPR (Lange, 2013), estos CRISPR no pertenecen a ninguna de las clases estructurales establecidas y son asignados a 3 de las 6 superclases. Solamente el arreglo CRISPR de *Listeria seeligeri* ha sido asignado la familia de secuencias 24, que generalmente está asociada a los sistemas del tipo II-C.

Mediante el análisis de los separadores en los loci de C2c2, fue posible identificar una región de 30 nucleótidos que fue idéntica a una secuencia genómica de *Listeria weihenstephanensis* y dos accesos imperfectos en los genomas de los bacteriófagos.

Debido a que se ha predicho que C2c2 comprende un único complejo efector, los sistemas de este tipo podrían ser sistemas CRISPR-Cas de un tipo potencial VI. Por otra parte, si se tiene en cuenta que todos los dominios HEPN activos que han sido caracterizados de acuerdo con procedimientos experimentales o mediante el uso de enzimas son RNAsas, es posible que los sistemas del tipo VI actúen a nivel del ARNm.

Los solicitantes aplicaron una estrategia de cálculo simple y sencilla para predecir nuevos sistemas CRISPR-Cas de la clase 2. Los sistemas de la clase 2 que habían sido descriptos con anterioridad, es decir, el tipo II y el tipo V potencial, consistieron en los genes cas1 y cas2 (y en algunos casos también cas4), que comprenden el módulo relacionado con la adaptación y una única proteína grane que comprende el módulo efector. Por lo tanto, los solicitantes conjeturaron que cualquier locus genómico que contuviera cas1 y una proteína grande podría ser un miembro potencial de un nuevo sistema de la clase 2, lo cual podría ameritar una investigación detallada. Mediante un análisis en el que usaron métodos sensibles para la comparación de las secuencias de diversas proteínas, fue posible identificar dos candidatos sólidos, uno de los cuales perteneció a un subtipo del tipo V que se describió con anterioridad, mientras que el otro fue calificado como un nuevo tipo VI potencial, en virtud de la presencia de una nueva proteína efectora predicha. Varios de estos sistemas novedosos se originan en el genoma de bacterias que no abarcan ningún otro loci propio de los sistemas CRISPR-Cas, sobre la base de lo cual puede llegarse a la conclusión de que los sistemas de los tipos VI

y V podrían operar de forma autónoma.

En combinación con los resultados de los análisis anteriores, (Chyliński, 2014; Makarova, 2011), mediante la identificación del tipo V-B potencial podría subrayarse el tema dominante en la evolución de los sistemas CRISPR-Cas de la clase 2. Las proteínas efectoras de todos los sistemas conocidos actualmente para esta clase parecen haber evolucionado a partir de un conjunto de elementos transponibles que codifican proteínas TnpB que contienen un dominio similar al de RuvC. Las secuencias de los dominios similares a los de RuvC de las proteínas TnpB y las secuencias de los dominios homólogos de las proteínas efectoras de la clase 2 son demasiado divergentes para poder llevar a cabo un análisis filogenético confiable. Sin embargo, en el caso de Cas9, la proteína efectora de los sistemas del tipo II, el antepasado específico parece ser fácilmente identificable, es decir, parece tratarse de una familia de proteínas similares TnpB, las cuales son particularmente abundantes en las cianobacterias, que comprenden secuencias con una similitud relativamente alta con la de Cas9 y que presentan dominios con una arquitectura similar, es decir, comprenden dominios del tipo de las nucleasa similares al de RuvC y dominios HNH, así como un puente de la hélice rico en arginina (Chylinski, 2014) (figura 11, "Organización del dominio de las familias de la clase 2"; figura 12, "Regiones de TnpB que presentan homología con las proteínas de la clase 2"). A diferencia de Cas9, fue imposible adjudicar Cpf1 y C2c1 a una familia TnpB específica; a pesar de la conservación de todos los motivos centrados en los residuos catalíticos de las nucleasas similares a RuvC, estas proteínas presentaron solamente una similitud limitada con los perfiles genéricos de TnpB. Sin embargo, debido a que C2c1p no comprende una secuencia con una similitud detectable con Cpf1 y contiene porciones insertadas diferentes entre los motivos propios de RuvC y extremos N claramente no relacionados, parece más probable que Cpf1 y C2c1 se hayan originado de manera independiente de las diferentes familias que abarcan los conjuntos de componentes que codifican TnpB.

Es interesante que las proteínas TnpB parecen haber sido "prediseñadas" para participar en los complejos efectores relacionados con los sistemas CRISPR-Cas de la clase 2, de manera tal que aparentemente han sido reclutadas en múltiples ocasiones diferentes. Es concebible, por ejemplo, que la utilidad de las proteínas TnpB esté relacionada con su capacidad prevista de cortar una molécula de ADN de cadena simple mientras está unida a una molécula de ARN a través de la hélice rica en R, de la cual en Cas9 se ha demostrado que puede unirse al ARNcr (Jinek, 2014; Nishimasu, 2014). Las funciones de TnpB son poco conocidas. Esta proteína no es necesaria para la transposición, y en un caso se ha demostrado que la transposición se encuentra regulada de manera negativa (Pasternak, 2013), aunque se desconoce su mecanismo de acción. Merced a un estudio experimental con TnpB, es probable que obtenga más información acerca de los aspectos mecánicos de los sistemas CRISPR-Cas de la clase 2. Cabe señalar que los mecanismos propios de Cpf1 y de C2c1 podrían ser similares, pero en cualquier caso deberían diferir sustancialmente de los de Cas9, debido a las dos primeras proteínas carecen del dominio HNH que en Cas9 es responsable de cortar una de las cadenas del ADN diana (Gasiunas, 2012) (Jinek, 2012) (Chen, 2014). Como consecuencia, mediante la explotación de Cpf1 y de C2c1, podrían obtenerse otros abordajes útiles para modificar el genoma.

En términos evolutivos, resulta sorprendente que los sistemas CRISPR-Cas de la clase 2 parezcan derivar de elementos transponibles completamente diferentes, debido a la evidencia reciente sobre acerca de la probabilidad de que los genes *cas1* se hayan originado en una familia de transposones diferente (Koonin, 2015; Krupovic, 2014). Por otra parte, debido a que es posible que las proteínas efectoras de las diversas familias de proteínas TnpB y las afinidades filogenéticas de las proteínas Cas1 respectivas son diferentes, podría llegarse a la conclusión de que los sistemas de la clase 2 han evolucionado en múltiples ocasiones como resultado la combinación de varios módulos relacionados con la adaptación y nucleasas derivadas de transposones que dieron lugar a las proteínas efectoras. Esta modalidad de evolución parecería ser la última manifestación de la modularidad, que es característica de la evolución de los sistemas CRISPR-Cas (Makarova, 2015), con la implicación de que es probable haya combinaciones adicionales de módulos relacionados con la adaptación y efectores en la naturaleza.

Los sistemas CRISPR-Cas del tipo VI potenciales abarcan una nueva proteína efectora predicha que contiene dos dominios HEPN predichos, los cuales podrían presentar actividad de ARNasas. Los dominios HEPN no forman parte de los complejos efectores en otros sistemas CRISPR-Cas, pero participan en una variedad de funciones de defensa que abarcan una función accesoria prevista en los diferentes sistemas CRISPR-Cas (Anantharaman, 2013) (Makarova, 2015). En función de la presencia de los dominios HEPN como la porción catalítica del módulo efector predicho, podría concluirse que los sistemas del tipo VI clivan el ARNm diana. Con anterioridad, se había indicado que determinados sistemas CRISPR-Cas del tipo III estaban dirigidos al ARNm (Hale, 2014; Hale, 2009) (Peng, 2015). Si bien hasta el momento no se han detectado dominios HEPN en los elementos de transposición bona fide, se caracterizan por una alta movilidad horizontal y son una parte integral de los elementos móviles, tales como las unidades de toxinas y antitoxinas (Anantharaman, 2013). Por lo tanto, los sistemas potenciales del tipo VI parecen encajar en el paradigma general de la evolución modular de los sistemas CRISPR-Cas de la clase 2 debido a la presencia de componentes móviles, y se espera que se descubran variantes adicionales y nuevos tipos mediante el análisis de la información genómica y metagenómica.

La evolución modular es una característica fundamental de los sistemas CRISPR-Cas. Esta modalidad de evolución parece ser más pronunciada en los sistemas de la clase 2 que evolucionan a través de la combinación de los módulos relacionados con la adaptación de varios otros sistemas CRISPR-Cas con proteínas efectoras que parecen haber sido reclutadas a partir de elementos móviles en múltiples ocasiones independientes. Debido a la diversidad extrema que se observa entre los elementos móviles en las bacterias, es probable que los módulos efectores de los sistemas CRISPR-Cas de la clase 2 también sean muy diversos. En la presente, los solicitantes emplearon un abordaje

informático sencillo para delimitar dos nuevas variantes de los sistemas CRISPR-Cas que probablemente existan en los genomas de bacterias que todavía no se han secuenciado. A pesar de que la mayoría, si no todos estos nuevos sistemas CRISPR-Cas probablemente sean poco frecuentes, podrían estar asociados a estrategias y mecanismos moleculares novedosos y podrían constituir un recurso importante en el contexto del desarrollo nuevas aplicaciones útiles en las áreas de la modificación del genoma y la biotecnología.

Se usó el programa TBLASTN para buscar el perfil Cas1 en la base de datos WGS del NCBI. En este contexto, se recuperaron secuencias de contigs o de particiones completas del genoma donde Cas1 había sido identificado con éxito a partir de la base de datos. La región alrededor del gen Cas1 se cortó y se procesó usando el programa GENMARK. Se buscaron las proteínas previstas en cada caso en una colección de perfiles de la base de datos DDC (Marchler-Bauer, 2009) y en los perfiles específicos de Cas disponibles en la base de datos FTP, con prioridad de acierto para las proteínas Cas. En cada locus, se empleó el procedimiento para identificar la integridad de los loci relacionados con los arreglos CRISPR que había sido desarrollado con anterioridad.

Se recurrió a un análisis de los arreglos CRISPR (Lange, 2013) para clasificar las repeticiones.

Se usaron las búsquedas iterativas de los perfiles que se realizaron con el programa PSI-BLAST (Altschul, 1997), las estadísticas basadas en la composición y condiciones relacionadas con una complejidad baja buscar secuencias similares pero alejadas en la base de datos de secuencias no redundantes (NR) del NCBI. Cada proteína no redundante identificada se comparó con las que se habían hallado en la base de datos WGS usando el programa TBLAST. Se usó el programa HHpred con los parámetros por omisión para identificar las secuencias con una similitud remota (Söding, 2005). Se construyeron múltiples alineamientos con las secuencias usando Muscle (Edgar, 2004). La estructura secundaria de las proteínas se predijo con el programa Jpred 4 (Drozdetskiy, 2015).

Genes candidatos seleccionados

Identificación del gen: A; Tipo de gen: C2C1; Organismo: 5. La bacteria Opitutaceae TAV5; Longitud del separador, modalidad (rango): 34 (entre 33 y 37); DR1: GCCGCAGCGAAUGCCGUUUCACGAAUCGUCAGGCGG (SEQ ID NO: 27); DR2: ninguno; ARNcrtrac1: GCUGGAGACGUUUUUUGAAACGGCGAGUGCUGCGGAUAGCGAGUUUCUCUUGGGGAGGCGCUCGCGGCCA CUUUU (SEQ ID NO: 28); ARNcrtrac2: ninguno; secuencia de la proteína:

MSLNRIYQGRVAAVETGTALAKGNVEWMPAAGGDEVLWQHHELFQAAINYYL
VALLALADKNNPVLGPLISQMDNPQSPYHVWGSFRRQGRQRTGLSQAVAPYITP
GNNAPTLDEVFRSILAGNPTDRATLDAALMQLLKACDGAGAIQQEGRSYWPKF
CDPDSTANFAGDPAMLRREQHRLLLPQVLHDPAITHDSPALGSFDTYSIATPDTR
TPQLTGPKARARLEQAITLWRVRLPESAADFDRLASSLKKIPDDDSRLNLQGYV
GSSAKGEVQARLFALLLFRHLERSSFTLGLLRSATPPPKNAETPPPAGVPLPAASA
ADPVRIARGKRSFVFRAFTSLPCWHGGDNIHPTWKSFDIAAFKYALTVINQIEEK
TKERQKECAELETDFDYMHGRLAKIPVKYTTGEAEPPPILANDLRIPLLRELLQNI
KVDTALTDGEAVSYGLQRRTIRGFRELRRIWRGHAPAGTVFSSELKEKLAGELR
QFQTDNSTTIGSVQLFNELIQNPKYWPIWQAPDVETARQWADAGFADDPLAAL
VQEAELQEDIDALKAPVKLTPADPEYSRRQYDFNAVSKFGAGSRSANRHEPGQT
ERGHNTFTTEIAARNAADGNRWRATHVRIHYSAPRLLRDGLRRPDTDGNEALE
AVPWLQPMMEALAPLPTLPQDLTGMPVFLMPDVTLSGERRILLNLPVTLEPAAL
VEQLGNAGRWQNQFFGSREDPFALRWPADGAVKTAKGKTHIPWHQDRDHFTV
LGVDLGTRDAGALALLNVTAQKPAKPVHRIIGEADGRTWYASLADARMIRLPG

EDARLFVRGKLVQEPYGERGRNASLLEWEDARNIILRLGQNPDELLGADPRRHS
YPEINDKLLVALRRAQARLARLQNRSWRLRDLAESDKALDEIHAERAGEKPSPL
PPLARDDAIKSTDEALLSQRDIIRRSFVQIANLILPLRGRRWEWRPHVEVPDCHIL
AQSDPGTDDTKRLVAGQRGISHERIEQIEELRRRCQSLNRALRHKPGERPVLGRP
AKGEEIADPCPALLEKINRLRDQRVDQTAHAILAAALGVRLRAPSKDRAERRHR
DIHGEYERFRAPADFVVIENLSRYLSSQDRARSENTRLMQWCHRQIVQKLRQLC
ETYGIPVLAVPAAYSSRFSSRDGSAGFRAVHLTPDHRHRMPWSRILARLKAHEE
DGKRLEKTVLDEARAVRGLFDRLDRFNAGHVPGKPWRTLLAPLPGGPVFVPLG
DATPMQADLNAAINIALRGIAAPDRHDIHHRLRAENKKRILSLRLGTQREKARW
PGGAPAVTLSTPNNGASPEDSDALPERVSNLFVDIAGVANFERVTIEGVSQKFAT
GRGLWASVKQRAWNRVARLNETVTDNNRNEEEDDIPM (SEQ ID NO: 29)

Identificación del gen: B; Tipo de gen: C2C1; Organismo: 7. Bacillus thermoamylovorans, cepa B4166; Longitud del separador, modalidad (rango): 37 (35-38); DR1: GUCCAAGAAAAAAGAAAUGAUACGAGGCAUUAGCAC (SEQ ID NO: 30); DR2: ninguno; ARNcrtrac1: CUGGACGAUGUCUCUUUUAUUUCUUUUUUCUUGGAUCUGAGUACGAGCACCCACAUUGGACAUUUCGCAUG GUGGGUGCUCGUACUAUAGGUAAAACAAACCUUUUU (SEQ ID NO: 31); ARNcrtrac2: ninguno; secuencia de la proteína:

MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYEHHEQDPK
NPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDVVFNILRELYEELVPSSVEK
KGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSWEEEKKKW
EEDKKKDPLAKILGKLAEYGLIPLFIPFTDSNEPIVKEIKWMEKSRNQSVRRLDKD
MFIQALERFLSWESWNLKVKEEYEKVEKEHKTLEERIKEDIQAFKSLEQYEKER
QEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEVFKDYQRK
HPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDAKQQATFTL
ADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLIYPTESGGW
EEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTLGGARVQFD
RDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPKFVNFKPKE
LTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQKPDIEGKLF
FPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQKLNFLRNVL
HFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYKDWVAFLK
QLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLLRWSLRPTE
PGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYDVRKKKWQ

AKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQGEIYGLQVG
EVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLDKIAVLKEG
DLYPDKGGEKFISLSKDRKLVTTHADINAAQNLQKRFWTRTHGFYKVYCKAYQ
VDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWGNAGKLKIKKGSSKQSSSE
LVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFGKLERILISK
LTNQYSISTIEDDSSKQSM (SEQ ID NO: 32)

Identificación del gen: C; Tipo de gen: C2C1; Organismo: 9. Bacillus sp. NSP2.1; Longitud del separador, modalidad (rango): 36 (35-42); DR1: GUUCGAAAGCUUAGUGGAAAGCUUCGUGGUUAGCAC (SEQ ID NO: 33); DR2: ninguno; ARNcrtrac1:
CACGGAUAAUCACGACUUUCCACUAAGCUUUCGAAUUUUAUGAUGCGAGCAUCCUCUCAGGUCAAAAAA (SEQ ID NO: 34); ARNcrtrac2: ninguno; secuencia de la proteína:

MAIRSIKLKLKTHTGPEAQNLRKGIWRTHRLLNEGVAYYMKMLLLFRQESTGER
PKEELQEELICHIREQQQRNQADKNTQALPLDKALEALRQLYELLVPSSVGQSG
DAQIISRKFLSPLVDPNSEGGKGTSKAGAKPTWQKKKEANDPTWEQDYEKWKK
RREEDPTASVITTLEEYGIRPIFPLYTNTVTDIAWLPLQSNQFVRTWDRDMLQQAI
ERLLSWESWNKRVQEEYAKLKEKMAQLNEQLEGGQEWISLLEQYEENRERELR
ENMTAANDKYRITKRQMKGWNELYELWSTFPASASHEQYKEALKRVQQRLRG
RFGDAHFFQYLMEEKNRLIWKGNPQRIHYFVARNELTKRLEEAKQSATMTLPN
ARKHPLWVRFDARGGNLQDYYLTAEADKPRSRRFVTFSQLIWPSESGWMEKKD
VEVELALSRQFYQQVKLLKNDKGKQKIEFKDKGSGSTFNGHLGGAKLQLERGD
LEKEEKNFEDGEIGSVYLNVVIDFEPLQEVKNGRVQAPYGQVLQLIRRPNEFPKV
TTYKSEQLVEWIKASPQHSAGVESLASGFRVMSIDLGLRAAAATSIFSVEESSDK
NAADFSYWIEGTPLVAVHQRSYMLRLPGEQVEKQVMEKRDERFQLHQRVKFQI
RVLAQIMRMANKQYGDRWDELDSLKQAVEQKKSPLDQTDRTFWEGIVCDLTK
VLPRNEADWEQAVVQIHRKAEEYVGKAVQAWRKRFAADERKGIAGLSMWNIE
ELEGLRKLLISWSRRTRNPQEVNRFERGHTSHQRLLTHIQNVKEDRLKQLSHAIV
MTALGYVYDERKQEWCAEYPACQVILFENLSQYRSNLDRSTKENSTLMKWAH
RSIPKYVHMQAEPYGIQIGDVRAEYSSRFYAKTGTPGIRCKKVRGQDLQGRRFE
NLQKRLVNEQFLTEEQVKQLRPGDIVPDDSGELFMTLTDGSGSKEVVFLQADIN
AAHNLQKRFWQRYNELFKVSCRVIVRDEEEYLVPKTKSVQAKLGKGLFVKKSD
TAWKDVYVWDSQAKLKGKTTFTEESESPEQLEDFQEIIEEAEEAKGTYRTLFRD
PSGVFFPESVWYPQKDFWGEVKRKLYGKLRERFLTKAR (SEQ ID NO: 35)

Identificación del gen: D; Tipo de gen: C2C2; Organismo: 4. la bacteria Lachnospiraceae NK4A144 G619; Longitud del separador, modalidad (rango): 35; DR1: GUUUUGAGAAUAGCCCGACAUAGAGGGCAAUAGAC (SEQ ID NO: 36); DR2: GUUAUGAAAACAGCCCGACAUAGAGGGCAAUAGACA (SEQ ID NO: 37); ARNcrtrac1: ninguno; ARNcrtrac2: ninguno; secuencia de la proteína:

MKISKVDHTRMAVAKGNQHRRDEISGILYKDPTKTGSIDFDERFKKLNCSAKILY
HVFNGIAEGSNKYKNIVDKVNNNLDRVLFTGKSYDRKSIIDIDTVLRNVEKINAF
DRISTEEREQIIDDLLEIQLRKGLRKGKAGLREVLLIGAGVIVRTDKKQEIADFLEI
LDEDFNKTNQAKNIKLSIENQGLVVSPVSRGEERIFDVSGAQKGKSSKKAQEKE
ALSAFLLDYADLDKNVRFEYLRKIRRLINLYFYVKNDDVMSLTEIPAEVNLEKDF
DIWRDHEQRKEENGDFVGCPDILLADRDVKKSNSKQVKIAERQLRESIREKNIKR
YRFSIKTIEKDDGTYFFANKQISVFWIHRIENAVERILGSINDKKLYRLRLGYLGE
KVWKDILNFLSIKYIAVGKAVFNFAMDDLQEKDRDIEPGKISENAVNGLTSFDYE
QIKADEMLQREVAVNVAFAANNLARVTVDIPQNGEKEDILLWNKSDIKKYKKN
SKKGILKSILQFFGGASTWNMKMFEIAYHDQPGDYEENYLYDIIQIIYSLRNKSFH
FKTYDHGDKNWNRELIGKMIEHDAERVISVEREKFHSNNLPMFYKDADLKKILD
LLYSDYAGRASQVPAFNTVLVRKNFPEFLRKDMGYKVHFNNPEVENQWHSAV
YYLYKEIYYNLFLRDKEVKNLFYTSLKNIRSEVSDKKQKLASDDFASRCEEIEDR
SLPEICQIIMTEYNAQNFGNRKVKSQRVIEKNKDIFRHYKMLLIKTLAGAFSLYL
KQERFAFIGKATPIPYETTDVKNFLPEWKSGMYASFVEEIKNNLDLQEWYIVGRF
LNGRMLNQLAGSLRSYIQYAEDIERRAAENRNKLFSKPDEKIEACKKAVRVLDL
CIKISTRISAEFTDYFDSEDDYADYLEKYLKYQDDAIKELSGSSYAALDHFCNKD
DLKFDIYVNAGQKPILQRNIVMAKLFGPDNILSEVMEKVTESAIREYYDYLKKVS
GYRVRGKCSTEKEQEDLLKFQRLKNAVEFRDVTEYAEVINELLGQLISWSYLRE
RDLLYFQLGFHYMCLKNKSFKPAEYVDIRRNNGTIIHNAILYQIVSMYINGLDFY
SCDKEGKTLKPIETGKGVGSKIGQFIKYSQYLYNDPSYKLEIYNAGLEVFENIDEH
DNITDLRKYVDHFKYYAYGNKMSLLDLYSEFFDRFFTYDMKYQKNVVNVLENI
LLRHFVIFYPKFGSGKKDVGIRDCKKERAQIEISEQSLTSEDFMFKLDDKAGEEA
KKFPARDERYLQTIAKLLYYPNEIEDMNRFMKKGETINKKVQFNRKKKITRKQK
NNSSNEVLSSTMGYLFKNIKL (SEQ ID NO: 38)

Identificación del gen: E; Tipo de gen: C2C2; Organismo: 8. Listeria seeligeri serovar. 1/2b str. SLCC3954; Longitud del separador, modalidad (rango): 30; DR1: GUUUUAGUCCUCUUUCAUAUAGAGGUAGUCUCUUAC (SEQ ID NO: 39); DR2: ninguno; ARNcrtrac1: AUGAAAAGAGGACUAAAACUGAAAGAGGACUAAAACACCAGAUGUGGAUAACUAUAUUAGUGGCUAUUAAAAA UUCGUCGAUAUUAGAGAGGAAACUUU (SEQ ID NO: 40); ARNcrtrac2: ninguno; secuencia de la proteína:

MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTMRITKVEVDRKKVLISRDKNG
GKLVYENEMQDNTEQIMHHKKSSFYKSVVNKTICRPEQKQMKKLVHGLLQENS
QEKIKVSDVTKLNISNFLNHRFKKSLYYFPENSPDKSEEYRIEINLSQLLEDSLKK
QQGTFICWESFSKDMELYINWAENYISSKTKLIKKSIRNNRIQSTESRSGQLMDR
YMKDILNKNKPFDIQSVSEKYQLEKLTSALKATFKEAKKNDKEINYKLKSTLQN
HERQIIEELKENSELNQFNIEIRKHLETYFPIKKTNRKVGDIRNLEIGEIQKIVNHRL
KNKIVQRILQEGKLASYEIESTVNSNSLQKIKIEEAFALKFINACLFASNNLRNMV
YPVCKKDILMIGEFKNSFKEIKHKKFIRQWSQFFSQEITVDDIELASWGLRGAIAPI
RNEIIHLKKHSWKKFFNNPTFKVKKSKIINGKTKDVTSEFLYKETLFKDYFYSEL
DSVPELIINKMESSKILDYYSSDQLNQVFTIPNFELSLLTSAVPFAPSFKRVYLKGF
DYQNQDEAQPDYNLKLNIYNEKAFNSEAFQAQYSLFKMVYYQVFLPQFTTNND
LFKSSVDFILTLNKERKGYAKAFQDIRKMNKDEKPSEYMSYIQSQLMLYQKKQE
EKEKINHFEKFINQVFIKGFNSFIEKNRLTYICHPTKNTVPENDNIEIPFHTDMDDS
NIAFWLMCKLLDAKQLSELRNEMIKFSCSLQSTEEISTFTKAREVIGLALLNGEK
GCNDWKELFDDKEAWKKNMSLYVSEELLQSLPYTQEDGQTPVINRSIDLVKKY
GTETILEKLFSSSDDYKVSAKDIAKLHEYDVTEKIAQQESLHKQWIEKPGLARDS
AWTKKYQNVINDISNYQWAKTKVELTQVRHLHQLTIDLLSRLAGYMSIADRDF
QFSSNYILERENSEYRVTSWILLSENKNKNKYNDYELYNLKNASIKVSSKNDPQL
KVDLKQLRLTLEYLELFDNRLKEKRNNISHFNYLNGQLGNSILELFDDARDVLS
YDRKLKNAVSKSLKEILSSHGMEVTFKPLYQTNHHLKIDKLQPKKIHHLGEKST
VSSNQVSNEYCQLVRTLLTMK (SEQ ID NO: 41)

Identificación del gen: F; Tipo de gen: C2C2; Organismo: 12. Leptotrichia Wadei F0279; Longitud del separador, modalidad (rango): 31; DR1: GUUUUAGUCCCCUUCGUUUUUGGGGUAGUCUAAAUC (SEQ ID NO: 42); DR2: ninguno; ARNcrtrac1: GAUUUAGAGCACCCCAAAAGUAAUGAAAAUUUGCAAUUAAAUAAGGAAUAUUAAAAAAAUGUGAUUUUAAAAAA AUUGAAGAAAUUAAAUGAAAAAUUGUCCAAGUAAAAAAA (SEQ ID NO: 43); ARNcrtrac2: AUUUAGAUUACCCCUUUAAUUUAUUUUACCAUAUUUUUCUCAUAAUGCAAACUAAUAUUCCAAAAUUUUU (SEQ ID NO: 44); secuencia de la proteína:

MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKIDNN
KFIRKYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYI
EAYGKSEKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNK
TLNDCSIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRYYEEHL
REKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVE
KILNINVDLTVEDIADFVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVL
LDKHEKFKIERENKKDKIVKFFVENIKNNSIKEKIEKILAEFKIDELIKKLEKELKK
GNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQKV
RLKKMEKIEIEKILNESILSEKILKRVKQYTLEHIMYLGKLRHNDIDMTTVNTDDF
SRLHAKEELDLELITFFASTNMELNKIFSRENINNDENIDFFGGDREKNYVLDKKI
LNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRILHAISKERDLQGTQDDYN
KVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENNNDIKYLPSFSKV
LPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLEENESKNIFL
QELKKTLGNIDEIDENIIENYYKNAQISASKGNNKAIKKYQKKVIECYIGYLRKN
YEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDKTIVINDDFEYIISIFALL
NSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLRNECITENWNLNLEE
FIQKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLEKKLEKI
VIFDDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKDKDQEIKSKILCRII
FNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKKNLFLNIGNP
NFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYK
EKYIKKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYL
IDINWKLAIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPKRNGSDGFYT
TTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHFYIVRNPFAD
YSIAEQIDRVSNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNNDI
LERLMKPKKVSVLELESYNSDYIKNLIIELLTKIENTNDTL (SEQ ID NO: 45)

Identificación del gen: G; Tipo de gen: C2C2; Organismo: 14. Leptotrichia shahii DSM 19757 B031; Longitud del separador, modalidad (rango): 30 (30-32); DR1: GUUUUAGUCCCCUUCGAUAUUGGGGUGGUCUAUAUC (SEQ ID NO: 46); DR2: ninguno; ARNcrtrac1: AUUGAUGUGGUAUACUAAAAAUGGAAAAUUGUAUUUUUGAUUAGAAAGAUGUAAAAUUGAUUUAAUUUAAAAA UAUUUUAUUAGAUUAAAGUAGA (SEQ ID NO: 47); ARNcrtrac2: ninguno; secuencia de la proteína:

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEII
KSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY
FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEK
AKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNK
EPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANK
MLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFY
KQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYR
KQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKS
SGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGN
DRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVL
RAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQE
FFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNH
NWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLML
LGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 48)

Identificación del gen: H; Tipo de gen: Cpf1; Organismo: Francisella ularensis subespecie novicida U112; Longitud del separador, modalidad (rango): 31; DR1: GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAU (SEQ ID NO: 49); DR2: ninguno; ARNcrtrac1: AUCUACAAAAUUAUAAACUAAAUAAAGAUUCUUAUAAUAACUUUAUAUAUAAUCGAAAUGUAGAGAAUUUU (SEQ ID NO: 50); ARNcrtrac2: ninguno; secuencia de la proteína:

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDK
YHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY

IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEII
KSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK
APEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITK
FNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKS
FVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIY
FKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEK
AKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISI
KYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKD
EHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNK
EPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANK
MLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFY
KQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYR
KQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKS
SGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGN
DRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEY
NAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVL
RAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQE
FFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNH
NWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQ
MRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLML
LGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 51)

Genes para la síntesis

En el caso de los genes A a H, se realiza una optimización para la expresión en los seres humanos y se agrega una secuencia de ADN al final de cada gen. Es necesario considerar que esta secuencia de ADN contiene un codón de terminación (subrayado), por lo que no es necesario agregar ningún codón de terminación a la secuencia del gen con los codones optimizados:

AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGggat
ccTACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTAT
GCATACCCATATGATGTCCCCGACTATGCCTAA (SEQ ID NO: 52)

En el contexto de la optimización, es necesario evitar los sitios de restricción para las enzimas BamHI, EcoRI, HindIII, BsmBI, BSAI, BbsI, AgeI, XhoI, NdeI, NotI, KpnI, BsrGI, SpeI, XbaI y NheI.

Es necesario clonar estos genes en un vector de expresión simple para los mamíferos:

> A

MSLNRIYQGRVAAVETGTALAKGNVEWMPAAGGDEVLWQHHELFQAAINYYL
VALLALADKNNPVLGPLISQMDNPQSPYHVWGSFRRQGRQRTGLSQAVAPYITP
GNNAPTLDEVFRSILAGNPTDRATLDAALMQLLKACDGAGAIQQEGRSYWPKF
CDPDSTANFAGDPAMLRREQHRLLLPQVLHDPAITHDSPALGSFDTYSIATPDTR
TPQLTGPKARARLEQAITLWRVRLPESAADFDRLASSLKKIPDDDSRLNLQGYV
GSSAKGEVQARLFALLLFRHLERSSFTLGLLRSATPPPKNAETPPPAGVPLPAASA
ADPVRIARGKRSFVFRAFTSLPCWHGGDNIHPTWKSFDIAAFKYALTVINQIEEK
TKERQKECAELETDFDYMHGRLAKIPVKYTTGEAEPPPILANDLRIPLLRELLQNI
KVDTALTDGEAVSYGLQRRTIRGFRELRRIWRGHAPAGTVFSSELKEKLAGELR
QFQTDNSTTIGSVQLFNELIQNPKYWPIWQAPDVETARQWADAGFADDPLAAL
VQEAELQEDIDALKAPVKLTPADPEYSRRQYDFNAVSKFGAGSRSANRHEPGQT
ERGHNTFTTEIAARNAADGNRWRATHVRIHYSAPRLLRDGLRRPDTDGNEALE
AVPWLQPMMEALAPLPTLPQDLTGMPVFLMPDVTLSGERRILLNLPVTLEPAAL
VEQLGNAGRWQNQFFGSREDPFALRWPADGAVKTAKGKTHIPWHQDRDHFTV
LGVDLGTRDAGALALLNVTAQKPAKPVHRIIGEADGRTWYASLADARMIRLPG
EDARLFVRGKLVQEPYGERGRNASLLEWEDARNIILRLGQNPDELLGADPRRHS
YPEINDKLLVALRRAQARLARLQNRSWRLRDLAESDKALDEIHAERAGEKPSPL
PPLARDDAIKSTDEALLSQRDIIRRSFVQIANLILPLRGRRWEWRPHVEVPDCHIL
AQSDPGTDDTKRLVAGQRGISHERIEQIEELRRRCQSLNRALRHKPGERPVLGRP
AKGEEIADPCPALLEKINRLRDQRVDQTAHAILAAALGVRLRAPSKDRAERRHR
DIHGEYERFRAPADFVVIENLSRYLSSQDRARSENTRLMQWCHRQIVQKLRQLC
ETYGIPVLAVPAAYSSRFSSRDGSAGFRAVHLTPDHRHRMPWSRILARLKAHEE
DGKRLEKTVLDEARAVRGLFDRLDRFNAGHVPGKPWRTLLAPLPGGPVFVPLG
DATPMQADLNAAINIALRGIAAPDRHDIHHRLRAENKKRILSLRLGTQREKARW
PGGAPAVTLSTPNNGASPEDSDALPERVSNLFVDIAGVANFERVTIEGVSQKFAT
GRGLWASVKQRAWNRVARLNETVTDNNRNEEEDDIPM (SEQ ID NO: 53)

> B

MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIYE
HHEQDPKNPKKVSKAEIQAELWDFVLKMQKCNSFTHEVDKDVVFNILRELYEE
LVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIAGDPSW
EEEKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPFTDSNEPIVKEIKWMEKSRNQ
SVRRLDKDMFIQALERFLSWESWNLKVKEEYEKVEKEHKTLEERIKEDIQAFKS
LEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREIIQKWLKMDENEPSEKYLEV

FKDYQRKHPREAGDYSVYEFLSKKENHFIWRNHPEYPYLYATFCEIDKKKKDA
KQQATFTLADPINHPLWVRFEERSGSNLNKYRILTEQLHTEKLKKKLTVQLDRLI
YPTESGGWEEKGKVDIVLLPSRQFYNQIFLDIEEKGKHAFTYKDESIKFPLKGTL
GGARVQFDRDHLRRYPHKVESGNVGRIYFNMTVNIEPTESPVSKSLKIHRDDFPK
FVNFKPKELTEWIKDSKGKKLKSGIESLEIGLRVMSIDLGQRQAAAASIFEVVDQ
KPDIEGKLFFPIKGTELYAVHRASFNIKLPGETLVKSREVLRKAREDNLKLMNQK
LNFLRNVLHFQQFEDITEREKRVTKWISRQENSDVPLVYQDELIQIRELMYKPYK
DWVAFLKQLHKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRTRKFLL
RWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTIIMHALGYCYD
VRKKKWQAKNPACQIILFEDLSNYNPYEERSRFENSKLMKWSRREIPRQVALQG
EIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKLQDNRFFKNLQREGRLTLD
KIAVLKEGDLYPDKGGEKFISLSKDRKLVTTHADINAAQNLQKRFWTRTHGFYK
VYCKAYQVDGQTVYIPESKDQKQKIIEEFGEGYFILKDGVYEWGNAGKLKIKKG
SSKQSSSELVDSDILKDSFDLASELKGEKLMLYRDPSGNVFPSDKWMAAGVFFG
KLERILISKLTNQYSISTIEDDSSKQSM (SEQ ID NO: 54)

> C

MAIRSIKLKLKTHTGPEAQNLRKGIWRTHRLLNEGVAYYMKMLLLFR
QESTGERPKEELQEELICHIREQQQRNQADKNTQALPLDKALEALRQLYELLVPS
SVGQSGDAQIISRKFLSPLVDPNSEGGKGTSKAGAKPTWQKKKEANDPTWEQD
YEKWKKRREEDPTASVITTLEEYGIRPIFPLYTNTVTDIAWLPLQSNQFVRTWDR
DMLQQAIERLLSWESWNKRVQEEYAKLKEKMAQLNEQLEGGQEWISLLEQYEE
NRERELRENMTAANDKYRITKRQMKGWNELYELWSTFPASASHEQYKEALKR
VQQRLRGRFGDAHFFQYLMEEKNRLIWKGNPQRIHYFVARNELTKRLEEAKQS
ATMTLPNARKHPLWVRFDARGGNLQDYYLTAEADKPRSRRFVTFSQLIWPSES
GWMEKKDVEVELALSRQFYQQVKLLKNDKGKQKIEFKDKGSGSTFNGHLGGA
KLQLERGDLEKEEKNFEDGEIGSVYLNVVIDFEPLQEVKNGRVQAPYGQVLQLI
RRPNEFPKVTTYKSEQLVEWIKASPQHSAGVESLASGFRVMSIDLGLRAAAATSI
FSVEESSDKNAADFSYWIEGTPLVAVHQRSYMLRLPGEQVEKQVMEKRDERFQ
LHQRVKFQIRVLAQIMRMANKQYGDRWDELDSLKQAVEQKKSPLDQTDRTFW
EGIVCDLTKVLPRNEADWEQAVVQIHRKAEEYVGKAVQAWRKRFAADERKGI
AGLSMWNIEELEGLRKLLISWSRRTRNPQEVNRFERGHTSHQRLLTHIQNVKED
RLKQLSHAIVMTALGYVYDERKQEWCAEYPACQVILFENLSQYRSNLDRSTKE
NSTLMKWAHRSIPKYVHMQAEPYGIQIGDVRAEYSSRFYAKTGTPGIRCKKVRG

QDLQGRRFENLQKRLVNEQFLTEEQVKQLRPGDIVPDDSGELFMTLTDGSGSKE
VVFLQADINAAHNLQKRFWQRYNELFKVSCRVIVRDEEEYLVPKTKSVQAKLG
KGLFVKKSDTAWKDVYVWDSQAKLKGKTTFTEESESPEQLEDFQEIIEEAEEAK
GTYRTLFRDPSGVFFPESVWYPQKDFWGEVKRKLYGKLRERFLTKAR (SEQ ID
NO: 55)

> D

MKISKVDHTRMAVAKGNQHRRDEISGILYKDPTKTGSIDFDERFKKL
NCSAKILYHVFNGIAEGSNKYKNIVDKVNNNLDRVLFTGKSYDRKSIIDIDTVLR
NVEKINAFDRISTEEREQIIDDLLEIQLRKGLRKGKAGLREVLLIGAGVIVRTDKK
QEIADFLEILDEDFNKTNQAKNIKLSIENQGLVVSPVSRGEERIFDVSGAQKGKSS
KKAQEKEALSAFLLDYADLDKNVRFEYLRKIRRLINLYFYVKNDDVMSLTEIPA
EVNLEKDFDIWRDHEQRKEENGDFVGCPDILLADRDVKKSNSKQVKIAERQLRE
SIREKNIKRYRFSIKTIEKDDGTYFFANKQISVFWIHRIENAVERILGSINDKKLYR
LRLGYLGEKVWKDILNFLSIKYIAVGKAVFNFAMDDLQEKDRDIEPGKISENAV
NGLTSFDYEQIKADEMLQREVAVNVAFAANNLARVTVDIPQNGEKEDILLWNK
SDIKKYKKNSKKGILKSILQFFGGASTWNMKMFEIAYHDQPGDYEENYLYDIIQII
YSLRNKSFHFKTYDHGDKNWNRELIGKMIEHDAERVISVEREKFHSNNLPMFYK
DADLKKILDLLYSDYAGRASQVPAFNTVLVRKNFPEFLRKDMGYKVHFNNPEV
ENQWHSAVYYLYKEIYYNLFLRDKEVKNLFYTSLKNIRSEVSDKKQKLASDDFA
SRCEEIEDRSLPEICQIIMTEYNAQNFGNRKVKSQRVIEKNKDIFRHYKMLLIKTL
AGAFSLYLKQERFAFIGKATPIPYETTDVKNFLPEWKSGMYASFVEEIKNNLDLQ
EWYIVGRFLNGRMLNQLAGSLRSYIQYAEDIERRAAENRNKLFSKPDEKIEACK
KAVRVLDLCIKISTRISAEFTDYFDSEDDYADYLEKYLKYQDDAIKELSGSSYAA
LDHFCNKDDLKFDIYVNAGQKPILQRNIVMAKLFGPDNILSEVMEKVTESAIREY
YDYLKKVSGYRVRGKCSTEKEQEDLLKFQRLKNAVEFRDVTEYAEVINELLGQ
LISWSYLRERDLLYFQLGFHYMCLKNKSFKPAEYVDIRRNNGTIIHNAILYQIVS
MYINGLDFYSCDKEGKTLKPIETGKGVGSKIGQFIKYSQYLYNDPSYKLEIYNAG
LEVFENIDEHDNITDLRKYVDHFKYYAYGNKMSLLDLYSEFFDRFFTYDMKYQ
KNVVNVLENILLRHFVIFYPKFGSGKKDVGIRDCKKERAQIEISEQSLTSEDFMFK
LDDKAGEEAKKFPARDERYLQTIAKLLYYPNEIEDMNRFMKKGETINKKVQFNR
KKKITRKQKNNSSNEVLSSTMGYLFKNIKL (SEQ ID NO: 56)

> E

MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTMRITKVEVDRKKVL
ISRDKNGGKLVYENEMQDNTEQIMHHKKSSFYKSVVNKTICRPEQKQMKKLVH
GLLQENSQEKIKVSDVTKLNISNFLNHRFKKSLYYFPENSPDKSEEYRIEINLSQL
LEDSLKKQQGTFICWESFSKDMELYINWAENYISSKTKLIKKSIRNNRIQSTESRS
GQLMDRYMKDILNKNKPFDIQSVSEKYQLEKLTSALKATFKEAKKNDKEINYKL
KSTLQNHERQIIEELKENSELNQFNIEIRKHLETYFPIKKTNRKVGDIRNLEIGEIQ
KIVNHRLKNKIVQRILQEGKLASYEIESTVNSNSLQKIKIEEAFALKFINACLFASN
NLRNMVYPVCKKDILMIGEFKNSFKEIKHKKFIRQWSQFFSQEITVDDIELASWG
LRGAIAPIRNEIIHLKKHSWKKFFNNPTFKVKKSKIINGKTKDVTSEFLYKETLFK
DYFYSELDSVPELIINKMESSKILDYYSSDQLNQVFTIPNFELSLLTSAVPFAPSFK
RVYLKGFDYQNQDEAQPDYNLKLNIYNEKAFNSEAFQAQYSLFKMVYYQVFLP
QFTTNNDLFKSSVDFILTLNKERKGYAKAFQDIRKMNKDEKPSEYMSYIQSQLM
LYQKKQEEKEKINHFEKFINQVFIKGFNSFIEKNRLTYICHPTKNTVPENDNIEIPF
HTDMDDSNIAFWLMCKLLDAKQLSELRNEMIKFSCSLQSTEEISTFTKAREVIGL
ALLNGEKGCNDWKELFDDKEAWKKNMSLYVSEELLQSLPYTQEDGQTPVINRS
IDLVKKYGTETILEKLFSSSDDYKVSAKDIAKLHEYDVTEKIAQQESLHKQWIEK
PGLARDSAWTKKYQNVINDISNYQWAKTKVELTQVRHLHQLTIDLLSRLAGYM
SIADRDFQFSSNYILERENSEYRVTSWILLSENKNKNKYNDYELYNLKNASIKVS
SKNDPQLKVDLKQLRLTLEYLELFDNRLKEKRNNISHFNYLNGQLGNSILELFDD
ARDVLSYDRKLKNAVSKSLKEILSSHGMEVTFKPLYQTNHHLKIDKLQPKKIHH
LGEKSTVSSNQVSNEYCQLVRTLLTMK (SEQ ID NO: 57)

> F

MKVTKVDGISHKKYIEEGKLVKSTSEENRTSERLSELLSIRLDIYIKNP
DNASEEENRIRRENLKKFFSNKVLHLKDSVLYLKNRKEKNAVQDKNYSEEDISE
YDLKNKNSFSVLKKILLNEDVNSEELEIFRKDVEAKLNKINSLKYSFEENKANYQ
KINENNVEKVGGKSKRNIIYDYYRESAKRNDYINNVQEAFDKLYKKEDIEKLFFL
IENSKKHEKYKIREYYHKIIGRKNDKENFAKIIYEEIQNVNNIKELIEKIPDMSELK
KSQVFYKYYLDKEELNDKNIKYAFCHFVEIEMSQLLKNYVYKRLSNISNDKIKRI
FEYQNLKKLIENKLLNKLDTYVRNCGKYNYYLQVGEIATSDFIARNRQNEAFLR
NIIGVSSVAYFSLRNILETENENDITGRMRGKTVKNNKGEEKYVSGEVDKIYNEN
KQNEVKENLKMFYSYDFNMDNKNEIEDFFANIDEAISSIRHGIVHFNLELEGKDIF
AFKNIAPSEISKKMFQNEINEKKLKLKIFKQLNSANVFNYYEKDVIIKYLKNTKFN
FVNKNIPFVPSFTKLYNKIEDLRNTLKFFWSVPKDKEEKDAQIYLLKNIYYGEFL

NKFVKNSKVFFKITNEVIKINKQRNQKTGHYKYQKFENIEKTVPVEYLAIIQSRE
MINNQDKEEKNTYIDFIQQIFLKGFIDYLNKNNLKYIESNNNNDNNDIFSKIKIKK
DNKEKYDKILKNYEKHNRNKEIPHEINEFVREIKLGKILKYTENLNMFYLILKLL
NHKELTNLKGSLEKYQSANKEETFSDELELINLLNLDNNRVTEDFELEANEIGKF
LDFNENKIKDRKELKKFDTNKIYFDGENIIKHRAFYNIKKYGMLNLLEKIADKAK
YKISLKELKEYSNKKNEIEKNYTMQQNLHRKYARPKKDEKFNDEDYKEYEKAI
GNIQKYTHLKNKVEFNELNLLQGLLLKILHRLVGYTSIWERDLRFRLKGEFPENH
YIEEIFNFDNSKNVKYKSGQIVEKYINFYKELYKDNVEKRSIYSDKKVKKLKQEK
KDLYIRNYIAHFNYIPHAEISLLEVLENLRKLLSYDRKLKNAIMKSIVDILKEYGF
VATFKIGADKKIEIQTLESEKIVHLKNLKKKKLMTDRNSEELCELVKVMFEYKA
LE (SEQ ID NO: 58)

> G

MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNN
KEKIDNNKFIRKYINYKKNDNILKEFTRKFHAGNILFKLKGKEGIIRIENNDDFLE
TEEVVLYIEAYGKSEKLKALGITKKKIIDEAIRQGITKDDKKIEIKRQENEEEIEIDI
RDEYTNKTLNDCSIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFEN
RYYEEHLREKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSK
NKKMLVEKILNINVDLTVEDIADFVIKELEFWNITKRIEKVKKVNNEFLEKRRNR
TYIKSYVLLDKHEKFKIERENKKDKIVKFFVENIKNNSIKEKIEKILAEFKIDELIK
KLEKELKKGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIE
KILVNEQKVRLKKMEKIEIEKILNESILSEKILKRVKQYTLEHIMYLGKLRHNDID
MTTVNTDDFSRLHAKEELDLELITFFASTNMELNKIFSRENINNDENIDFFGGDRE
KNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRILHAISKERDL
QGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENNNDI
KYLPSFSKVLPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLE
ENESKNIFLQELKKTLGNIDEIDENIIENYYKNAQISASKGNNKAIKKYQKKVIEC
YIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDKTIVINDDF
EYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLRNECITE
NWNLNLEEFIQKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCD
VLEKKLEKIVIFDDETKFEIDKKSNILQDEQRKLSNINKKDLKKKVDQYIKDKDQ
EIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNELYIYKK
NLFLNIGNPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKL
NGYSKEYKEKYIKKLKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFN

YLNKIESYLIDINWKLAIQMARFERDMHYIVNGLRELGIIKLSGYNTGISRAYPKR
NGSDGFYTTTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHFY
IVRNPFADYSIAEQIDRVSNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKF
KLIGNNDILERLMKPKKVSVLELESYNSDYIKNLIIELLTKIENTNDTL (SEQ ID
NO: 59)

> H

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYK
KAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDT
IKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDIT
DIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAK
YESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNY
LNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQI
LSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQE
LIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNK
DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDK
ANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLAN
GWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIED
CRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISES
YIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLN
GEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFF
HCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNII
KQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVV
HEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDN
EFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPK
YESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLIN
FRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKL
TSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAY
HIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 60)

En el caso de los locus A a G, es necesario clonarlos e insertarlos en un plásmido con una cantidad baja de copias. También es necesario usar un vector que no confiera resistencia a la ampicilina.

> Locus A

TATCCGGTCGAATCGAGAATGACGACCGCTACGTCTTGGACTACG
AAGCCGTGGCCCTTGCCGATGCTCTCGGTGTGGATGTTGCCGACCTGTTCCGC
AAGATCGATTGCCCCAAGAACCTGCTGCGCAGGCGGGCAGGGTAGGGGAGC
GGTTTCCGGCGGAGATTTTCGGAGGCGCCGGTAACGTTATGTCGGGGAATTT
GCTATACATCGACGATAATTAGTTTTGTTGATTCAGGATCGAAATGCGCTCA
AACAAAGAACGTTCCGCGTTTCCCTCATGCGCTACTACGCCCACACCGCCAT
CTTTCGGCACGCAAACAAAGCAGATGGGTTGCCTGTCAATGGGTGATCATTG
CCTGAAGTTACCATCCATCAATAATATAAATCATCCTTACTCCGAATGTCCCT
CAATCGCATCTATCAAGGCCGCGTGGCGGCCGTCGAAACAGGAACGGCCTTA
GCGAAAGGTAATGTCGAATGGATGCCTGCCGCAGGAGGCGACGAAGTTCTC
TGGCAGCACCACGAACTTTTCCAAGCTGCCATCAACTACTATCTCGTCGCCCT
GCTCGCACTCGCCGACAAAAACAATCCCGTACTTGGCCCGCTGATCAGCCAG
ATGGATAATCCCCAAAGCCCTTACCATGTCTGGGGAAGTTTCCGCCGCCAAG
GACGTCAGCGCACAGGTCTCAGTCAAGCCGTTGCACCTTATATCACGCCGGG
CAATAACGCTCCCACCCTTGACGAAGTTTTCCGCTCCATTCTTGCGGGCAACC
CAACCGACCGCGCAACTTTGGACGCTGCACTCATGCAATTGCTCAAGGCTTG
TGACGGCGCGGGCGCTATCCAGCAGGAAGGTCGTTCCTACTGGCCCAAATTC
TGCGATCCTGACTCCACTGCCAACTTCGCGGGAGATCCGGCCATGCTCCGGC
GTGAACAACACCGCCTCCTCCTTCCGCAAGTTCTCCACGATCCGGCGATTACT
CACGACAGTCCTGCCCTTGGCTCGTTCGACACTTATTCGATTGCTACCCCCGA
CACCAGAACTCCTCAACTCACCGGCCCCAAGGCACGCGCCCGTCTTGAGCAG
GCGATCACCCTCTGGCGCGTCCGTCTTCCCGAATCGGCTGCTGACTTCGATCG
CCTTGCCAGTTCCCTCAAAAAAATTCCGGACGACGATTCTCGCCTTAACCTTC
AGGGCTACGTCGGCAGCAGTGCGAAAGGCGAAGTTCAGGCCCGTCTTTTCGC
CCTTCTGCTATTCCGTCACCTGGAGCGTTCCTCCTTTACGCTTGGCCTTCTCCG
TTCCGCCACCCCGCCGCCCAAGAACGCTGAAACACCTCCTCCCGCCGGCGTT
CCTTTACCTGCGGCGTCCGCAGCCGATCCGGTGCGGATAGCCCGTGGCAAAC
GCAGTTTTGTTTTTCGCGCATTCACCAGTCTCCCCTGCTGGCATGGCGGTGAT
AACATCCATCCCACCTGGAAGTCATTCGACATCGCAGCGTTCAAATATGCCC
TCACGGTCATCAACCAGATCGAGGAAAAGACGAAAGAACGCCAAAAAGAAT
GTGCGGAACTTGAAACTGATTTCGACTACATGCACGGACGGCTCGCCAAGAT
TCCGGTAAAATACACGACCGGCGAAGCCGAACCGCCCCCCATTCTCGCAAAC
GATCTCCGCATCCCCCTCCTCCGCGAACTTCTCCAGAATATCAAGGTCGACA

CCGCACTCACCGATGGCGAAGCCGTCTCCTATGGTCTCCAACGCCGCACCAT
TCGCGGTTTCCGCGAGCTGCGCCGCATCTGGCGCGGCCATGCCCCCGCTGGC
ACGGTCTTTTCCAGCGAGTTGAAAGAAAAACTAGCCGGCGAACTCCGCCAGT
TCCAGACCGACAACTCCACCACCATCGGCAGCGTCCAACTCTTCAACGAACT
CATCCAAAACCCGAAATACTGGCCCATCTGGCAGGCTCCTGACGTCGAAACC
GCCCGCCAATGGGCCGATGCCGGTTTTGCCGACGATCCGCTCGCCGCCCTTG
TGCAAGAAGCCGAACTCCAGGAAGACATCGACGCCCTCAAGGCTCCAGTCA
AACTCACTCCGGCCGATCCTGAGTATTCAAGAAGGCAATACGATTTCAATGC
CGTCAGCAAATTCGGGGCCGGCTCCCGCTCCGCCAATCGCCACGAACCCGGG
CAGACGGAGCGCGGCCACAACACCTTTACCACCGAAATCGCCGCCCGTAAC
GCGGCGGACGGGAACCGCTGGCGGGCAACCCACGTCCGCATCCATTACTCCG
CTCCCCGCCTTCTTCGTGACGGACTCCGCCGACCTGACACCGACGGCAACGA
AGCCCTGGAAGCCGTCCCTTGGCTCCAGCCCATGATGGAAGCCCTCGCCCCT
CTCCCGACGCTTCCGCAAGACCTCACAGGCATGCCGGTCTTCCTCATGCCCG
ACGTCACCCTTTCCGGTGAGCGTCGCATCCTCCTCAATCTTCCTGTCACCCTC
GAACCAGCCGCTCTTGTCGAACAACTGGGCAACGCCGGTCGCTGGCAAAACC
AGTTCTTCGGCTCCCGCGAAGATCCATTCGCTCTCCGATGGCCCGCCGACGG
TGCTGTAAAAACCGCCAAGGGGAAAACCCACATACCTTGGCACCAGGACCG
CGATCACTTCACCGTACTCGGCGTGGATCTCGGCACGCGCGATGCCGGGGCG
CTCGCTCTTCTCAACGTCACTGCGCAAAAACCGGCCAAGCCGGTCCACCGCA
TCATTGGTGAGGCCGACGGACGCACCTGGTATGCCAGCCTTGCCGACGCTCG
CATGATCCGCCTGCCCGGGGAGGATGCCCGGCTCTTTGTCCGGGGAAAACTC
GTTCAGGAACCCTATGGTGAACGCGGGCGAAACGCGTCTCTTCTCGAATGGG
AAGACGCCCGCAATATCATCCTTCGCCTTGGCCAAAATCCCGACGAACTCCT
CGGCGCCGATCCCCGGCGCCATTCGTATCCGGAAATAAACGATAAACTTCTC
GTCGCCCTTCGCCGCGCTCAGGCCCGTCTTGCCCGTCTCCAGAACCGGAGCT
GGCGGTTGCGCGACCTTGCAGAATCGGACAAGGCCCTTGATGAAATCCATGC
CGAGCGTGCCGGGGAGAAGCCTTCTCCGCTTCCGCCCTTGGCTCGCGACGAT
GCCATCAAAAGCACCGACGAAGCCCTCCTTTCCCAGCGTGACATCATCCGGC
GATCCTTCGTTCAGATCGCCAACTTGATCCTTCCCCTTCGCGGACGCCGATGG
GAATGGCGGCCCCATGTCGAGGTCCCGGATTGCCACATCCTTGCGCAGAGCG
ATCCCGGTACGGATGACACCAAGCGTCTTGTCGCCGGACAACGCGGCATCTC
TCACGAGCGTATCGAGCAAATCGAAGAACTCCGTCGTCGCTGCCAATCCCTC

AACCGTGCCCTGCGTCACAAACCCGGAGAGCGTCCCGTGCTCGGACGCCCCG
CCAAGGGCGAGGAAATCGCCGATCCCTGTCCCGCGCTCCTCGAAAAGATCAA
CCGTCTCCGGGACCAGCGCGTTGACCAAACCGCGCATGCCATCCTCGCCGCC
GCTCTCGGTGTTCGACTCCGCGCCCCCTCAAAAGACCGCGCCGAACGCCGCC
ATCGCGACATCCATGGCGAATACGAACGCTTTCGTGCGCCCGCTGATTTTGT
CGTCATCGAAAACCTCTCCCGTTATCTCAGCTCGCAGGATCGTGCTCGTAGTG
AAAACACCCGTCTCATGCAGTGGTGCCATCGCCAGATCGTGCAAAAACTCCG
TCAGCTCTGCGAGACCTACGGCATCCCCGTCCTCGCCGTCCCGGCGGCCTAC
TCATCGCGTTTTTCTTCCCGGGACGGCTCGGCCGGATTCCGGGCCGTCCATCT
GACACCGGACCACCGTCACCGGATGCCATGGAGCCGCATCCTCGCCCGCCTC
AAGGCCCACGAGGAAGACGGAAAAAGACTCGAAAAGACGGTGCTCGACGA
GGCTCGCGCCGTCCGGGGACTCTTTGACCGGCTCGACCGGTTCAACGCCGGG
CATGTCCCGGGAAAACCTTGGCGCACGCTCCTCGCGCCGCTCCCCGGCGGCC
CTGTGTTTGTCCCCCTCGGGGACGCCACACCCATGCAGGCCGATCTGAACGC
CGCCATCAACATCGCCCTCCGGGGCATCGCGGCTCCCGACCGCCACGACATC
CATCACCGGCTCCGTGCCGAAAACAAAAAACGCATCCTGAGCTTGCGTCTCG
GCACTCAGCGCGAGAAAGCCCGCTGGCCTGGAGGAGCTCCGGCGGTGACAC
TCTCCACTCCGAACAACGGCGCCTCTCCCGAAGATTCCGATGCGTTGCCCGA
ACGGGTATCCAACCTGTTTGTGGACATCGCCGGTGTCGCCAACTTCGAGCGA
GTCACGATCGAAGGAGTCTCGCAAAAATTCGCCACCGGGCGTGGCCTTTGGG
CCTCCGTCAAGCAACGTGCATGGAACCGCGTTGCCAGACTCAACGAGACAGT
AACAGATAACAACAGGAACGAAGAGGAGGACGACATTCCGATGTAACCATT
GCTTCATTACATCTGAGTCTCCCCTCAATCCCTCTGCCCCATGCGTGATATAA
CCTCCACCTCATGTCCCGGATCGGCGCCGGCAACCTGTAGTTCCCTTCCATCC
TCCAACACTCCCGCAGATCGCGATCCGCTGCCGCCGATGCCGGTGCGCCGCC
TTCACAACTATCTCTACTGTCCGCGGCTTTTTTATCTCCAGTGGGTCGAGAAT
CTCTTTGAGGAAAATGCCGACACCATTGCCGGCAGCGCCGTGCATCGTCACG
CCGACAAACCTACGCGTTACGATGATGAAAAAGCCGAGGCACTTCGCACTG
GTCTCCCTGAAGGCGCGCACATACGCAGCCTTCGCCTGGAAAACGCCCAACT
CGGTCTCGTTGGCGTGGTGGATATCGTGGAGGGAGGCCCCGACGGACTCGAA
CTCGTCGACTACAAAAAAGGTTCCGCCTTCCGCCTCGACGACGGCACGCTCG
CTCCCAAGGAAAACGACACCGTGCAACTTGCCGCCTACGCTCTTCTCCTGGC
TGCCGATGGTGCGCGCGTTGCGCCCATGGCGACGGTCTATTACGCTGCCGAT

CGCCGGCGTGTCACCTTCCCGCTCGATGACGCCCTCTACGCCCGCACCCGTTC
CGCCCTCGAAGAGGCCCGCGCCGTTGCAACCTCGGGGCGCATACCTCCGCCG
CTCGTCTCTGACGTCCGCTGCCTCCATTGTTCCTCCTATGCGCTTTGCCTTCCC
CGCGAGTCCGCCTGGTGGTGCCGCCATCGCAGCACGCCGCGGGGAGCCGGC
CACACCCCCATGTTGCCGGGCTTTGAGGATGACGCCGCCGCCATTCACCAAA
TCTCCGAACCTGACACCGAGCCACCACCCGATCTTGCCAGCCAGCCTCCCCG
TCCCCCGCGGCTCGATGGAGAATTGTTGGTTGTCCAGACTCCGGGAGCGATG
ATCGGACAAAGCGGCGGTGAGTTTACCGTGTCCGTCAAGGGTGAGGTTTTGC
GCAAGCTTCCGGTTCATCAACTCCGGGCCATTTACGTTTACGGAGCCGTGCA
ACTCACGGCGCATGCTGTGCAGACCGCCCTTGAGGAGGATATCGACGTCTCC
TATTTTGCGCCCAGCGGCCGCTTTCTTGGCCTCCTCCGCGGCCTGCCCGCATC
CGGCGTGGATGCGCGTCTCGGGCAATACACCCTGTTTCGCGAACCCTTTGGC
CGTCTCCGTCTCGCCTGCGAGGCGATTCGGGCCAAGATCCATAACCAGCGCG
TCCTCCTCATGCGTAACGGCGAGCCCCGGGGAGGGCGTCTTGCGCGAACTCGC
CCGTCTGCGCGACGCCACCAGTGAGGCGACTTCGCTCGACGAACTCCTCGGC
ATCGAGGGCATCGCCGCGCATTTCTATTTCCAGTATTTTCCCACCATGCTGAA
AGAACGGGCGGCCTGGGCCTTTGATTTTTCCGGACGCAATCGCCGCCCGCCG
CGCGACCCGGTCAACGCCCTGCTTTCGTTCGGTTACAGCGTGTTGTCCAAGG
AACTTGCCGGCGTCTGCCACGCTGTTGGCCTAGACCCGTTTTTCGGCTTCATG
CACCAGCCGCGTTACGGGCGCCCCGCACTCGCTCTCGATCTGATGGAGGAGT
TTCGCCCTCTCATCGCCGACAGTGTTGCCCTGAATCTCATCAACCGTGGCGAA
CTCGACGAAGGGGACTTTATCCGGTCGGCCAATGGCACCGCGCTCAATGATC
GGGGCCGCCGGCGTTTTTGGGAGGCATGGTTCCGGCGTCTCGACAGCGAAGT
CAGCCATCCTGAATTTGGTTACAAGATGAGCTATCGACGGATGCTTGAAGTG
CAGGCGCGCCAGCTATGGCGCTATGTGCGCGGTGACGCCTTCCGCTACCACG
GATTCACCACCCGTTGATTCCGATGTCAGATCCCCGCCGCCGTTATCTTGTGT
GTTACGACATCGCCAATCCGAAGCGATTGCGCCAAGTGGCCAAGCTGCTGGA
GAGCTATGGCACGCGTCTGCAATACTCGGTTTTCGAATGTCCTTTGGACGATC
TTCGTCTTGAACAGGCGAAGGCTGATTTGCGCGACACGATTAATGCCGACCA
AGACCAGGTGTTATTTGTTTCGCTTGGCCCCGAAGCCAACGATGCCACGTTG
ATCATCGCCACGCTTGGGCTCCCTTATACCGTGCGCTCGCGAGTGACGATTAT
CTGACCCATAACCCACGTGTTGAAGAGGCTGAAAACAGACGGACCTCTATGA
AGAACAATTGACGTTTTGGCCGAACTCAGCAGACCTTTATGCGGCTAAGGCC

AATGATCATCCATCCTACCGCCATTGGGCTGGAGACGTTTTTTGAAACGGCG
AGTGCTGCGGATAGCGAGTTTCTCTTGGGGAGGCGCTCGCGGCCACTTTTAC
AGAGGAGATGTTCGGGCGAACTGGCCGACCTAACAAGGCGTACCCGGCTCA
AAATCGAGGCACGCTCGCACGGGATGATGTAATTCGTTGTTTTTCAGCATAC
CGTGCGAGCACGGGCCGCAGCGAATGCCGTTTCACGAATCGTCAGGCGGCG
GGGAGAAGTCATTTAATAAGGCCACTGTTAAAAGCCGCAGCGAATGCCGTTT
CACGAATCGTCAGGCGGGCAGTGGATGTTTTTCCATGAGGCGAAGAATTTCA
TCGCCGCAGTGAATGCCGTTTCACCATTGATGAAGAATGCGAGGTGAAAACA
GAGAAATTGGGTCAACTCTATCACTCTTATTCAGCCATCGTTTCAAGAAAGG
ATACCTCGTATTGGATACAACACAGCTCGTTCGTTCTCTCTACCTCCCTCGAC
AATCTCAAGGA (SEQ ID NO: 61)

> Locus B

TAATAAAATTGAAATATCACTATGGATTATTGTAATATTACCATAA
AGATAGGTGACGTTTTTTTGAAAATTGTAAACCTAATTTGAAGAAAACCAAT
TAAAAATCGCTTCGGCTTTTTTTTAAGTGCCAGGTAGCATTGATGCTAACCCA
TGTGTAATAAAGGTTTGTTTTCCTTCGGGGCACGAACACATTATAAGGGAAA
CCTAAAGATTCCCTTTCTTGTTTAATATTATAACCAGTGAAAATAAGAATAAT
GCACCTAAAACTAATATACAGAAAATAAGAATTAAAAGTACTAATATATAC
ATCATATGTTATCCTCCAATGCTTTATTTTTTAATAATTGATGTTAGTATTAGT
TTTATTTTAATTTCTAAACATAAGAATTTGAAAAGGATGTGTTTATTATGGCG
ACACGCAGTTTTATTTTAAAAATTGAACCAAATGAAGAAGTTAAAAAGGGAT
TATGGAAGACGCATGAGGTATTGAATCATGGAATTGCCTACTACATGAATAT
TCTGAAACTAATTAGACAGGAAGCTATTTATGAACATCATGAACAAGATCCT
AAAAATCCGAAAAAAGTTTCAAAAGCAGAAATACAAGCCGAGTTATGGGAT
TTTGTTTTAAAAATGCAAAAATGTAATAGTTTTACACATGAAGTTGACAAAG
ATGTTGTTTTTAACATCCTGCGTGAACTATATGAAGAGTTGGTCCCTAGTTCA
GTCGAGAAAAAGGGTGAAGCCAATCAATTATCGAATAAGTTTCTGTACCCGC
TAGTTGATCCGAACAGTCAAAGTGGGAAAGGGACGGCATCATCCGGACGTA
AACCTCGGTGGTATAATTTAAAAATAGCAGGCGACCCATCGTGGGAGGAAG
AAAAGAAAAAATGGGAAGAGGATAAAAAGAAAGATCCCCTTGCTAAAATCT
TAGGTAAGTTAGCAGAATATGGGCTTATTCCGCTATTTATTCCATTTACTGAC
AGCAACGAACCAATTGTAAAAGAAATTAAATGGATGGAAAAAAGTCGTAAT
CAAAGTGTCCGGCGACTTGATAAGGATATGTTTATCCAAGCATTAGAGCGTT

TTCTTTCATGGGAAAGCTGGAACCTTAAAGTAAAGGAAGAGTATGAAAAAG
TTGAAAAGGAACACAAAACACTAGAGGAAAGGATAAAAGAGGACATTCAA
GCATTTAAATCCCTTGAACAATATGAAAAAGAACGGCAGGAGCAACTTCTTA
GAGATACATTGAATACAAATGAATACCGATTAAGCAAAAGAGGATTACGTG
GTTGGCGTGAAATTATCCAAAAATGGCTAAAGATGGATGAAAATGAACCAT
CAGAAAAATATTTAGAAGTATTTAAAGATTATCAACGGAAACATCCACGAG
AAGCCGGGGACTATTCTGTCTATGAATTTTTAAGCAAGAAAGAAAATCATTT
TATTTGGCGAAATCATCCTGAATATCCTTATTTGTATGCTACATTTTGTGAAA
TTGACAAAAAAAAAGAAAGACGCTAAGCAACAGGCAACTTTTACTTTGGCTG
ACCCGATTAACCATCCGTTATGGGTACGATTTGAAGAAAGAAGCGGTTCGAA
CTTAAACAAATATCGAATTTTAACAGAGCAATTACACACTGAAAAGTTAAAA
AAGAAATTAACAGTTCAACTTGATCGTTTAATTTATCCAACTGAATCCGGCG
GTTGGGAGGAAAAAGGTAAAGTAGATATCGTTTTGTTGCCGTCAAGACAATT
TTATAATCAAATCTTCCTTGATATAGAAGAAAAGGGGAAACATGCTTTTACT
TATAAGGATGAAAGTATTAAATTCCCCCTTAAAGGTACACTTGGTGGTGCAA
GAGTGCAGTTTGACCGTGACCATTTGCGGAGATATCCGCATAAAGTAGAATC
AGGAAATGTTGGACGGATTTATTTTAACATGACAGTAAATATTGAACCAACT
GAGAGCCCTGTTAGTAAGTCTTTGAAAATACATAGGGACGATTTCCCCAAGT
TCGTTAATTTTAAACCGAAAGAGCTCACCGAATGGATAAAAGATAGTAAAG
GGAAAAAATTAAAAAGTGGTATAGAATCCCTTGAAATTGGTCTACGGGTGAT
GAGTATCGACTTAGGTCAACGTCAAGCGGCTGCTGCATCGATTTTTGAAGTA
GTTGATCAGAAACCGGATATTGAAGGGAAGTTATTTTTTCCAATCAAAGGAA
CTGAGCTTTATGCTGTTCACCGGGCAAGTTTTAACATTAAATTACCGGGTGA
AACATTAGTAAAATCACGGGAAGTATTGCGGAAAGCTCGGGAGGACAACTT
AAAATTAATGAATCAAAAGTTAAACTTTCTAAGAAATGTTCTACATTTCCAA
CAGTTTGAAGATATCACAGAAAGAGAGAAGCGTGTAACTAAATGGATTTCTA
GACAAGAAAATAGTGATGTTCCTCTTGTATATCAAGATGAGCTAATTCAAAT
TCGTGAATTAATGTATAAACCCTATAAAGATTGGGTTGCCTTTTTAAAACAA
CTCCATAAACGGCTAGAAGTCGAGATTGGCAAAGAGGTTAAGCATTGGCGA
AAATCATTAAGTGACGGGAGAAAAGGTCTTTACGGAATCTCCCTAAAAAATA
TTGATGAAATTGATCGAACAAGGAAATTCCTTTTAAGATGGAGCTTACGTCC
AACAGAACCTGGGGAAGTAAGACGCTTGGAACCAGGACAGCGTTTTGCGAT
TGATCAATTAAACCACCTAAATGCATTAAAAGAAGATCGATTAAAAAAGAT

GGCAAATACGATTATCATGCATGCCTTAGGTTACTGTTATGATGTAAGAAAG
AAAAAGTGGCAGGCAAAAAATCCAGCATGTCAAATTATTTTATTTGAAGATT
TATCTAACTACAATCCTTACGAGGAAAGGTCCCGTTTTGAAAACTCAAAACT
GATGAAGTGGTCACGGAGAGAAATTCCACGACAAGTCGCCTTACAAGGTGA
AATTTACGGATTACAAGTTGGGGAAGTAGGTGCCCAATTCAGTTCAAGATTC
CATGCGAAAACCGGGTCGCCGGGAATTCGTTGCAGTGTTGTAACGAAAGAA
AAATTGCAGGATAATCGCTTTTTTAAAAATTTACAAAGAGAAGGACGACTTA
CTCTTGATAAAATCGCAGTTTTAAAAGAAGGAGACTTATATCCAGATAAAGG
TGGAGAAAAGTTTATTTCTTTATCAAAGGATCGAAAGTTGGTAACTACGCAT
GCTGATATTAACGCGGCCCAAAATTTACAGAAGCGTTTTTGGACAAGAACAC
ATGGATTTTATAAAGTTTACTGCAAAGCCTATCAGGTTGATGGACAAACTGT
TTATATTCCGGAGAGCAAGGACCAAAAACAAAAAATAATTGAAGAATTTGG
GGAAGGCTATTTTATTTTAAAAGATGGTGTATATGAATGGGGTAATGCGGGG
AAACTAAAAATTAAAAAAGGTTCCTCTAAACAATCATCGAGTGAATTAGTAG
ATTCGGACATACTGAAAGATTCATTTGATTTAGCAAGTGAACTTAAGGGAGA
GAAACTCATGTTATATCGAGATCCGAGTGGAAACGTATTTCCTTCCGACAAG
TGGATGGCAGCAGGAGTATTTTTTGGCAAATTAGAAAGAATATTGATTTCTA
AGTTAACAAATCAATACTCAATATCAACAATAGAAGATGATTCTTCAAAACA
ATCAATGTAAAAGTTTGCCCGTATAAGAACTTAATTAATTAGGATGGTAGGA
TGTTACTAAATATGTCTGTAGGCATCATTCCTACTATCCGTTTTGTCCGAATA
TCAGAGCATTAGGTGAGGAATGGTAAGAAAGGAAAATTTATATGAACCAAC
CGATTCCTATTCGAATGTTAAATGAAATACAATATTGTGAGCGACTTTTTTAC
TTTATGCATGTCCAAAAGCTATTTGATGAGAATGCAGATACAGTTGAAGGAA
GTGCACAGCATGAGCGGGCAGAAAGAAGCAAAAGACCAAGTAAAATGGGA
CCAAAGGAATTATGGGGTGAGGCGCCAAGAAGTCTTAAGCTTGGTGATGAG
CTGTTAAATATTACCGGTGTTCTTGATGCCATAAGTCATGAAGAGAACAGTT
GGATCCCGGTTGAATCAAAACACAGTTCCGCACCGGATGGATTGAACCCTTT
TAAAGTAGATGGCTTTCTACTTGACGGGTCTGCATGGCCAAACGATCAAATT
CAACTTTGTGCACAAGGCTTGCTCTTGAATGCCAATGGATACCCGTGTGATT
ATGGGTATTTATTTTATCGTGGTAATAAGAAAAAGGTGAAAATTTATTTTACT
GAAGATTTAATCGCTGCCACAAAGTACTATATTAAAAAAGCACACGAGATAC
TAGTATTATCTGGTGATGAATCAGCTATTCCTAAGCCTTTAATTGATTCTAAT
AAGTGTTTTCGCTGTTCTTTAAACTATATCTGTCTTCCGGATGAAACGAACTA

TCTATTAGGGGCAAGTTCAACAATTCGTAAAATTGTGCCTTCAAGGACAGAT
GGTGGCGTTTTATATGTATCAGAGTCTGGTACAAAATTAGGAAAATCGGGTG
AGGAGTTAATCATTCAGTATAAAGATGGCCAAAAGCAGGGTGTTCCTATAAA
AGATATTATTCAAGTTTCGTTAATTGGAAATGTTCAATGCTCAACGCAATTAC
TTCATTTTTTAATGCAATCAAATATTCCTGTAAGTTATTTATCATCCCACGGT
CGTTTGATTGGTGTCAGTTCATCTTTAGTTACAAAAAATGTTTTAACAAGGCA
GCAACAGTTCATTAAATTTACAAATCCTGAGTTTGGACTAAATCTAGCAAAA
CAAATTGTTTATGCCAAGATTCGAAATCAACGAACTTTACTTAGAAGAAATG
GGGGGAGTGAGGTAAAGGAGATTTTAACAGATTTAAAATCTTTAAGTGACA
GTGCACTGAACGCAATATCAATAGAACAATTACGGGGTATTGAAGGGATTTC
TGCAAAACATTATTTCGCAGGATTTCCGTTTATGTTGAAAAATGAATTACGTG
AATTGAATTTAATGAAAGGGCGTAATAGGAGACCGCCAAAAGATCCTGTAA
ATGTACTTCTTTCTCTTGGTTATACTTTATTGACACGTGATATTCATGCTGCGT
GTGGTTCAGTCGGATTGGATCCGATGTTTGGTTGTTACCATCGTCCAGAAGC
AGGTCGACCGGCTCTAGTATTAGATGTTATGGAAACATTTCGACCACTTATT
GTAGACAGTATTGTCATCCGAGCTTTGAATACGGGTGAAATCTCATTAAAAG
ATTTTTATATAGGAAAAGATAGTTGTCAATTATTAAAACATGGCCGCGATTC
CTTTTTTGCCATTTATGAAAGAAGAATGCATGAAACTATTACCGATCCAATTT
TCGGCTATAAGATTAGCTATCGCCGTATGCTCGATTTGCACATTCGAATGCTT
GCAAGGTTTATTGAAGGGGAACTGCCGGAATATAAACCATTAATGACCCGGT
GAGTTTGTTTATTAGGTTAAAAGAAGGTGAAGACATGCAGCAATACGTCCTT
GTTTCTTATGATATTTCGGACCAAAAAAGATGGAGAAAAGTATTTAAACTGA
TGAAAGGATACGGAGAACATGTTCAATATTCCGTATTCATATGCCAGTTAAC
TGAATTACAGAAGGCAAAATTACAAGCCTCTTTAGAAGACATTATCCATCAT
AAGAATGACCAAGTAATGTTTGTTCACATCGGGCCAGTGAAAGATGGTCAAC
TATCTAAAAAAATCTCAACAATTGGGAAAGAATTTGTTCCATTGGATTTAAA
GCGGCTTATATTTTGAAAAGATATAGCAAAGAAATCTTATGAAAAAAATACA
AAAATATATTGTTAAAAAATAGGGAATATTATATAATGGACTTACGAGGTTC
TGTCTTTTGGTCAGGACAACCGTCTAGCTATAAGTGCTGCAGGGGTGTGAGA
AACTCCTATTGCTGGACGATGTCTCTTTTATTTCTTTTTTCTTGGATCTGAGTA
CGAGCACCCACATTGGACATTTCGCATGGTGGGTGCTCGTACTATAGGTAAA
ACAAACCTTTTTAAGAAGAATACAAAAATAACCACAATATTTTTTAAAAGGA
ATTTTGATGGATTTACATAACCTCTCGCAACATGCTTCTAAAACCCAAGCCCA

CCATAGCCCAAAACCCCCTGCGGTCCAAGAAAAAAGAAATGATACGAGGCA
TTAGCACCGGGGAGAAGTCATTTAATAAGGCCACTGTTAAAAGTCCAAGAA
AAAAGAAATGATACGAGGCATTAGCACAACAATATAAACGACTACTTTACC
GTGTTCAAGAAAAAAGAAATGATATGAGGCATTAGCACGATGGGATGGGAG
AGAGAGGACAGTTCTACTCTTGCTGTATCCAGCTTCTTTTACTTTATCCGGTA
TCATTTCTTCACTTCTTTCTGCACATAAAAAAGCACCTAACTATTTGGATAAG
TTAAGTGCTTTTATTTCCGTTTGAAGTTGTCTATTGCTTTTTTCTTCATATCTTC
AAATTTTTTCTGTTTCTCAGAGTCAACTTTACCAACTGTAATCCCTTTTCTTTT
TGGCATTGGGGTATCTTTCCACCTTAGTGTGTTCATAAGGCTTATATTTATCA
CTCATTGTATTCCTCCAACACAATTATAATTTTTCCGTCATCCTCAATCCAAC
CGTCAACTGTGACAAAAGACGAATCTCTCTTAT (SEQ ID NO: 62)

> Locus C

GTTTCATTTGGAAAGGGAGAGCATTGGCTTTTCTCTTTGTAAATAA
AGTGCAAGCTTTGTAATAAGCTTCTAGTGGAGAAGTGATTGTTTGAATCACC
CAATGCACACGCACTAAAGTTAGACGAACCTATAATTCGTATTAGTAAGTAT
AGTACATGAAGAAAAATGCAACAAGCATTTACTCTCTTTTAAATAAAGAATT
GATAGCTGTTAATATTGATAGTATATTATACCTTATAGATGTTCGATTTTTTTT
GAAATTCAAAAATCATACTTAGTAAAGAAAGGAAATAACGTCATGGACAAG
CGAAAGCGTAGAAGTTACGAGTTTAGGTGGGAAGCGGGAGGCACCAGTCAT
GGCAATCCGTAGCATAAAACTAAAACTAAAAACCCACACAGGCCCGGAAGC
GCAAAACCTCCGAAAAGGAATATGGCGGACGCATCGGTTGTTAAATGAAGG
CGTCGCCTATTACATGAAAATGCTCCTGCTCTTTCGTCAGGAAAGCACTGGT
GAACGGCCAAAAGAAGAACTACAGGAAGAACTGATTTGTCACATACGCGAA
CAGCAACAACGAAATCAGGCAGATAAAAATACGCAAGCGCTTCCGCTAGAT
AAGGCACTGGAAGCTTTGCGCCAACTATATGAACTGCTTGTCCCCTCCTCGG
TCGGACAAAGTGGCGACGCCCAGATCATCAGCCGAAAGTTTCTCAGCCCGCT
CGTCGATCCGAACAGCGAAGGCGGCAAAGGTACTTCGAAGGCAGGGGCAAA
ACCCACTTGGCAGAAGAAAAAAGAAGCGAACGACCCAACCTGGGAACAGGA
TTACGAAAAATGGAAAAAAAGACGCGAGGAAGACCCAACCGCTTCTGTGAT
TACTACTTTGGAGGAATACGGCATTAGACCGATCTTTCCCCTGTACACGAAC
ACCGTAACAGATATCGCGTGGTTGCCACTTCAATCCAATCAGTTTGTGCGAA
CCTGGGACAGAGACATGCTTCAACAAGCGATTGAAAGACTGCTCAGTTGGG
AGAGCTGGAACAAACGTGTCCAGGAAGAGTATGCCAAGCTGAAAGAAAAAA

TGGCTCAACTGAACGAGCAACTCGAAGGCGGTCAGGAATGGATCAGCTTGCT
AGAGCAGTACGAAGAAAACCGAGAGCGAGAGCTTAGGGAAAACATGACCG
CTGCCAATGACAAGTATCGGATTACCAAGCGGCAAATGAAAGGCTGGAACG
AGCTGTACGAGCTATGGTCAACCTTTCCCGCCAGTGCCAGTCACGAGCAATA
CAAAGAGGCGCTCAAGCGTGTGCAGCAGCGACTGAGAGGGCGGTTTGGGGA
TGCTCATTTCTTCCAGTATCTGATGGAAGAGAAGAACCGCCTGATCTGGAAG
GGGAATCCGCAGCGTATCCATTATTTTGTCGCGCGCAACGAACTGACGAAAC
GGCTGGAGGAAGCCAAGCAAAGCGCCACGATGACGTTGCCCAATGCCAGGA
AGCATCCATTGTGGGTGCGCTTCGATGCACGGGGAGGAAATTTGCAAGACTA
CTACTTGACGGCTGAAGCGGACAAACCGAGAAGCAGACGTTTTGTAACGTTT
AGTCAGTTGATATGGCCAAGCGAATCGGGATGGATGGAAAAGAAAGACGTC
GAGGTCGAGCTAGCTTTGTCCAGGCAGTTTTACCAGCAGGTGAAGTTGCTGA
AAAATGACAAAGGCAAGCAGAAAATCGAGTTCAAGGATAAAGGTTCGGGCT
CGACGTTTAACGGACACTTGGGGGGAGCAAAGCTACAACTGGAGCGGGGCG
ATTTGGAGAAGGAAGAAAAAAACTTCGAGGACGGGGAAATCGGCAGCGTTT
ACCTTAACGTTGTCATTGATTTCGAACCTTTGCAAGAAGTGAAAAATGGCCG
CGTGCAGGCGCCGTATGGACAAGTACTGCAACTCATTCGTCGCCCCAACGAG
TTTCCCAAGGTCACTACCTATAAGTCGGAGCAACTTGTTGAATGGATAAAAG
CTTCGCCACAACACTCGGCTGGGGTGGAGTCGCTGGCATCCGGTTTTCGTGT
AATGAGCATAGACCTTGGGCTGCGCGCGGCTGCAGCGACTTCTATTTTTTCTG
TAGAAGAGAGTAGCGATAAAAATGCGGCTGATTTTTCCTACTGGATTGAAGG
AACGCCGCTGGTCGCTGTCCATCAGCGGAGCTATATGCTCAGGTTGCCTGGT
GAACAGGTAGAAAAACAGGTGATGGAAAAACGGGACGAGCGGTTCCAGCTA
CACCAACGTGTGAAGTTTCAAATCAGAGTGCTCGCCCAAATCATGCGTATGG
CAAATAAGCAGTATGGAGATCGCTGGGATGAACTCGACAGCCTGAAACAAG
CGGTTGAGCAGAAAAAGTCGCCGCTCGATCAAACAGACCGGACATTTTGGG
AGGGGATTGTCTGCGACTTAACAAAGGTTTTGCCTCGAAACGAAGCGGACTG
GGAACAAGCGGTAGTGCAAATACACCGAAAAGCAGAGGAATACGTCGGAAA
AGCCGTTCAGGCATGGCGCAAGCGCTTTGCTGCTGACGAGCGAAAAGGCATC
GCAGGTCTGAGCATGTGGAACATAGAAGAATTGGAGGGCTTGCGCAAGCTG
TTGATTTCCTGGAGCCGCAGGACGAGGAATCCGCAGGAGGTTAATCGCTTTG
AGCGAGGCCATACCAGCCACCAGCGTCTGTTGACCCATATCCAAAACGTCAA
AGAGGATCGCCTGAAGCAGTTAAGTCACGCCATTGTCATGACTGCCTTGGGG

TATGTTTACGACGAGCGGAAACAAGAGTGGTGCGCCGAATACCCGGCTTGCC
AGGTCATTCTGTTTGAAAATCTGAGCCAGTACCGTTCTAACCTGGATCGCTCG
ACCAAAGAAAACTCCACCTTGATGAAGTGGGCGCATCGCAGCATTCCGAAAT
ACGTCCACATGCAGGCGGAGCCATACGGGATTCAGATTGGCGATGTCCGGGC
GGAATATTCCTCTCGTTTTTACGCCAAGACAGGAACGCCAGGCATTCGTTGT
AAAAAGGTGAGAGGCCAAGACCTGCAGGGCAGACGGTTTGAGAACTTGCAG
AAGAGGTTAGTCAACGAGCAATTTTTGACGGAAGAACAAGTGAAACAGCTA
AGGCCCGGCGACATTGTCCCGGATGATAGCGGAGAACTGTTCATGACCTTGA
CAGACGGAAGCGGAAGCAAGGAGGTCGTGTTTCTCCAGGCCGATATTAACG
CGGCGCACAATCTGCAAAAACGTTTTTGGCAGCGATACAATGAACTGTTCAA
GGTTAGCTGCCGCGTCATCGTCCGAGACGAGGAAGAGTATCTCGTTCCCAAG
ACAAAATCGGTGCAGGCAAAGCTGGGCAAAGGGCTTTTTGTGAAAAAATCG
GATACAGCCTGGAAAGATGTATATGTGTGGGACAGCCAGGCAAAGCTTAAA
GGTAAAACAACCTTTACAGAAGAGTCTGAGTCGCCCGAACAACTGGAAGAC
TTTCAGGAGATCATCGAGGAAGCAGAAGAGGCGAAAGGAACATACCGTACA
CTGTTCCGCGATCCTAGCGGAGTCTTTTTTCCCGAATCCGTATGGTATCCCCA
AAAAGATTTTTGGGGCGAGGTGAAAAGGAAGCTGTACGGAAAATTGCGGGA
ACGGTTTTTGACAAAGGCTCGGTAAGGGTGTGCAAGGAGAGTGAATGGCTTG
TCCTGGATACCTGTCCGCATGCTAAATGAAATTCAGTATTGTGAGCGACTGT
ACCATATTATGCATGTGCAGGGGCTGTTTGAGGAAAGCGCAGACACGGTCGA
AGGAGCAGCACAACACAAGCGTGCAGAGACACATCTGCGCAAAAGCAAGGC
AGCGCCGGAAGAGATGTGGGGGGACGCTCCGTTTAGCTTGCAGCTCGGCGA
CCCTGTGCTTGGCATTACGGGAAAGCTGGATGCCGTCTGTCTGGAAGAAGGT
AAGCAGTGGATTCCGGTAGAAGGAAAGCATTCGGCGTCGCCAGAAGGCGGG
CAGATGTTCACTGTAGGCGTGTATTCGCTGGACGGTTCTGCCTGGCCCAACG
ACCAAATCCAATTGTGTGCGCAAGGCTTGCTGCTTCGCGCGAATGGATATGA
ATCCGATTATGGCTACTTATACTACCGTGGCAATAAAAAGAAGGTTCGCATT
CCTTTTTCGCAGGAACTCATAGCGGCTACTCACGCCTGCATTCAAAAAGCTC
ATCAGCTTCGGGAAGCCGAAATTCCCCCTCCGTTGCAGGAGTCGAAAAAGTG
CTTTCGATGCTCGTTAAATTACGTATGCATGCCTGACGAGACGAATTACATGT
TGGGGTTGAGCGCAAACATCAGAAAGATTGTGCCCAGTCGTCCAGATGGCG
GGGTACTGTATGTTACAGAGCAGGGGGCAAAACTGGGCAGAAGCGGAGAAA
GCTTGACCATCACCTGCCGGGGCGAAAAGATAGACGAAATCCCGATCAAAG

ACTTGATTCACGTGAGCTTGATGGGGCATGTGCAATGCTCTACGCAGCTTCT
GCACACCTTGATGAACTGTGGCGTCCACGTCAGCTACTTGACTACGCATGGC
ACATTGACAGGAATAATGACTCCCCCTTTATCGAAAAACATTCGAACAAGAG
CCAAGCAGTTTATCAAATTTCAGCACGCGGAGATCGCCCTTGGAATCGCGAG
AAGGGTCGTGTATGCGAAAATTTCCAATCAGCGCACGATGCTGCGCCGCAAT
GGCTCACCAGATAAAGCAGTTTTAAAAGAGTTAAAAGAGCTTAGAGATCGC
GCGTGGGAGGCGCCATCACTGGAAATAGTGAGAGGTATCGAGGGACGTGCA
GCACAGTTGTACATGCAGTTTTTCCCTACCATGTTAAAGCACCCAGTAGTAG
ACGGTATGGCGATCATGAACGGTCGCAACCGTCGCCCGCCCAAAGATCCGGT
CAATGCGCTGCTCTCCCTCGGCTATACGCTTCTTTCACGGGATGTTTACTCCG
CATGTGCCAATGTCGGACTCGATCCACTGTTCGGCTTTTTCCATACGATGGAG
CCGGGCAGACCAGCTTTGGCACTCGATCTGATGGAACCGTTCCGCGCCTTGA
TTGCCGATAGCGTAGCGATACGTACCTTGAATACGGAGGAACTCACCCTCGG
GGACTTTTATTGGGGAAAAGACAGTTGTTATTTGAAAAAGGCAGGAAGACA
AACGTATTTCGCTGCCTATGAAAGACGGATGAACGAGACGCTGACGCATCCG
CAATTTGGGTATAAGCTCAGCTATCGCCGTATGCTGGAGCTGGAAGCAAGGT
TTTTGGCCCGGTATCTGGATGGAGAGCTGGTGGAATATACGCCGCTCATGAC
AAGGTAGGAAATGACCATGCGACAATTTGTTCTGGTAAGCTATGATATTGCC
GATCAAAAACGTTGGAGAAAAGTATTCAAGCTGATGAAGGGGCAAGGCGAG
CACGTCCAGTACTCGGTGTTTCTGTGCCAACTCACCGAGATTCAGCAAGCCA
AGCTAAAGGTAAGCCTGGCGGAGCTGGTTCACCATGGAGAAGACCAGGTCA
TGTTTGTAAAAATCGGCCCAGTGACGAGAGATCAACTGGACAAGCGGATATC
TACTGTTGGCAGGGAGTTTCTGCCTCGCGATTTGACCAAATTTATCTATTAAG
GAATGAAGAAAGCTAGTTGTAACAAAAGTGGAAAAAGAGTAAAATAAAGGT
GTCAGTCGCACGCTATAGGCCATAAGTCGACTTACATATCCGTGCGTGTGCA
TTATGGGCCCATCCACAGGTCTATTCCCACGGATAATCACGACTTTCCACTAA
GCTTTCGAATTTTATGATGCGAGCATCCTCTCAGGTCAAAAAAGCCGGGGGA
TGCTCGAACTCTTTGTGGGCGTAGGCTTTCCAGAGTTTTTTAGGGGAAGAGG
CAGCCGATGGATAAGAGGAATGGCGATTGAATTTTGGCTTGCTCGAAAAACG
GGTCTGTAAGGCTTGCGGCTGTAGGGGTTGAGTGGGAAGGAGTTCGAAAGCT
TAGTGGAAAGCTTCGTGGTTAGCACCGGGGAGAAGTCATTTAATAAGGCCAC
TGTTAAAAGTTCGAAAGCTTAGTGGAAAGCTTCGTGGTTAGCACGCTAAAGT

CCGTCTAAACTACTGAGATCTTAAATCGGCGCTCAAATAAAAAACCTCGCTA
ATGCGAGGTTTCAGC (SEQ ID NO: 63)

> Locus D

GAAGTTATGTTGATAAAATGGTTTATGAAAACGTGAGTCTGTGGT
AGTATTATAAACAATGATGGAATAAAGTGTTTTTTGCGCCGCACGGCATGAA
TTCAGGGGTTAGCTTGGTTTTGTGTATAAATAAATGTTCTACATATTTATTTT
GTTTTTTGCGCCGCAAAATGCAACTGAAAGCCGCATCTAGAGCACCCTGTAG
AAGACAGGGTTTTGAGAATAGCCCGACATAGAGGGCAATAGACACGGGGAG
AAGTCATTTAATAAGGCCACTGTTAAAAGTTTTGAGAATAGCCCGACATAGA
GGGCAATAGACTTTTGCTTCGTCACGGATGGACTTCACAATGGCAACAACGT
TTTGAGAATAGCCCGACATAGTTATAGAGATGTATAAATATAACCGATAAAC
ATTGACTAATTTGTTGAAGTCAGTGTTTATCGGTTTTTTGTGTAAATATAGGA
GTTGTTAGAATGATACTTTTTGCCTAATTTTGGAACTTTATGAGGATATAAGA
TAGACTTGATAAAAAGGTAAAAGAAAGGTTAAAGAGCATGGCAGGAATAGT
GACCTGTGATGAAGATGATGGTAGAATTAAAAGTGTTCTTAAAGAAAAACA
ATATTGGATAAGGAAAATAATTCAATAGATAAAAAATTTAGGGGGAAAAAT
GAAAATATCAAAAGTCGATCATACCAGAATGGCGGTTGCTAAAGGTAATCA
ACACAGGAGAGATGAGATTAGTGGGATTCTCTATAAGGATCCGACAAAGAC
AGGAAGTATAGATTTTGATGAACGATTCAAAAAACTGAATTGTTCGGCGAAG
ATACTTTATCATGTATTCAATGGAATTGCTGAGGGAAGCAATAAATACAAAA
ATATTGTTGATAAAGTAAATAACAATTTAGATAGGGTCTTATTTACAGGTAA
GAGCTATGATCGAAAATCTATCATAGACATAGATACTGTTCTTAGAAATGTT
GAGAAAATTAATGCATTTGATCGAATTTCAACAGAGGAAAGAGAACAAATA
ATTGACGATTTGTTAGAAATACAATTGAGGAAGGGGTTAAGGAAAGGAAAA
GCTGGATTAAGAGAGGTATTACTAATTGGTGCTGGTGTAATAGTTAGAACCG
ATAAGAAGCAGGAAATAGCTGATTTTCTGGAGATTTTAGATGAAGATTTCAA
TAAGACGAATCAGGCTAAGAACATAAAATTGTCTATTGAGAATCAGGGGTTG
GTGGTCTCGCCTGTATCAAGGGGAGAGGAACGGATTTTTGATGTCAGTGGCG
CACAAAAGGGAAAAAGCAGCAAAAAAGCGCAGGAGAAAGAGGCACTATCT
GCATTTCTGTTAGATTATGCTGATCTTGATAAGAATGTCAGGTTTGAGTATTT
ACGTAAAATTAGAAGACTGATAAATCTATATTTCTATGTCAAAAATGATGAT
GTTATGTCTTTAACTGAAATTCCGGCAGAAGTGAATCTGGAAAAAGATTTTG
ATATCTGGAGAGATCACGAACAAAGAAAGGAAGAGAATGGAGATTTTGTTG

GATGTCCGGACATACTTTTGGCAGATCGTGATGTGAAGAAAAGTAACAGTAA
GCAGGTAAAAATTGCAGAGAGGCAATTAAGGGAGTCAATACGTGAAAAAAA
TATAAAACGATATAGATTTAGCATAAAAACGATTGAAAAGGATGATGGAAC
ATACTTTTTTGCAAATAAGCAGATAAGTGTATTTTGGATTCATCGCATTGAAA
ATGCTGTAGAACGTATATTAGGATCTATTAATGATAAAAAACTGTATAGATT
ACGTTTAGGATATCTAGGAGAAAAAGTATGGAAGGACATACTCAATTTTCTC
AGCATAAAATACATTGCAGTAGGCAAGGCAGTATTCAATTTTGCAATGGATG
ATCTGCAGGAGAAGGATAGAGATATAGAACCCGGCAAGATATCAGAAAATG
CAGTAAATGGATTGACTTCGTTTGATTATGAGCAAATAAAGGCAGATGAGAT
GCTGCAGAGAGAAGTTGCTGTTAATGTAGCATTCGCAGCAAATAATCTTGCT
AGAGTAACTGTAGATATTCCGCAAAATGGAGAAAAAGAGGATATCCTTCTTT
GGAATAAAAGTGACATAAAAAAATACAAAAAGAATTCAAAGAAAGGTATTC
TGAAATCTATACTTCAGTTTTTTGGTGGTGCTTCAACTTGGAATATGAAAATG
TTTGAGATTGCATATCATGATCAGCCAGGTGATTACGAAGAAAACTACCTAT
ATGACATTATTCAGATCATTTACTCGCTCAGAAATAAGAGCTTTCATTTCAAG
ACATATGATCATGGGGATAAGAATTGGAATAGAGAACTGATAGGAAAGATG
ATTGAGCATGATGCTGAAAGAGTCATTTCTGTTGAGAGGGAAAAGTTTCATT
CCAATAACCTGCCGATGTTTTATAAAGACGCTGATCTAAAGAAAATATTGGA
TCTCTTGTATAGCGATTATGCAGGACGTGCATCTCAGGTTCCGGCATTTAACA
CTGTCTTGGTTCGAAAGAACTTTCCGGAATTTCTTAGGAAAGATATGGGCTA
CAAGGTTCATTTTAACAATCCTGAAGTAGAGAATCAGTGGCACAGTGCGGTG
TATTACCTATATAAAGAGATTTATTACAATCTATTTTTGAGAGATAAAGAGG
TAAAGAATCTTTTTTATACTTCATTAAAAAATATAAGAAGTGAAGTTTCGGA
CAAAAAACAAAAGTTAGCTTCAGATGATTTTGCATCCAGGTGTGAAGAAATA
GAGGATAGAAGTCTTCCGGAAATTTGTCAGATAATAATGACAGAATACAATG
CGCAGAACTTTGGTAATAGAAAAGTTAAATCTCAGCGTGTTATTGAAAAAAA
TAAGGATATTTTCAGACATTATAAAATGCTTTTGATAAAGACTTTAGCAGGT
GCTTTTTCTCTTTATTTGAAGCAGGAAAGATTTGCATTTATTGGTAAGGCAAC
ACCTATACCATACGAAACAACCGATGTTAAGAATTTTTTGCCTGAATGGAAA
TCCGGAATGTATGCATCGTTTGTAGAGGAGATAAAGAATAATCTTGATCTTC
AAGAATGGTATATCGTCGGACGATTCCTTAATGGGAGGATGCTCAATCAATT
GGCAGGAAGCCTGCGGTCATACATACAGTATGCGGAAGATATAGAACGTCG
TGCTGCAGAAAATAGGAATAAGCTTTTCTCCAAGCCTGATGAAAAGATTGAA

GCATGTAAAAAAGCGGTCAGAGTGCTTGATTTGTGTATAAAAATTTCAACTA
GAATATCTGCGGAATTTACTGACTATTTTGATAGTGAAGATGATTATGCAGA
TTATCTTGAAAAATATCTCAAGTATCAGGATGATGCCATTAAGGAATTGTCA
GGATCTTCGTATGCTGCGTTGGATCATTTTTGCAACAAGGATGATCTGAAATT
TGATATCTATGTAAATGCCGGACAGAAGCCTATCTTACAGAGAAATATCGTG
ATGGCAAAGCTTTTTGGACCAGATAACATTTTGTCTGAAGTTATGGAAAAGG
TAACAGAAAGTGCCATACGAGAATACTATGACTATCTGAAGAAAGTTTCAGG
ATATCGGGTAAGGGGAAAATGTAGTACAGAGAAAGAACAGGAAGATCTGCT
AAAGTTCCAAAGATTGAAAAACGCAGTAGAATTCCGGGATGTTACTGAATAT
GCTGAGGTTATTAATGAGCTTTTAGGACAGTTGATAAGTTGGTCATATCTTAG
GGAGAGGGATCTATTATATTTCCAGCTGGGATTCCATTACATGTGTCTGAAA
AACAAATCTTTCAAACCGGCAGAATATGTGGATATTCGTAGAAATAATGGTA
CGATTATACATAATGCGATACTTTACCAGATTGTTTCGATGTATATTAATGGA
CTGGATTTCTATAGTTGTGATAAAGAAGGGAAAACGCTCAAACCAATTGAAA
CAGGAAAGGGCGTAGGAAGTAAGATAGGACAATTTATAAAGTATTCCCAGT
ATTTATACAATGATCCGTCATATAAGCTTGAGATCTATAATGCAGGATTAGA
AGTTTTTGAAAACATTGATGAACATGATAATATTACAGATCTTAGAAAGTAT
GTGGATCATTTTAAGTATTATGCATATGGTAATAAAATGAGCCTGCTTGATCT
GTATAGTGAATTCTTCGATCGTTTCTTTACATATGATATGAAGTATCAGAAGA
ATGTAGTGAATGTGTTGGAGAATATCCTTTTAAGGCATTTTGTAATTTTCTAT
CCGAAGTTTGGATCAGGAAAAAAAGATGTTGGAATTAGGGATTGTAAAAAA
GAAAGAGCTCAGATTGAAATAAGTGAGCAGAGCCTCACATCGGAAGACTTC
ATGTTTAAGCTTGACGACAAAGCAGGAGAAGAAGCAAAGAAGTTTCCGGCA
AGGGATGAACGTTATCTCCAGACAATAGCCAAGTTGCTCTATTATCCTAACG
AAATTGAGGATATGAACAGATTCATGAAGAAAGGAGAAACGATAAATAAAA
AAGTTCAGTTTAATAGAAAAAAGAAGATAACCAGGAAACAAAAGAATAATT
CATCAAACGAGGTATTGTCTTCAACTATGGGTTATTTATTTAAGAACATTAAA
TTGTAAAAAAGATTCGTTGTAGATAATTGATAGGTAAAAGCTGACCGGAGCC
TTTGGCTCCGGACAGTTGTATATAAGAGGATATTAATGACTGAAAATGATTT
TTGTTGGAAGTCAGTTTTTTCTGTGGAAAGCGAAATCGAATATGATGAGTAT
GCATATGGCAGAAGAGCTGTAGAAGGCGAGAATACATATGATTACATTACT
AAGGAAGAAAGACCGGAACTTAATGACGAATATGTAGCGAGACGTTGCATT
TTCGGTAAAAAAGCAGGAAAAATATCCAGGTCGGATTTTAGTAGGATAAGA

TCTGCGTTGGATCATGCGATGATAAATAATACACATACAGCATTTGCCAGAT
TTATCACTGAAAATCTGACGAGACTCAATCACAAAGAACATTTTCTGAATGT
GACACGTGCATATTCTAAACCTGATTCTGAAAAATTGATACAACCGAGATAC
TGGCAGTCGCCTGTAGTTCCAAAGGATAAACAAATATATTATAGCAAGAATG
CGATTAAAAAATGGTGTGGTTACGAAGATGATATTCCGCCTCGTTCTGTGAT
AGTTCAGATGTGTCTATTGTGGGGGACTGATCATGAAGAGGCAGATCATATC
CTTCGCAGTTCAGGATACGCGGCGCTTAGTCCTGTTGTACTTCGAGATCTTAT
CTATATGTATTATCTGGATCATCAGGATTTGCAAAAAAATGAGTTGATATGG
GAAGTAAAAAAGCAGTTGGATCACTTCGATTTGACAAATAGAAATTATGATA
CAAATCCTTTTGATGTAGGGGGCAGCGTAAATGATCATATCTGTGAACTGAG
CGAGCATATAGCGAAGGCTCATTATATTTATGAGAGGGCTAAGGAAGGACC
ATTGCAAAATGTAATTCGGGATATTTTGGGAGATACACCTGCCCTTTATTCTG
AAATGGCATTTCCTCAGCTAGCATCTATAAACAGGTGTGCTTGCAATTCGCTT
TCTTCATATCAAAAAAATATTTTTGATACTGACATAGCTATATATGCAGATGA
AAAGGACACAAGAGGTAAATCAGACCGTATCCTTGTTGAGGGCGCATCTTCG
AAATGGTATGAATTGAAGAAACGCGATGCTAATAATGTCAAAATTTCTGAAA
AGCTGAGTATACTCAATACTATTCTTAAATTTAATAGTGTTTTTTGGGAAGAA
TGTTACCTTGATGGAAATATAAAACAATCGAGCGGAAAGCGATCTGAGGCA
GGAAAAATTCTTTATGGTCGCGACAACGGAAAAGAAAATGTCGGAGTTTCA
AAATTGGAATTGGTGCGGTATATGATAGCTGCAGGTCAGGAACAAAATCTGG
GAAATTACCTGGTGAGTTCAGGATTTTGGAGAAAAAATCATATGCTGTCATT
TATACAAGGCAATGATATAGCGCTTGATGAGATGGATGAATTGGATCTCTTA
GACTATATTCTGATATATGCATGGGGATTTAGGGAAAATATCATTAAAAAGA
ACAGTAATGTGAATTCTTTGGATGAAAAGACTAGAAAAGTGCAGTTTCCGTT
TATAAAGTTACTCATGGCAATTGCAAGAGATATCCAGATACTTATATGTTCA
GCACATGAAAAAACAGTCGATGAGTCATCTCGAAATGCAGCAAAGAAGATA
GATATATTGGGAAATTATATTCCTTTTCAGATTCATCTTCAGAGAACTAAAAA
AGATGGTGGAAGAGTGGTAATGGATACATTGTGTGCTGATTGGATTGCGGAT
TATGAATGGTACATTGATCTTGAGAAAGGAACACTTGGATGAGCAGTGATGA
AAGGATATTTAAAAAATTTTTGGAAAAAGGATCGATTTCTGAGCAGAAAAA
GATGCTTTTAGAAGAAAAGAAATGTTCGGATAAACTAACTGCACTGCTTGGG
AATTACTGCATACCGATAGACAATATTTCAGAGTCAGACGGAAAAATATATG
CGGTCTATAAGCTTCCAAAAAATGTTAAACCTTTGTCCGAAATCATTAATGA

TGTATCCTTTTCTGATTGTACGATGAGAGTACGTTTGCTTCTCATAAAGAGAA
TTCTGGAACTCGTGTGTGCTTTTCACGAAAAAAAATGGTATTGTCTCAGTATT
TCACCGGGAATGCTCATGGTTGAAGATTTTGATATACCGATGGGAAATGTCG
GAAAAGTATTGATATATGATTTCAGAAATCCTGTTCCGTTCGAGTCAGTAAA
TGAAAGACATAATTTTAACGTTTCAAATAAATACACTTCACCGGAGCTGCTC
ATCCATTCAAGATATGACGAGTCGAAATCTGTGAGTGAAAAATCAGATTTGT
ATTCTGTTGCAAAAATTGCGGAAACAATAATAGGAGATTTTAACAGTATTAT
TGCAAATGGAAATTTGATACTACTTGCAATGCTTAGAGTTTTTATCAGTACAG
GGAAAAGTCCGGAACCTGAGTATCGGTTTGAATCGTCGGAAAATATGCTTTC
AGTATTTGAAAATTTGATCAAAGAAAATTGTTTTTTTGAAAAAAACGATTAT
ACATCTATGTTTCATCAGGCGTATGACAATTTTTTTGAATGGCAGGAATGTTT
GATATCACCGGATCACTTGGATAAAAATATGTTCGAGGCAGCTTTATCAAAT
CTTGAGGATCAGCTGCTTAGGGTTGATATTGATAAGTATAGAGCAGAGTACT
TCTATAAGCTTCTCCGAGAGTTGTCTAATAAATATAAAAATACAATTACTGA
TGAACAAAAGGTAAGGTTGGCAATACTTGGAATCAGAGCGAAAAATAATCT
GGGAAAAAGTTTTGATGCATTGGAAATATATGAGTCAGTACGTGATTTAGAA
ACTATGTTGGAGGAGATGGCAGAGCTTAGTCCTGTCATTGCTTCGACATATA
TGGATTGCTACCGATATGCAGATGCGCAGAAAGTGGCGGAAGAAAACATTA
TCAGGCTTCATAATAGTAATATTCGTATGGAGAAAAAAAGAATACTGCTTGG
AAGGTCATATAGTTCAAAAGGGTGCAGCATGGGGTTTCAGCATATTCTTGGT
GCGGATGAGTCATTTGAACAGGCTTTATATTTCTTTAACGAAAAGGACAATT
TTTGGAAAGAAATATTTGAGAGCAGAAATTTAGAGGACAGCGATAGACTTAT
AAAGTCTTTACGAAGCAATACGCATATTACGCTGTTTCATTACATGCAATAT
GCATGTGAAACAAGGAGAAAGGAATTATATGGAGCACTTTCAGACAAATAT
TTTATAGGTAAAGAATGGACAGAAAGACTCAAAGCATATATAAGCAACAAG
GATATATGGAAAAACTATTATGAGATATATATTCTGCTAAAGGGTATTTATT
GCTTCTATCCAGAAGTCATGTGTTCGTCTGCGTTTTATGATGAAATCCAAAAA
ATGTACGATCTTGAATTTGAAAAGGAAAAAATGTTTTACCCATTGAGTCTGA
TAGAACTGTATCTTGCTCTGATAGAGATAAAAGTTAATGGGAGTCTGACGGA
GAATGCCGAGAAGTTGTTTAAACAGGCATTGACACATGACAATGAAGTCAA
AAAAGGAAATATGAATATTCAGACCGCCATTTGGTATCGAATATATGCACTG
TATAACGATGTAAAAGATGAAACTGATAAGAATAAAAGGCTTTTAAAACGG
CTTATGATTCTTTGCCGACGATTTGGTTGGGCGGATATGTATAGTGCTTTGGA

GAAGGATGGGAAGTTAATTGATTTTTTGAGATTTGAGGTATGTTAAATGATA
ACACTTGCATTAGATGAAAATGGCAAATTTGAAGATGCTTTTTCTAAAAAAA
ATGAAAAACCGATAATGATTGCGGGGATAATCTATGATGACAAGGGGAAAG
AGTATGATGCTGAGAATGAACGCTACAGGATATCCAGTTATCTGCGAGCAGT
ATGTGACAGTTTGGGTGCGAAATACCCTCAGGATCTACATTCAAATAGTAAT
GGAAATAAGGCGACTGTTGGGAAAGTAAAATGTAAAATTGGTGAAACACTA
AAGGAATTCTTGAGAGAAGGAACCTATGAAAAAAAGGAATTGCCGACAAAG
AACGGTTATTTAAATAAGAGATCTGGAAAATATGTAATGTTTGCAGAACTCA
GGAGTAGTCAGGGAGTTAAAAAGCGTGTTAGTGGTTGGAATGACAATGATCT
GACTCAGGATGAAAAGGTCAGCAATCTGTACCTTCATATGGCAGAAAATGCC
GTTGTCAGAATGCTCTTCCATAATCCTATATATGAAGATGTAACAGATGTAA
ATCTCTATTTTCCCACGCGAAAAGTTGTTCTGAAAGATAGAGATAGAGAATA
CGATAAACAAGATTTCAAAATATATGGTGATAAGGACAAGTGCGAAGCAGA
AAGCGGGAGATTGGTGCATTATGATATCGTGTCATCGGATTTTTACCGTACG
ATAATGGAGAACGAATGTACAAGAATTAATAAAAAGCAATTAAATGTTCATT
ATATGAACACAAGCCCAATTTCGTACTGGGAGAAAAATGAAAAATATAATA
CATTTTTATATTTGGCTGACATAGTTTGTTCTATGCTGGATTATTACAAAAAG
GGTTCGAGTCCGGCAGAGTGGATGGATTCTTTTGCCGAATGGGGAAACAAAT
ATTTTGGTGATGATCAGATAATCTTATTTGGGTATGATGATATAGATGACAA
ATACATGGAGGCTGTAGATGCAGTAGGACAGGGAGAGTATTTTCATGCGCTG
GATATTATATATGATGCGGAATGTAGTGGAAGTGAATTTGAGAAGCACTACA
AAGATTATTGGTTTCCAAAGCTTATAAAAAAGATACGAATAACAGCAACTGT
GGATAATTTATGCAGATCGATCTCAGATCTGGAGAGTTTTACATATCGAAGT
AATCTTGATCAGCAGAAACTTTTGTGGATTTTTGAGGAAATCAAAGCTATCG
TCGATAAGGGAGATTTTGGAAAGAAATATCATACAGATCAGGTTATGTTTGA
TATGTGTAATGCCGGTATTGCTGTGTACAATCATATCGGAGATTTTGGGACTG
CAAAGGAATACTATGATGAGTGCATGAAACACACTGGGGATGTGGATCTGG
TAAAGATACTTCGTGCATCAAATAAAATGGTGGTCTTTCTTGACGATGCTTTT
AGGTATGGTGACGCGACAGAACGTGCCAGGAAGAATGTTGAATACCAAAAA
GCTTTGCACGATATAAAGAGTGAGATTTGTCCGGAAAAGAAAGATGAAGAC
TTGAACTATGCCATATCGCTCAGTCAATTTGGACAGGCGCTTGCGTGTGAAA
AAAATTCTGATGCAGAGAGTGTTTTCCTAGAGTCGTTGCGGCATATGAGGAA
AGGGACTGCCAATTATCAGATTACTCTTTCATATTTACTCCATTTTTATCTGG

ATATGGGAATGACAGATTCTTATCGAGAAAAAACAAAGGACTATTTTGGAA
GTGAAAAACCAAAGGAACAGCTGAAAGAATTGCTGAAGTTATCGGGAAAGG
ATGATAGTATAGTTACTTTCAAATTTGCAATGTATGTCTATTTACGTGCACTT
TGGGTATTACAGGAACCGCTTACTGATTTTATCAGAACAAGATTAGAGGACA
TACGTGAGACTCTTGTAAAGAAGAAAATGAGTGAACATATGGTTGGACATCC
GTGGGAGTTGATTTATAAATATCTGGCATTTCTTTTTTATCGTGATGGAAATT
GTGAAGCTGCTGAAAAATATATTCATAAAAGTGAAGAGTGCTTGGAAACAC
AAGGACTGACTATAGATGCGATTATTCATAATGGTAAGTATGAATATGCAGA
ATTGTCAGGTGACGAGGAGATGATGGCAAGAGAGAAAGCGTACTTTGATGA
AAAAGGGATAGATAGAAAAAATGTTTGTACTTTTATGTATCATTGATGTTTA
ATAAGATTTGACCGAGGAGTGACAGGTAATCGCCGGTATATCTGGTATTACC
TGTCATTTTTTGATGAAATAAGCTACTTTTTGCCTAAAAAACGAAACTGTTGG
TGTTTTATGATGATTGTGTCAACAAAAGAGAGCAAAAGAAGAGGAGAAAAG
TAATGTCAATGATTTCATGTCCGAATTGTGGTGGAGAGATATCTGAAAGGTC
AAAGAAATGTGTTCATTGTGGATATGTGTTAGTCGAAGAAGCTAAAGTAGTG
TGCACAGAATGTGGAACTGAGGTAGAGAGTGGCGCTGCTGTATGTCCGAAGT
GCGGCTGTCCTGTAAATGATAGTGAGACGCCTCAGAAAGTTGAAGTGACTAG
GGTAAATGTATCTTCCGTAATCAGCAAAAAAGTCGTTGTAAGCATACTGATC
GCAGTGATTACAATTGCAGGTTTTTTCTATGGAGTGAAGTATTCGCAGGAAA
AGAAAGCAATTGAAGAGTCAGTAAAGCAGAAGGAAGACTATCAAAGTACGC
TAGAGCTTGCTTCGCTAATGATGCTTCAAGGAGCTTCGGATGCAGAAACTTG
TGGGAATTTGGTTAGGAAAGTGTGGAGCAACTGCATTTATAAGGAGAGGGA
TGAAGAAACCGACAAGTATACGTGTGATAGCAGGGGTGCAGGATGGTTTTAT
GATGATTTTAATGATGCATTAATGGCTCTTTACAGTGACAGCAGTTTTGGCAA
GAAGATAAATGAAATCAAAAACGGTCAGGAAACCGTTGCGGCGATGATGAA
AGATCTGAAAAATCCGCCGGATGAGATGGCAGATGCCTATGAGGATATTCA
AAATTTTTATGTGTCCTATCTAACGCTGACAGAAATGGTTGTGAATCCAACTG
GAAGTTTGAGTTCTTTTTCATCTGATTTTTCCGATGCGGATACGGAGGTGTCC
AATGCCTATAGCCGGATGAAGTTGTATTTAGATTAAACTATTGAGGAAAAAA
TGGAGGTGCTTTAATGCGGGGGAGAAACTGTGGAGGGTCATCAGGCGACGG
ACTGCTGGTACTTCTCGTACTGCTTGTCCTTTTTTATAAAATCATGCCATTCAT
AGGTTTATGGATTTTAATTTTTGGTGATGCTGAACGTAAAGATCTGGGTATGG
GTATGATTATTGTCGGGATAGTTCTATATGTATTATTAGAGGTTTTTTAATGT

GAGTTTCTGTGGTAAACTATAAAAGTACAAGCTTTTGCGCCGCACCGCATAA
ATAGCGGATTTATGACCATTATTTGGTGAAAAAAATGGTGTACACCTGTGTT
TTTTTGTTTTGCGCCGCAAAATGCGCCACGGAACCGCATGCAGAGCACCCTG
CAAGAGACAGGGTTATGAAAACAGCCCGACATAGAGGGCAATAGACACGGG
GAGAAGTCATTTAATAAGGCCACTGTTAAAAGTTATGAAAACAGCCCGACAT
AGAGGGCAATAGACATAAAGACCAAAAACAGGTCATCTGCATACTGTGTTA
TGAAAACAGCCCGATATAGAGGGTGTGAGAGATATAGTTCTCGTCACAGTGC
AGAAAATGACCTATTATGTGCCGAAAAACAAAATGAAAAAAGAATGGAAAG
GCGTATTTAATGAAATGCTGATCTGTTGATTTGAATTAACAAAAAAAGGTCG
CCCCACGGATGACAAAAACATCCGGGGGCGACCCTTTT (SEQ ID NO: 64)

> Locus E

TACTGTGTGCATAAGTCTTCCTTAGATCCATAGGTACAGCAGTTTT
ATTTATTAGCCTTAGAAAATGGAAAATAGAGCTTATAAATGATATGATATTT
ATGAATAAAATGATTGCATTCTCGTGCAAACTTTAAATATATTGATTATATCC
TTTACATTGGTTGTTTTAATTACTATTATTAAGTAGGAATACGATATACCTCT
AAATGAAAGAGGACTAAAACCCGCCAAAAGTATCAGAAAATGTTATTGCAG
TAAGAGACTACCTCTATATGAAAGAGGACTAAAACTTTTAACAGTGGCCTTA
TTAAATGACTTCTGTAAGAGACTACCTCTATATGAAAGAGGACTAAAACGTC
TAATGTGGATAAGTATAAAAACGCTTATCCATCATTTAGGTGTTTTATTTTTT
TGTGATTATATGTACAATAGAAGAGAGAAAAAAATCATTGAGGTGAAAACT
ATGAGAATTACTAAAGTAGAGGTTGATAGAAAAAAAGTACTAATTTCTAGG
GATAAAAACGGGGGCAAGTTAGTTTATGAAAATGAAATGCAAGATAATACA
GAACAAATCATGCATCACAAAAAAAGTTCTTTTTACAAAAGTGTGGTAAACA
AAACTATTTGTCGTCCTGAACAAAAACAAATGAAAAAATTAGTTCATGGATT
ATTACAAGAAAATAGTCAAGAAAAAATAAAAGTTTCAGATGTCACTAAACTT
AATATCTCAAATTTCTTAAATCATCGTTTCAAAAAAAGTTTATATTATTTTCC
TGAAAATAGTCCTGACAAAAGCGAAGAATACAGAATAGAAATAAATCTCTC
CCAATTGTTAGAAGATAGCTTAAAAAAACAGCAAGGGACATTTATATGTTGG
GAATCTTTTAGCAAAGACATGGAATTATACATTAATTGGGCGGAAAATTATA
TTTCATCAAAAACGAAGCTAATAAAAAAATCCATTCGAAACAATAGAATTCA
ATCTACTGAATCAAGAAGTGGACAACTAATGGATAGATATATGAAAGACATT
TTAAATAAAAACAAACCTTTCGATATCCAATCAGTTAGCGAAAAGTACCAAC
TTGAAAAATTGACTAGTGCTTTAAAAGCTACTTTTAAAGAAGCGAAGAAAAA

CGACAAAGAGATTAACTATAAGCTTAAGTCCACTCTCCAAAACCATGAAAGA
CAAATAATAGAAGAATTGAAGGAAAATTCCGAACTGAACCAATTTAATATA
GAAATAAGAAAACATCTTGAAACTTATTTCCTATTAAGAAAACAAACAGAA
AAGTTGGAGATATAAGGAATTTAGAAATAGGAGAAATCCAAAAAATAGTAA
ATCATCGGTTGAAAAATAAAATAGTTCAACGCATTCTCCAAGAAGGGAAATT
AGCTTCTTATGAGATTGAATCAACAGTTAACTCTAATTCCTTACAAAAAATTA
AAATTGAAGAAGCATTTGCCTTAAAGTTTATCAATGCTTGTTTATTTGCTTCT
AACAATTTAAGGAATATGGTATATCCTGTTTGCAAAAAGGATATATTAATGA
TAGGTGAATTTAAAAATAGTTTTAAAGAAATAAAACACAAAAAATTCATTCG
TCAATGGTCGCAATTCTTCTCTCAAGAAATAACTGTTGATGACATTGAATTAG
CTTCATGGGGGCTGAGAGGAGCCATTGCACCAATAAGAAATGAAATAATTC
ATTTAAAGAAGCATAGCTGGAAAAAATTTTTTAATAACCCTACTTTCAAAGT
GAAAAAAAGTAAAATAATAAATGGGAAAACGAAAGATGTTACATCTGAATT
CCTTTATAAAGAAACTTTATTTAAGGATTATTTCTATAGTGAGTTAGATTCTG
TTCCAGAATTGATTATTAATAAAATGGAAAGTAGCAAAATTTTAGATTATTA
TTCCAGTGACCAGCTTAACCAAGTTTTTACAATTCCGAATTTCGAATTATCTT
TACTGACTTCGGCCGTTCCCTTTGCACCTAGCTTTAAACGAGTTTATTTGAAA
GGCTTTGATTATCAGAATCAAGATGAAGCACAACCGGATTATAATCTTAAAT
TAAATATCTATAACGAAAAAGCCTTTAATTCGGAGGCATTTCAGGCGCAATA
TTCATTATTTAAAATGGTTTATTATCAAGTCTTTTTACCGCAATTCACTACAA
ATAACGATTTATTTAAGTCAAGTGTGGATTTTATTTTAACATTAAACAAAGA
ACGGAAAGGTTACGCCAAAGCATTTCAAGATATTCGAAAGATGAATAAAGA
TGAAAAGCCCTCAGAATATATGAGTTACATTCAGAGTCAATTAATGCTCTAT
CAAAAAAAGCAAGAAGAAAAAGAGAAAATTAATCATTTTGAAAAATTTATA
AATCAAGTGTTTATTAAAGGTTTCAATTCTTTTATAGAAAAGAATAGATTAA
CCTATATTTGCCATCCAACCAAAAACACAGTGCCAGAAAATGATAATATAGA
AATACCTTTCCACACGGATATGGATGATTCCAATATTGCATTTTGGCTTATGT
GTAAATTATTAGATGCTAAACAACTTAGCGAATTACGTAATGAAATGATAAA
ATTCAGTTGTTCCTTACAATCAACTGAAGAAATAAGCACATTTACCAAGGCG
CGAGAAGTGATTGGTTTAGCTCTTTTAAATGGCGAAAAAGGATGTAATGATT
GGAAAGAACTTTTTGATGATAAAGAAGCTTGGAAAAAGAACATGTCCTTATA
TGTTTCCGAGGAATTGCTTCAATCATTGCCGTACACACAAGAAGATGGTCAA
ACACCTGTAATTAATCGAAGTATCGATTTAGTAAAAAAATACGGTACAGAAA

CAATACTAGAGAAATTATTTTCCTCCTCAGATGATTATAAAGTTTCAGCTAAA
GATATCGCAAAATTACATGAATATGATGTAACGGAGAAAATAGCACAGCAA
GAGAGTCTACATAAGCAATGGATAGAAAAGCCCGGTTTAGCCCGTGACTCA
GCATGGACAAAAAAATACCAAAATGTGATTAATGATATTAGTAATTACCAAT
GGGCTAAGACAAAGGTCGAATTAACACAAGTAAGGCATCTTCATCAATTAAC
TATTGATTTGCTTTCAAGGTTAGCAGGATATATGTCTATCGCTGACCGTGATT
TCCAGTTTTCTAGTAATTATATTTTAGAAAGAGAGAACTCTGAGTATAGAGTT
ACAAGTTGGATATTATTAAGTGAAAATAAAAATAAAAATAAATATAACGAC
TACGAATTGTATAATCTAAAAAATGCCTCTATAAAAGTATCATCAAAAAATG
ATCCCCAGTTAAAAGTTGATCTTAAGCAATTACGATTAACCTTAGAGTACTT
AGAACTTTTTGATAACCGATTGAAAGAAAAACGAAATAACATTTCACATTTT
AATTACCTTAACGGACAGTTAGGGAACTCTATTTTAGAATTATTTGACGATG
CTCGAGATGTACTTTCCTATGATCGTAAACTAAAGAATGCGGTGTCTAAATC
TTTGAAAGAAATTTTAAGCTCTCATGGAATGGAAGTGACATTTAAACCACTA
TATCAAACCAATCATCATTTAAAAATTGATAAACTCCAACCTAAAAAAATAC
ACCACTTAGGTGAAAAAAGTACTGTTTCTTCAAATCAAGTTTCTAATGAATA
CTGTCAACTAGTAAGAACGCTATTAACGATGAAGTAATTCTTTTAAAGCACA
TTAATTACCTCTAAATGAAAAGAGGACTAAAACTGAAAGAGGACTAAAACA
CCAGATGTGGATAACTATATTAGTGGCTATTAAAAATTCGTCGATATTAGAG
AGGAAACTTTAGATGAAGATGAAATGGAAATTAAAAGAAAATGACGTTCGC
AAAGGGGTGGTGGTCATTGAGTAAAATTGACATCGGAGAAGTAACCCACTTT
TTACAAGGTCTAAAGAAAAGTAACGAAAACGCCCGAAAAATGATAGAAGAC
ATTCAATCGGCTGTCAAAGCCTACGCTGATGATACAACTTTAAAAGGAAAAG
CAGTGGATTCTTCACAAAGATACTTTGATGAAACGTATACTGTTATTTGTAAA
AGTATCATAGAAGCATTAGATGAAAGCGAAGAGAGATTACAACAATATATT
CATGATTTTGGAGATCAAGTGGATTCTTCACCTAACGCACGAATTGATGCGG
AATTACTACAAGAAGCAATGAGTAGGTTAGCTGACATAAAGCGGAAGCAAG
AAGCACTTATGCAATCCTTATCTTCTTCTACAGCAACGCTTTACGAAGGCAA
GCAACAAGCGTTACACACTCAATTCACGGATGCGCTGGAGCAAGAAAAAAT
ATTGGAACGCTATATTACTTTTGAACAAACTCACGGGAATTTTTTTGACTCAT
TTGGAGAACTTGTCTATCGAACGGGACAAGCAGTGCGTGAATTAGCTAATAA
CGTCACATTCGAGAGCCAAACAGGAAGCTATCATTTTGATAAAATAGATGCT
TCTAGATTCCAAACTTTGCAAGAAATGTTGCCAAAGGCAAAGAAAAAAGCA

TTTAATTTTAATGACTACCAAATAACATGGAATGGCACCACGCACCTTTTATG
GAAAAATGGTAAAGTGGATGCAGAAGCAACCAAAGCTTATAACGAGGCGAA
ACTGAATGGAAAGCTACCAAAGGAAGGTAATGTAGCAACACAAGATGCAGA
ACTATTAAAAGGCATTTTGGCTTCACTGAAAAACAAGAAAGATCCTATCACT
GGAGCAGATATAAGCAGTGTGCATGTATTATCTATCCTTAGCGGGCTCGCAT
TCTCCTATACAGCTGGGAATTATAAGGGAAGAAAACTTACTGTTCCAAAAAG
TTTCTTAGACAAATTAAAGAAAAACCGAAAATCTAAAGTACCTAAACTATCT
AGTTTATCAGAAAAACAACAACTAAAACTCGCAAATAAATACAAGAAAAAA
TCACCTATTCCAATTCCAGATGATGCTAAAATCAAAGCTCAGACGAAAAAGG
CTGGTTATGAACAAATATCTTATAAATGGAAAGAGAATGGGATAACCTTTGA
AGTTAGATGGCATACTAGGACACCAGGTGCACCAAAGGAACAAGGAAATAC
GTTTGTTATAGAAAGAAAAATTCAGGGTACAGCAGAAGGGAAAACAAAAGT
TCAACAAATATTGGTTGGAGATAATAAGTGGGTGAGTAAAAGTGAGTGGCA
AAAGGCTATAACTGATAAGAAAAATGGTGTAAGTACCTCGGAGCAAAATAA
AATGTTGTCTGATGGACATTGGAAAGAATAGAAAGGAGCAAAATGATGGAA
GATTATTATAAAGGTTTTGAGGGATATCCAGAGATAGATTTTTATACGTATAT
AGATGATATGAAATTGGGTATAGCAATGTGGGAAGGATACTTTGACAACATT
ATGAAAGAAATTAATCCAAGTAACGGAAGATGGACTTCATTAGCGTATTATT
ATCATTTAGATGAGGGGTGGTATGATGAAAGTCCTTGGGAAATACCAAGTAA
TACAGAAGCATTAGAATTATTGGAAACAATCCATATATCTAATCTAGATACT
ATCACACAAGAGATATTACTTAAATTAATAAATTTATTAAAGAAGAATATAA
ATAGACAAGTTTATATTGAATACTCATAAAAAAGATGATTATGATATATTAT
AGAACAAACGAACAAGCCCCAAATACGAGGTTTGTTCGTTTGTTTTCAATAT
AATTATTTGCCACCAAGTGAGATATTACGGTTTTAAATAGCTTATTTGACGAT
ACCAAACCCTGATAAGAGAAAGAAGAAAGAGAAAGCTGGTGTAGTTGTTTT
AAGTGAACTAGATAAAAAATTAATAGCAAAACTTGAAAAAGATGGTGTGAA
AATATCAAAAGAAGATGTTATAGGAATAAAATAATTGCCAGATGATGAGAA
ATCGTTTGGCTGGAAAAAGGAAATCCATCCGCTGGATTTGAGCATATTCTTA
TTGAACATGGTGAACAATTTGCTAAATAGGGAATTTCAAAAGCTGAGTTACC
TGATTTTTTGATGACTGCTTTAGAAAAGGAAA (SEQ ID NO: 65)

> Locus F

ATTCTTTAAAAATATCTAATAATTTATTTACTATATACTCTAATACA
TCTTTTAACCTATCTAAAACATCATCACCTACAACATCCCAAAAATCATCTAA

AAAGTTAAAAAAATCCATCTTTATCAACTCCTATATCTATTTTTTATTGTGTA
ATTCCTGAGTTACAAAACCATTATAACACGTATTACACACGTAGTCAATACT
TCAAAAAAATTTTTTGTATATTTTTTTGAATAAGTAAATAAAAAGAGCTGTGT
AGCTCTTTATTAAAATCAATATTTTTATTTTGTTAACAAACTTAGACAACATT
AAATTTAGAAACCTATATATATTTCAGTACTTTTCATTTTTAGGTAGTCTAAA
TCAGAAATGGTTTTGTCTAAATGATGTATGTAAGTTTTAGTCCCCTTCGTTTT
TAGGGTAGTCTAAATCAGAAGTCATTTAATAAGGCCACTGTTAAAAGTTTTA
GTCCCCTTCGTTTTTAGGGTAGTCTAAATCCCATCCAAATTATGGGATAATAT
GTTACTTTTTATTTTAATATTTGATTATTTATTGTTTTTTTTACTGATTTAGATTA
CCCCTTTAATTTATTTTACCATATTTTTCTCATAATGCAAACTAATATTCCAAA
ATTTTTGTTTCTTTTCTTATGATCTTTTCTCCGATAGTTATTTCTCCAGATAAG
ATTTTCATTTTTTTGAATTGATCTTCTGTTAGAATTAATGTTCTTACTGATGAA
TTTTCTGGAACTATCATTGACAACTGATTTTCATAGGAAATTATTTTTTCTTTT
GTGCTAGAACTTACAATGTATACTGATTTTTGTACCTGATAATATCCTTTTCT
TATAATTTCTTTTCTAAATTTTGCATATTCTTTTTTTTCTTTTCCTGTTTGCATT
GGAAAATCATACATTAGAATCCCTACATAATTAGTACTCATAATCCTCTATCC
TTAACTCAGGAATTTCTACTTCTGACATTTCTCCTGTAAAATAATTTCTAATA
TTATCTAAAAAATAATCAATCACTTGAGCCAATTCATATTTTTTATTTTTCCA
ATAAACTTTTTGTGTTAATACCAATAACAATTTTTGTCTTAATGATTTATTCA
AACTTACTTCTTCCTGTTGATTAAAATATACGATATAATCTACCATTGGACGA
AATATTTCAATAATATCATCTGCAAAATTATAATTATTAAATTGTGAACTGTG
ATGTATTCCCAAACTTGGATGAAATCCTTTAGCCACAATTTTTGAAGAGATTA
AGCTTCTCAAAACCATATACCCATAATTTAATGCCGAATTTGTCCCGTCTTCA
CCAAATCTCTTAAATTTTTTCCCAAAAAGTTCACCAAAATACATTCTTGCAGC
AATTGCTTCCTGATGTTCCGCTTCTTTTCCTTTTAATCTAATATTATTTTCATA
TGCTTCCAACTTATATGATACTTCCTGAGATTTTTTCAAAAACTGCAATAAAT
TTCTTTGATTTTCTATTTTTCTCATTACAATTTTTCTCCAGATTTCTTCTTTTTT
ATCGTCAATCCAGCTCACTTGCTCATTAATTCTTGTTGTTACTTGAAAATGAT
TATACAGTCCTAATGAATGTAAAACTGGCTGATGTTTTTCATTACAAATTATC
AGTGGAATATTATGTTCTGATAATCTTAACTGTAATATTCCGCTAATTTTACA
TCTGCAATTTTCAACTACAATTGCCATGATATCATTTAAAGATACTTTATCAG
CCTTATTTTCATCATCTTCATTTATCATCACAAGCTGGTTATTTAAAACTGAT
AATTCATTGACTCTTGTTACATGGATAATATTAGACATTTTTATTACTCCTTTA

CTCTAAAGCTTTATATTCAAACATAACTTTCACAAGTTCACACAATTCTTCTG
AATTTCTATCAGTCATTAATTTTTTCTTTTTTAAATTTTTCAAATGTACAATTT
TTTCCGATTCTAAAGTCTGAATTTCTATTTTCTTATCTGCTCCTATTTTAAATG
TTGCTACAAAACCATATTCCTTTAATATATCCACTATTGATTTCATAATTGCA
TTTTTAAGTTTTCTATCATAAGAAAGTAATTTTCTTAAATTTTCCAGCACTTCT
AAAAGTGAAATTTCAGCATGCGGAATATAGTTAAAATGTGCAATATAGTTTC
GTATATACAAATCTTTTTTCTCTTGTTTTAATTTTTTTACTTTTTTATCAGAAT
AGATGCTTCTTTTTTCTACATTATCTTTGTATAATTCTTTATAAAAATTTATAT
ATTTTTCAACAATTTGCCCACTTTTATATTTTACATTTTTACTGTTATCAAAAT
TAAATATTTCTTCAATATAATGATTTTCAGGAAATTCACCTTTCAATCTAAAT
CTTAAGTCCCTTTCCCAGATCGAAGTATATCCCACAAGTCTGTGGAGTATTTT
TAATAACAAGCCTTGCAACAAGTTTAATTCATTAAATTCCACTTTATTTTTCA
AATGAGTATATTTTTGTATATTTCCAATTGCTTTTTCATATTCTTTATAATCTT
CATCATTAAATTTTTCATCTTTTTTAGGTCTTGCATATTTTCTATGTAAATTTT
GCTGCATTGTATAATTTTTTTCTATTTCATTTTTTTTTATTGCTGTATTCTTTCAA
TTCTTTTAAACTTATTTTATACTTCGCTTTATCAGCTATTTTTTCAAGTAAATT
TAACATCCCATATTTTTTTATATTATAAAAAGCTCTATGCTTTATAATATTTTC
TCCATCAAAATATATTTTATTTGTGTCAAATTTCTTCAATTCTTTCCTATCTTT
TATTTTATTTTCATTAAAATCTAAAAATTTTCCAATTTCATTCGCTTCTAATTC
AAAATCTTCTGTTACTCTATTATTATCTAAATTTAAAAGATTTATAAGTTCAA
GTTCATCTGAAAAAGTTTCTTCTTTATTTGCACTCTGATATTTTTCAAGACTTC
CCTTCAAATTAGTCAATTCTTTATGATTAAGCAATTTTAAAATTAAATAAAAC
ATATTCAAATTTTCAGTGTATTTTAATATCTTTCCTAATTTTATCTCTCTTACA
AATTCATTTATTTCATGTGGAATTTCTTTATTCCTATTATGTTTTTCATAATTT
TTTAAAATTTTATCATATTTTTCTTTATTATCTTTTTTTATTTTTATTTTAGAAA
ATATATCATTATTATCATTGTTATTATTACTTTCTATATATTTTAAATTATTTT
TATTCAAATAATCTATAAAACCTTTTAAAAATATTTGTTGTATAAAATCAATG
TATGTATTTTTTTCTTCTTTATCTTGATTATTAATCATCTCCCTACTTTGTATAA
TAGCAAGATATTCTACTGGTACAGTTTTTTCTATATTTTCAAATTTTTGATATT
TATAATGTCCTGTTTTTTGATTTCTTTGTTTATTTATTTTTATTACTTCATTAGT
TATTTTAAAAAAAACTTTACTATTTTTAACAAATTTATTAAGAAATTCACCAT
AATAAATATTTTTCAAAAGATATATTTGAGCATCTTTTTCTTCTTTATCCTTAG
GAACACTCCAAAAAAATTTTAAAGTATTTCTTAAATCTTCTATTTTATTATAT

AATTTCGTAAAAGAAGGAACAAAAGGAATATTCTTATTTACAAAATTAAATT
TTGTATTTTTTAAATATTTAATTATCACATCCTTTTCATAATAATTAAATACAT
TTGCACTATTTAACTGCTTAAATATCTTCAATTTCAATTTTTTCTCATTTATTT
CATTTTGAAACATTTTTTTTGAAATTTCAGAAGGAGCTATATTTTTAAATGCA
AATATATCTTTCCCTTCTAATTCCAAATTAAAATGCACAATCCCATGTCTAAT
ACTGCTAATAGCTTCATCAATATTTGCAAAAAAATCTTCTATCTCATTTTTAT
TATCCATATTAAAATCATAACTATAGAACATTTTTAAATTTTCTTTTACTTCAT
TTTGCTTGTTTTCATTATATATTTTATCAACTTCTCCAGAAACATATTTTTCTT
CGCCCTTATTATTTTTTACAGTTTTTCCTCTCATTCTACCTGTAATATCATTCT
CATTTTCAGTTTCAAGAATATTTCTCAATGAAAAATATGCAACCGAAGAAAC
TCCAATTATATTTCGTAAAAATGCTTCATTTTGTCTATTCCTAGCAATAAAAT
CACTTGTTGCAATCTCTCCAACTTGTAAATAATAATTGTATTTCCCACAATTT
CTTACATAAGTATCCAATTTATTTAGTAATTTGTTTTCAATTAATTTTTTTAAA
TTTTGATATTCAAATATTCTCTTAATTTTATCGTTACTTATGTTACTCAGTCTT
TTATACACATAATTTTTCAAAAGCTGACTCATTTCAATTTCCACAAAATGACA
AAAAGCATATTTTATATTTTTATCATTAAGTTCTTCTTTATCCAAATAATATTT
ATAAAACACTTGTGATTTTTTTAATTCACTCATATCCGGAATTTTTTCAATTA
ATTCTTTTATATTATTTACATTTTGTATTTCTTCGTAAATAATTTTAGCAAAAT
TTTCTTTATCATTTTTTCTTCCAATTATTTTGTGATAGTATTCTCTTATTTTATA
TTTTTCATGTTTTTTTGAATTTTCTATTAAAAAAAATAACTTCTCAATATCTTC
TTTTTTATACAATTTATCAAATGCTTCCTGTACATTATTTATATAATCATTACG
CTTTGCTGATTCTCTATAATAATCATAAATAATATTTCTTTTGCTCTTCCCTCC
AACTTTTTCAACATTATTTTCATTAATTTTCTGATAATTAGCCTTATTTTCTTC
AAATGAATATTTTAAAGAATTTATCTTATTCAATTTTGCCTCAACATCTTTTCT
AAATATTTCTAATTCTTCAGAGTTCACATCTTCATTTAACAATATTTTCTTTAA
AACTGAAAAACTATTTTTATTTTTTAAATCATATTCTGAAATATCTTCTTCAG
AATAATTTTTATCCTGTACTGCATTTTTCTCTTTCCTATTCTTTAAATACAGAA
CACTATCTTTTAGATGCAATACTTTATTTGAAAAAAACTTTTTTAAATTTTCTC
TTCTTATTCTATTTTCTTCTTCACTTGCATTATCAGGATTTTTTATATATATATC
CAGTCTTATACTTAAAAGCTCTGACAATCTCTCACTAGTCCTATTTTCTTCGC
TCGTACTTTTTACTAATTTTCCCTCTTCAATATATTTTTTATGCGAAATTCCAT
CAACTTTTGTAACTTTCATATATAAAAACCTCCTAATATCTATATTTTTTACTC
AATACCTAATTCTTTTTTCAATGCTTTTTGTAAAATTTGTGAAAAATTCAGAT

TTTTTTCCTGTGCCAATATATCTAACCAAACAGGAATTGTTAAAGTTTTCTTT
TTAAGTGCATTTGTAACTTTTGCCACTTCATACACTGGATCAACAGATAAAAT
ATACAAATACTGATTTTCTTTCAGTTTCACATCCTCCACTTTTGAAGGCTCAG
GAAATTTTTTTCTTACATCCAAAAAATCAGCCAAATGCAGACCCAATGTCTCT
CTCAAATTGGAAACAGCCTCCTCCATGCTATCTCCAAATGTAGCATAATAAT
TTATCTCTCCATCTTCAAACTTATCAAAATCAACAATACAACCATAATAAGTC
CCATCTTCCTTAGTTACCACTGCTGGATAAAATACATCCATTTTAATTATCTC
CAATCTATACCACGTGTTAAATACGTGTTTAAAAATATTTATAAAATTTTTTA
GCATCTCTGCTAAAATAAAACAATTATTTCAAATTTTTCTATTCCTTAATCAC
TCATTGTTAGTGATTCTTTTTTTACTTGGACAATTTTTCATTTAATTTCTTCAA
TTTTTTTAAAATCACATTTTTTTTAATATTCCTTATTTAATTGCAAATTTTCATT
ACTTTTGGGGTGCTCTAAATCCCATCCAAATTATGGGATAATAATTTTTAGTG
AAAGCAAGAAGGGACTAGAATTTAATCCCAACTTGTTTTTCAATACTTCTTA
ATGTTCCTACAGGTATATCTTTTGAATATGGTACTGTGACCACACCTTCCACA
CCTGGGATCATCCATTGATAATGACTACCTCTTATACGCACAACTTTTCCGCC
TAATTTTCTAAATCTTTTTTCGAT (SEQ ID NO: 66)

> Locus G

CTTTCTATCTTTTTCAAATAAAATTAGGCTCTAGTTAGCCTAATCGC
ATAATTATTTATTATAGTATAATTCTTATTTTTTTTCAACCTAAAAATTTAAAA
CATCTCCAAAAATTTTCGTTTCAGAACAACCAAGCAACCATATTCAAAAAAC
AATAAAAAAATGAGCAAGAATTGAAATTTTATTCTCACTCAGAAGTTATTTTT
ATTAAATATCACTTTTCGATATTGGGGTGGTCTATATCAATTTAAAAGACAG
AATAGATAATTCTTTAGAGTTTTAGTCCCCTTCGATATTGGGGTGGTCTATAT
CAGAAGTCATTTAATAAGGCCACTGTTAAAAGTTTTAGTCCCCTTCGATATTG
GGGTGGTCTATATCCCATCCTAATTTCTTGCTGATGAGATATTTATTTCTAAT
TTTTCTATTTTGTCTTTATTTTCAATACTTTCAATCCTATTTTTCTCTTTATTAA
TAATATAGAACCACCCTATACTATTATACCATATTTTTTGATTTTTCAAAATT
CCAATATTTTGTTTTGTGAAATTTTTTCTCCCATTGTCACTTCTCCTGCAAGTA
CCTTCATTTTTTGAAACTGATCTTCTGTCAGGATAATGGAACGGATTGATGAA
TTTTCTGGAGCGAGCATTGATAACTGTTTTTCTGCCAGTTCGATTTTTTCTTTT
GTTTTCGACCTCATTATATATACCGATTTTTGAAGCTGATAATATCCCTTTTCT
ATCAATTTTTTCCTAAAAGTCCTATATTCAAATCTCTCAACATCTGTCTGCAT
AGGAAAATCATACATAAGCAGACCAAAATACTCAATACTCATAGTCCATCAC

GCTCAATGTCGGAATTATCACTTCTTCATCTTTTACAAAATAATTTCGTATAC
TATCCAAATAATAGTCTACCGCTTGGAAAAAATCATATTTCTTATTGTTAAAT
AATACCTTCTGCTGTGCTACAAGAAGTATTTTTTGCCTTATTTCCTTACTTAAT
TTCACTTCATTCAAAATATCCTTGTACATATAAACAAGATAATCCACCATAG
GACGAAAAACCTCTATTATATCATCAGAAAAATTATAGGCATTAAACTGTGA
CTTATGATGTAATCCTAAACTTGGATGAAATCCTTTTGCTACAATCTTTGATG
ATATTATAGCTCTTAAAATCATATATCCATAATTAAGTGCAGAATTCACTCCA
TCTTCATCAAATCTTTTAAAACTATTACTATACAATTCCTGAAAATATATCCT
TGAAGCTATTGCTTCCTGATGTTCTGCACTCGCATCATCTTTTTTCAAGTTTTC
CTTATATGTTTTCAGTCTTTCAATGGAAATATCACTTTTTTCAAGATACTCTA
ACAATGCTCTTTGATTTTCAATCTTATTCTCCACTATCCTGCTCCACAATTTTT
CCTTTTTCTCTTTTTCCCACTCAATCTGCTCATTTATTCGTAAAGTCACTTGAA
AATGATTAAATAATCCCAGCGAATGAATTTCAGGCTGATGTTTCTCGTTGCA
AATAATAATCGGAATGTTATTTTCCACCAGCCTCAACTGCAAAATCGCACTA
ATCTTACAATAGCAGTTTTCAATAACTATCGCAGATATATCATTCAAAGAAA
TCTTATTTTTCTCATCATTATTGTCTTCATCAACCATTATAAGCTGATTATTCG
ATATTGACAAATCATCAGCCCTTGTTATGTGAATTATATTGGGCATTTTAATC
ATACTCCTTATAAATTTCATTCTTATAACGTATCATTCGTATTTTCTATTTTTG
TTAAAAGTTCTATTATCAAGTTTTTAATATAATCAGAATTATAACTTTCTAAT
TCTAAAACAGAAACTTTTTTAGGTTTCATTAATCTTTCAAGTATATCATTATT
ACCGATAAGTTTAAATTTTTTCTTTAATTCATCATAATCTAAATTCACATCTTT
TTTAAATACTTCAAATACACTTGCATAAGTTGAATTATTATAACGTGTACTAT
ATGATAATAAATTAGAAACTCTATCAATTTGTTCTGCAATACTGTAATCAGC
AAACGGATTTCTTACAATATAGAAATGTGAAATATAGTTTCTAATACTTTCAT
TTTCCGGCTTATTAATTTCAGAATTTTCAGACAAATCAATTCCAAATCCATAA
CATATTTTCTCAAATTTTTTATAAGATTCTTCATCAAAAAATTTATAGTATGC
TGTTGTTGTATAAAAGCCATCAGATCCATTACGCTTAGGATAAGCTCTACTTA
TTCCAGTATTGTAGCCACTTAACTTAATAATTCCTAATTCTCTTAGCCCATTT
ACAATATAGTGCATATCTCTTTCAAATCTAGCCATTTGAATAGCAAGTTTCCA
ATTTATATCTATCAAATAACTTTCTATTTTATTCAAATAATTAAATTCTACCA
AATCTCTAATTTTTTTGTATTCAGAAACTCTATTATAATCTTTTTCAAATGATT
TATAGTTTTTATTTTGTATATTTTTTGCAAAAAAGTCATCATTTTCTTTCAATT
TTTTTATATACTTCTCTTTGTATTCTTTAGAATATCCATTTAGTTTATCATTTA

GATTTTTCAATATTGCATCAATTTCAGATATTTTATTTTTTCTAATATTTTTAC
CATCAATATTAAATAAAAATTTTGCATCAGCCATTTTAATATCATTTGAAATT
AATCCATAAATTTTATCAAAATTTGGATTTCCAATATTTAAAAATAAATTCTT
TTTATAAATATATAATTCATTCTTACGTTCTTTAGGATAATATATTTCTTGAA
ATTTATTTTCATTCTCTGATTCCATATCTTCTATTAAATTATCTATTTCTTTTTT
GTATTTTTTTAAAAAATCAGAATTAAATATTATTCTACACAATATTTTACTCT
TTATTTCCTGATCTTTATCTTTTATATACTGATCAACCTTTTTTTTCAAATCCTT
TTTATTTATGTTTGATAACTTTCTTTGTTCATCTTGTAATATATTCGATTTTTTA
TCTATCTCAAATTTAGTTTCATCATCAAAAATTACAATTTTTTCTAATTTTTTC
TCTAAAACATCACAACCATTAATATCATCTTTAAATTCAGTTAATATATTATT
TTTTATATCCTCATAATAATTATTAAAAATTTCTTTTTTAGTTTGTATTTTAAA
ATCATCAAAGTCTTTTTCTATCTCTTTCATTTTTTGAATAAATTCTTCTAAATT
AAGATTCCAATTTTCAGTTATACATTCATTTCTCAAAGTATTTAATTGCATTA
TTTCATCTAAAATATCTATAATATTTTGATATTCTGAAGTATTTAACCAAACT
GATGTTGCAAAAAATCTATTTCTAATTTTATTTATAACCGCATTACTATTTAA
CAGTGCAAATATTGAAATTATATATTCAAAATCATCATTTATTACTATAGTTT
TATCACTAGTCTTTACAGTTATTCTTTCGTAAGTTTTATTATCATTAATGTCTT
TTATTTGTTTCTTAATTTCTTGAATATTCATTTTAAAATCTGAAAAATCAAAA
AGTTCCTCATAATTTTTTCTCAAATATCCAATATAACATTCTATTACTTTTTTC
TGATATTTTTTAATAGCTTTATTATTACCTTTTGAAGCAGAAATCTGAGCATT
TTTATAATAATTTTCTATAATATTTTCATCTATTTCATCAATGTTTCCTAAAGT
TTTCTTTAATTCTTGTAAAAATATATTCTTACTTTCATTTTCTTCTAAATCATC
TTCTAAAATTAATTTCTTATACAATTCTTTATTCACATATATTAAAGCATTTA
ATACTATTTTTTCTGTTTCTATAGTATCAAATGGTTCATTCTTAGGATTATTCC
TATATAAATTTAATATTTCAGGAAGTACTTTAGAAAAGGATGGTAAATATTT
AATATCATTATTATTTTCTTCTGAAATTTTAATATCATTTATTTTAGTAATTAT
ATTTTTTTTATCTTTAAATACTACATCTAAATTTAATGCTTTTGACACTTCTTC
ATCTGATATTTTTAAATTTTGAATTATATTTATGACTTTATTATAGTCATCTTG
CGTTCCTTGTAAATCTCTTTCCTTGCTAATCGCATGTAATATCCTGTTTCTTTC
ATTTGTTCCTATCTTTGTAAATTTCCTAATAAAATTATTTGTAATGTTATTTTT
ATTATCTATAAAATCTAAGTCTCTTATTATTTTTATTTTTGAATTTAAAATTTT
TTTATCAAGTACGTAATTTTTTTCTCGATCTCCTCCAAAGAAATCTATATTTTC
ATCATTATTTATATTTTCTCTAGAAAAAATCTTATTTAATTCCATATTGGTAG

AAGCAAAAAAAGTAATCAATTCTAAATCCAATTCCTCTTTAGCGTGAAGTCT
AGAAAAATCATCAGTATTTACTGTTGTCATATCTATATCATTATGTCTTAATT
TCCCTAAATACATAATATGCTCTAACGTATATTGCTTAACTCTTTTTAAAATT
TTTTCAGATAATATACTTTCATTTAAAATTTTTTCTATTTCTATTTTTTCCATTT
TCTTTAATCTGACTTTTTGTTCATTTACCAATATTTTTTCAATTCTTCCTTTCAA
ATATCGATATATGATTTTATATAGTTCTTTTTCTTCATCAGATTTCTTTGAAAA
TTTTTTCGAATCAAAATTAACTTTATAATGTTTTTTAAATATTCCAAAAATTTC
TGTATCACAATTTCCTTTTTTTAGTTCTTTTTCTAATTTTTTTATTAATTCATCT
ATTTTAAATTCTGCTAAAATTTTTTCTATTTTTTCTTTTATACTATTATTTTTTA
TATTTTCTACAAAAAATTTTACAATTTTATCTTTTTTATTTTCTCTTTCTATTTT
AAATTTTTCGTGCTTATCTAATAGTACATAAGATTTTATATATGTTCTATTTCT
TCTCTTTTCAAGAAATTCATTATTAACTTTTTTTACTTTTTCAATTCTTTTAGT
AATATTCCAAAATTCTAACTCTTTTATAACAAAATCAGCTATATCTTCTACTG
TTAAATCTACATTTATATTTAAAATTTTTTCAACAAGCATTTTTTTATTTTTAG
ATTTCTTTTTATCACCACCAACATTAAGATAAAATTTTACAAAACCCAGAATT
TCTAAATTACTTTTTATTTTTTCTCTTATTTCCATAAAATTAGTCAAAATAACA
TCTATTTTATCATCTTTCAATAATTTTTCTCTTAAATGTTCTTCATAATATCGA
TTTTCAAATACTTTTTCTGTTTCATTTTCAATTATTTTTTCTATAATCTTATATA
AACTCATGTTAATATTTTTAAAAATTTCGTAAATTGATTTTTTTGTTTCTAATT
CATCATTTTCTATTATTCTTAATATTATTGAACAATCATTTAGTGTTTTATTAG
TATACTCATCTCTGATATCTATCTCTATTTCTTCTTCATTCTCTTGTCTCTTTAT
TTCTATTTTTTTTATCATCTTTAGTTATTCCTTGCCTAATTGCTTCATCTATTATT
TTCTTTTTTGTAATCCCCAATGCTTTCAATTTCTCAGATTTTCCATATGCTTCT
ATATATAATACAACTTCTTCTGTTTCCAAAAAATCATCATTATTTTCTATTCTT
ATGATTCCTTCTTTACCTTTCAACTTAAATAGAATATTTCCTGCATGAAATTTT
CTTGTAAATTCTTTAAGAATATTATCATTTTTTTTGTAATTAATATATTTTCTA
ATAAATTTATTATTATCAATTTTTTCTTTATTATTATTTTCATTAATATTTAAA
ATGTATTTGTTTCCATCATAGTTCCTTTTAACTTTTACTTTCCGTTTTATTTTAA
AATCTTTTTTATCACGAACTTCATACCATCTCTTATGTCCAAATAAATTTCCC
ATTCCAATCTCCTCGTTTCTACTTTAATCTAATAAAATATTTTTAAATTAAATC
AATTTTACATCTTTCTAATCAAAAATACAATTTTCCATTTTTAGTATACCACA
TCAATATTAAATCTCAAAAAAATAAGGAGCCGTCAAACATAGCTCCCTACTT
CTATTTACTCATAATCCCCATCTATCCTTACTTTTCGTAAAATCAATCCTTCTT

TCGCCTTTAGATCCAACTTAATTTTCCCATTTGAACCTGTTCTAAATGTTCTGC
CTTCTGTTACCAAATCAATAAATCTTTCATCCTGATAATTTGTTTCAAATTCC
ACATTTTCCCAGCTGTTAAACGAATTATTTATTACAACAATAATTAAATGATC
CTCGATTACTCTTTCATACACAATTATTT (SEQ ID NO: 67)

Ejemplo 3. Una evaluación más profunda de Cpf1 y los componentes asociados

Los solicitantes llevaron a cabo alineamientos de secuencias con ortólogos de Cas-Cpf1 y compararon la estructura y la organización de los dominios (figuras 38A-N). En la figura 39 se provee una visión general del alineamiento de los loci relacionados con Cpf1.

Las secuencias de los loci relacionados con Cpf1 en los diversos ortólogos se enumeran a continuación.

> KKP36646_(con modificaciones), proteína hipotética UR27_C0015G0004 [la bacteria Peregrinibacteria GW2011_GWA2_33_10]

MSNFFKNFTNLYELSKTLRFELKPVGDTLTNMKDHLEYDEKLQTFLK
DQNIDDAYQALKPQFDEIHEEFITDSLESKKAKEIDFSEYLDLFQEKKELNDSEKK
LRNKIGETFNKAGEKWKKEKYPQYEWKKGSKIANGADILSCQDMLQFIKYKNP
EDEKIKNYIDDTLKGFFTYFGGFNQNRANYYETKKEASTAVATRIVHENLPKFC
DNVIQFKHIIKRKKDGTVEKTERKTEYLNAYQYLKNNNKITQIKDAETEKMIEST
PIAEKIFDVYYFSSCLSQKQIEEYNRIIGHYNLLINLYNQAKRSEGKHLSANEKKY
KDLPKFKTLYKQIGCGKKKDLFYTIKCDTEEEANKSRNEGKESHSVEEIINKAQE
AINKYFKSNNDCENINTVPDFINYILTKENYEGVYWSKAAMNTISDKYFANYHD
LQDRLKEAKVFQKADKKSEDDIKIPEAIELSGLFGVLDSLADWQTTLFKSSILSNE
DKLKIITDSQTPSEALLKMIFNDIEKNMESFLKETNDIITLKKYKGNKEGTEKIKQ
WFDYTLAINRMLKYFLVKENKIKGNSLDTNISEALKTLIYSDDAEWFKWYDALR
NYLTQKPQDEAKENKLKLNFDNPSLAGGWDVNKECSNFCVILKDKNEKKYLAI
MKKGENTLFQKEWTEGRGKNLTKKSNPLFEINNCEILSKMEYDFWADVSKMIP
KCSTQLKAVVNHFKQSDNEFIFPIGYKVTSGEKFREECKISKQDFELNNKVFNKN
ELSVTAMRYDLSSTQEKQYIKAFQKEYWELLFKQEKRDTKLTNNEIFNEWINFC
NKKYSELLSWERKYKDALTNWINFCKYFLSKYPKTTLFNYSFKESENYNSLDEF
YRDVDICSYKLNINTTINKSILDRLVEEGKLYLFEIKNQDSNDGKSIGHKNNLHTI
YWNAIFENFDNRPKLNGEAEIFYRKAISKDKLGIVKGKKTKNGTEHKNYRFSKE
KFILHVPITLNFCSNNEYVNDIVNTKFYNFSNLHFLGIDRGEKHLAYYSLVNKNG
EIVDQGTLNLPFTDKDGNQRSIKKEKYFYNKQEDKWEAKEVDCWNYNDLLDA
MASNRDMARKNWQRIGTIKEAKNGYVSLVIRKIADLAVNNERPAFIVLEDLNTG
FKRSRQKIDKSVYQKFELALAKKLNFLVDKNAKRDEIGSPTKALQLTPPVNNYG

DIENKKQAGIMLYTRANYTSQTDPATGWRKTIYLKAGPEETTYKKDGKIKNKSV
KDQIIETFTDIGFDGKDYYFEYDKGEFVDEKTGEIKPKKWRLYSGENGKSLDRFR
GEREKDKYEWKIDKIDIVKILDDLFVNFDKNISLLKQLKEGVELTRNNEHGTGES
LRFAINLIQQIRNTGNNERDNDFILSPVRDENGKHFDSREYWDKETKGEKISMPS
SGDANGAFNIARKGIIMNAHILANSDSKDLSLFVSDEEWDLHLNNKTEWKKQLN
IFSSRKAMAKRKK (SEQ ID NO: 68)

> KKR91555_(con modificaciones), proteína hipotética UU43_C0004G0003 [la bacteria Parcubacteria (Falkowbacteria) GW2011_GWA2_41_14]

MLFFMSTDITNKPREKGVFDNFTNLYEFSKTLTFGLIPLKWDDNKKMI
VEDEDFSVLRKYGVIEEDKRIAESIKIAKFYLNILHRELIGKVLGSLKFEKKNLEN
YDRLLGEIEKNNKNENISEDKKKEIRKNFKKELSIAQDILLKKVGEVFESNGSGIL
SSKNCLDELTKRFTRQEVDKLRRENKDIGVEYPDVAYREKDGKEETKSFFAMD
VGYLDDFHKNRKQLYSVKGKKNSLGRRILDNFEIFCKNKKLYEKYKNLDIDFSEI
ERNFNLTLEKVFDFDNYNERLTQEGLDEYAKILGGESNKQERTANIHGLNQIINL
YIQKKQSEQKAEQKETGKKKIKFNKKDYPTFTCLQKQILSQVFRKEIIIESDRDLI
RELKFFVEESKEKVDKARGIIEFLLNHEENDIDLAMVYLPKSKINSFVYKVFKEP
QDFLSVFQDGASNLDFVSFDKIKTHLENNKLTYKIFFKTLIKENHDFESFLILLQQ
EIDLLIDGGETVTLGGKKESITSLDEKKNRLKEKLGWFEGKVRENEKMKDEEEG
EFCSTVLAYSQAVLNITKRAEIFWLNEKQDAKVGEDNKDMIFYKKFDEFADDGF
APFFYFDKFGNYLKRRSRNTTKEIKLHFGNDDLLEGWDMNKEPEYWSFILRDRN
QYYLGIGKKDGEIFHKKLGNSVEAVKEAYELENEADFYEKIDYKQLNIDRFEGIA
FPKKTKTEEAFRQVCKKRADEFLGGDTYEFKILLAIKKEYDDFKARRQKEKDW
DSKFSKEKMSKLIEYYITCLGKRDDWKRFNLNFRQPKEYEDRSDFVRHIQRQAY
WIDPRKVSKDYVDKKVAEGEMFLFKVHNKDFYDFERKSEDKKNHTANLFTQY
LLELFSCENIKNIKSKDLIESIFELDGKAEIRFRPKTDDVKLKIYQKKGKDVTYAD
KRDGNKEKEVIQHRRFAKDALTLHLKIRLNFGKHVNLFDFNKLVNTELFAKVPV
KILGMDRGENNLIYYCFLDEHGEIENGKCGSLNRVGEQIITLEDDKKVKEPVDYF
QLLVDREGQRDWEQKNWQKMTRIKDLKKAYLGNVVSWISKEMLSGIKEGVVT
IGVLEDLNSNFKRTRFFRERQVYQGFEKALVNKLGYLVDKKYDNYRNVYQFAPI
VDSVEEMEKNKQIGTLVYVPASYTSKICPHPKCGWRERLYMKNSASKEKIVGLL
KSDGIKISYDQKNDRFYFEYQWEQEHKSDGKKKKYSGVDKVFSNVSRMRWDV
EQKKSIDFVDGTDGSITNKLKSLLKGKGIELDNINQQIVNQQKELGVEFFQSIIFYF
NLIMQIRNYDKEKSGSEADYIQCPSCLFDSRKPEMNGKLSAITNGDANGAYNIAR

KGFMQLCRIRENPQEPMKLITNREWDEAVREWDIYSAAQKIPVLSEEN (SEQ ID NO: 69)

> KDN25524_(con modificaciones), proteína hipotética MBO_03467 [Moraxella bovoculi 237]

MLFQDFTHLYPLSKTVRFELKPIDRTLEHIHAKNFLSQDETMADMHQ
KVKVILDDYHRDFIADMMGEVKLTKLAEFYDVYLKFRKNPKDDELQKQLKDL
QAVLRKEIVKPIGNGGKYKAGYDRLFGAKLFKDGKELGDLAKFVIAQEGESSPK
LAHLAHFEKFSTYFTGFHDNRKNMYSDEDKHTAIAYRLIHENLPRFIDNLQILTTI
KQKHSALYDQIINELTASGLDVSLASHLDGYHKLLTQEGITAYNTLLGGISGEAG
SPKIQGINELINSHHNQHCHKSERIAKLRPLHKQILSDGMSVSFLPSKFADDSEMC
QAVNEFYRHYADVFAKVQSLFDGFDDHQKDGIYVEHKNLNELSKQAFGDFALL
GRVLDGYYVDVVNPEFNERFAKAKTDNAKAKLTKEKDKFIKGVHSLASLEQAI
EHYTARHDDESVQAGKLGQYFKHGLAGVDNPIQKIHNNHSTIKGFLERERPAGE
RALPKIKSGKNPEMTQLRQLKELLDNALNVAHFAKLLTTKTTLDNQDGNFYGEF
GVLYDELAKIPTLYNKVRDYLSQKPFSTEKYKLNFGNPTLLNGWDLNKEKDNF
GVILQKDGCYYLALLDKAHKKVFDNAPNTGKSIYQKMIYKYLEVRKQFPKVFFS
KEAIAINYHPSKELVEIKDKGRQRSDDERLKLYRFILECLKIHPKYDKKFEGAIGD
IQLFKKDKKGREVPISEKDLFDKINGIFSSKPKLEMEDFFIGEFKRYNPSQDLVDQ
YNIYKKIDSNDNRKKENFYNNHPKFKKDLVRYYYESMCKHEEWEESFEFSKKL
QDIGCYVDVNELFTEIETRRLNYKISFCNINADYIDELVEQGQLYLFQIYNKDFSP
KAHGKPNLHTLYFKALFSEDNLADPIYKLNGEAQIFYRKASLDMNETTIHRAGE
VLENKNPDNPKKRQFVYDIIKDKRYTQDKFMLHVPITMNFGVQGMTIKEFNKK
VNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILEQCSLNDITTASANGTQMTT
PYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVVHQISQLMLKYNAIVVLE
DLNFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDKADDEIGSYKNALQLTNN
FTDLKSIGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAFFGKFDKIC
YNADKDYFEFHIDYAKFTDKAKNSRQIWTICSHGDKRYVYDKTANQNKGAAK
GINVNDELKSLFARHHINEKQPNLVMDICQNNDKEFHKSLMYLLKTLLALRYSN
ASSDEDFILSPVANDEGVFFNSALADDTQPQNADANGAYHIALKGLWLLNELKN
SDDLNKVKLAIDNQTWLNFAQNR (SEQ ID NO: 70)

> KKT48220_(con modificaciones), proteína hipotética UW39_C0001G0044 [la bacteria Parcubacteria GW2011_GWC2_44_17]

MENIFDQFIGKYSLSKTLRFELKPVGKTEDFLKINKVFEKDQTIDDSYN
QAKFYFDSLHQKFIDAALASDKTSELSFQNFADVLEKQNKIILDKKREMGALRK
RDKNAVGIDRLQKEINDAEDIIQKEKEKIYKDVRTLFDNEAESWKTYYQEREVD
GKKITFSKADLKQKGADFLTAAGILKVLKYEFPEEKEKEFQAKNQPSLFVEEKEN
PGQKRYIFDSFDKFAGYLTKFQQTKKNLYAADGTSTAVATRIADNFIIFHQNTKV
FRDKYKNNHTDLGFDEENIFEIERYKNCLLQREIEHIKNENSYNKIIGRINKKIKEY
RDQKAKDTKLTKSDFPFFKNLDKQILGEVEKEKQLIEKTREKTEEDVLIERFKEFI
ENNEERFTAAKKLMNAFCNGEFESEYEGIYLKNKAINTISRRWFVSDRDFELKLP
QQKSKNKSEKNEPKVKKFISIAEIKNAVEELDGDIFKAVFYDKKIIAQGGSKLEQF
LVIWKYEFEYLFRDIERENGEKLLGYDSCLKIAKQLGIFPQEKEAREKATAVIKN
YADAGLGIFQMMKYFSLDDKDRKNTPGQLSTNFYAEYDGYYKDFEFIKYYNEF
RNFITKKPFDEDKIKLNFENGALLKGWDENKEYDFMGVILKKEGRLYLGIMHKN
HRKLFQSMGNAKGDNANRYQKMIYKQIADASKDVPRLLLTSKKAMEKFKPSQE
ILRIKKEKTFKRESKNFSLRDLHALIEYYRNCIPQYSNWSFYDFQFQDTGKYQNIK
EFTDDVQKYGYKISFRDIDDEYINQALNEGKMYLFEVVNKDIYNTKNGSKNLHT
LYFEHILSAENLNDPVFKLSGMAEIFQRQPSVNEREKITTQKNQCILDKGDRAYK
YRRYTEKKIMFHMSLVLNTGKGEIKQVQFNKIINQRISSSDNEMRVNVIGIDRGE
KNLLYYSVVKQNGEIIEQASLNEINGVNYRDKLIEREKERLKNRQSWKPVVKIK
DLKKGYISHVIHKICQLIEKYSAIVVLEDLNMRFKQIRGGIERSVYQQFEKALIDK
LGYLVFKDNRDLRAPGGVLNGYQLSAPFVSFEKMRKQTGILFYTQAEYTSKTDP
ITGFRKNVYISNSASLDKIKEAVKKFDAIGWDGKEQSYFFKYNPYNLADEKYKN
STVSKEWAIFASAPRIRRQKGEDGYWKYDRVKVNEEFEKLLKVWNFVNPKATD
IKQEIIKKEKAGDLQGEKELDGRLRNFWHSFIYLFNLVLELRNSFSLQIKIKAGEVI
AVDEGVDFIASPVKPFFTTPNPYIPSNLCWLAVENADANGAYNIARKGVMILKKI
REHAKKDPEFKKLPNLFISNAEWDEAARDWGKYAGTTALNLDH (SEQ ID NO:
71)

> WP_031492824_(con modificaciones), proteína hipotética [Succinivibrio dextrinosolvens]

MSSLTKFTNKYSKQLTIKNELIPVGKTLENIKENGLIDGDEQLNENYQ
KAKIIVDDFLRDFINKALNNTQIGNWRELADALNKEDEDNIEKLQDKIRGIIVSKF
ETFDLFSSYSIKKDEKIIDDDNDVEEEELDLGKKTSSFKYIFKKNLFKLVLPSYLK
TTNQDKLKIISSFDNFSTYFRGFFENRKNIFTKKPISTSIAYRIVHDNFPKFLDNIRC
FNVWQTECPQLIVKADNYLKSKNVIAKDKSLANYFTVGAYDYFLSQNGIDFYN

NIIGGLPAFAGHEKIQGLNEFINQECQKDSELKSKLKNRHAFKMAVLFKQILSDR
EKSFVIDEFESDAQVIDAVKNFYAEQCKDNNVIFNLLNLIKNIAFLSDDELDGIFIE
GKYLSSVSQKLYSDWSKLRNDIEDSANSKQGNKELAKKIKTNKGDVEKAISKYE
FSLSELNSIVHDNTKFSDLLSCTLHKVASEKLVKVNEGDWPKHLKNNEEKQKIK
EPLDALLEIYNTLLIFNCKSFNKNGNFYVDYDRCINELSSVVYLYNKTRNYCTKK
PYNTDKFKLNFNSPQLGEGFSKSKENDCLTLLFKKDDNYYVGIIRKGAKINFDDT
QAIADNTDNCIFKMNYFLLKDAKKFIPKCSIQLKEVKAHFKKSEDDYILSDKEKF
ASPLVIKKSTFLLATAHVKGKKGNIKKFQKEYSKENPTEYRNSLNEWIAFCKEFL
KTYKAATIFDITTLKKAEEYADIVEFYKDVDNLCYKLEFCPIKTSFIENLIDNGDL
YLFRINNKDFSSKSTGTKNLHTLYLQAIFDERNLNNPTIMLNGGAELFYRKESIEQ
KNRITHKAGSILVNKVCKDGTSLDDKIRNEIYQYENKFIDTLSDEAKKVLPNVIK
KEATHDITKDKRFTSDKFFFHCPLTINYKEGDTKQFNNEVLSFLRGNPDINIIGIDR
GERNLIYVTVINQKGEILDSVSFNTVTNKSSKIEQTVDYEEKLAVREKERIEAKRS
WDSISKIATLKEGYLSAIVHEICLLMIKHNAIVVLENLNAGFKRIRGGLSEKSVYQ
KFEKMLINKLNYFVSKKESDWNKPSGLLNGLQLSDQFESFEKLGIQSGFIFYVPA
AYTSKIDPTTGFANVLNLSKVRNVDAIKSFFSNFNEISYSKKEALFKFSFDLDSLS
KKGFSSFVKFSKSKWNVYTFGERIIKPKNKQGYREDKRINLTFEMKKLLNEYKV
SFDLENNLIPNLTSANLKDTFWKELFFIFKTTLQLRNSVTNGKEDVLISPVKNAK
GEFFVSGTHNKTLPQDCDANGAYHIALKGLMILERNNLVREEKDTKKIMAISNV
DWFEYVQKRRGVL (SEQ ID NO: 72)

> KKT50231_(con modificaciones), proteína hipotética UW40_C0007G0006 [la bacteria Parcubacteria GW2011_GWF2_44_17]

MKPVGKTEDFLKINKVFEKDQTIDDSYNQAKFYFDSLHQKFIDAALA
SDKTSELSFQNFADVLEKQNKIILDKKREMGALRKRDKNAVGIDRLQKEINDAE
DIIQKEKEKIYKDVRTLFDNEAESWKTYYQEREVDGKKITFSKADLKQKGADFL
TAAGILKVLKYEFPEEKEKEFQAKNQPSLFVEEKENPGQKRYIFDSFDKFAGYLT
KFQQTKKNLYAADGTSTAVATRIADNFIIFHQNTKVFRDKYKNNHTDLGFDEEN
IFEIERYKNCLLQREIEHIKNENSYNKIIGRINKKIKEYRDQKAKDTKLTKSDFPFF
KNLDKQILGEVEKEKQLIEKTREKTEEDVLIERFKEFIENNEERFTAAKKLMNAF
CNGEFESEYEGIYLKNKAINTISRRWFVSDRDFELKLPQQKSKNKSEKNEPKVKK
FISIAEIKNAVEELDGDIFKAVFYDKKIIAQGGSKLEQFLVIWKYEFEYLFRDIERE
NGEKLLGYDSCLKIAKQLGIFPQEKEAREKATAVIKNYADAGLGIFQMMKYFSL
DDKDRKNTPGQLSTNFYAEYDGYYKDFEFIKYYNEFRNFITKKPFDEDKIKLNFE

NGALLKGWDENKEYDFMGVILKKEGRLYLGIMHKNHRKLFQSMGNAKGDNA
NRYQKMIYKQIADASKDVPRLLLTSKKAMEKFKPSQEILRIKKEKTFKRESKNFS
LRDLHALIEYYRNCIPQYSNWSFYDFQFQDTGKYQNIKEFTDDVQKYGYKISFR
DIDDEYINQALNEGKMYLFEVVNKDIYNTKNGSKNLHTLYFEHILSAENLNDPV
FKLSGMAEIFQRQPSVNEREKITTQKNQCILDKGDRAYKYRRYTEKKIMFHMSL
VLNTGKGEIKQVQFNKIINQRISSSDNEMRVNVIGIDRGEKNLLYYSVVKQNGEII
EQASLNEINGVNYRDKLIEREKERLKNRQSWKPVVKIKDLKKGYISHVIHKICQL
IEKYSAIVVLEDLNMRFKQIRGGIERSVYQQFEKALIDKLGYLVFKDNRDLRAPG
GVLNGYQLSAPFVSFEKMRKQTGILFYTQAEYTSKTDPITGFRKNVYISNSASLD
KIKEAVKKFDAIGWDGKEQSYFFKYNPYNLADEKYKNSTVSKEWAIFASAPRIR
RQKGEDGYWKYDRVKVNEEFEKLLKVWNFVNPKATDIKQEIIKKEKAGDLQGE
KELDGRLRNFWHSFIYLFNLVLELRNSFSLQIKIKAGEVIAVDEGVDFIASPVKPF
FTTPNPYIPSNLCWLAVENADANGAYNIARKGVMILKKIREHAKKDPEFKKLPN
LFISNAEWDEAARDWGKYAGTTALNLDH (SEQ ID NO: 73)

> WP_004356401_(con modificaciones), proteína hipotética [Prevotella disiens]

MENYQEFTNLFQLNKTLRFELKPIGKTCELLEEGKIFASGSFLEKDKV
RADNVSYVKKEIDKKHKIFIEETLSSFSISNDLLKQYFDCYNELKAFKKDCKSDE
EEVKKTALRNKCTSIQRAMREAISQAFLKSPQKKLLAIKNLIENVFKADENVQHF
SEFTSYFSGFETNRENFYSDEEKSTSIAYRLVHDNLPIFIKNIYIFEKLKEQFDAKT
LSEIFENYKLYVAGSSLDEVFSLEYFNNTLTQKGIDNYNAVIGKIVKEDKQEIQGL
NEHINLYNQKHKDRRLPFFISLKKQILSDREALSWLPDMFKNDSEVIKALKGFYI
EDGFENNVLTPLATLLSSLDKYNLNGIFIRNNEALSSLSQNVYRNFSIDEAIDANA
ELQTFNNYELIANALRAKIKKETKQGRKSFEKYEEYIDKKVKAIDSLSIQEINELV
ENYVSEFNSNSGNMPRKVEDYFSLMRKGDFGSNDLIENIKTKLSAAEKLLGTKY
QETAKDIFKKDENSKLIKELLDATKQFQHFIKPLLGTGEEADRDLVFYGDFLPLY
EKFEELTLLYNKVRNRLTQKPYSKDKIRLCFNKPKLMTGWVDSKTEKSDNGTQ
YGGYLFRKKNEIGEYDYFLGISSKAQLFRKNEAVIGDYERLDYYQPKANTIYGS
AYEGENSYKEDKKRLNKVIIAYIEQIKQTNIKKSIIESISKYPNISDDDKVTPSSLLE
KIKKVSIDSYNGILSFKSFQSVNKEVIDNLLKTISPLKNKAEFLDLINKDYQIFTEV
QAVIDEICKQKTFIYFPISNVELEKEMGDKDKPLCLFQISNKDLSFAKTFSANLRK
KRGAENLHTMLFKALMEGNQDNLDLGSGAIFYRAKSLDGNKPTHPANEAIKCR
NVANKDKVSLFTYDIYKNRRYMENKFLFHLSIVQNYKAANDSAQLNSSATEYIR
KADDLHIIGIDRGERNLLYYSVIDMKGNIVEQDSLNIIRNNDLETDYHDLLDKRE

KERKANRQNWEAVEGIKDLKKGYLSQAVHQIAQLMLKYNAIIALEDLGQMFVT
RGQKIEKAVYQQFEKSLVDKLSYLVDKKRPYNELGGILKAYQLASSITKNNSDK
QNGFLFYVPAWNTSKIDPVTGFTDLLRPKAMTIKEAQDFFGAFDNISYNDKGYF
EFETNYDKFKIRMKSAQTRWTICTFGNRIKRKKDKNYWNYEEVELTEEFKKLFK
DSNIDYENCNLKEEIQNKDNRKFFDDLIKLLQLTLQMRNSDDKGNDYIISPVANA
EGQFFDSRNGDKKLPLDADANGAYNIARKGLWNIRQIKQTKNDKKLNLSISSTE
WLDFVREKPYLK (SEQ ID NO: 74)

> CCB70584_(con modificaciones), proteína con una función desconocida [Flavobacterium branchiophilum FL-15]

MTNKFTNQYSLSKTLRFELIPQGKTLEFIQEKGLLSQDKQRAESYQEM
KKTIDKFHKYFIDLALSNAKLTHLETYLELYNKSAETKKEQKFKDDLKKVQDNL
RKEIVKSFSDGDAKSIFAILDKKELITVELEKWFENNEQKDIYFDEKFKTFTTYFT
GFHQNRKNMYSVEPNSTAIAYRLIHENLPKFLENAKAFEKIKQVESLQVNFREL
MGEFGDEGLIFVNELEEMFQINYYNDVLSQNGITIYNSIISGFTKNDIKYKGLNEY
INNYNQTKDKKDRLPKLKQLYKQILSDRISLSFLPDAFTDGKQVLKAIFDFYKINL
LSYTIEGQEESQNLLLLIRQTIENLSSFDTQKIYLKNDTHLTTISQQVFGDFSVFST
ALNYWYETKVNPKFETEYSKANEKKREILDKAKAVFTKQDYFSIAFLQEVLSEY
ILTLDHTSDIVKKHSSNCIADYFKNHFVAKKENETDKTFDFIANITAKYQCIQGIL
ENADQYEDELKQDQKLIDNLKFFLDAILELLHFIKPLHLKSESITEKDTAFYDVFE
NYYEALSLLTPLYNMVRNYVTQKPYSTEKIKLNFENAQLLNGWDANKEGDYLT
TILKKDGNYFLAIMDKKHNKAFQKFPEGKENYEKMVYKLLPGVNKMLPKVFFS
NKNIAYFNPSKELLENYKKETHKKGDTFNLEHCHTLIDFFKDSLNKHEDWKYFD
FQFSETKSYQDLSGFYREVEHQGYKINFKNIDSEYIDGLVNEGKLFLFQIYSKDFS
PFSKGKPNMHTLYWKALFEEQNLQNVIYKLNGQAEIFFRKASIKPKNIILHKKKI
KIAKKHFIDKKTKTSEIVPVQTIKNLNMYYQGKISEKELTQDDLRYIDNFSIFNEK
NKTIDIIKDKRFTVDKFQFHVPITMNFKATGGSYINQTVLEYLQNNPEVKIIGLDR
GERHLVYLTLIDQQGNILKQESLNTITDSKISTPYHKLLDNKENERDLARKNWGT
VENIKELKEGYISQVVHKIATLMLEENAIVVMEDLNFGFKRGRFKVEKQIYQKLE
KMLIDKLNYLVLKDKQPQELGGLYNALQLTNKFESFQKMGKQSGFLFYVPAWN
TSKIDPTTGFVNYFYTKYENVDKAKAFFEKFEAIRFNAEKKYFEFEVKKYSDFNP
KAEGTQQAWTICTYGERIETKRQKDQNNKFVSTPINLTEKIEDFLGKNQIVYGDG
NCIKSQIASKDDKAFFETLLYWFKMTLQMRNSETRTDIDYLISPVMNDNGTFYN
SRDYEKLENPTLPKDADANGAYHIAKKGLMLLNKIDQADLTKKVDLSISNRDW
LQFVQKNK (SEQ ID NO: 75)

> WP_005398606_(con modificaciones), proteína hipotética [Helcococcus kunzii]

MFEKLSNIVSISKTIRFKLIPVGKTLENIEKLGKLEKDFERSDFYPILKNI
SDDYYRQYIKEKLSDLNLDWQKLYDAHELLDSSKKESQKNLEMIQAQYRKVLF
NILSGELDKSGEKNSKDLIKNNKALYGKLFKKQFILEVLPDFVNNNDSYSEEDLE
GLNLYSKFTTRLKNFWETRKNVFTDKDIVTAIPFRAVNENFGFYYDNIKIFNKNI
EYLENKIPNLENELKEADILDDNRSVKDYFTPNGFNYVITQDGIDVYQAIRGGFT
KENGEKVQGINEILNLTQQQLRRKPETKNVKLGVLTKLRKQILEYSESTSFLIDQI
EDDNDLVDRINKFNVSFFESTEVSPSLFEQIERLYNALKSIKKEEVYIDARNTQKF
SQMLFGQWDVIRRGYTVKITEGSKEEKKKYKEYLELDETSKAKRYLNIREIEELV
NLVEGFEEVDVFSVLLEKFKMNNIERSEFEAPIYGSPIKLEAIKEYLEKHLEEYHK
WKLLLIGNDDLDTDETFYPLLNEVISDYYIIPLYNLTRNYLTRKHSDKDKIKVNF
DFPTLADGWSESKISDNRSIILRKGGYYYLGILIDNKLLINKKNKSKKIYEILIYNQ
IPEFSKSIPNYPFTKKVKEHFKNNVSDFQLIDGYVSPLIITKEIYDIKKEKKYKKDF
YKDNNTNKNYLYTIYKWIEFCKQFLYKYKGPNKESYKEMYDFSTLKDTSLYVN
LNDFYADVNSCAYRVLFNKIDENTIDNAVEDGKLLLFQIYNKDFSPESKGKKNL
HTLYWLSMFSEENLRTRKLKLNGQAEIFYRKKLEKKPIIHKEGSILLNKIDKEGN
TIPENIYHECYRYLNKKIGREDLSDEAIALFNKDVLKYKEARFDIIKDRRYSESQF
FFHVPITFNWDIKTNKNVNQIVQGMIKDGEIKHIIGIDRGERHLLYYSVIDLEGNI
VEQGSLNTLEQNRFDNSTVKVDYQNKLRTREEDRDRARKNWTNINKIKELKDG
YLSHVVHKLSRLIIKYEAIVIMENLNQGFKRGRFKVERQVYQKFELALMNKLSA
LSFKEKYDERKNLEPSGILNPIQACYPVDAYQELQGQNGIVFYLPAAYTSVIDPV
TGFTNLFRLKSINSSKYEEFIKKFKNIYFDNEEEDFKFIFNYKDFAKANLVILNNIK
SKDWKISTRGERISYNSKKKEYFYVQPTEFLINKLKELNIDYENIDIIPLIDNLEEK
AKRKILKALFDTFKYSVQLRNYDFENDYIISPTADDNGNYYNSNEIDIDKTNLPN
NGDANGAFNIARKGLLLKDRIVNSNESKVDLKIKNEDWINFIIS (SEQ ID NO: 76)

> WP_021736722_proteína Cpf1 asociada a los arreglos CRISPR (con modificaciones), subtipo PREFRAN [Acidaminococcus sp. BV3L6]

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYK
ELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYR
NAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENAL
LRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAV

PSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTE
KIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVI
QSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLR
NALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFS
ARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEK
NNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAK
MIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYA
KKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAEL
NPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGL
FSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTL
YQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLN
YQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTI
QQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQA
VVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLN
PYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKH
FLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAK
GTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLEND
DSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM
DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN (SEQ ID
NO: 77)

> WP_004339290_(con modificaciones), proteína hipotética [Francisella tularensis]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYK
KAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDT
IKKQISKYINDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDIT
DIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAK
YESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNY
LNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQI
LSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQVAPKNLDNPSKKEQ
DLIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILSNFAAIPMIFDEIAQN
KDNLAQISIKYQNQGKKDLLQASAEEDVKAIKDLLDQTNNLLHRLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLA

SGWDKNKESANTAILFIKDDKYYLGIMDKKHNKIFSDKAIEENKGEGYKKIVYK
QIADASKDIQNLMIIDGKTVCKKGRKDRNGVNRQLLSLKRKHLPENIYRIKETKS
YLKNEARFSRKDLYDFIDYYKDRLDYYDFEFELKPSNEYSDFNDFTNHIGSQGY
KLTFENISQDYINSLVNEGKLYLFQIYSKDFSAYSKGRPNLHTLYWKALFDERNL
QDVVYKLNGEAELFYRKQSIPKKITHPAKETIANKNKDNPKKESVFEYDLIKDKR
FTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTL
VDGKGNIIKQDNFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKE
GYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLN
YLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGF
VNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIA
SFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESD
KKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLDRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID
NO: 78)

> WP_022501477_(con modificaciones), proteína hipotética [Eubacterium sp. CAG: 76]

MNKAADNYTGGNYDEFIALSKVQKTLRNELKPTPFTAEHIKQRGIISE
DEYRAQQSLELKKIADEYYRNYITHKLNDINNLDFYNLFDAIEEKYKKNDKDNR
DKLDLVEKSKRGEIAKMLSADDNFKSMFEAKLITKLLPDYVERNYTGEDKEKAL
ETLALFKGFTTYFKGYFKTRKNMFSGEGGASSICHRIVNVNASIFYDNLKTFMRI
QEKAGDEIALIEEELTEKLDGWRLEHIFSRDYYNEVLAQKGIDYYNQICGDINKH
MNLYCQQNKFKANIFKMMKIQKQIMGISEKAFEIPPMYQNDEEVYASFNEFISRL
EEVKLTDRLINILQNINIYNTAKIYINARYYTNVSSYVYGGWGVIDSAIERYLYNT
IAGKGQSKVKKIENAKKDNKFMSVKELDSIVAEYEPDYFNAPYIDDDDNAVKAF
GGQGVLGYFNKMSELLADVSLYTIDYNSDDSLIENKESALRIKKQLDDIMSLYH
WLQTFIIDEVVEKDNAFYAELEDICCELENVVTLYDRIRNYVTKKPYSTQKFKLN
FASPTLAAGWSRSKEFDNNAIILLRNNKYYIAIFNVNNKPDKQIIKGSEEQRLSTD
YKKMVYNLLPGPNKMLPKVFIKSDTGKRDYNPSSYILEGYEKNRHIKSSGNFDIN
YCHDLIDYYKACINKHPEWKNYGFKFKETNQYNDIGQFYKDVEKQGYSISWAY
ISEEDINKLDEEGKIYLFEIYNKDLSAHSTGRDNLHTMYLKNIFSEDNLKNICIELN
GEAELFYRKSSMKSNITHKKDTILVNKTYINETGVRVSLSDEDYMKVYNYYNN
NYVIDTENDKNLIDIIEKIGHRKSKIDIVKDKRYTEDKYFLYLPITINYGIEDENVN
SKIIEYIAKQDNMNVIGIDRGERNLIYISVIDNKGNIIEQKSFNLVNNYDYKNKLK
NMEKTRDNARKNWQEIGKIKDVKSGYLSGVISKIARMVIDYNAIIVMEDLNKGF

KRGRFKVERQVYQKFENMLISKLNYLVFKERKADENGGILRGYQLTYIPKSIKN
VGKQCGCIFYVPAAYTSKIDPATGFINIFDFKKYSGSGINAKVKDKKEFLMSMNS
IRYINECSEEYEKIGHRELFAFSFDYNNFKTYNVSSPVNEWTAYTYGERIKKLYK
DGRWLRSEVLNLTENLIKLMEQYNIEYKDGHDIREDISHMDETRNADFICSLFEE
LKYTVQLRNSKSEAEDENYDRLVSPILNSSNGFYDSSDYMENENNTTHTMPKDA
DANGAYCIALKGLYEINKIKQNWSDDKKFKENELYINVTEWLDYIQNRRFE
(SEQ ID NO: 79)

> WP_014550095_(con modificaciones), proteína hipotética [Francisella tularensis]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYK
KAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDT
IKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDIT
DIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAK
YESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNY
LNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQI
LSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQ
KLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQVAPKNLDNPSKKEQ
DLIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQN
KDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHRLKIFHISQSED
KANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLA
NGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVY
KLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGNPQKGYEKFEFNIE
DCRKFIDFYKESISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISE
SYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKL
NGEAELFYRKKSIPKKITHPAKEAIANKNKDNPKKESFFEYDLIKDKRFTEDKFFF
HCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNII
KQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVV
HEIAKLVIEHNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDN
EFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPK
YESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLIN
FRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKL
TSILNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADANGAYH
IGLKGLMLLDRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO: 80)

> WP_003034647_(con modificaciones), proteína hipotética [Francisella tularensis]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKA
KQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKG
WTTYFKGFHENRKNVYSSDDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAV
LEYITQQVAPKNLDNPSKKEQDLIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEI
LANFAAIPMIFDEIAQNKDNLAQISLKYQNQGKKDLLQASAEEDVKAIKDLLDQTNNLL
HRLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKL
NFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKK
IVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGNPQKGYEKFEFNIEDC
RKFIDFYKESISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVV
NQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQS
IPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFN
DEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKL
AAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEHNAIVVFEDLNFGFKRGR
FKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIY
YVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGD
KAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAA
ICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQD
ADANGAYHIGLKGLMLLDRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO:
81)

> FnCpf1 Francisella tularensis subespecie novicida U112, genoma completo

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKD
YKKAKQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSA
KDTIKKQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANS
DITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLEN
KAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANF
NNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVL
FKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDD
LKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSK
KEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEI
AQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHIS
QSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFE

NSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGY
KKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEK
FEFNIEDCRKFIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLT
FENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQD
VVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFT
EDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLAYYTLVD
GKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGY
LSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYL
VFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFV
NQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAAKGKWTIAS
FGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDK
KFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDAD
ANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN (SEQ ID NO:
82)

> KKQ38174_(con modificaciones), proteína hipotética US54_C0016G0015 [la bacteria Microgenomates (Roizmanbacteria) GW2011_GWA2_37_7]

MKSFDSFTNLYSLSKTLKFEMRPVGNTQKMLDNAGVFEKDKLIQ
KKYGKTKPYFDRLHREFIEEALTGVELIGLDENFRTLVDWQKDKKNNVAMKAY
ENSLQRLRTEIGKIFNLKAEDWVKNKYPILGLKNKNTDILFEEAVFGILKARYGE
EKDTFIEVEEIDKTGKSKINQISIFDSWKGFTGYFKKFFETRKNFYKNDGTSTAIA
TRIIDQNLKRFIDNLSIVESVRQKVDLAETEKSFSISLSQFFSIDFYNKCLLQDGIDY
YNKIIGGETLKNGEKLIGLNELINQYRQNNKDQKIPFFKLLDKQILSEKILFLDEIK
NDTELIEALSQFAKTAEEKTKIVKKLFADFVENNSKYDLAQIYISQEAFNTISNK
WTSETETFAKYLFEAMKSGKLAKYEKKDNSYKFPDFIALSQMKSALLSISLEGHF
WKEKYYKISKFQEKTNWEQFLAIFLYEFNSLFSDKINTKDGETKQVGYYLFAKD
LHNLILSEQIDIPKDSKVTIKDFADSVLTIYQMAKYFAVEKKRAWLAEYELDSFY
TQPDTGYLQFYDNAYEDIVQVYNKLRNYLTKKPYSEEKWKLNFENSTLANGW
DKNKESDNSAVILQKGGKYYLGLITKGHNKIFDDRFQEKFIVGIEGGKYEKIVYK
FFPDQAKMFPKVCFSAKGLEFFRPSEEILRIYNNAEFKKGETYSIDSMQKLIDFYK
DCLTKYEGWACYTFRHLKPTEEYQNNIGEFFRDVAEDGYRIDFQGISDQYIHEK
NEKGELHLFEIHNKDWNLDKARDGKSKTTQKNLHTLYFESLFSNDNVVQNFPIK
LNGQAEIFYRPKTEKDKLESKKDKKGNKVIDHKRYSENKIFFHVPLTLNRTKND
SYRFNAQINNFLANNKDINIIGVDRGEKHLVYYSVITQASDILESGSLNELNGVN

YAEKLGKKAENREQARRDWQDVQGIKDLKKGYISQVVRKLADLAIKHNAIIILE
DLNMRFKQVRGGIEKSIYQQLEKALIDKLSFLVDKGEKNPEQAGHLLKAYQLSA
PFETFQKMGKQTGIIFYTQASYTSKSDPVTGWRPHLYLKYFSAKKAKDDIAKFT
KIEFVNDRFELTYDIKDFQQAKEYPNKTVWKVCSNVERFRWDKNLNQNKGGYT
HYTNITENIQELFTKYGIDITKDLLTQISTIDEKQNTSFFRDFIFYFNLICQIRNTDDS
EIAKKNGKDDFILSPVEPFFDSRKDNGNKLPENGDDNGAYNIARKGIVILNKISQ
YSEKNENCEKMKWGDLYVSNIDWDNFVTQANARH (SEQ ID NO: 83)

> WP_022097749_proteína hipotética (con modificaciones) [eubacteria seleccionada CAG: 72]

MNGNRSIVYREFVGVTPVAKTLRNELRPVGHTQEHIIQNGLIQEDE
LRQEKSTELKNIMDDYYREYIDKSLSGLTDLDFTLLFELMNSVQSSLSKDNKKAL
EKEHNKMREQICTHLQSDSDYKNMFNAKLFKEILPDFIKNYNQYDVKDKAGKL
ETLALFNGFSTYFTDFFEKRKNVFTKEAVSTSIAYRIVHENSLIFLANMTSYKKIS
EKALDEIEVIEKNNQDKMGDWELNQIFNPDFYNMVLIQSGIDFYNEICGVVNAH
MNLYCQQTKNNYNLFKMRKLHKQILAYTSTSFEVPKMFEDDMSVYNAVNAFID
ETEKGNIIGKLKDIVNKYDELDEKRIYISKDFYETLSCFMSGNWNLITGCVENFY
DENIHAKGKSKEEKVKKAVKEDKYKSINDVNDLVEKYIDEKERNEFKNSNAKQ
YIREISNIITDTETAHLEYDEHISLIESEEKADEIKKRLDMYMNMYHWVKAFIVDE
VLDRDEMFYSDIDDIYNILENIVPLYNRVRNYVTQKPYTSKKIKLNFQSPTLANG
WSQSKEFDNNAIILIRDNKYYLAIFNAKNKPDKKIIQGNSDKKNDNDYKKMVYN
LLPGANKMLPKVFLSKKGIETFKPSDYIISGYNAHKHIKTSENFDISFCRDLIDYFK
NSIEKHAEWRKYEFKFSATDSYNDISEFYREVEMQGYRIDWTYISEADINKLDEE
GKIYLFQIYNKDFAENSTGKENLHTMYFKNIFSEENLKNIVIKLNGQAELFYRKA
SVKNPVKHKKDSVLVNKTYKNQLDNGDVVRIPIPDDIYNEIYKMYNGYIKESDL
SEAAKEYLDKVEVRTAQKDIVKDYRYTVDKYFIHTPITINYKVTARNNVNDMA
VKYIAQNDDIHVIGIDRGERNLIYISVIDSHGNIVKQKSYNILNNYDYKKKLVEKE
KTREYARKNWKSIGNIKELKEGYISGVVHEIAMLMVEYNAIIAMEDLNYGFKRG
RFKVERQVYQKFESMLINKLNYFASKGKSVDEPGGLLKGYQLTYVPDNIKNLG
KQCGVIFYVPAAFTSKIDPSTGFISAFNFKSISTNASRKQFFMQFDEIRYCAEKDM
FSFGFDYNNFDTYNITMGKTQWTVYTNGERLQSEFNNARRTGKTKSINLTETIKL
LLEDNEINYADGHDVRIDMEKMYEDKNSEFFAQLLSLYKLTVQMRNSYTEAEE
QEKGISYDKIISPVINDEGEFFDSDNYKESDDKECKMPKDADANGAYCIALKGLY
EVLKIKSEWTEDGFDRNCLKLPHAEWLDFIQNKRYE (SEQ ID NO: 84)

> WP_012739647_proteína hipotética (con modificaciones) [eubacteria seleccionada]

MNGNRSIVYREFVGVIPVAKTLRNELRPVGHTQEHIIQNGLIQEDE
LRQEKSTELKNIMDDYYREYIDKSLSGVTDLDFTLLFELMNLVQSSPSKDNKKA
LEKEQSKMREQICTHLQSDSNYKNIFNAKLLKEILPDFIKNYNQYDVKDKAGKL
ETLALFNGFSTYFTDFFEKRKNVFTKEAVSTSIAYRIVHENSLIFLANMTSYKKIS
EKALDEIEVIEKNNQDKMGDWELNQIFNPDFYNMVLIQSGIDFYNEICGVVNAH
MNLYCQQTKNNYNLFKMRKLHKQILAYTSTSFEVPKMFEDDMSVYNAVNAFID
ETEKGNIIGKLKDIVNKYDELDEKRIYISKDFYETLSCFMSGNWNLITGCVENFY
DENIHAKGKSKEEKVKKAVKEDKYKSINDVNDLVEKYIDEKERNEFKNSNAKQ
YIREISNIITDTETAHLEYDDHISLIESEEKADEMKKRLDMYMNMYHWAKAFIVD
EVLDRDEMFYSDIDDIYNILENIVPLYNRVRNYVTQKPYNSKKIKLNFQSPTLAN
GWSQSKEFDNNAIILIRDNKYYLAIFNAKNKPDKKIIQGNSDKKNDNDYKKMVY
NLLPGANKMLPKVFLSKKGIETFKPSDYIISGYNAHKHIKTSENFDISFCRDLIDYF
KNSIEKHAEWRKYEFKFSATDSYSDISEFYREVEMQGYRIDWTYISEADINKLDE
EGKIYLFQIYNKDFAENSTGKENLHTMYFKNIFSEENLKDIIIKLNGQAELFYRRA
SVKNPVKHKKDSVLVNKTYKNQLDNGDVVRIPIPDDIYNEIYKMYNGYIKESDL
SEAAKEYLDKVEVRTAQKDIVKDYRYTVDKYFIHTPITINYKVTARNNVNDMV
VKYIAQNDDIHVIGIDRGERNLIYISVIDSHGNIVKQKSYNILNNYDYKKKLVEKE
KTREYARKNWKSIGNIKELKEGYISGVVHEIAMLIVEYNAIIAMEDLNYGFKRGR
FKVERQVYQKFESMLINKLNYFASKEKSVDEPGGLLKGYQLTYVPDNIKNLGKQ
CGVIFYVPAAFTSKIDPSTGFISAFNFKSISTNASRKQFFMQFDEIRYCAEKDMFSF
GFDYNNFDTYNITMGKTQWTVYTNGERLQSEFNNARRTGKTKSINLTETIKLLL
EDNEINYADGHDIRIDMEKMDEDKKSEFFAQLLSLYKLTVQMRNSYTEAEEQEN
GISYDKIISPVINDEGEFFDSDNYKESDDKECKMPKDADANGAYCIALKGLYEVL
KIKSEWTEDGFDRNCLKLPHAEWLDFIQNKRYE (SEQ ID NO: 85)

> WP_045971446_(con modificaciones), proteína hipotética [Flavobacterium sp. 316]

MKNFSNLYQVSKTVRFELKPIGNTLENIKNKSLLKNDSIRAESYQK
MKKTIDEFHKYFIDLALNNKKLSYLNEYIALYTQSAEAKKEDKFKADFKKVQDN
LRKEIVSSFTEGEAKAIFSVLDKKELITIELEKWKNENNLAVYLDESFKSFTTYFT
GFHQNRKNMYSAEANSTAIAYRLIHENLPKFIENSKAFEKSSQIAELQPKIEKLYK
EFEAYLNVNSISELFEIDYFNEVLTQKGITVYNNIIGGRTATEGKQKIQGLNEIINL
YNQTKPKNERLPKLKQLYKQILSDRISLSFLPDAFTEGKQVLKAVFEFYKINLLSY

KQDGVEESQNLLELIQQVVKNLGNQDVNKIYLKNDTSLTTIAQQLFGDFSVFSA
ALQYRYETVVNPKYTAEYQKANEAKQEKLDKEKIKFVKQDYFSIAFLQEVVAD
YVKTLDENLDWKQKYTPSCIADYFTTHFIAKKENEADKTFNFIANIKAKYQCIQG
ILEQADDYEDELKQDQKLIDNIKFFLDAILEVVHFIKPLHLKSESITEKDNAFYDV
FENYYEALNVVTPLYNMVRNYVTQKPYSTEKIKLNFENAQLLNGWDANKEKD
YLTTILKRDGNYFLAIMDKKHNKTFQQFTEDDENYEKIVYKLLPGVNKMLPKVF
FSNKNIAFFNPSKEILDNYKNNTHKKGATFNLKDCHALIDFFKDSLNKHEDWKY
FDFQFSETKTYQDLSGFYKEVEHQGYKINFKKVSVSQIDTLIEEGKMYLFQIYNK
DFSPYAKGKPNMHTLYWKALFETQNLENVIYKLNGQAEIFFRKASIKKKNITHK
AHQPIAAKNPLTPTAKNTFAYDLIKDKRYTVDKFQFHVPITMNFKATGNSYINQ
DVLAYLKDNPEVNIIGLDRGERHLVYLTLIDQKGTILLQESLNVIQDEKTHTPYH
TLLDNKEIARDKARKNWGSIESIKELKEGYISQVVHKITKMMIEHNAIVVMEDLN
FGFKRGRFKVEKQIYQKLEKMLIDKLNYLVLKDKQPHELGGLYNALQLTNKFES
FQKMGKQSGFLFYVPAWNTSKIDPTTGFVNYFYTKYENVEKAKTFFSKFDSILY
NKTKGYFEFVVKNYSDFNPKAADTRQEWTICTHGERIETKRQKEQNNNFVSTTI
QLTEQFVNFFEKVGLDLSKELKTQLIAQNEKSFFEELFHLLKLTLQMRNSESHTEI
DYLISPVANEKGIFYDSRKATASLPIDADANGAYHIAKKGLWIMEQINKTNSEDD
LKKVKLAISNREWLQYVQQVQKK (SEQ ID NO: 86)

> WP_044110123_(con modificaciones), proteína hipotética [Prevotella brevis]

MKQFTNLYQLSKTLRFELKPIGKTLEHINANGFIDNDAHRAESYK
KVKKLIDDYHKDYIENVLNNFKLNGEYLQAYFDLYSQDTKDKQFKDIQDKLRK
SIASALKGDDRYKTIDKKELIRQDMKTFLKKDTDKALLDEFYEFTTYFTGYHEN
RKNMYSDEAKSTAIAYRLIHDNLPKFIDNIAVFKKIANTSVADNFSTIYKNFEEYL
NVNSIDEIFSLDYYNIVLTQTQIEVYNSIIGGRTLEDDTKIQGINEFVNLYNQQLAN
KKDRLPKLKPLFKQILSDRVQLSWLQEEFNTGADVLNAVKEYCTSYFDNVEESV
KVLLTGISDYDLSKIYITNDLALTDVSQRMFGEWSIIPNAIEQRLRSDNPKKTNEK
EEKYSDRISKLKKLPKSYSLGYINECISELNGIDIADYYATLGAINTESKQEPSIPTS
IQVHYNALKPILDTDYPREKNLSQDKLTVMQLKDLLDDFKALQHFIKPLLGNGD
EAEKDEKFYGELMQLWEVIDSITPLYNKVRNYCTRKPFSTEKIKVNFENAQLLD
GWDENKESTNASIILRKNGMYYLGIMKKEYRNILTKPMPSDGDCYDKVVYKFF
KDITTMVPKCTTQMKSVKEHFSNSNDDYTLFEKDKFIAPVVITKEIFDLNNVLYN
GVKKFQIGYLNNTGDSFGYNHAVEIWKSFCLKFLKAYKSTSIYDFSSIEKNIGCY
NDLNSFYGAVNLLLYNLTYRKVSVDYIHQLVDEDKMYLFMIYNKDFSTYSKGT

PNMHTLYWKMLFDESNLNDVVYKLNGQAEVFYRKKSITYQHPTHPANKPIDNK
NVNNPKKQSNFEYDLIKDKRYTVDKFMFHVPITLNFKGMGNGDINMQVREYIK
TTDDLHFIGIDRGERHLLYICVINGKGEIVEQYSLNEIVNNYKGTEYKTDYHTLLS
ERDKKRKEERSSWQTIEGIKELKSGYLSQVIHKITQLMIKYNAIVLLEDLNMGFK
RGRQKVESSVYQQFEKALIDKLNYLVDKNKDANEIGGLLHAYQLTNDPKLPNK
NSKQSGFLFYVPAWNTSKIDPVTGFVNLLDTRYENVAKAQAFFKKFDSIRYNKE
YDRFEFKFDYSNFTAKAEDTRTQWTLCTYGTRIETFRNAEKNSNWDSREIDLTT
EWKTLFTQHNIPLNANLKEAILLQANKNFYTDILHLMKLTLQMRNSVTGTDIDY
MVSPVANECGEFFDSRKVKEGLPVNADANGAYNIARKGLWLAQQIKNANDLSD
VKLAITNKEWLQFAQKKQYLKD (SEQ ID NO: 87)

> WP_036388671_(con modificaciones), proteína hipotética [Moraxella caprae]

MLFQDFTHLYPLSKTMRFELKPIGKTLEHIHAKNFLSQDETMADM
YQKVKAILDDYHRDFIADMMGEVKLTKLAEFYDVYLKFRKNPKDDGLQKQLK
DLQAVLRKEIVKPIGNGGKYKAGYDRLFGAKLFKDGKELGDLAKFVIAQEGESS
PKLAHLAHFEKFSTYFTGFHDNRKNMYSDEDKHTAITYRLIHENLPRFIDNLQIL
ATIKQKHSALYDQIINELTASGLDVSLASHLDGYHKLLTQEGITAYNTLLGGISGE
AGSRKIQGINELINSHHNQHCHKSERIAKLRPLHKQILSDGMGVSFLPSKFADDSE
MCQAVNEFYRHYADVFAKVQSLFDGFDDHQKDGIYVEHKNLNELSKQAFGDF
ALLGRVLDGYYVDVVNPEFNERFAKAKTDNAKAKLTKEKDKFIKGVHSLASLE
QAIEHYTARHDDESVQAGKLGQYFKHGLAGVDNPIQKIHNNHSTIKGFLERERP
AGERALPKIKSGKNPEMTQLRQLKELLDNALNVAHFAKLLTTKTTLDNQDGNF
YGEFGALYDELAKIPTLYNKVRDYLSQKPFSTEKYKLNFGNPTLLNGWDLNKEK
DNFGIILQKDGCYYLALLDKAHKKVFDNAPNTGKNVYQKMIYKLLPGPNKMLP
KVFFAKSNLDYYNPSAELLDKYAQGTHKKGNNFNLKDCHALIDFFKAGINKHPE
WQHFGFKFSPTSSYQDLSDFYREVEPQGYQVKFVDINADYINELVEQGQLYLFQI
YNKDFSPKAHGKPNLHTLYFKALFSKDNLANPIYKLNGEAQIFYRKASLDMNET
TIHRAGEVLENKNPDNPKKRQFVYDIIKDKRYTQDKFMLHVPITMNFGVQGMTI
KEFNKKVNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILEQRSLNDITTASAN
GTQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVVHQISQLMLKY
NAIVVLEDLNFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDEADDEIGSYKN
ALQLTNNFTDLKSIGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAF
FGKFDKICYNADKDYFEFHIDYAKFTDKAKNSRQIWKICSHGDKRYVYDKTAN
QNKGATKGINVNDELKSLFARHHINDKQPNLVMDICQNNDKEFHKSLIYLLKTL

LALRYSNASSDEDFILSPVANDEGMFFNSALADDTQPQNADANGAYHIALKGL
WVLEQIKNSDDLNKVKLAIDNQTWLNFAQNR (SEQ ID NO: 88)

> WP_020988726_proteína Cpf1 asociada a los arreglos CRISPR (con modificaciones), subtipo PREFRAN [Leptospira inadai]

MEDYSGFVNIYSIQKTLRFELKPVGKTLEHIEKKGFLKKDKIRAED
YKAVKKIIDKYHRAYIEEVFDSVLHQKKKKDKTRFSTQFIKEIKEFSELYYKTEK
NIPDKERLEALSEKLRKMLVGAFKGEFSEEVAEKYKNLFSKELIRNEIEKFCETDE
ERKQVSNFKSFTTYFTGFHSNRQNIYSDEKKSTAIGYRIIHQNLPKFLDNLKIIESI
QRRFKDFPWSDLKKNLKKIDKNIKLTEYFSIDGFVNVLNQKGIDAYNTILGGKSE
ESGEKIQGLNEYINLYRQKNNIDRKNLPNVKILFKQILGDRETKSFIPEAFPDDQS
VLNSITEFAKYLKLDKKKKSIIAELKKFLSSFNRYELDGIYLANDNSLASISTFLFD
DWSFIKKSVSFKYDESVGDPKKKIKSPLKYEKEKEKWLKQKYYTISFLNDAIESY
SKSQDEKRVKIRLEAYFAEFKSKDDAKKQFDLLERIEEAYAIVEPLLGAEYPRDR
NLKADKKEVGKIKDFLDSIKSLQFFLKPLLSAEIFDEKDLGFYNQLEGYYEEIDSI
GHLYNKVRNYLTGKIYSKEKFKLNFENSTLLKGWDENREVANLCVIFREDQKY
YLGVMDKENNTILSDIPKVKPNELFYEKMVYKLIPTPHMQLPRIIFSSDNLSIYNP
SKSILKIREAKSFKEGKNFKLKDCHKFIDFYKESISKNEDWSRFDFKFSKTSSYENI
SEFYREVERQGYNLDFKKVSKFYIDSLVEDGKLYLFQIYNKDFSIFSKGKPNLHTI
YFRSLFSKENLKDVCLKLNGEAEMFFRKKSINYDEKKKREGHIHPELFEKLKYPIL
KDKRYSEDKFQFHLPISLNFKSKERLNFNLKVNEFLKRNKDINIIGIDRGERNLLY
LVMINQKGEILKQTLLDSMQSGKGRPEINYKEKLQEKEIERDKARKSWGTVENI
KELKEGYLSIVIHQISKLMVENNAIVVLEDLNIGFKRGRQKVERQVYQKFEKMLI
DKLNFLVFKENKPTEPGGVLKAYQLTDEFQSFEKLSKQTGFLFYVPSWNTSKIDP
RTGFIDFLHPAYENIEKAKQWINKFDSIRFNSKMDWFEFTADTRKFSENLMLGK
NRVWVICTTNVERYFTSKTANSSIQYNSIQITEKLKELFVDIPFSNGQDLKPEILRK
NDAVFFKSLLFYIKTTLSLRQNNGKKGEEEKDFILSPVVDSKGRFFNSLEASDDEP
KDADANGAYHIALKGLMNLLVLNETKEENLSRPKWKIKNKDWLEFVWERNR
(SEQ ID NO: 89)

> WP_023936172_(con modificaciones), exonucleasa SbcC [Porphyromonas crevioricanis]

MPWIDLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRA
ESYRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCELYKKDHRTEGED
KALDKIRAVLRGLIVGAFTGVCGRRENTVQNEKYESLFKEKLIKEILPDFVLSTE

AESLPFSVEEATRSLKEFDSFTSYFAGFYENRKNIYSTKPQSTAIAYRLIHENLPKF
IDNILVFQKIKEPIAKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEK
YNALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQILSDREQLSY
LPESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSISEYDLSRIYVRNDSQLTDIS
KKMLGDWNAIYMARERAYDHEQAPKRITAKYERDRIKALKGEESISLANLNSCI
AFLDNVRDCRVDTYLSTLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENN
LIQDKDNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIRGALDQ
VIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRNKEKDNSCVILRKGQNF
YLAIMNNRHKRSFENKVLPEYKEGEPYFEKMDYKFLPDPNKMLPKVFLSKKGIE
IYEPSPKLLEQYGHGTHKKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDT
ATYENVSSFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPCSKG
TPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKNDHPTHPAGKPIKKK
SRQKKGEESLFEYDLVKDRRYTMDKFQFHVPITMNFKCSAGSKVNDMVNAHIR
EAKDMHVIGIDRGERNLLYICVIDSRGTILDQISLNTINDIDYHDLLESRDKDRQQ
ERRNWQTIEGIKELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQK
VESSVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFKEMGKQNG
FLFYIPAWNTSNIDPTTGFVNLFHAQYENVDKAKSFFQKFDSISYNPKKDWFEFA
FDYKNFTKKAEGSRSMWILCTHGSRIKNFRNSQKNGQWDSEEFALTEAFKSLFV
RYEIDYTADLKTAIVDEKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAG
ADGRFFDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKLAISNK
EWLQFVQERSYEKD (SEQ ID NO: 90)

> WP_009217842_(con modificaciones), proteína hipotética [un taxón oral de Bacteroidetes 274]

MRKFNEFVGLYPISKTLRFELKPIGKTLEHIQRNKLLEHDAVRADD
YVKVKKIIDKYHKCLIDEALSGFTFDTEADGRSNNSLSEYYLYYNLKKRNEQEQ
KTFKTIQNNLRKQIVNKLTQSEKYKRIDKKELITTDLPDFLTNESEKELVEKFKNF
TTYFTEFHKNRKNMYSKEEKSTAIAFRLINENLPKFVDNIAAFEKVVSSPLAEKIN
ALYEDFKEYLNVEEISRVFRLDYYDELLTQKQIDLYNAIVGGRTEEDNKIQIKGL
NQYINEYNQQQTDRSNRLPKLKPLYKQILSDRESVSWLPPKFDSDKNLLIKIKEC
YDALSEKEKVFDKLESILKSLSTYDLSKIYISNDSQLSYISQKMFGRWDIISKAIRE
DCAKRNPQKSRESLEKFAERIDKKLKTIDSISIGDVDECLAQLGETYVKRVEDYF
VAMGESEIDDEQTDTTSFKKNIEGAYESVKELLNNADNITDNNLMQDKGNVEKI
KTLLDAIKDLQRFIKPLLGKGDEADKDGVFYGEFTSLWTKLDQVTPLYNMVRN

YLTSKPYSTKKIKLNFENSTLMDGWDLNKEPDNTTVIFCKDGLYYLGIMGKKYN
RVFVDREDLPHDGECYDKMEYKLLPGANKMLPKVFFSETGIQRFLPSEELLGKY
ERGTHKKGAGFDLGDCRALIDFFKKSIERHDDWKKFDFKFSDTSTYQDISEFYRE
VEQQGYKMSFRKVSVDYIKSLVEEGKLYLFQIYNKDFSAHSKGTPNMHTLYWK
MLFDEENLKDVVYKLNGEAEVFFRKSSITVQSPTHPANSPIKNKNKDNQKKESK
FEYDLIKDRRYTVDKFLFHVPITMNFKSVGGSNINQLVKRHIRSATDLHIIGIDRG
ERHLLYLTVIDSRGNIKEQFSLNEIVNEYNGNTYRTDYHELLDTREGERTEARRN
WQTIQNIRELKEGYLSQVIHKISELAIKYNAVIVLEDLNFGFMRSRQKVEKQVYQ
KFEKMLIDKLNYLVDKKKPVAETGGLLRAYQLTGEFESFKTLGKQSGILFYVPA
WNTSKIDPVTGFVNLFDTHYENIEKAKVFFDKFKSIRYNSDKDWFEFVVDDYTR
FSPKAEGTRRDWTICTQGKRIQICRNHQRNNEWEGQEIDLTKAFKEHFEAYGVDI
SKDLREQINTQNKKEFFEELLRLLRLTLQMRNSMPSSDIDYLISPVANDTGCFFDS
RKQAELKENAVLPMNADANGAYNIARKGLLAIRKMKQEENDSAKISLAISNKE
WLKFAQTKPYLED (SEQ ID NO: 91)

> WP_036890108_(con modificaciones), proteína hipotética [Porphyromonas crevioricanis]

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRAES
YRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCELYKKDHRTEGEDK
ALDKIRAVLRGLIVGAFTGVCGRRENTVQNEKYESLFKEKLIKEILPDFVLSTEAE
SLPFSVEEATRSLKEFDSFTSYFAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFID
NILVFQKIKEPIAKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYN
ALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQILSDREQLSYLP
ESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSISEYDLSRIYVRNDSQLTDISK
KMLGDWNAIYMARERAYDHEQAPKRITAKYERDRIKALKGEESISLANLNSCIA
FLDNVRDCRVDTYLSTLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLI
QDKDNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIRGALDQVI
PLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRNKEKDNSCVILRKGQNFY
LAIMNNRHKRSFENKMLPEYKEGEPYFEKMDYKFLPDPNKMLPKVFLSKKGIEI
YKPSPKLLEQYGHGTHKKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDT
ATYENVSSFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPCSKG
TPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKNDHPTHPAGKPIKKK
SRQKKGEESLFEYDLVKDRRYTMDKFQFHVPITMNFKCSAGSKVNDMVNAHIR
EAKDMHVIGIDRGERNLLYICVIDSRGTILDQISLNTINDIDYHDLLESRDKDRQQ

EHRNWQTIEGIKELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQK
VESSVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFKEMGKQNG
FLFYIPAWNTSNIDPTTGFVNLFHVQYENVDKAKSFFQKFDSISYNPKKDWFEFA
FDYKNFTKKAEGSRSMWILCTHGSRIKNFRNSQKNGQWDSEEFALTEAFKSLFV
RYEIDYTADLKTAIVDEKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAG
ADGRFFDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKLAISNK
EWLQFVQERSYEKD (SEQ ID NO: 92)

> WP_036887416_(con modificaciones), proteína hipotética [Porphyromonas crevioricanis]

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRAES
YRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCELYKKDHRTEGEDK
ALDKIRAVLRGLIVGAFTGVCGRRENTVQNEKYESLFKEKLIKEILPDFVLSTEAE
SLPFSVEEATRSLKEFDSFTSYFAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFID
NILVFQKIKEPIAKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYN
ALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQILSDREQLSYLP
ESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSISEYDLSRIYVRNDSQLTDISK
KMLGDWNAIYMARERAYDHEQAPKRITAKYERDRIKALKGEESISLANLNSCIA
FLDNVRDCRVDTYLSTLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLI
QDKDNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIRGALDQVI
PLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRNKEKDNSCVILRKGQNFY
LAIMNNRHKRSFENKVLPEYKEGEPYFEKMDYKFLPDPNKMLPKVFLSKKGIEI
YKPSPKLLEQYGHGTHKKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDT
ATYENVSSFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPCSKG
TPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKNDHPTHPAGKPIKKK
SRQKKGEESLFEYDLVKDRHYTMDKFQFHVPITMNFKCSAGSKVNDMVNAHIR
EAKDMHVIGIDRGERNLLYICVIDSRGTILDQISLNTINDIDYHDLLESRDKDRQQ
ERRNWQTIEGIKELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQK
VESSVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFKEMGKQNG
FLFYIPAWNTSNIDPTTGFVNLFHAQYENVDKAKSFFQKFDSISYNPKKDWFEFA
FDYKNFTKKAEGSRSMWILCTHGSRIKNFRNSQKNGQWDSEEFALTEAFKSLFV
RYEIDYTADLKTAIVDEKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAG
ADGRFFDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKLAISNK
EWLQFVQERSYEKD (SEQ ID NO: 93)

> WP_023941260_(con modificaciones), exonucleasa SbcC [Porphyromonas cansulci]

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRAES
YRRVKKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCELYKKDHRTEGEDK
ALDKIRAVLRGLIVGAFTGVCGRRENTVQNEKYESLFKEKLIKEILPDFVLSTEAE
SLPFSVEEATRSLKEFDSFTSYFAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFID
NILVFQKIKEPIAKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYN
ALIGKIVTEGDGEMKGLNEHINLYNQQRGREDRLPLFRPLYKQILSDREQLSYLP
ESFEKDEELLRALKEFYDHIAEDILGRTQQLMTSISEYDLSRIYVRNDSQLTDISK
KMLGDWNAIYMARERAYDHEQAPKRITAKYERDRIKALKGEESISLANLNSCIA
FLDNVRDCRVDTYLSTLGQKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLI
QDKDNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYNYIRGALDQVI
PLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRNKEKDNSCVILRKGQNFY
LAIMNNRHKRSFENKVLPEYKEGEPYFEKMDYKFLPDPNKMLPKVFLSKKGIEI
YKPSPKLLEQYGHGTHKKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDT
ATYENVSSFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPCSKG
TPNLHTLYWRMLFDERNLADVIYKLDGKAEIFFREKSLKNDHPTHPAGKPIKKK
SRQKKGEESLFEYDLVKDRRYTMDKFQFHVPITMNFKCSAGSKVNDMVNAHIR
EAKDMHVIGIDRGERNLLYICVIDSRGTILDQISLNTINDIDYHDLLESRDKDRQQ
ERRNWQTIEGIKELKQGYLSQAVHRIAELMVAYKAVVALEDLNMGFKRGRQK
VESSVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQFTAPFKSFKEMGKQNG
FLFYIPAWNTSNIDPTTGFVNLFHAQYENVDKAKSFFQKFDSISYNPKKDWFEFA
FDYKNFTKKAEGSRSMWILCTHGSRIKNFRNSQKNGQWDSEEFALTEAFKSLFV
RYEIDYTADLKTAIVDEKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAG
ADGRFFDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKLAISNK
EWLQFVQERSYEKD (SEQ ID NO: 94)

> WP_037975888_(con modificaciones), proteína hipotética [Synergistes jonesii]

MANSLKDFTNIYQLSKTLRFELKPIGKTEEHINRKLIIMHDEKRGE
DYKSVTKLIDDYHRKFIHETLDPAHFDWNPLAEALIQSGSKNNKALPAEQKEMR
EKIISMFTSQAVYKKLFKKELFSELLPEMIKSELVSDLEKQAQLDAVKSFDKFSTY
FTGFHENRKNIYSKKDTSTSIAFRIVHQNFPKFLANVRAYTLIKERAPEVIDKAQK
ELSGILGGKTLDDIFSIESFNNVLTQDKIDYYNQIIGGVSGKAGDKKLRGVNEFSN
LYRQQHPEVASLRIKMVPLYKQILSDRTTLSFVPEALKDDEQAINAVDGLRSELE

RNDIFNRIKRLFGKNNLYSLDKIWIKNSSISAFSNELFKNWSFIEDALKEFKENEF
NGARSAGKKAEKWLKSKYFSFADIDAAVKSYSEQVSADISSAPSASYFAKFTNLI
ETAAENGRKFSYFAAESKAFRGDDGKTEIIKAYLDSLNDILHCLKPFETEDISDID
TEFYSAFAEIYDSVKDVIPVYNAVRNYTTQKPFSTEKFKLNFENPALAKGWDKN
KEQNNTAIILMKDGKYYLGVIDKNNKLRADDLADDGSAYGYMKMNYKFIPTPH
MELPKVFLPKRAPKRYNPSREILLIKENKTFIKDKNFNRTDCHKLIDFFKDSINKH
KDWRTFGFDFSDTDSYEDISDFYMEVQDQGYKLTFTRLSAEKIDKWVEEGRLFL
FQIYNKDFADGAQGSPNLHTLYWKAIFSEENLKDVVLKLNGEAELFFRRKSIDKP
AVHAKGSMKVNRRDIDGNPIDEGTYVEICGYANGKRDMASLNAGARGLIESGL
VRITEVKHELVKDKRYTIDKYFFHVPFTINFKAQGQGNINSDVNLFLRNNKDVNI
IGIDRGERNLVYVSLIDRDGHIKLQKDFNIIGGMDYHAKLNQKEKERDTARKSW
KTIGTIKELKEGYLSQVVHEIVRLAVDNNAVIVMEDLNIGFKRGRFKVEKQVYQ
KFEKMLIDKLNYLVFKDAGYDAPCGILKGLQLTEKFESFTKLGKQCGIIFYIPAG
YTSKIDPTTGFVNLFNINDVSSKEKQKDFIGKLDSIRFDAKRDMFTFEFDYDKFRT
YQTSYRKKWAVWTNGKRIVREKDKDGKFRMNDRLLTEDMKNILNKYALAYK
AGEDILPDVISRDKSLASEIFYVFKNTLQMRNSKRDTGEDFIISPVLNAKGRFFDS
RKTDAALPIDADANGAYHIALKGSLVLDAIDEKLKEDGRIDYKDMAVSNPKWF
EFMQTRKFDF (SEQ ID NO: 95)

> EFI70750_(con modificaciones), proteína hipotética conservada [Prevotella bryantii B14]

MQINNLKIIYMKFTDFTGLYSLSKTLRFELKPIGKTLENIKKAGLLE
QDQHRADSYKKVKKIIDEYHKAFIEKSLSNFELKYQSEDKLDSLEEYLMYYSMK
RIEKTEKDKFAKIQDNLRKQIADHLKGDESYKTIFSKDLIRKNLPDFVKSDEERTL
IKEFKDFTTYFKGFYENRENMYSAEDKSTAISHRIIHENLPKFVDNINAFSKIILIPE
LREKLNQIYQDFEEYLNVESIDEIFHLDYFSMVMTQKQIEVYNAIIGGKSTNDKKI
QGLNEYINLYNQKHKDCKLPKLKLLFKQILSDRIAISWLPDNFKDDQEALDSIDT
CYKNLLNDGNVLGEGNLKLLLENIDTYNLKGIFIRNDLQLTDISQKMYASWNVI
QDAVILDLKKQVSRKKKESAEDYNDRLKKLYTSQESFSIQYLNDCLRAYGKTEN
IQDYFAKLGAVNNEHEQTINLFAQVRNAYTSVQAILTTPYPENANLAQDKETVA
LIKNLLDSLKRLQRFIKPLLGKGDESDKDERFYGDFTPLWETLNQITPLYNMVRN
YMTRKPYSQEKIKLNFENSTLLGGWDLNKEHDNTAIILRKNGLYYLAIMKKSAN
KIFDKDKLDNSGDCYEKMVYKLLPGANKMLPKVFFSKSRIDEFKPSENIIENYKK
GTHKKGANFNLADCHNLIDFFKSSISKHEDWSKFNFHFSDTSSYEDLSDFYREVE

QQGYSISFCDVSVEYINKMVEKGDLYLFQIYNKDFSEFSKGTPNMHTLYWNSLF
SKENLNNIIYKLNGQAEIFFRKKSLNYKRPTHPAHQAIKNKNKCNEKKESIFDYD
LVKDKRYTVDKFQFHVPITMNFKSTGNTNINQQVIDYLRTEDDTHIIGIDRGERH
LLYLVVIDSHGKIVEQFTLNEIVNEYGGNIYRTNYHDLLDTREQNREKARESWQ
TIENIKELKEGYISQVIHKITDLMQKYHAVVVLEDLNMGFMRGRQKVEKQVYQ
KFEEMLINKLNYLVNKKADQNSAGGLLHAYQLTSKFESFQKLGKQSGFLFYIPA
WNTSKIDPVTGFVNLFDTRYESIDKAKAFFGKFDSIRYNADKDWFEFAFDYNNF
TTKAEGTRTNWTICTYGSRIRTFRNQAKNSQWDNEEIDLTKAYKAFFAKHGINIY
DNIKEAIAMETEKSFFEDLLHLLKLTLQMRNSITGTTTDYLISPVHDSKGNFYDSR
ICDNSLPANADANGAYNIARKGLMLIQQIKDSTSSNRFKFSPITNKDWLIFAQEKP
YLND (SEQ ID NO: 96)

> WP_024988992_(con modificaciones), proteína hipotética [Prevotella albensis]

MNIKNFTGLYPLSKTLRFELKPIGKTKENIEKNGILTKDEQRAKDY
LIVKGFIDEYHKQFIKDRLWDFKLPLESEGEKNSLEEYQELYELTKRNDAQEADF
TEIKDNLRSSITEQLTKSGSAYDRIFKKEFIREDLVNFLEDEKDKNIVKQFEDFTTY
FTGFYENRKNMYSSEEKSTAIAYRLIHQNLPKFMDNMRSFAKIANSSVSEHFSDI
YESWKEYLNVNSIEEIFQLDYFSETLTQPHIEVYNYIIGKKVLEDGTEIKGINEYV
NLYNQQQKDKSKRLPFLVPLYKQILSDREKLSWIAEEFDSDKKMLSAITESYNHL
HNVLMGNENESLRNLLLNIKDYNLEKINITNDLSLTEISQNLFGRYDVFTNGIKN
KLRVLTPRKKKETDENFEDRINKIFKTQKSFSIAFLNKLPQPEMEDGKPRNIEDYF
ITQGAINTKSIQKEDIFAQIENAYEDAQVFLQIKDTDNKLSQNKTAVEKIKTLLDA
LKELQHFIKPLLGSGEENEKDELFYGSFLAIWDELDTITPLYNKVRNWLTRKPYS
TEKIKLNFDNAQLLGGWDVNKEHDCAGILLRKNDSYYLGIINKKTNHIFDTDITP
SDGECYDKIDYKLLPGANKMLPKVFFSKSRIKEFEPSEAIINCYKKGTHKKGKNF
NLTDCHRLINFFKTSIEKHEDWSKFGFKFSDTETYEDISGFYREVEQQGYRLTSHP
VSASYIHSLVKEGKLYLFQIWNKDFSQFSKGTPNLHTLYWKMLFDKRNLSDVV
YKLNGQAEVFYRKSSIEHQNRIIHPAQHPITNKNELNKKHTSTFKYDIIKDRRYTV
DKFQFHVPITINFKATGQNNINPIVQEVIRQNGITHIIGIDRGERHLLYLSLIDLKGN
IIKQMTLNEIINEYKGVTYKTNYHNLLEKREKERTEARHSWSSIESIKELKDGYM
SQVIHKITDMMVKYNAIVVLEDLNGGFMRGRQKVEKQVYQKFEKKLIDKLNYL
VDKKLDANEVGGVLNAYQLTNKFESFKKIGKQSGFLFYIPAWNTSKIDPITGFVN
LFNTRYESIKETKVFWSKFDIIRYNKEKNWFEFVFDYNTFTTKAEGTRTKWTLCT
HGTRIQTFRNPEKNAQWDNKEINLTESFKALFEKYKIDITSNLKESIMQETEKKFF

QELHNLLHLTLQMRNSVTGTDIDYLISPVADEDGNFYDSRINGKNFPENADANG
AYNIARKGLMLIRQIKQADPQKKFKFETITNKDWLKFAQDKPYLKD (SEQ ID
NO: 97)

> WP_039658684_(con modificaciones), proteína hipotética [Smithella sp. SC_K08D17]

MQTLFENFTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKDEDRAE
KYKKVKNIIDEYHKDFIEKSLNGLKLDGLEKYKTLYLKQEKDDKDKKAFDKEK
ENLRKQIANAFRNNEKFKTLFAKELIKNDLMSFACEEDKKNVKEFEAFTTYFTGF
HQNRANMYVADEKRTAIASRLIHENLPKFIDNIKIFEKMKKEAPELLSPFNQTLK
DMKDVIKGTTLEEIFSLDYFNKTLTQSGIDIYNSVIGGRTPEEGKTKIKGLNEYINT
DFNQKQTDKKKRQPKFKQLYKQILSDRQSLSFIAEAFKNDTEILEAIEKFYVNEL
LHFSNEGKSTNVLDAIKNAVSNLESFNLTKMYFRSGASLTDVSRKVFGEWSIINR
ALDNYYATTYPIKPREKSEKYEERKEKWLKQDFNVSLIQTAIDEYDNETVKGKN
SGKVIADYFAKFCDDKETDLIQKVNEGYIAVKDLLNTPCPENEKLGSNKDQVKQ
IKAFMDSIMDIMHFVRPLSLKDTDKEKDETFYSLFTPLYDHLTQTIALYNKVRNY
LTQKPYSTEKIKLNFENSTLLGGWDLNKETDNTAIILRKDNLYYLGIMDKRHNRI
FRNVPKADKKDFCYEKMVYKLLPGANKMLPKVFFSQSRIQEFTPSAKLLENYAN
ETHKKGDNFNLNHCHKLIDFFKDSINKHEDWKNFDFRFSATSTYADLSGFYHEV
EHQGYKISFQSVADSFIDDLVNEGKLYLFQIYNKDFSPFSKGKPNLHTLYWKMLF
DENNLKDVVYKLNGEAEVFYRKKSIAEKNTTIHKANESIINKNPDNPKATSTFNY
DIVKDKRYTIDKFQFHIPITMNFKAEGIFNMNQRVNQFLKANPDINIIGIDRGERH
LLYYALINQKGKILKQDTLNVIANEKQKVDYHNLLDKKEGDRATARQEWGVIE
TIKELKEGYLSQVIHKLTDLMIENNAIIVMEDLNFGFKRGRQKVEKQVYQKFEK
MLIDKLNYLVDKNKKANELGGLLNAFQLANKFESFQKMGKQNGFIFYVPAWNT
SKTDPATGFIDFLKPRYENLNQAKDFFEKFDSIRLNSKADYFEFAFDFKNFTEKA
DGGRTKWTVCTTNEDRYAWNRALNNNRGSQEKYDITAELKSLFDGKVDYKSG
KDLKQQIASQESADFFKALMKNLSITLSLRHNNGEKGDNEQDYILSPVADSKGR
FFDSRKADDDMPKNADANGAYHIALKGLWCLEQISKTDDLKKVKLAISNKEWL
EFVQTLKG (SEQ ID NO: 98)

> WP_037385181_(con modificaciones), proteína hipotética [Smithella sp. SCADC]

MQTLFENFTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKDEDRAE
KYKKVKNIIDEYHKDFIEKSLNGLKLDGLEEYKTLYLKQEKDDKDKKAFDKEKE
NLRKQIANAFRNNEKFKTLFAKELIKNDLMSFACEEDKKNVKEFEAFTTYFTGF

HQNRANMYVADEKRTAIASRLIHENLPKFIDNIKIFEKMKKEAPELLSPFNQTLK
DMKDVIKGTTLEEIFSLDYFNKTLTQSGIDIYNSVIGGRTPEEGKTKIKGLNEYINT
DFNQKQTDKKKRQPKFKQLYKQILSDRQSLSFIAEAFKNDTEILEAIEKFYVNEL
LHFSNEGKSTNVLDAIKNAVSNLESFNLTKIYFRSGTSLTDVSRKVFGEWSIINRA
LDNYYATTYPIKPREKSEKYEERKEKWLKQDFNVSLIQTAIDEYDNETVKGKNS
GKVIVDYFAKFCDDKETDLIQKVNEGYIAVKDLLNTPYPENEKLGSNKDQVKQI
KAFMDSIMDIMHFVRPLSLKDTDKEKDETFYSLFTPLYDHLTQTIALYNKVRNY
LTQKPYSTEKIKLNFENSTLLGGWDLNKETDNTAIILRKENLYYLGIMDKRHNRI
FRNVPKADKKDSCYEKMVYKLLPGANKMLPKVFFSQSRIQEFTPSAKLLENYEN
ETHKKGDNFNLNHCHQLIDFFKDSINKHEDWKNFDFRFSATSTYADLSGFYHEV
EHQGYKISFQSIADSFIDDLVNEGKLYLFQIYNKDFSPFSKGKPNLHTLYWKMLF
DENNLKDVVYKLNGEAEVFYRKKSIAEKNTTIHKANESIINKNPDNPKATSTFNY
DIVKDKRYTIDKFQFHVPITMNFKAEGIFNMNQRVNQFLKANPDINIIGIDRGERH
LLYYTLINQKGKILKQDTLNVIANEKQKVDYHNLLDKKEGDRATARQEWGVIE
TIKELKEGYLSQVIHKLTDLMIENNAIIVMEDLNFGFKRGRQKVEKQVYQKFEK
MLIDKLNYLVDKNKKANELGGLLNAFQLANKFESFQKMGKQNGFIFYVPAWNT
SKTDPATGFIDFLKPRYENLKQAKDFFEKFDSIRLNSKADYFEFAFDFKNFTGKA
DGGRTKWTVCTTNEDRYAWNRALNNNRGSQEKYDITAELKSLFDGKVDYKSG
KDLKQQIASQELADFFRTLMKYLSVTLSLRHNNGEKGETEQDYILSPVADSMGK
FFDSRKAGDDMPKNADANGAYHIALKGLWCLEQISKTDDLKKVKLAISNKEWL
EFMQTLKG (SEQ ID NO: 99)

> WP_039871282_(con modificaciones), proteína hipotética [Prevotella bryantii]

MKFTDFTGLYSLSKTLRFELKPIGKTLENIKKAGLLEQDQHRADSY
KKVKKIIDEYHKAFIEKSLSNFELKYQSEDKLDSLEEYLMYYSMKRIEKTEKDKF
AKIQDNLRKQIADHLKGDESYKTIFSKDLIRKNLPDFVKSDEERTLIKEFKDFTTY
FKGFYENRENMYSAEDKSTAISHRIIHENLPKFVDNINAFSKIILIPELREKLNQIY
QDFEEYLNVESIDEIFHLDYFSMVMTQKQIEVYNAIIGGKSTNDKKIQGLNEYINL
YNQKHKDCKLPKLKLLFKQILSDRIAISWLPDNFKDDQEALDSIDTCYKNLLND
GNVLGEGNLKLLLENIDTYNLKGIFIRNDLQLTDISQKMYASWNVIQDAVILDLK
KQVSRKKKESAEDYNDRLKKLYTSQESFSIQYLNDCLRAYGKTENIQDYFAKLG
AVNNEHEQTINLFAQVRNAYTSVQAILTTPYPENANLAQDKETVALIKNLLDSL
KRLQRFIKPLLGKGDESDKDERFYGDFTPLWETLNQITPLYNMVRNYMTRKPYS
QEKIKLNFENSTLLGGWDLNKEHDNTAIILRKNGLYYLAIMKKSANKIFDKDKL

DNSGDCYEKMVYKLLPGANKMLPKVFFSKSRIDEFKPSENIIENYKKGTHKKGA
NFNLADCHNLIDFFKSSISKHEDWSKFNFHFSDTSSYEDLSDFYREVEQQGYSISF
CDVSVEYINKMVEKGDLYLFQIYNKDFSEFSKGTPNMHTLYWNSLFSKENLNNII
YKLNGQAEIFFRKKSLNYKRPTHPAHQAIKNKNKCNEKKESIFDYDLVKDKRYT
VDKFQFHVPITMNFKSTGNTNINQQVIDYLRTEDDTHIIGIDRGERHLLYLVVIDS
HGKIVEQFTLNEIVNEYGGNIYRTNYHDLLDTREQNREKARESWQTIENIKELKE
GYISQVIHKITDLMQKYHAVVVLEDLNMGFMRGRQKVEKQVYQKFEEMLINKL
NYLVNKKADQNSAGGLLHAYQLTSKFESFQKLGKQSGFLFYIPAWNTSKIDPVT
GFVNLFDTRYESIDKAKAFFGKFDSIRYNADKDWFEFAFDYNNFTTKAEGTRTN
WTICTYGSRIRTFRNQAKNSQWDNEEIDLTKAYKAFFAKHGINIYDNIKEAIAME
TEKSFFEDLLHLLKLTLQMRNSITGTTTDYLISPVHDSKGNFYDSRICDNSLPANA
DANGAYNIARKGLMLIQQIKDSTSSNRFKFSPITNKDWLIFAQEKPYLND (SEQ
ID NO: 100)

> EKE28449_(con modificaciones), proteína hipotética ACD_3C00058G0015 [bacteria no cultivada (gcode 4)]

MFKGDAFTGLYEVQKTLRFELVPIGLTQSYLENDWVIQKDKEVEE
NYGKIKAYFDLIHKEFVRQSLENAWLCQLDDFYEKYIELHNSLETRKDKNLAKQ
FEKVMKSLKKEFVSFFDAKWNEWKQKFSFLKKWWIDVLNEKEVLDLMAEFYP
DEKELFDKFDKFFTYFSNFKESRKNFYADDGRAWAIATRAIDENLITFIKNIEDFK
KLNSSFREFVNDNFSEEDKQIFEIDFYNNCLLQPWIDKYNKIVWWYSLENWEKV
QWLNEKINNFKQNQNKSNSKDLKFPRMKLLYKQILGDKEKKVYIDEIRDDKNLI
DLIDNSKRRNQIKIDNANDIINDFINNNAKFELDKIYLTRQSINTISSKYFSSWDYI
RWYFWTGELQEFVSFYDLKETFWKIEYETLENIFKDCYVKGINTESQNNIVFETQ
GIYENFLNIFKFEFNQNISQISLLEWELDKIQNEDIKKNEKQVEVIKNYFDSVMSV
YKMTKYFSLEKWKKRVELDTDNNFYNDFNEYLEGFEIWKDYNLVRNYITKKQ
VNTDKIKLNFDNSQFLTWWDKDKENERLGIILRREWKYYLWILKKWNTLNFGD
YLQKEWEIFYEKMNYKQLNNVYRQLPRLLFPLTKKLNELKWDELKKYLSKYIQ
NFWYNEEIAQIKIEFDIFQESKEKWEKFDIDKLRKLIEYYKKWVLALYSDLYDLE
FIKYKNYDDLSIFYSDVEKKMYNLNFTKIDKSLIDGKVKSWELYLFQIYNKDFSE
SKKEWSTENIHTKYFKLLFNEKNLQNLVVKLSWWADIFFRDKTENLKFKKDKN
GQEILDHRRFSQDKIMFHISITLNANCWDKYWFNQYVNEYMNKERDIKIIWIDR
WEKHLAYYCVIDKSWKIFNNEIWTLNELNWVNYLEKLEKIESSRKDSRISWWEI
ENIKELKNGYISQVINKLTELIVKYNAIIVFEDLNIWFKRWRQKIEKQIYQKLELA

LAKKLNYLTQKDKKDDEILWNLKALQLVPKVNDYQDIWNYKQSWIMFYVRAN
YTSVTCPNCWLRKNLYISNSATKENQKKSLNSIAIKYNDWKFSFSYEIDDKSWK
QKQSLNKKKFIVYSDIERFVYSPLEKLTKVIDVNKKLLELFRDFNLSLDINKQIQE
KDLDSVFFKSLTHLFNLILQLRNSDSKDNKDYISCPSCYYHSNNWLQWFEFNWD
ANWAYNIARKGIILLDRIRKNQEKPDLYVSDIDWDNFVQSNQFPNTIIPIQNIEKQ
VPLNIKI (SEQ ID NO: 101)

> WP_018359861_(con modificaciones), proteína hipotética [Porphyromonas macacae]

MKTQHFFEDFTSLYSLSKTIRFELKPIGKTLENIKKNGLIRRDEQRL
DDYEKLKKVIDEYHEDFIANILSSFSFSEEILQSYIQNLSESEARAKIEKTMRDTLA
KAFSEDERYKSIFKKELVKKDIPVWCPAYKSLCKKFDNFTTSLVPFHENRKNLYT
SNEITASIPYRIVHVNLPKFIQNIEALCELQKKMGADLYLEMMENLRNVWPSFVK
TPDDLCNLKTYNHLMVQSSISEYNRFVGGYSTEDGTKHQGINEWINIYRQRNKE
MRLPGLVFLHKQILAKVDSSSFISDTLENDDQVFCVLRQFRKLFWNTVSSKEDD
AASLKDLFCGLSGYDPEAIYVSDAHLATISKNIFDRWNYISDAIRRKTEVLMPRK
KESVERYAEKISKQIKKRQSYSLAELDDLLAHYSEESLPAGFSLLSYFTSLGGQK
YLVSDGEVILYEEGSNIWDEVLIAFRDLQVILDKDFTEKKLGKDEEAVSVIKKAL
DSALRLRKFFDLLSGTGAEIRRDSSFYALYTDRMDKLKGLLKMYDKVRNYLTK
KPYSIEKFKLHFDNPSLLSGWDKNKELNNLSVIFRQNGYYYLGIMTPKGKNLFK
TLPKLGAEEMFYEKMEYKQIAEPMLMLPKVFFPKKTKPAFAPDQSVVDIYNKKT
FKTGQKGFNKKDLYRLIDFYKEALTVHEWKLFNFSFSPTEQYRNIGEFFDEVREQ
AYKVSMVNVPASYIDEAVENGKLYLFQIYNKDFSPYSKGIPNLHTLYWKALFSE
QNQSRVYKLCGGGELFYRKASLHMQDTTVHPKGISIHKKNLNKKGETSLFNYD
LVKDKRFTEDKFFFHVPISINYKNKKITNVNQMVRDYIAQNDDLQIIGIDRGERN
LLYISRIDTRGNLLEQFSLNVIESDKGDLRTDYQKILGDREQERLRRRQEWKSIES
IKDLKDGYMSQVVHKICNMVVEHKAIVVLENLNLSFMKGRKKVEKSVYEKFER
MLVDKLNYLVVDKKNLSNEPGGLYAAYQLTNPLFSFEELHRYPQSGILFFVDPW
NTSLTDPSTGFVNLLGRINYTNVGDARKFFDRFNAIRYDGKGNILFDLDLSRFDV
RVETQRKLWTLTTFGSRIAKSKKSGKWMVERIENLSLCFLELFEQFNIGYRVEKD
LKKAILSQDRKEFYVRLIYLFNLMMQIRNSDGEEDYILSPALNEKNLQFDSRLIEA
KDLPVDADANGAYNVARKGLMVVQRIKRGDHESIHRIGRAQWLRYVQEGIVE
(SEQ ID NO: 102)

> WP_013282991_(con modificaciones), proteína hipotética [Butyrivibrio proteoclasticus]

MLLYENYTKRNQITKSLRLELRPQGKTLRNIKELNLLEQDKAIYAL
LERLKPVIDEGIKDIARDTLKNCELSFEKLYEHFLSGDKKAYAKESERLKKEIVKT
LIKNLPEGIGKISEINSAKYLNGVLYDFIDKTHKDSEEKQNILSDILETKGYLALFS
KFLTSRITTLEQSMPKRVIENFEIYAANIPKMQDALERGAVSFAIEYESICSVDYY
NQILSQEDIDSYNRLISGIMDEDGAKEKGINQTISEKNIKIKSEHLEEKPFRILKQL
HKQILEEREKAFTIDHIDSDEEVVQVTKEAFEQTKEQWENIKKINGFYAKDPGDI
TLFIVVGPNQTHVLSQLIYGEHDRIRLLLEEYEKNTLEVLPRRTKSEKARYDKFV
NAVPKKVAKESHTFDGLQKMTGDDRLFILYRDELARNYMRIKEAYGTFERDILK
SRRGIKGNRDVQESLVSFYDELTKFRSALRIINSGNDEKADPIFYNTFDGIFEKAN
RTYKAENLCRNYVTKSPADDARIMASCLGTPARLRTHWWNGEENFAINDVAMI
RRGDEYYYFVLTPDVKPVDLKTKDETDAQIFVQRKGAKSFLGLPKALFKCILEP
YFESPEHKNDKNCVIEEYVSKPLTIDRRAYDIFKNGTFKKTNIGIDGLTEEKFKDD
CRYLIDVYKEFIAVYTRYSCFNMSGLKRADEYNDIGEFFSDVDTRLCTMEWIPVS
FERINDMVDKKEGLLFLVRSMFLYNRPRKPYERTFIQLFSDSNMEHTSMLLNSR
AMIQYRAASLPRRVTHKKGSILVALRDSNGEHIPMHIREAIYKMKNNFDISSEDFI
MAKAYLAEHDVAIKKANEDHRNRRYTEDKFFLSLSYTKNADISARTLDYINDKV
EEDTQDSRMAVIVTRNLKDLTYVAVVDEKNNVLEEKSLNEIDGVNYRELLKER
TKIKYHDKTRLWQYDVSSKGLKEAYVELAVTQISKLATKYNAVVVVESMSSTF
KDKFSFLDEQIFKAFEARLCARMSDLSFNTIKEGEAGSISNPIQVSNNNGNSYQD
GVIYFLNNAYTRTLCPDTGFVDVFDKTRLITMQSKRQFFAKMKDIRIDDGEMLF
TFNLEEYPTKRLLDRKEWTVKIAGDGSYFDKDKGEYVYVNDIVREQIIPALLED
KAVFDGNMAEKFLDKTAISGKSVELIYKWFANALYGIITKKDGEKIYRSPITGTEI
DVSKNTTYNFGKKFMFKQEYRGDGDFLDAFLNYMQAQDIAV (SEQ ID NO: 103)

> AIZ56868_(con modificaciones), proteína hipotética Mpt1_c09950 [la bacteria Methanoplasma termitum candidata]

MNNYDEFTKLYPIQKTIRFELKPQGRTMEHLETFNFFEEDRDRAE
KYKILKEAIDEYHKKFIDEHLTNMSLDWNSLKQISEKYYKSREEKDKKVFLSEQ
KRMRQEIVSEFKKDDRFKDLFSKKLFSELLKEEIYKKGNHQEIDALKSFDKFSGY
FIGLHENRKNMYSDGDEITAISNRIVNENFPKFLDNLQKYQEARKKYPEWIIKAE
SALVAHNIKMDEVFSLEYFNKVLNQEGIQRYNLALGGYVTKSGEKMMGLNDAL
NLAHQSEKSSKGRIHMTPLFKQILSEKESFSYIPDVFTEDSQLLPSIGGFFAQIEND

KDGNIFDRALELISSYAEYDTERIYIRQADINRVSNVIFGEWGTLGGLMREYKAD
SINDINLERTCKKVDKWLDSKEFALSDVLEAIKRTGNNDAFNEYISKMRTAREKI
DAARKEMKFISEKISGDEESIHIIKTLLDSVQQFLHFFNLFKARQDIPLDGAFYAEF
DEVHSKLFAIVPLYNKVRNYLTKNNLNTKKIKLNFKNPTLANGWDQNKVYDYA
SLIFLRDGNYYLGIINPKRKKNIKFEQGSGNGPFYRKMVYKQIPGPNKNLPRVFL
TSTKGKKEYKPSKEIIEGYEADKHIRGDKFDLDFCHKLIDFFKESIEKHKDWSKF
NFYFSPTESYGDISEFYLDVEKQGYRMHFENISAETIDEYVEKGDLFLFQIYNKDF
VKAATGKKDMHTIYWNAAFSPENLQDVVVKLNGEAELFYRDKSDIKEIVHREG
EILVNRTYNGRTPVPDKIHKKLTDYHNGRTKDLGEAKEYLDKVRYFKAHYDITK
DRRYLNDKIYFHVPLTLNFKANGKKNLNKMVIEKFLSDEKAHIIGIDRGERNLLY
YSIIDRSGKIIDQQSLNVIDGFDYREKLNQREIEMKDARQSWNAIGKIKDLKEGYL
SKAVHEITKMAIQYNAIVVMEELNYGFKRGRFKVEKQIYQKFENMLIDKMNYL
VFKDAPDESPGGVLNAYQLTNPLESFAKLGKQTGILFYVPAAYTSKIDPTTGFVN
LFNTSSKTNAQERKEFLQKFESISYSAKDGGIFAFAFDYRKFGTSKTDHKNVWTA
YTNGERMRYIKEKKRNELFDPSKEIKEALTSSGIKYDGGQNILPDILRSNNNGLIY
TMYSSFIAAIQMRVYDGKEDYIISPIKNSKGEFFRTDPKRRELPIDADANGAYNIA
LRGELTMRAIAEKFDPDSEKMAKLELKHKDWFEFMQTRGD (SEQ ID NO: 104)

> WP_027407524_(con modificaciones), proteína hipotética [Anaerovibrio sp. RM50]

MVAFIDEFVGQYPVSKTLRFEARPVPETKKWLESDQCSVLFNDQK
RNEYYGVLKELLDDYYRAYIEDALTSFTLDKALLENAYDLYCNRDTNAFSSCCE
KLRKDLVKAFGNLKDYLLGSDQLKDLVKLKAKVDAPAGKGKKKIEVDSRLIN
WLNNNAKYSAEDREKYIKAIESFEGFVTYLTNYKQARENMFSSEDKSTAIAFRVI
DQNMVTYFGNIRIYEKIKAKYPELYSALKGFEKFFSPTAYSEILSQSKIDEYNYQC
IGRPIDDADFKGVNSLINEYRQKNGIKARELPVMSMLYKQILSDRDNSFMSEVIN
RNEEAIECAKNGYKVSYALFNELLQLYKKIFTEDNYGNIYVKTQPLTELSQALFG
DWSILRNALDNGKYDKDIINLAELEKYFSEYCKVLDADDAAKIQDKFNLKDYFI
QKNALDATLPDLDKITQYKPHLDAMLQAIRKYKLFSMYNGRKKMDVPENGIDF
SNEFNAIYDKLSEFSILYDRIRNFATKKPYSDEKMKLSFNMPTMLAGWDYNNET
ANGCFLFIKDGKYFLGVADSKSKNIFDFKKNPHLLDKYSSKDIYYKVKYKQVSG
SAKMLPKVVFAGSNEKIFGHLISKRILEIREKKLYTAAAGDRKAVAEWIDFMKS
AIAIHPEWNEYFKFKFKNTAEYDNANKFYEDIDKQTYSLEKVEIPTEYIDEMVSQ
HKLYLFQLYTKDFSDKKKKKGTDNLHTMYWHGVFSDENLKAVTEGTQPIIKLN

GEAEMFMRNPSIEFQVTHEHNKPIANKNPLNTKKESVFNYDLIKDKRYTERKFY
FHCPITLNFRADKPIKYNEKINRFVENNPDVCHGIDRGERHLLYYTVINQTGDILE
QGSLNKISGSYTNDKGEKVNKETDYHDLLDRKEKGKHVAQQAWETIENIKELK
AGYLSQVVYKLTQLMLQYNAVIVLENLNVGFKRGRTKVEKQVYQKFEKAMID
KLNYLVFKDRGYEMNGSYAKGLQLTDKFESFDKIGKQTGCIYYVIPSYTSHIDPK
TGFVNLLNAKLRYENITKAQDTIRKFDSISYNAKADYFEFAFDYRSFGVDMARN
EWVVCTCGDLRWEYSAKTRETKAYSVTDRLKELFKAHGIDYVGGENLVSHITE
VADKHFLSTLLFYLRLVLKMRYTVSGTENENDFILSPVEYAPGKFFDSREATSTE
PMNADANGAYHIALKGLMTIRGIEDGKLHNYGKGGENAAWFKFMQNQEYKNN
G (SEQ ID NO: 105)

> WP_044910712_(con modificaciones), proteína hipotética [la bacteria Lachnospiraceae MC2017]

MDYGNGQFERRAPLTKTITLRLKPIGETRETIREQKLLEQDAAFRK
LVETVTPIVDDCIRKIADNALCHFGTEYDFSCLGNAISKNDSKAIKKETEKVEKLL
AKVLTENLPDGLRKVNDINSAAFIQDTLTSFVQDDADKRVLIQELKGKTVLMQR
FLTTRITALTVWLPDRVFENFNIFIENAEKMRILLDSPLNEKIMKFDPDAEQYASL
EFYGQCLSQKDIDSYNLIISGIYADDEVKNPGINEIVKEYNQQIRGDKDESPLPKL
KKLHKQILMPVEKAFFVRVLSNDSDARSILEKILKDTEMLPSKIIEAMKEADAGD
IAVYGSRLHELSHVIYGDHGKLSQIIYDKESKRISELMETLSPKERKESKKRLEGL
EEHIRKSTYTFDELNRYAEKNVMAAYIAAVEESCAEIMRKEKDLRTLLSKEDVKI
RGNRHNTLIVKNYFNAWTVFRNLIRILRRKSEAEIDSDFYDVLDDSVEVLSLTYK
GENLCRSYITKKIGSDLKPEIATYGSALRPNSRWWSPGEKFNVKFHTIVRRDGRL
YYFILPKGAKPVELEDMDGDIECLQMRKIPNPTIFLPKLVFKDPEAFFRDNPEADE
FVFLSGMKAPVTITRETYEAYRYKLYTVGKLRDGEVSEEEYKRALLQVLTAYKE
FLENRMIYADLNFGFKDLEEYKDSSEFIKQVETHNTFMCWAKVSSSQLDDLVKS
GNGLLFEIWSERLESYYKYGNEKVLRGYEGVLLSILKDENLVSMRTLLNSRPML
VYRPKESSKPMVVHRDGSRVVDRFDKDGKYIPPEVHDELYRFFNNLLIKEKLGE
KARKILDNKKVKVKVLESERVKWSKFYDEQFAVTFSVKKNADCLDTTKDLNAE
VMEQYSESNRLILIRNTTDILYYLVLDKNGKVLKQRSLNIINDGARDVDWKERF
RQVTKDRNEGYNEWDYSRTSNDLKEVYLNYALKEIAEAVIEYNAILIIEKMSNA
FKDKYSFLDDVTFKGFETKLLAKLSDLHFRGIKDGEPCSFTNPLQLCQNDSNKIL
QDGVIFMVPNSMTRSLDPDTGFIFAINDHNIRTKKAKLNFLSKFDQLKVSSEGCLI
MKYSGDSLPTHNTDNRVWNCCCNHPITNYDRETKKVEFIEEPVEELSRVLEENGI

ETDTELNKLNERENVPGKVVDAIYSLVLNYLRGTVSGVAGQRAVYYSPVTGKK
YDISFIQAMNLNRKCDYYRIGSKERGEWTDFVAQLIN (SEQ ID NO: 106)

> WP_027216152_(con modificaciones), proteína hipotética [Butyrivibrio fibrisolvens]

MYYESLTKLYPIKKTIRNELVPIGKTLENIKKNNILEADEDRKIAYI
RVKAIMDDYHKRLINEALSGFALIDLDKAANLYLSRSKSADDIESFSRFQDKLRK
AIAKRLREHENFGKIGNKDIIPLLQKLSENEDDYNALESFKNFYTYFESYNDVRL
NLYSDKEKSSTVAYRLINENLPRFLDNIRAYDAVQKAGITSEELSSEAQDGLFLV
NTFNNVLIQDGINTYNEDIGKLNVAINLYNQKNASVQGFRKVPKMKVLYKQILS
DREESFIDEFESDTELLDSLESHYANLAKYFGSNKVQLLFTALRESKGVNVYVKN
DIAKTSFSNVVFGSWSRIDELINGEYDDNNNRKKDEKYYDKRQKELKKNKSYTI
EKIITLSTEDVDVIGKYIEKLESDIDDIRFKGKNFYEAVLCGHDRSKKLSKNKGAV
EAIKGYLDSVKDFERDLKLINGSGQELEKNLVVYGEQEAVLSELSGIDSLYNMT
RNYLTKKPFSTEKIKLNFNKPTFLDGWDYGNEEAYLGFFMIKEGNYFLAVMDA
NWNKEFRNIPSVDKSDCYKKVIYKQISSPEKSIQNLMVIDGKTVKKNGRKEKEGI
HSGENLILEELKNTYLPKKINDIRKRRSYLNGDTFSKKDLTEFIGYYKQRVIEYYN
GYSFYFKSDDDYASFKEFQEDVGRQAYQISYVDVPVSFVDDLINSGKLYLFRVY
NKDFSEYSKGRLNLHTLYFKMLFDERNLKNVVYKLNGQAEVFYRPSSIKKEELI
VHRAGEEIKNKNPKRAAQKPTRRLDYDIVKDRRYSQDKFMLHTSIIMNFGAEEN
VSFNDIVNGVLRNEDKVNVIGIDRGERNLLYVVVIDPEGKILEQRSLNCITDSNLD
IETDYHRLLDEKESDRKIARRDWTTIENIKELKAGYLSQVVHIVAELVLKYNAIIC
LEDLNFGFKRGRQKVEKQVYQKFEKMLIDKLNYLVMDKSREQLSPEKISGALN
ALQLTPDFKSFKVLGKQTGIIYYVPAYLTSKIDPMTGFANLFYVKYENVDKAKE
FFSKFDSIKYNKDGKNWNTKGYFEFAFDYKKFTDRAYGRVSEWTVCTVGERIIK
FKNKEKNNSYDDKVIDLTNSLKELFDSYKVTYESEVDLKDAILAIDDPAFYRDLT
RRLQQTLQMRNSSCDGSRDYIISPVKNSKGEFFCSDNNDDTTPNDADANGAFNI
ARKGLWVLNEIRNSEEGSKINLAMSNAQWLEYAQDNTI (SEQ ID NO: 107)

> WP_016301126_(con modificaciones), proteína hipotética [la bacteria Lachnospiraceae COE1]

MHENNGKIADNFIGIYPVSKTLRFELKPVGKTQEYIEKHGILDEDL
KRAGDYKSVKKIIDAYHKYFIDEALNGIQLDGLKNYYELYEKKRDNNEEKEFQK
IQMSLRKQIVKRFSEHPQYKYLFKKELIKNVLPEFTKDNAEEQTLVKSFQEFTTY
FEGFHQNRKNMYSDEEKSTAIAYRVVHQNLPKYIDNMRIFSMILNTDIRSDLTEL

FNNLKTKMDITIVEEYFAIDGFNKVVNQKGIDVYNTILGAFSTDDNTKIKGLNEY
INLYNQKNKAKLPKLKPLFKQILSDRDKISFIPEQFDSDTEVLEAVDMFYNRLLQF
VIENEGQITISKLLTNFSAYDLNKIYVKNDTTISAISNDLFDDWSYISKAVRENYD
SENVDKNKRAAAYEEKKEKALSKIKMYSIEELNFFVKKYSCNECHIEGYFERRIL
EILDKMRYAYESCKILHDKGLINNISLCQDRQAISELKDFLDSIKEVQWLLKPLMI
GQEQADKEEAFYTELLRIWEELEPITLLYNKVRNYVTKKPYTLEKVKLNFYKST
LLDGWDKNKEKDNLGIILLKDGQYYLGIMNRRNNKIADDAPLAKTDNVYRKM
EYKLLTKVSANLPRIFLKDKYNPSEEMLEKYEKGTHLKGENFCIDDCRELIDFFK
KGIKQYEDWGQFDFKFSDTESYDDISAFYKEVEHQGYKITFRDIDETYIDSLVNE
GKLYLFQIYNKDFSPYSKGTKNLHTLYWEMLFSQQNLQNIVYKLNGNAEIFYRK
ASINQKDVVVHKADLPIKNKDPQNSKKESMFDYDIIKDKRFTCDKYQFHVPITM
NFKALGENHFNRKVNRLIHDAENMHIIGIDRGERNLIYLCMIDMKGNIVKQISLN
EIISYDKNKLEHKRNYHQLLKTREDENKSARQSWQTIHTIKELKEGYLSQVIHVI
TDLMVEYNAIVVLEDLNFGFKQGRQKFERQVYQKFEKMLIDKLNYLVDKSKG
MDEDGGLLHAYQLTDEFKSFKQLGKQSGFLYYIPAWNTSKLDPTTGFVNLFYTK
YESVEKSKEFINNFTSILYNQEREYFEFLFDYSAFTSKAEGSRLKWTVCSKGERV
ETYRNPKKNNEWDTQKIDLTFELKKLFNDYSISLLDGDLREQMGKIDKADFYKK
FMKLFALIVQMRNSDEREDKLISPVLNKYGAFFETGKNERMPLDADANGAYNIA
RKGLWIIEKIKNTDVEQLDKVKLTISNKEWLQYAQEHIL (SEQ ID NO. 108)

> WP_035635841_(con modificaciones), proteína hipotética [la bacteria Lachnospiraceae ND2006]

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAE
DYKGVKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEIN
LRKEIAKAFKGNEGYKSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDN
RENMFSEEAKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDY
DVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKL
PKFKPLYKQVLSDRESLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFK
NFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLKKKAVVTEKY
EDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFD
ADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDFV
LAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRA
TILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFF
SKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSN

AYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIY
NKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVV
HPANSPIANKNPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINT
EVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKT
DYHSLLDKKEKERFEARQNWTSIENIKELKAGYISQVVHKICELVEKYDAVIALE
DLNSGFKNSRVKVEKQVYQKFEKMLIDKLNYMVDKKSNPCATGGALKGYQIT
NKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRI
MYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKKNNVFDWEE
VCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNSIT
GRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIG
QFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH (SEQ ID NO: 109)

> WP_015504779_(con modificaciones), exonucleasa SbcC [la bacteria Methanomethylophilus alvus candidata]

MDAKEFTGQYPLSKTLRFELRPIGRTWDNLEASGYLAEDRHRAEC
YPRAKELLDDNHRAFLNRVLPQIDMDWHPIAEAFCKVHKNPGNKELAQDYNLQ
LSKRRKEISAYLQDADGYKGLFAKPALDEAMKIAKENGNESDIEVLEAFNGFSV
YFTGYHESRENIYSDEDMVSVAYRITEDNFPRFVSNALIFDKLNESHPDIISEVSG
NLGVDDIGKYFDVSNYNNFLSQAGIDDYNHIIGGHTTEDGLIQAFNVVLNLRHQ
KDPGFEKIQFKQLYKQILSVRTSKSYIPKQFDNSKEMVDCICDYVSKIEKSETVER
ALKLVRNISSFDLRGIFVNKKNLRILSNKLIGDWDAIETALMHSSSSENDKKSVY
DSAEAFTLDDIFSSVKKFSDASAEDIGNRAEDICRVISETAPFINDLRAVDLDSLN
DDGYEAAVSKIRESLEPYMDLFHELEIFSVGDEFPKCAAFYSELEEVSEQLIEIIPL
FNKARSFCTRKRYSTDKIKVNLKFPTLADGWDLNKERDNKAAILRKDGKYYLAI
LDMKKDLSSIRTSDEDESSFEKMEYKLLPSPVKMLPKIFVKSKAAKEKYGLTDR
MLECYDKGMHKSGSAFDLGFCHELIDYYKRCIAEYPGWDVFDFKFRETSDYGS
MKEFNEDVAGAGYYMSLRKIPCSEVYRLLDEKSIYLFQIYNKDYSENAHGNKN
MHTMYWEGLFSPQNLESPVFKLSGGAELFFRKSSIPNDAKTVHPKGSVLVPRND
VNGRRIPDSIYRELTRYFNRGDCRISDEAKSYLDKVKTKKADHDIVKDRRFTVD
KMMFHVPIAMNFKAISKPNLNKKVIDGIIDDQDLKIIGIDRGERNLIYVTMVDRK
GNILYQDSLNILNGYDYRKALDVREYDNKEARRNWTKVEGIRKMKEGYLSLAV
SKLADMIIENNAIIVMEDLNHGFKAGRSKIEKQVYQKFESMLINKLGYMVLKDK
SIDQSGGALHGYQLANHVTTLASVGKQCGVIFYIPAAFTSKIDPTTGFADLFALS
NVKNVASMREFFSKMKSVIYDKAEGKFAFTFDYLDYNVKSECGRTLWTVYTVG

ERFTYSRVNREYVRKVPTDIIYDALQKAGISVEGDLRDRIAESDGDTLKSIFYAFK
YALDMRVENREEDYIQSPVKNASGEFFCSKNAGKSLPQDSDANGAYNIALKGIL
QLRMLSEQYDPNAESIRLPLITNKAWLTFMQSGMKTWKN (SEQ ID NO: 110)

> WP_044910713_(con modificaciones), proteína hipotética [la bacteria Lachnospiraceae MC2017]

MGLYDGFVNRYSVSKTLRFELIPQGRTREYIETNGILSDDEERAKD
YKTIKRLIDEYHKDYISRCLKNVNISCLEEYYHLYNSSNRDKRHEELDALSDQMR
GEIASFLTGNDEYKEQKSRDIIINERIINFASTDEELAAVKRFRKFTSYFTGFFTNR
ENMYSAEKKSTAIAHRIIDVNLPKYVDNIKAFNTAIEAGVFDIAEFESNFKAITDE
HEVSDLLDITKYSRFIRNEDIIIYNTLLGGISMKDEKIQGLNELINLHNQKHPGKK
VPLLKVLYKQILGDSQTHSFVDDQFEDDQQVINAVKAVTDTFSETLLGSLKIIINN
IGHYDLDRIYIKAGQDITTLSKRALNDWHIITECLESEYDDKFPKNKKSDTYEEM
RNRYVKSFKSFSIGRLNSLVTTYTEQACFLENYLGSFGGDTDKNCLTDFTNSLME
VEHLLNSEYPVTNRLITDYESVRILKRLLDSEMEVIHFLKPLLGNGNESDKDLVF
YGEFEAEYEKLLPVIKVYNRVRNYLTRKPFSTEKIKLNFNSPTLLCGWSQSKEKE
YMGVILRKDGQYYLGIMTPSNKKIFSEAPKPDEDCYEKMVLRYIPHPYQMLPKV
FFSKSNIAFFNPSDEILRIKKQESFKKGKSFNRDDCHKFIDFYKDSINRHEEWRKF
NFKFSDTDSYEDISRFYKEVENQAFSMSFTKIPTVYIDSLVDEGKLYLFKLHNKD
FSEHSKGKPNLHTVYWNALFSEYNLQNTVYQLNGSAEIFFRKASIPENERVIHKK
NVPITRKVAELNGKKEVSVFPYDIIKNRRYTVDKFQFHVPLKMNFKADEKKRIN
DDVIEAIRSNKGIHVIGIDRGERNLLYLSLINEEGRIIEQRSLNIIDSGEGHTQNYRD
LLDSREKDREKARENWQEIQEIKDLKTGYLSQAIHTITKWMKEYNAIIVLEDLND
RFTNGRKKVEKQVYQKFEKMLIDKLNYYVDKDEEFDRMGGTHRALQLTEKFES
FQKLGRQTGFIFYVPAWNTSKLDPTTGFVDLLYPKYKSVDATKDFIKKFDFIRFN
SEKNYFEFGLHYSNFTERAIGCRDEWILCSYGNRIVNFRNAAKNNSWDYKEIDIT
KQLLDLFEKNGIDVKQENLIDSICEMKDKPFFKSLIANIKLILQIRNSASGTDIDYM
ISPAMNDRGEFFDTRKGLQQLPLDADANGAYNIAKKGLWIVDQIRNTTGNNVK
MAMSNREWMHFAQESRLA (SEQ ID NO: 111)

> KKQ36153_(con modificaciones), proteína hipotética US52_C0007G0008 [la bacteria candidata de la división WS6, GW2011_GWA2_37_6]

MKNVFGGFTNLYSLTKTLRFELKPTSKTQKLMKRNNVIQTDEEID
KLYHDEMKPILDEIHRRFINDALAQKIFISASLDNFLKVVKNYKVESAKKNIKQN
QVKLLQKEITIKTLGLRREVVSGFITVSKKWKDKYVGLGIKLKGDGYKVLTEQA

VLDILKIEFPNKAKYIDKFRGFWTYFSGFNENRKNYYSEEDKATSIANRIVNENLS
RYIDNIIAFEEILQKIPNLKKFKQDLDITSYNYYLNQAGIDKYNKIIGGYIVDKDKK
IQGINEKVNLYTQQTKKKLPKLKFLFKQIGSERKGFGIFEIKEGKEWEQLGDLFK
LQRTKINSNGREKGLFDSLRTMYREFFDEIKRDSNSQARYSLDKIYFNKASVNTIS
NSWFTNWNKFAELLNIKEDKKNGEKKIPEQISIEDIKDSLSIIPKENLEELFKLTNR
EKHDRTRFFGSNAWVTFLNIWQNEIEESFNKLEEKEKDFKKNAAIKFQKNNLVQ
KNYIKEVCDRMLAIERMAKYHLPKDSNLSREEDFYWIIDNLSEQREIYKYYNAF
RNYISKKPYNKSKMKLNFENGNLLGGWSDGQERNKAGVILRNGNKYYLGVLIN
RGIFRTDKINNEIYRTGSSKWERLILSNLKFQTLAGKGFLGKHGVSYGNMNPEKS
VPSLQKFIRENYLKKYPQLTEVSNTKFLSKKDFDAAIKEALKECFTMNFINIAEN
KLLEAEDKGDLYLFEITNKDFSGKKSGKDNIHTIYWKYLFSESNCKSPIIGLNGG
AEIFFREGQKDKLHTKLDKKGKKVFDAKRYSEDKLFFHVSITINYGKPKNIKFRD
IINQLITSMNVNIIGIDRGEKHLLYYSVIDSNGIILKQGSLNKIRVGDKEVDFNKKL
TERANEMKKARQSWEQIGNIKNFKEGYLSQAIHEIYQLMIKYNAIIVLEDLNTEF
KAKRLSKVEKSVYKKFELKLARKLNHLILKDRNTNEIGGVLKAYQLTPTIGGGD
VSKFEKAKQWGMMFYVRANYTSTTDPVTGWRKHLYISNFSNNSVIKSFFDPTN
RDTGIEIFYSGKYRSWGFRYVQKETGKKWELFATKELERFKYNQTTKLCEKINL
YDKFEELFKGIDKSADIYSQLCNVLDFRWKSLVYLWNLLNQIRNVDKNAEGNK
NDFIQSPVYPFFDSRKTDGKTEPINGDANGALNIARKGLMLVERIKNNPEKYEQL
IRDTEWDAWIQNFNKVN (SEQ ID NO: 112)

> WP_044919442_(con modificaciones), proteína hipotética [la bacteria Lachnospiraceae MA2020]

MYYESLTKQYPVSKTIRNELIPIGKTLDNIRQNNILESDVKRKQNY
EHVKGILDEYHKQLINEALDNCTLPSLKIAAEIYLKNQKEVSDREDFNKTQDLLR
KEVVEKLKAHENFTKIGKKDILDLLEKLPSISEDDYNALESFRNFYTYFTSYNKV
RENLYSDKEKSSTVAYRLINENFPKFLDNVKSYREVKTAGILADGLGEEEQDSLF
IVETFNKTLTQDGIDTYNSQVGKINSSINLYNQKNQKANGFRKIPKMKMLYKQIL
SDREESFIDEFQSDEVLIDNVESYGSVLIESLKSSKVSAFFDALRESKGKNVYVKN
DLAKTAMSNIVFENWRTFDDLLNQEYDLANENKKKDDKYFEKRQKELKKNKS
YSLEHLCNLSEDSCNLIENYIHQISDDIENIINNETFLRIVINEHDRSRKLAKNRKA
VKAIKDFLDSIKVLERELKLINSSGQELEKDLIVYSAHEELLVELKQVDSLYNMT
RNYLTKKPFSTEKVKLNFNRSTLLNGWDRNKETDNLGVLLLKDGKYYLGIMNT
SANKAFVNPPVAKTEKVFKKVDYKLLPVPNQMLPKVFFAKSNIDFYNPSSEIYSN

YKKGTHKKGNMFSLEDCHNLIDFFKESISKHEDWSKFGFKFSDTASYNDISEFYR
EVEKQGYKLTYTDIDETYINDLIERNELYLFQIYNKDFSMYSKGKLNLHTLYFM
MLFDQRNIDDVVYKLNGEAEVFYRPASISEDELIIHKAGEEIKNKNPNRARTKET
STFSYDIVKDKRYSKDKFTLHIPITMNFGVDEVKRFNDAVNSAIRIDENVNVIGID
RGERNLLYVVVIDSKGNILEQISLNSIINKEYDIETDYHALLDEREGGRDKARKD
WNTVENIRDLKAGYLSQVVNVVAKLVLKYNAIICLEDLNFGFKRGRQKVEKQV
YQKFEKMLIDKLNYLVIDKSREQTSPKELGGALNALQLTSKFKSFKELGKQSGVI
YYVPAYLTSKIDPTTGFANLFYMKCENVEKSKRFFDGFDFIRFNALENVFEFGFD
YRSFTQRACGINSKWTVCTNGERIIKYRNPDKNNMFDEKVVVVTDEMKNLFEQ
YKIPYEDGRNVKDMIISNEEAEFYRRLYRLLQQTLQMRNSTSDGTRDYIISPVKN
KREAYFNSELSDGSVPKDADANGAYNIARKGLWVLEQIRQKSEGEKINLAMTN
AEWLEY AQTHLL (SEQ ID NO: 113)

> WP_035798880_(con modificaciones), proteína hipotética [Butyrivibrio sp. NC3005]

MYYQNLTKKYPVSKTIRNELIPIGKTLENIRKNNILESDVKRKQDY
EHVKGIMDEYHKQLINEALDNYMLPSLNQAAEIYLKKHVDVEDREEFKKTQDL
LRREVTGRLKEHENYTKIGKKDILDLLEKLPSISEEDYNALESFRNFYTYFTSYNK
VRENLYSDEEKSSTVAYRLINENLPKFLDNIKSYAFVKAAGVLADCIEEEEQDAL
FMVETFNMTLTQEGIDMYNYQIGKVNSAINLYNQKNHKVEEFKKIPKMKVLYK
QILSDREEVFIGEFKDDETLLSSIGAYGNVLMTYLKSEKINIFFDALRESEGKNVY
VKNDLSKTTMSNIVFGSWSAFDELLNQEYDLANENKKKDDKYFEKRQKELKKN
KSYTLEQMSNLSKEDISPIENYIERISEDIEKICIYNGEFEKIVVNEHDSSRKLSKNI
KAVKVIKDYLDSIKELEHDIKLINGSGQELEKNLVVYVGQEEALEQLRPVDSLYN
LTRNYLTKKPFSTEKVKLNFNKSTLLNGWDKNKETDNLGILFFKDGKYYLGIMN
TTANKAFVNPPAAKTENVFKKVDYKLLPGSNKMLPKVFFAKSNIGYYNPSTELY
SNYKKGTHKKGPSFSIDDCHNLIDFFKESIKKHEDWSKFGFEFSDTADYRDISEFY
REVEKQGYKLTFTDIDESYINDLIEKNELYLFQIYNKDFSEYSKGKLNLHTLYFM
MLFDQRNLDNVVYKLNGEAEVFYRPASIAENELVIHKAGEGIKNKNPNRAKVK
ETSTFSYDIVKDKRYSKYKFTLHIPITMNFGVDEVRRFNDVINNALRTDDNVNVI
GIDRGERNLLYVVVINSEGKILEQISLNSIINKEYDIETNYHALLDEREDDRNKAR
KDWNTIENIKELKTGYLSQVVNVVAKLVLKYNAIICLEDLNFGFKRGRQKVEKQ
VYQKFEKMLIEKLNYLVIDKSREQVSPEKMGGALNALQLTSKFKSFAELGKQSG
IIYYVPAYLTSKIDPTTGFVNLFYIKYENIEKAKQFFDGFDFIRFNKKDDMFEFSFD

YKSFTQKACGIRSKWIVYTNGERIIKYPNPEKNNLFDEKVINVTDEIKGLFKQYRI
PYENGEDIKEIIISKAEADFYKRLFRLLHQTLQMRNSTSDGTRDYIISPVKNDRGE
FFCSEFSEGTMPKDADANGAYNIARKGLWVLEQIRQKDEGEKVNLSMTNAEWL
KYAQLHLL (SEQ ID NO: 114)

> WP_027109509_(con modificaciones), proteína hipotética [la bacteria Lachnospiraceae NC2008]

MENYYDSLTRQYPVTKTIRQELKPVGKTLENIKNAEIIEADKQKKE
AYVKVKELMDEFHKSIIEKSLVGIKLDGLSEFEKLYKIKTKTDEDKNRISELFYY
MRKQIADALKNSRDYGYVDNKDLIEKILPERVKDENSLNALSCFKGFTTYFTDY
YKNRKNIYSDEEKHSTVGYRCINENLLIFMSNIEVYQIYKKANIKNDNYDEETLD
KTFMIESFNECLTQSGVEAYNSVVASIKTATNLYIQKNNKEENFVRVPKMKVLF
KQILSDRTSLFDGLIIESDDELLDKLCSFSAEVDKFLPINIDRYIKTLMDSNNGTGI
YVKNDSSLTTLSNYLTDSWSSIRNAFNENYDAKYTGKVNDKYEEKREKAYKSN
DSFELNYIQNLLGINVIDKYIERINFDIKEICEAYKEMTKNCFEDHDKTKKLQKNI
KAVASIKSYLDSLKNIERDIKLLNGTGLESRNEFFYGEQSTVLEEITKVDELYNIT
RNYLTKKPFSTEKMKLNFNNPQLLGGWDVNKERDCYGVILIKDNNYYLGIMDK
SANKSFLNIKESKNENAYKKVNCKLLPGPNKMFPKVFFAKSNIDYYDPTHEIKKL
YDKGTFKKGNSFNLEDCHKLIDFYKESIKKNDDWKNFNFNFSDTKDYEDISGFF
REVEAQNYKITYTNVSCDFIESLVDEGKLYLFQIYNKDFSEYATGNLNLHTLYLK
MLFDERNLKDLCIKMNGEAEVFYRPASILDEDKVVHKANQKITNKNTNSKKKES
IFSYDIVKDKRYTVDKFFIHLPITLNYKEQNVSRFNDYIREILKKSKNIRVIGIDRG
ERNLLYVVVCDSDGSILYQRSINEIVSGSHKTDYHKLLDNKEKERLSSRRDWKTI
ENIKDLKAGYMSQVVNEIYNLILKYNAIVVLEDLNIGFKNGRKKVEKQVYQNFE
KALIDKLNYLCIDKTREQLSPSSPGGVLNAYQLTAKFESFEKIGKQTGCIFYVPAY
LTSQIDPTTGFVNLFYQKDTSKQGLQLFFRKFKKINFDKVASNFEFVFDYNDFTN
KAEGTKTNWTISTQGTRIAKYRSDDANGKWISRTVHPTDIIKEALNREKINYNDG
HDLIDEIVSIEKSAVLKEIYYGFKLTLQLRNSTLANEEEQEDYIISPVKNSSGNYFD
SRITSKELPCDADANGAYNIARKGLWALEQIRNSENVSKVKLAISNKEWFEYTQ
NNIPSL (SEQ ID NO: 1581)

> WP_029202018_(con modificaciones), proteína hipotética [Oribacterium sp. NK2B42]

MYYDGLTKQYALSKTIRNELVPIGKTLDNIKKNRILEADIKRKSDY
EHVKKLMDMYHKKIINEALDNFKLSVLEDAADIYFNKQNDERDIDAFLKIQDKL

RKEIVEQLKGHTDYSKVGNKDFLGLLKAASTEEDRILIESFDNFYTYFTSYNKVR
SNLYSAEDKSSTVAYRLINENLPKFFDNIKAYRTVRNAGVISGDMSIVEQDELFE
VDTFNHTLTQYGIDTYNHMIGQLNSAINLYNQKMHGAGSFKKLPKMKELYKQL
LTEREEEFIEEYTDDEVLITSVHNYVSYLIDYLNSDKVESFFDTLRKSDGKEVFIK
NDVSKTTMSNILFDNWSTIDDLINHEYDSAPENVKKTKDDKYFEKRQKDLKKN
KSYSLSKIAALCRDTTILEKYIRRLVDDIEKIYTSNNVFSDIVLSKHDRSKKLSKNT
NAVQAIKNMLDSIKDFEHDVMLINGSGQEIKKNLNVYSEQEALAGILRQVDHIY
NLTRNYLTKKPFSTEKIKLNFNRPTFLDGWDKNKEEANLGILLIKDNRYYLGIMN
TSSNKAFVNPPKAISNDIYKKVDYKLLPGPNKMLPKVFFATKNIAYYAPSEELLS
KYRKGTHKKGDSFSIDDCRNLIDFFKSSINKNTDWSTFGFNFSDTNSYNDISDFY
REVEKQGYKLSFTDIDACYIKDLVDNNELYLFQIYNKDFSPYSKGKLNLHTLYFK
MLFDQRNLDNVVYKLNGEAEVFYRPASIESDEQIIHKSGQNIKNKNQKRSNCKK
TSTFDYDIVKDRRYCKDKFMLHLPITVNFGTNESGKFNELVNNAIRADKDVNVI
GIDRGERNLLYVVVVDPCGKIIEQISLNTIVDKEYDIETDYHQLLDEKEGSRDKA
RKDWNTIENIKELKEGYLSQVVNIIAKLVLKYDAIICLEDLNFGFKRGRQKVEKQ
VYQKFEKMLIDKMNYLVLDKSRKQESPQKPGGALNALQLTSAFKSFKELGKQT
GIIYYVPAYLTSKIDPTTGFANLFYIKYESVDKARDFFSKFDFIRYNQMDNYFEFG
FDYKSFTERASGCKSKWIACTNGERIVKYRNSDKNNSFDDKTVILTDEYRSLFDK
YLQNYIDEDDLKDQILQIDSADFYKNLIKLFQLTLQMRNSSSDGKRDYIISPVKN
YREEFFCSEFSDDTFPRDADANGAYNIARKGLWVIKQIRETKSGTKINLAMSNSE
WLEYAQCNLL (SEQ ID NO: 115)

> WP_028248456_(con modificaciones), proteína hipotética [Pseudobutyrivibrio ruminis]

MYYQNLTKMYPISKTLRNELIPVGKTLENIRKNGILEADIQRKADY
EHVKKLMDNYHKQLINEALQGVHLSDLSDAYDLY
FNLSKEKNSVDAFSKCQDKLRKEIVSLLKNHENFPKIGNKEIIKLL
QSLYDNDTDYKALDSFSNFYTYFSSYNEVRKNLYSDEEKSSTVAYRLINENLPKF
LDNIKAYAIAKKAGVRAEGLSEEDQDCLFIIETFERTLTQDGIDNYNAAIGKLNT
AINLFNQQNKKQEGFRKVPQMKCLYKQILSDREEAFIDEFSDDEDLITNIESFAEN
MNVFLNSEIITDFKIALVESDGSLVYIKNDVSKTSFSNIVFGSWNAIDEKLSDEYD
LANSKKKKDEKYYEKRQKELKKNKSYDLETIIGLFDDNSDVIGKYIEKLESDITAI
AEAKNDFDEIVLRKHDKNKSLRKNTNAVEAIKSYLDTVKDFERDIKLINGSGQE
VEKNLVVYAEQENILAEIKNVDSLYNMSRNYLTQKPFSTEKFKLNFNRATLLNG

WDKNKETDNLGILFEKDGMYYLGIMNTKANKIFVNIPKATSNDVYHKVNYKLL
PGPNKMLPKVFFAQSNLDYYKPSEELLAKYKAGTHKKGDNFSLEDCHALIDFFK
ASIEKHPDWSSFGFEFSETCTYEDLSGFYREVEKQGYKITYTDVDADYITSLVER
DELYLFQIYNKDFSPYSKGNLNLHTIYLQMLFDQRNLNNVVYKLNGEAEVFYRP
ASINDEEVIIHKAGEEIKNKNSKRAVDKPTSKFGYDIIKDRRYSKDKFMLHIPVTM
NFGVDETRRFNDVVNDALRNDEKVRVIGIDRGERNLLYVVVVDTDGTILEQISL
NSIINNEYSIETDYHKLLDEKEGDRDRARKNWTTIENIKELKEGYLSQVVNVIAK
LVLKYNAIICLEDLNFGFKRGRQKVEKQVYQKFEKMLIDKLNYLVIDKSRKQDK
PEEFGGALNALQLTSKFTSFKDMGKQTGIIYYVPAYLTSKIDPTTGFANLFYVKY
ENVEKAKEFFSRFDSISYNNESGYFEFAFDYKKFTDRACGARSQWTVCTYGERII
KFRNTEKNNSFDDKTIVLSEEFKELFSIYGISYEDGAELKNKIMSVDEADFFRSLT
RLFQQTMQMRNSSNDVTRDYIISPIMNDRGEFFNSEACDASKPKDADANGAFNI
ARKGLWVLEQIRNTPSGDKLNLAMSNAEWLEYAQRNQI (SEQ ID NO: 116)

> WP_028830240_(con modificaciones), proteína hipotética [Proteocatella sphenisci]

MENFKNLYPINKTLRFELRPYGKTLENFKKSGLLEKDAFKANSRR
SMQAIIDEKFKETIEERLKYTEFSECDLGNMTSKDKKITDKAATNLKKQVILSFD
DEIFNNYLKPDKNIDALFKNDPSNPVISTFKGFTTYFVNFFEIRKHIFKGESSGSMA
YRIIDENLTTYLNNIEKIKKLPEELKSQLEGIDQIDKLNNYNEFITQSGITHYNEIIG
GISKSENVKIQGINEGINLYCQKNKVKLPRLTPLYKMILSDRVSNSFVLDTIENDT
ELIEMISDLINKTEISQDVIMSDIQNIFIKYKQLGNLPGISYSSIVNAICSDYDNNFG
DGKRKKSYENDRKKHLETNVYSINYISELLTDTDVSSNIKMRYKELEQNYQVCK
ENFNATNWMNIKNIKQSEKTNLIKDLLDILKSIQRFYDLFDIVDEDKNPSAEFYT
WLSKNAEKLDFEFNSVYNKSRNYLTRKQYSDKKIKLNFDSPTLAKGWDANKEI
DNSTIIMRKFNNDRGDYDYFLGIWNKSTPANEKIIPLEDNGLFEKMQYKLYPDPS
KMLPKQFLSKIWKAKHPTTPEFDKKYKEGRHKKGPDFEKEFLHELIDCFKHGLV
NHDEKYQDVFGFNLRNTEDYNSYTEFLEDVERCNYNLSFNKIADTSNLINDGKL
YVFQIWSKDFSIDSKGTKNLNTIYFESLFSEENMIEKMFKLSGEAEIFYRPASLNY
CEDIIKKGHHHAELKDKFDYPIIKDKRYSQDKFFFHVPMVINYKSEKLNSKSLNN
RTNENLGQFTHIIGIDRGERHLIYLTVVDVSTGEIVEQKHLDEIINTDTKGVEHKT
HYLNKLEEKSKTRDNERKSWEAIETIKELKEGYISHVINEIQKLQEKYNALIVME
NLNYGFKNSRIKVEKQVYQKFETALIKKFNYIIDKKDPETYIIHGYQLTNPITTLDK
IGNQSGIVLYIPAWNTSKIDPVTGFVNLLYADDLKYKNQEQAKSFIQKIDNIYFEN

GEFKFDIDFSKWNNRYSISKTKWTLTSYGTRIQTFRNPQKNNKWDSAEYDLTEE
FKLILNIDGTLKSQDVETYKKFMSLFKLMLQLRNSVTGTDIDYMISPVTDKTGTH
FDSRENIKNLPADADANGAYNIARKGIMAIENIMNGISDPLKISNEDYLKYIQNQ
QE (SEQ ID NO: 117)

Los solicitantes generaron construcciones con vectores como los que se representan en las figuras 40A-L (por ejemplo, PACYC184 fnCpf1 (PY001)) y en las figuras 41A-E (por ejemplo, PaCpf1).

Estudio con una exposición a PAM para la detección de las secuencias de PAM potenciales en FnCpf1 (figura 42). Los solicitantes aislaron los loci relacionados con Cpf1 de *Francisella novicida* (Fn) (figura 43) y los introdujeron en *E. coli* por medio de un procedimiento de transformación. Los loci fueron expresados en *E. coli* a partir de pACYC184, de manera similar al experimento que describieron Sapranauskas et al.

E. coli con el locus pACYC-FnCpf1 = positiva para Cpf1

E. coli con vector de vacío pACYC184 = control

Los solicitantes transformaron tanto las bacterias *E. coli* positivas para Cpf1 como las bacterias *E. coli* de control con los plásmidos apropiados para generar las bibliotecas de PAM. De esta manera, se obtuvieron dos bibliotecas de PAM (figura 44). Las bibliotecas de PAM comprendieron plásmidos pUC19 que contenían una secuencia correspondiente a un protoseparador de 31 pares de bases que coincidía con el separador 1 en el locus FnCpf1. La biblioteca de PAM de la izquierda comprendía 8 nucleótidos degenerados provenientes de los PAM en el extremo 5' del protoseparador. La biblioteca de PAM de la derecha comprendía 7 nucleótidos degenerados provenientes de los PAM en el extremo 3' del protoseparador. Los solicitantes sembraron las bacterias *E. coli* positivas para Cpf1 y las bacterias *E. coli* de control y recolectaron todas las colonias después de ~12 horas. Cada colonia representó un evento de transformación de PAM-pUC19 que no resultó en el corte ni la interferencia por parte de Cpf1. Estos plásmidos PAM-pUC19 no comprendieron PAM reconocibles. A partir de la secuenciación de todas las colonias, los solicitantes determinaron que los plásmidos PAM-pUC19 ya no estaban presentes, en comparación con el control, y comprobaron que estos plásmidos contenían PAM reconocibles.

Clonación de pY0001. pY0001 es la cadena principal de pACYC184 (de NEB) y comprende un locus FnCpf1 parcial. pY0001 contiene el locus endógeno FnCpf1, que comprende una secuencia de 255 pares de bases, que proviene del extremo 3' de una acetiltransferasa, y se extiende hasta la cuarta secuencia separadora. Solamente los separadores 1-3 son potencialmente activos, ya que el separador 4 no está rodeado por repeticiones directas.

Los solicitantes amplificaron por medio de una PCR el locus FnCpf1 en 3 trozos y lo clonaron en pACYC184 cortado con Xba1 y Hind3 de acuerdo con el método de Gibson.

Análisis informático para buscar PAM Cpf1

Después de realizar la búsqueda con el ADN, los solicitantes extrajeron las regiones correspondientes a la izquierda o la derecha de los PAM. Para cada muestra, se comparó la cantidad de PAM presentes en la biblioteca secuenciada con la cantidad de PAM esperada para la biblioteca (4^8 para la biblioteca de la izquierda y 4^7 para la biblioteca de la derecha).

En la biblioteca de la izquierda, fue posible observar el agotamiento de los PAM. Para cuantificar este agotamiento, los solicitantes calcularon una proporción de enriquecimiento. Bajo dos condiciones diferentes (en presencia del plásmido de control pACYC solo o en presencia de pACYC en combinación con FnCpf1), los solicitantes calcularon la proporción de cada PAM en la biblioteca según se detalla a continuación:

$$\text{Proporción} = -\log_2 \frac{\text{muestra} + 0{,}01}{\text{biblioteca inicial} + 0{,}01}$$

A partir de la representación de la distribución, los solicitantes determinaron que la muestra de control se caracterizó por un enriquecimiento, mientras que el enriquecimiento en las dos réplicas biológicas fue evidente. Los solicitantes recolectaron todos los PAM con una proporción de 8 o más y representaron la distribución de la frecuencia en la que estuvieron presentes, con lo pudieron establecer la presencia de PAM 5'-TTN (figuras 45A-E). Los solicitantes confirmaron que el PAM perteneció al tipo TTN, donde N es A/C/G o T.

Los solicitantes realizaron una secuenciación del ARN en el locus Cpf1 de *Francisella tularensis*, y mediante el análisis de la secuencia del ARN, demostraron que el locus CRISPR se expresó de manera activa (figura 46). En la figura 86 se provee una representación adicional del análisis de la secuencia del ARN del locus FnCpf1. Además de los genes Cpf1 y Cas, se observaron dos transcriptos no codificantes pequeños con una frecuencia alta, y los solicitantes conjeturaron eran ARNcrtrac potenciales. También se expresó un arreglo CRISPR. Tanto los ARNcrtrac potenciales como el arreglo CRISPR se transcribieron en la misma dirección que los genes Cpf1 y Cas. En este caso, se determinó

que todos los tipos de ARN transcripto que se identificaron a partir del experimento de secuenciación del ARN provinieron de los loci que se analizaron. Al centrarse en el arreglo CRISPR de Cpf1, los solicitantes pudieron identificar diversos transcriptos cortos. En este conjunto, se comprobó que todos los tipos de ARN transcripto que se identificaron provinieron de los loci relacionados con Cpf1 (figura 47). Después de seleccionar los transcriptos que tenían una longitud inferior a 85 nucleótidos, los solicitantes identificaron dos tipos de ARNcrtrac potencial (figura 48). En la figura 49 se provee una representación centrada en el ARNcrtrac potencial 1 y en el arreglo CRISPR. La figura 50 se provee una representación centrada en el ARNcrtrac potencial 2. Las secuencias potenciales del ARNcr se representan en la figura 51.

Los solicitantes evaluaron la función en células de mamíferos usando productos de una PCR que comprendían un separador U6 (DR-separador-DR) (en ciertos aspectos, los separadores pueden ser conocidos como ARNcr, ARN de guía o con términos análogos como los que se usan en la presente) e identificaron otros loci Cpf1.

Ejemplo 4. Otros experimentos para convalidar FnCpf1

In vivo, los solicitantes confirmaron que el PAM predicho para FnCpf1 fue del tipo TTN de acuerdo con el estudio que se representa en la figura 52. Los solicitantes introdujeron FnCpf1 por medio de un procedimiento de transformación en diversas células y como control usaron pUC19 con un separador endógeno 1 y un PAM 5'-TTN (figura 53). En pocas palabras, en el estudio para confirmar los PAM in vivo, se transformaron 50 µl de *E. coli* competentes con el locus FnCpf1 (en la cepa de prueba) o con pACYC184 vacío (en la cepa de control), en combinación con 10 ng de un plásmido que contenía un protoseparador 1. Antes de la secuencia de protoseparador, se colocaron las secuencias de predichas para los PAM (TTC, TTG, TTA y TTT). Después, las células transformadas se sometieron a una dilución de 1:2000 y se sembraron en placas con agar LB que contenían ampicilina y cloranfenicol. Solamente las células con el plásmido y el protoseparador intactos pudieron formar colonias. Las placas con las colonias se fotografiaron ~14 h después de la siembra, y las colonias se contaron usando el software ImageJ.

Los solicitantes realizaron el lisado de las células para convalidar con mayor detalle el locus FnCpf1. El protocolo para el estudio basado en el lisado de las células se provee a continuación.

Reacción de clivaje in vitro. Amortiguador de disociación: HEPES 100 mM, pH 7,5, KCl 500 mM, $MgCl_2$ 25 mM, DTT 5 mM, 25% de glicerol. La reacción puede llevarse a cabo sin DTT.

Elaboración del lisado de células

Amortiguador de lisis: Hepes 20 mM, pH 7,5, cloruro de potasio [KCl] 100 mM, cloruro de magnesio [$MgCl_2$] 5 mM, ditiotreitol [DTT] 1 mM, 5% de glicerol, 0,1% de Triton X-100, con la adición del cóctel inhibidor de proteasas de Roche 10X. La solución madre concentrada del amortiguador de lisis puede mantenerse sin el inhibidor de proteasas de Roche y sin DTT. El almacenamiento debe realizarse a -20 ºC.

Las células HEK se transfectaron con la cantidad recomendada de ADN con el reactivo Lipofectamine 2000

- 500 ng por 24 cavidades
- 2000 ng por 6 cavidades

Las células se cosechan con el amortiguador de lisis 24-72 horas después de la transfección

- Se aspiran los medios
- Se realiza un lavado suave con DPBS
- Se aspira el DPBS
- Se usan 50 µl de un amortiguador de lisis por 24 cavidades o bien por 250 µl por 6 cavidades
- Se deja reposar el producto en hielo durante 5 minutos
- Se coloca el producto en un tubo Eppendorf
- Se coloca el producto en hielo durante 15 minutos
- Se aplica ultrasonido a una potencia elevada, con un ciclo de trabajo de 50% durante 5-10 minutos
- Se centrifuga en frío a la velocidad máxima durante 20 minutos
- Se transfiere el sobrenadante a un nuevo tubo
- Se toman alícuotas de 10 µl en tubos para PCR y se las congela a -80 ºC

Transcripción *in vitro* del ARN de guía

Protocolo del conjunto de elementos. La información puede consultarse en la página de Internet www.neb.com/products/e2030-hiscribe-t7-in-vitro-transcription-kit.

Tomar solución madre de los oligonucleótidos 100 µM

Hibridar en 10 ul de reacción:

1 µl de T7 la cadena "directa" = "XRP2649"

1 µl de T7 del oligonucleótido "inverso"

1 µl de un amortiguador TaqB

7 µl de agua

Se ejecuta el programa de la PCR sin PNK, con una incubación a 37 ºC (básicamente se calienta hasta 95 ºC durante 5 minutos y se enfría lentamente hasta 4 ºC, aunque no tan lentamente como podría hacérselo si se usara la nucleasa Surveyor). En el caso de los oligonucleótidos NanoDrop alineados: se realiza una normalización con agua a razón de 500 ng/µl (por lo general, se usan 1000-2000 ng/µl para un oligonucleótido de 120 nucleótidos)

La transcripción de T7 se monitorea de acuerdo con las instrucciones del conjunto de elementos (pero con un tamaño reducido al cuarto)

Reacción de 10 ul

1 µl de un amortiguador 10x

1 µl de la transcriptasa T7

0,5 µl de rNTP

0,5 µl de una mezcla HMW

1 µl del ADN que se usa como molde (alineado)

6 µl de agua

Se lleva a cabo una transcripción a 42 ºC (preferiblemente en un termociclador) durante al menos 2-3 horas, se la deja proceder durante la noche. El rendimiento debería ser de aproximadamente 1000-2000 ng/µl de ARN. Es normal que se formen residuos blancos.

Preparación del ADN

En el caso de pUC19, linealizar con HindIII y purificar en una columna

→ se necesitarán 300-400 ng del plásmido por reacción, por lo cual debe separarse la cantidad necesaria

Para obtener el ADNg, se amplifica el ADN a partir de las células por medio de una PCR

→ se ponen en práctica varias PCR, se combina el producto y se lo purifica en una columna

→ se concentra el producto hasta obtener aproximadamente 100-200 ng/µl

Se mantiene a -20 ºC

Reacción de 20 µl

10 µl del lisado (dividido con anterioridad en alícuotas)

2 µl del amortiguador de disociación (el amortiguador NEB 3)

1 µl del ARN (directamente desde el paso anterior, no es necesario purificarlo)

1 µl del ADN (desde el paso anterior)

6 µl de agua

Se incuba a 37 ºC durante 1-2 horas (30 minutos son suficientes)

Reacción de purificación en una columna

Se lleva a cabo en un gel E al 2%

En el estudio basado en el lisado de las células, el ARNcrtrac se usó en las posiciones 1, 2, 3, 4 y 5, según se indica en la figura 54. En el estudio basado en el lisado de las células (1) (figura 55), se usó un gel en el que se observó que el fragmento que se obtuvo por medio de la PCR era un PAM TTa y que había una secuencia propia de un protoseparador 1 una vez incubado en el lisado de células. En el estudio basado en el lisado de las células (2) (figura 56), se usó un gel en el que se observó pUC en combinación con el separador 1 y diversos PAM una vez incubado en el lisado de células. En el estudio basado en el lisado de las células (3) (figura 57), se usó un gel que fue sometido a una digestión con BasI una vez incubado en el lisado de células. En el gel que se obtuvo a partir del estudio basado en el lisado de las células (4) (figura 58), fue posible observar la digestión de tres secuencias de ARNcr potenciales.

Los solicitantes también determinaron el efecto de la longitud del separador sobre la eficiencia del clivaje. Los solicitantes evaluaron separadores con longitudes diferentes sobre un fragmento del ADN diana que contenía el sitio diana: 5'-TTAgagaagtcatttaataaggccactgttaaaa-3' (SEQ ID NO: 119). Para este experimento, el plásmido pUC19 que contenía el separador (5'-TTcgagaagucauuuaauaaggccacuguuaaaa-3' (SEQ ID NO: 120)) fue sometido a un tratamiento bajo las condiciones que se detallan a continuación:

2 µl de lisado de células que contenía Cpf1

2 µl de ADN de pUC19 con un separador (300 ng)

1 µl del ARNcr (500 ng)

2 µl del amortiguador NE 3

2 µl de DTT 40 mM

0,3 µl de BsaI

10,7 µl de ddH2O.

Se llevó a cabo una incubación a 37 ºC durante 30 minutos, a lo que siguió un tratamiento con una RNasa durante 5 minutos. A continuación, la reacción se depuró usando el conjunto de elementos para llevar a cabo una purificación por medio de una PCR de Qiagen y se analizó en un gel E EX al 2% de Invitrogen. En la figura 59 se representa un gel en el que se observa que el ARNcr de los tipos 1 a 7 dio como resultado el clivaje exitoso del ADN diana in vitro con FnCpf1, mientras que el ARNcr de los tipos 8-13 no dio como resultado el clivaje exitoso del ADN diana.

Los solicitantes obtuvieron el locus mínimo de Fn Cpf1 (figura 60) y también obtuvieron la guía mínima para Cpf1 (figura 61). Los solicitantes también clivaron un amplicón por medio de una PCR a partir del locus EMX1 humano (figura 81). El amplicón EMX fue sometido a un tratamiento bajo las condiciones que se detallan a continuación:

2 µl de lisado de células que contenía Cpf1

3 µl de ADN de pUC19 con un separador (300 ng)

1 µl del ARNcr (500 ng)

2 µl del amortiguador NE 3

2 µl de DTT40 mM

0,3 µl de BsaI

9,7 µl de ddH2O.

Se llevó a cabo una incubación a 37 ºC durante 30 minutos, a lo que siguió un tratamiento con una RNasa durante 5 minutos. A continuación, la reacción se depuró usando el conjunto de elementos para llevar a cabo una purificación por medio de una PCR de Qiagen y se analizó en un gel E EX al 2% de Invitrogen.

Los solicitantes estudiaron más a fondo el efecto del clivaje en la porción DR 5' sobre la actividad de clivaje (figura 82A-B). Para este experimento, el plásmido pUC19 que contenía el separador (5'-TTcgagaagucauuuaauaaggccacuguuaaaa-3' (SEQ ID NO: 121)) fue tratado bajo las siguientes condiciones:

2 µl de lisado de células que contenía Cpf1

2 µl de ADN de pUC19 con un separador (300 ng)

1 µl del ARNcr (500 ng)

2 µl del amortiguador NE 3

2 µl de DTT40 mM

0,3 µl de BsaI

10,7 µl de ddH2O.

Se llevó a cabo una incubación a 37 ºC durante 30 minutos, a lo que siguió un tratamiento con una RNasa durante 5 minutos. A continuación, la reacción se depuró usando el conjunto de elementos para llevar a cabo una purificación por medio de una PCR de Qiagen y se analizó en un gel E EX al 2% de Invitrogen. Los solicitantes determinaron que el ADNcr deltaDR5 interrumpió el rizo del tallo en el extremo 5', y a partir de esto concluyeron que la de horquilla en el extremo 5' es esencial para la actividad de clivaje (figura 82B).

Los solicitantes investigaron el efecto del desajuste entre el ARNcr y el ADN diana sobre la eficacia del clivaje (figura 83). Para este experimento, el plásmido pUC19 que contenía el separador (5'-TTcgagaagucauuuaauaaggccacuguuaaaa-3' (SEQ ID NO: 122)) fue sometido a un tratamiento bajo las condiciones que se detallan a continuación:

2 µl de lisado de células que contenía Cpf1

2 µl de ADN de pUC19 con un separador (300 ng)

1 µl del ARNcr (500 ng)

2 µl del amortiguador NE 3

2 µl de DTT40 mM

0,3 µl de BsaI

10,7 µl de ddH2O.

Se llevó a cabo una incubación a 37 ºC durante 30 minutos, a lo que siguió un tratamiento con una RNasa durante 5 minutos. A continuación, la reacción se depuró usando el conjunto de elementos para llevar a cabo una purificación por medio de una PCR de Qiagen y se analizó en un gel E EX al 2% de Invitrogen. Cada carril del gel se representa en la figura 83, y consistió en un lisado de células que contenía Cpf1, pUC19 con el protoseparador TTc y el ARNcr correspondiente, que ha sido indicado con los números 1 a 11.

Los solicitantes estudiaron el dominio RuvC de FnCpf1p e identificaron mutaciones en los aminoácidos que podían convertir la proteína efectora FnCpf1 en una enzima generadora de cortes, por lo que la proteína efectora presentó una actividad de nucleasa sustancialmente menor y solamente pudo cortar o clivar una cadena del ADN. Las posiciones de los aminoácidos en el dominio RuvC de FnCpf1p abarcaron, sin limitaciones, D917A, E1006A, E1028A, D1227A, D1255A, N1257A, D917A, E1006A, E1028A, D1227A, D1255A y N1257A. Las posiciones de los aminoácidos en AsCpf1 correspondieron a AsD908A, a AsE993A o a AsD1263A. Las posiciones de los aminoácidos en LbCpf1 correspondieron a LbD832A

Los solicitantes también identificaron o un segundo dominio de nucleasa potencial similar al de las nucleasas de la superfamilia PD-(D/E) XK y al de las endonucleasas HincII. Las mutaciones puntuales que puedan generarse en este dominio nucleasa potencial para disminuir sustancialmente la actividad de la nucleasa podrán abarcar, sin limitaciones, N580A, N584A, T587A, W609A, D610A, K613A, E614A, D616A, K624A, D625A, K627A y Y629A.

Los solicitantes realizarán experimentos de clivaje con el plásmido FnCpf1p y secuenciarán los plásmidos resultantes para obtener información acerca de la naturaleza del sitio de clivaje, es decir, si hay un extremo saliente o romo. Los solicitantes deberán proveer más detalles acerca de los diversos dominios de FnCpf1p, particularmente acerca de la estructura cristalina de la proteína en un complejo apropiado. Con el propósito de efectuar la optimización de los componentes de los loci relacionados con FnCpf1 en el contexto de la evaluación de la actividad en las células humanas, los solicitantes evaluarán diversas arquitecturas para el ARNcr y tratarán los blancos adicionales que se describen en la presente.

Los solicitantes clivaron el ADN purificado usando Cpf1 de *Francisella* y de *Prevotella* (figura 84). Para este experimento, el plásmido pUC19 que contenía el separador (5'-TTcgagaagucauuuaauaaggccacuguuaaaa-3' (SEQ ID NO: 123)) fue sometido a un tratamiento bajo las condiciones que se detallan a continuación.

2 µl de una solución de proteína purificada

2 µl de ADN de pUC19 con un separador (300 ng)

1 µl del ARNcr (500 ng)

2 µl del amortiguador NE 3

2 µl de DTT40 mM

0,3 µl de BsaI

10,7 µl de ddH2O.

Se llevó a cabo una incubación a 37 ºC durante 30 minutos, a lo que siguió un tratamiento con una RNasa durante 5 minutos. A continuación, la reacción se depuró usando el conjunto de elementos para llevar a cabo una purificación por medio de una PCR de Qiagen y se analizó en un gel E EX al 2% de Invitrogen. Por medio del análisis del gel, que se representa en la figura 84, se determinó que PaCpf1 puede operar con el ARNcr FnCpf1, aunque la actividad no es tan alta como con FnCpf1. Los solicitantes concluyeron que esto tiene sentido si se tiene en cuenta que las secuencias del tallo y el rizo para PaCpf1 y para FnCpf1 son casi idénticas (la diferencia es de apenas 1 base) (véanse las figuras 85A-B). Esto resulta incluso más evidente en las secuencias del ARNcr maduro correspondientes a FnCpf1 y a PaCpf1, que se representan en las figuras 87A-B. En las formas de realización preferidas de la invención, para poner en práctica un clivaje bioquímico o in vitro en el cual un arreglo CRISPR del tipo Cpf1p opere de manera eficaz, puede no ser necesaria una secuencia tracr. La inclusión de un tallo o un rizo optimizado puede resultar importante para la actividad de clivaje.

Clivaje del ADN humano con una secuencia de FnCpf1p de *Francisella novicida* con codones optimizados.

Los solicitantes también demostraron que el ADN de FnCpf1p se cliva en las células humanas. Se transfectaron 400 ng de FnCpf1p con codones optimizados y 100 ng del ARNcr U6 por cavidad en células HEK293T (~240000 células), en placas con 24 cavidades. Se usaron cinco tipos de ARNcr que comprendían secuencias separadoras con una longitud de 20-24 nucleótidos, sobre la base de 5'-ctgatggtccatgtctgttactcg-3' (SEQ ID NO: 124) (es decir, los primeros 20, 21, 22 o 23 nucleótidos o la totalidad de los 24 nucleótido). El ARNcr de 20 nucleótidos también comprendió una secuencia repetida PaCpf1 en el extremo 5' del separador. Los solicitantes habían determinado con anterioridad que la secuencia repetida PaCpf1 podía ser reconocida por FnCpf1.

El ADN fue cosechado después de ~60 horas y fue analizado en un estudio con una nucleasa Surveyor. Los cebadores para la nucleasa Surveyor DNMT1 fueron 5'-ctgggactcaggcgggtcac-3' (SEQ ID NO: 125) (directo) y 5'-cctcacacaacagcttcatgtcagc-3' (SEQ ID NO: 126) (inverso). Los fragmentos de ADN clivados coincidieron con los productos esperados del clivaje, de ~345 pb y ~261 pb, y pudieron ser observados para los cinco tipos de ARNcr (longitudes de los separadores 20-24 nucleótidos) (figura 88).

Ejemplo 5. Otros experimentos para convalidar PaCpf1

Se llevó a cabo una búsqueda informática de los PAM en Cpf1 de *Prevotella albensis* (PaCpf1) de una manera similar a la que se describió para FnCpf1 en el ejemplo 3. Después de secuenciar el ADN hallado, se extrajeron las regiones correspondientes a los PAM hacia la izquierda o hacia la derecha. Para cada muestra, se comparó la cantidad de PAM presentes en la biblioteca secuenciada con la cantidad de PAM esperada para la biblioteca (4^8 para la biblioteca de la izquierda y 4^7 para la biblioteca de la derecha). Para cuantificar el agotamiento, los solicitantes calcularon una proporción de enriquecimiento. Bajo dos condiciones diferentes (en presencia del plásmido de control pACYC solo o en presencia de pACYC en combinación con PaCpf1), los solicitantes calcularon la proporción de cada PAM en la biblioteca según se detalla a continuación

$$\text{Proporción} = -\log_2 \frac{\text{muestra} + 0{,}01}{\text{biblioteca inicial} + 0{,}01}$$

A partir de la representación de la distribución, los solicitantes determinaron que la muestra de control se caracterizó por un enriquecimiento, mientras que el enriquecimiento en las dos réplicas biológicas fue evidente. Los solicitantes recolectaron todos los PAM con una proporción de 4,5 o más y representaron la distribución de la frecuencia en la que estuvieron presentes, con lo pudieron establecer la presencia de PAM 5'-TTTV, donde V es A o C o G (figuras 62A-E).

Los solicitantes deberán proveer más detalles acerca los diversos dominios de PaCpf1p, particularmente acerca de la estructura cristalina de la proteína en un complejo apropiado. Con el propósito de efectuar la optimización de los componentes de los loci relacionados con PaCpf1 en el contexto de la evaluación de la actividad en las células humanas, los solicitantes evaluarán diversas arquitecturas para el ARNcr (de Guiderna) y diversas proteínas efectoras PaCpf1 optimizadas. Los solicitantes usarán una secuencia de PaCpf1 con codones optimizados como la que se detalla a continuación.

NLS (<u>subrayado</u>)

conector GS (**en negrita**)

Marca 3xHA (*en cursiva*)

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCA
GCAGCCggtagtAACATCAAAAACTTTACCGGGCTCTACCCCCTCAGCAAAACT
TTGCGCTTTGAACTCAAGCCTATTGGCAAAACCAAGGAAAACATCGAGAAA
AATGGCATCCTGACCAAGGACGAGCAACGGGCTAAAGACTACCTCATAGTC
AAAGGCTTTATTGACGAGTATCACAAGCAGTTCATCAAAGACAGGCTTTGGG
ACTTTAAATTGCCTCTCGAAAGTGAGGGGGAGAAGAACAGTCTCGAAGAAT
ACCAGGAACTGTACGAGCTCACTAAGCGCAACGATGCCCAGGAGGCCGACT
TCACCGAGATTAAAGATAACCTTCGCAGCTCTATTACCGAACAGCTCACGAA
GTCTGGATCTGCGTACGATCGGATTTTTAAAAAAGAGTTCATTAGAGAAGAC
CTGGTCAACTTCCTCGAAGATGAAAAAGATAAAAATATCGTGAAACAGTTCG
AGGACTTTACTACATATTTTACGGGTTTTTATGAAAATAGGAAGAACATGTA
CTCTAGCGAAGAGAAGTCCACGGCCATCGCATACCGGCTTATCCATCAGAAT
CTGCCAAAATTCATGGACAACATGAGAAGTTTTGCCAAAATTGCAAATTCCA
GTGTTTCCGAGCACTTTAGCGACATCTATGAAAGCTGGAAGGAATATCTGAA
TGTAAATAGCATCGAGGAAATCTTCCAGCTCGACTATTTTAGCGAAACCTTG

ACTCAGCCACATATTGAGGTGTATAACTATATTATCGGGAAGAAAGTCCTGG
AAGACGGAACCGAGATAAAGGGCATCAACGAGTATGTGAACCTCTACAATC
AGCAGCAGAAAGATAAGAGTAAACGACTGCCTTTCCTGGTGCCACTGTATAA
GCAAATTTTGTCTGATAGGGAAAAACTCTCCTGGATTGCTGAAGAGTTCGAC
AGCGACAAGAAGATGCTGAGCGCTATCACCGAGTCTTACAACCACCTGCACA
ACGTGTTGATGGGTAACGAGAACGAAAGCCTGCGAAATCTGCTGCTGAATAT
TAAGGACTATAACCTGGAGAAAATTAATATCACAAACGACTTGTCTCTCACC
GAAATCTCCCAGAATCTTTTTGGCCGATATGATGTATTCACAAATGGGATCA
AAAACAAGCTGAGAGTGTTGACTCCAAGGAAGAAAAAGGAGACGGACGAA
AATTTTGAGGACCGCATTAACAAAATTTTTAAGACCCAGAAGTCCTTCAGCA
TCGCTTTTCTGAACAAGCTGCCTCAGCCCGAAATGGAGGATGGGAAGCCCCG
GAACATTGAGGACTATTTCATTACACAGGGGGCGATTAACACCAAATCTATA
CAGAAAGAAGATATCTTCGCCCAAATTGAGAATGCATACGAGGATGCACAG
GTGTTCCTGCAAATTAAGGACACCGACAACAAACTTAGCCAGAACAAGACG
GCGGTGGAAAAGATCAAAACTTTGCTGGACGCCTTGAAGGAACTCCAGCACT
TCATCAAACCGCTGCTGGGCTCTGGGGAGGAGAACGAGAAAGACGAACTGT
TCTACGGTTCCTTCCTGGCCATCTGGGACGAACTGGACACCATTACACCACTT
TATAACAAAGTGAGAAATTGGCTGACCCGAAAACCATATTCAACAGAAAAA
ATCAAATTGAATTTCGACAACGCTCAGCTGCTGGGAGGGTGGGATGTCAATA
AAGAACACGACTGTGCAGGTATCTTGTTGCGGAAAAACGATAGCTACTATCT
CGGAATTATCAATAAGAAAACCAACCACATCTTTGATACGGATATTACGCCA
TCAGATGGCGAGTGCTATGACAAAATCGACTACAAGCTCCTTCCCGGGGCGA
ACAAAATGCTTCCAAAGGTGTTTTTTAGTAAGTCCCGAATCAAAGAGTTCGA
GCCATCAGAGGCCATAATCAATTGCTATAAGAAGGGGACACACAAAAAAGG
AAAAAACTTTAACCTGACGGACTGTCACCGCCTGATCAACTTTTTTAAGACC
TCAATCGAGAAACACGAGGATTGGTCAAAATTCGGATTCAAGTTCTCCGATA
CCGAAACGTATGAGGATATTAGCGGTTTTTATAGAGAGGTCGAGCAGCAGG
GATACAGGCTGACGAGCCATCCAGTCAGTGCCAGCTATATACATAGTCTGGT
CAAGGAAGGAAAACTGTACCTCTTCCAAATCTGGAACAAGGACTTTTCTCAA
TTCTCCAAGGGGACCCCTAACTTGCACACTCTCTATTGGAAGATGCTGTTTGA
CAAACGGAATCTTAGCGATGTGGTTTATAAGCTGAATGGCCAGGCTGAAGTG
TTCTATAGAAAGAGCTCCATTGAACACCAGAACCGAATTATCCACCCCGCTC
AGCATCCCATCACAAATAAGAATGAGCTTAACAAAAAGCACACTAGCACCTT

CAAATACGATATCATCAAAGATCGCAGATACACGGTGGATAAATTCCAGTTC
CATGTGCCCATTACTATAAATTTTAAGGCGACCGGGCAGAACAACATCAACC
CAATCGTCCAAGAGGTGATTCGCCAAAACGGTATCACCCACATCATAGGCAT
CGATCGAGGTGAACGCCATCTTCTGTACCTCTCTCTCATCGATTTGAAAGGCA
ACATCATCAAGCAGATGACTCTCAACGAAATTATTAATGAGTATAAGGGTGT
GACCTATAAGACCAACTACCATAACCTCCTGGAGAAGAGGGAGAAGGAGCG
GACCGAGGCCAGACACTCCTGGAGTAGTATTGAAAGCATAAAAGAACTGAA
GGATGGATACATGTCACAGGTGATTCACAAAATTACGGACATGATGGTTAAG
TACAATGCGATTGTGGTCCTGGAGGACCTCAACGGGGGGTTTATGCGAGGCC
GCCAGAAGGTCGAGAAGCAGGTGTACCAGAAATTTGAAAAAAAGTTGATCG
ACAAGCTGAACTATCTCGTTGACAAGAAACTCGACGCTAACGAGGTCGGCG
GAGTACTGAATGCTTATCAGCTGACCAACAAGTTCGAGTCTTTCAAGAAGAT
TGGGAAACAAAGCGGATTTTTGTTCTACATCCCCGCCTGGAACACAAGCAAA
ATCGATCCTATAACAGGGTTCGTTAATCTGTTCAACACCAGGTACGAGTCTA
TCAAGGAGACAAAAGTTTTTTGGTCTAAGTTTGATATTATCCGATACAATAA
AGAGAAGAATTGGTTCGAGTTCGTCTTCGATTACAATACCTTTACGACTAAA
GCGGAGGGAACACGCACTAAGTGGACTCTGTGCACCCACGGCACTCGCATCC
AGACATTCCGGAACCCAGAAAAGAATGCCCAGTGGGACAATAAAGAGATCA
ATTTGACTGAGTCCTTCAAAGCTCTGTTTGAAAAGTACAAGATCGATATCAC
CAGTAATCTCAAGGAATCCATCATGCAGGAAACCGAGAAGAAGTTCTTCCAG
GAACTGCATAATCTGCTCCACCTGACCCTGCAGATGAGGAATAGCGTTACTG
GAACCGACATAGACTATTTGATCAGCCCCGTTGCCGATGAGGATGGAAATTT
CTATGATAGTCGCATAAATGGCAAAAATTTTCCGGAGAATGCCGATGCCAAT
GGCGCGTACAACATCGCACGAAAGGGTCTGATGCTTATTCGGCAGATCAAGC
AAGCAGATCCACAGAAGAAATTCAAGTTTGAGACAATCACCAATAAAGACT
GGCTGAAATTCGCCCAAGACAAGCCCTATCTTAAAGAT**ggcagcgggAAAAGGC**
CGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG**ggatcc***TACCCA*
TACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATA
CGATGTCCCCGACTATGCCTAA (SEQ ID NO: 127)

El mapa de la secuencia del vector que contiene PaCpf1 humano con codones optimizados se provee en la figura 63.

Ejemplo 6. Ortólogos de Cpf1

Los solicitantes analizaron un lote de ortólogos de Cpf1 en expansión (figura 64). Las secuencias humanas con codones optimizados se obtuvieron a partir de varios componentes de loci relacionados con Cpf1 (figuras 65-79). Los solicitantes también obtuvieron las secuencias de las repeticiones directas (DR) de cada ortólogo y determinaron la estructura prevista para el plegamiento (figuras 80A-I).

Los solicitantes estudiaron más ortólogos de Cpf1 sobre la base del tamaño de las proteínas efectoras, donde las proteínas efectoras más pequeñas posibilitarían una introducción más sencilla en los vectores, y en función de la composición de los PAM. Todos los aspectos posibilitaron una optimización adicional en las células procariotas o eucariotas, preferiblemente de modo tal de obtener una actividad eficaz en las células de mamíferos, en especial en las células humanas.

Los solicitantes demostraron que los ortólogos de la proteína efectora en los siguientes loci presentaron actividad en el estudio del clivaje in vitro: la bacteria Peregrinibacteria GW2011_GWA2_33_10, Cpf1; Acidaminococcus sp. BV3L6, Cpf1; Francisalla tularensis, Cpf1 1; Moraxella bovoculi, Cpf1 237; la bacteria Lachnospiraceae ND2006, Cpf1, la bacteria Lachnospiraceaa MA2020, Cpf1; Porphyromonas macacee, Cpf1; Porphyromonas crevioricanis, Cpf1 3; Prevotella albensis, Cpf1 (figura 64).

En el estudio del clivaje in vitro con los ortólogos, se cosecharon las células HEK293 en las que se expresaban los ortólogos de Cpf1, se incubó el lisado con el ARNcr maduro predicho y con un separador artificial y se lo clonó en plásmidos pUC19. El separador fue precedido por 8 bases degeneradas para permitir la identificación de los PAM por medio de una secuenciación. Las bandas inferiores representan el clivaje con la enzima Cpf1 (figura 89).

Con el abordaje informático que se ha descripto, los solicitantes identificaron los PAM que se obtuvieron a partir del estudio del clivaje in vitro (figura 90). Se extrajo el ADN no procesado que se observa en la figura 89 (la banda superior) y se lo amplificó para llevar a cabo un procedimiento de secuenciación de la siguiente generación. Se calculó la abundancia de cada 8-mero y se usó la proporción logarítmica con relación a la biblioteca de referencia para cuantificar el enriquecimiento. Se agruparon los 8-meros individuales con una proporción logarítmica superior 4 y se los usó para determinar el PAM de consenso con el programa WebLogo.

Los solicitantes también evaluaron las proteínas efectoras Cpf1p clivadas de manera escalonada con un dominio 5'. Se recolectó una proteína FnCpf1 purificada, se la incubó con el ARNcr y se clono el diana correspondiente en pUC19. El producto clivado fue extraído en un gel y fue sometido a una secuenciación de Sanger. En función de la lectura asimétrica, se determinó que hubo un clivaje escalonado (figura 91). En una forma de realización preferida de la invención, los solicitantes pudieron hallar un clivaje escalonado in vivo en el molde (por ejemplo, en un molde exógeno).

Los solicitantes también determinaron el efecto de la longitud del separador sobre la capacidad de clivaje de la proteína efectora (figura 92). Se recolectó una proteína FnCpf1 purificada, se la incubó con el ARNcr y se clono el diana correspondiente en pUC19. Los separadores con longitudes superiores a 17 nucleótidos fueron clivados por completo, mientras que los separadores con una longitud de 17 nucleótidos presentaron una actividad menor y los separadores con longitudes inferiores a 17 nucleótidos no estuvieron activos.

Los solicitantes demostraron que FnCpf1 media en la formación de indels en las células HEK293T.

Se transfectaron 280000 células HEK/24 con 350 ng de un plásmido huFnCpf1 y 150 ng del ARNcr U6. Las células se recolectaron tres días después de la transfección y se analizaron con un nucleasa Surveyor. El tamaño de los fragmentos no clivados en la PCR fue de 606 pares de bases. Los tamaños de los fragmentos esperados fueron de ~418 pb y ~188 pb para el ARNcr DNMT1-1 y de ~362 pb y ~244 pb para el ARNcr DNMT1-3 (figura 93).

DNMT1-1, secuencia separadora: cctcactcctgctcggtgaattt (SEQ ID NO: 128)

DNMT1-3, secuencia separadora: ctgatggtccatgtctgttactc (SEQ ID NO: 129)

Los solicitantes identificaron los componentes del sistema de Cpf1 que serían necesarios para obtener el clivaje mediante la determinación de la naturaleza del procesamiento de los transcriptos como resultado de la supresión de las secuencias de determinados loci (figura 94A-F). Las secuencias eliminadas pudieron abarcar, sin limitaciones, el gen Cas1, el gen Cas2 y la secuencia tracr. Por lo tanto, en una forma de realización preferida de la invención, los solicitantes demostraron que la secuencia tracr no es un componente de un sistema o un complejo funcional Cpf1 que sea necesario para obtener el clivaje.

Ejemplo 7. Procedimientos

Generación de los plásmidos heterólogos

Con el fin de generar un locus FnCpf1 que fuera apropiado para una expresión heteróloga, el ADN genómico de *Francisella novicida* fue amplificado por medio de PCR con una polimerasa Herculase II (de Agilent Technologies) y fue clonado en pACYC-184 usando un dispositivo de clonación Gibson (de New England Biolabs). Las células que albergaban los plásmidos se tornaron competentes con el conjunto de elementos Z (de Zymo).

Secuenciación del ARN de las bacterias

Se aisló el ARN a partir de las bacterias en una fase estacionaria primero resuspendiendo *F. novicida* (provista generosamente por David Weiss) o *E. coli* en TRIzol y luego homogenizando las bacterias con esferas de circonio/sílice (de BioSpec Products) en un dispositivo BeadBeater (de BioSpec Products) durante 3 ciclos de un minuto. El ARN total se purificó a partir de las muestras homogenizadas de acuerdo con el protocolo del conjunto de elementos Zol

Direct-RNA Miniprep (de Zymo), donde se usó una ADNasa Turbo (de Life Technologies) y una quinasa de polinucleótidos T4 (de New England Biolabs) para desfosforilar los extremos 3'. El ARNr se eliminó con el conjunto de elementos Ribo-Zero (de Illumina). Las bibliotecas de ARN se prepararon a partir del ARNr libre de ARN con el conjunto de elementos NEBNext® para Illumina (de New England Biolabs), y el tamaño se seleccionó con el conjunto de elementos Pippin Prep (de Sage Science).

Para efectuar una expresión heteróloga en *E. coli* del locus FnCpf1, se prepararon bibliotecas con el ARN secuenciado a partir del ARNr sin ARN de acuerdo con un método para secuenciar el ARN asociado a los arreglos CRISPR que fue similar a uno que había sido descripto con anterioridad (Heidrich y col., 2015). En resumen, se realizaron transcripciones donde se unió una cola de poli-A proveniente de *E. coli*, una poli polimerasa A (de New England Biolabs), los adaptadores 5' apropiados para el ARN y una ligasa de ARN de T4 1 (que actúa sobre el ARN de cadena simple) en una concentración alta (de New England Biolabs), a lo que siguió una transcripción inversa con una transcriptasa inversa capaz de actuar a diversas temperaturas AffinityScript (de Agilent Technologies). El ADNc se amplificó por medio de una PCR con los cebadores apropiados en la que se usó una polimerasa Herculase II (de Agilent Technologies), a lo que siguió la secuenciación del ARN.

Las bibliotecas de ADNc preparadas se secuenciaron con un conjunto de elementos MiSeq (de Illumina). Las muestras leídas fueron identificadas sobre la base de sus características distintivas y fueron alineadas con la secuencia genómica de referencia apropiada mediante el uso del programa BWA (de Li y Durbin, 2009). Sobre la base de los alineamientos, se extrajeron las secuencias transcriptas completas con las herramientas de Picard (http://broadinstitute.github.io/picard) y se las analizó con el programa Geneious 8.1.5.

Análisis de los PAM relacionados con FnCpf1 in vivo

Se construyeron bibliotecas de plásmidos que contenían PAM al azar con oligonucleótidos sintetizados (IDT) que comprendían 7 nucleótidos al azar hacia el extremo 5' o 3' del separador 1 del diana (tabla suplementaria S8). Los oligonucleótidos compuestos por ADN de cadena simple al azar fueron convertidos en moléculas de cadena doble al formar híbridos con un cebador corto, mediante el uso de un fragmento de Klenow grande (de New England Biolabs), de manera tal de sintetizar la segunda cadena. El ADN de cadena doble que se produjo fue montado en un plásmido pUC19 linealizado con un dispositivo de clonación Gibson (de New England Biolabs). Se transformaron bacterias *E. coli* Stbl3 competentes (de Invitrogen) con los productos clonados, se las recolectó y se las combinó hasta obtener una cantidad superior a 10^7. El ADN del plásmido se cosechó usando un conjunto de elementos Maxi-prep (de Qiagen). En la transformación, se introdujeron 360 ng de la biblioteca combinada en células de *E. coli* que comprendían un locus FnCpf1 o un control pACYC184. Una vez terminada la transformación, las células se sembraron en placas con ampicilina. Después de 16 horas de cultivo, se recolectaron más de 4×10^6 células y se extrajo el ADN del plásmido usando un conjunto de elementos Maxi-prep (de Qiagen). La región del PAM diana se amplificó y se secuenció usando un conjunto de elementos MiSeq (de Illumina) con un solo extremo, durante 150 ciclos.

Abordaje informático para descubrir los PAM

Se extrajeron las regiones de los PAM, se las contó y se las normalizó para determinar la cantidad total de lecturas en cada una de las muestras. Para cualquiera de los PAM, el enriquecimiento se midió como la proporción logarítmica en comparación con el plásmido de control pACYC184, con un ajuste de 0,01 pseudoconteos. Los PAM con valores superiores al umbral de enriquecimiento, 3,5, se recolectaron y se usaron para generar las secuencias de referencia (Crooks y col., 2004).

Convalidación de los PAM

Las secuencias correspondientes a los PAM y las secuencias diferentes de ellos se clonaron en un plásmido pUC19 digerido y se unieron con la ligasa T4 (de Enzymatics). Se transformaron bacterias *E. coli* competentes que contenían un plásmido con un locus FnCpf1 o un plásmido de control pACYC184 con 20 ng de un plásmido que contenía los PAM y se las sembró en placas con agar LB, ampicilina y cloranfenicol. Se contaron las colonias después de 18 horas.

Síntesis del ARNcr y del ARNg

El ARNcr y el ARNg que se usó in vitro fueron sintetizados usando el conjunto de elementos de HiScribe™ T7 para sintetizar ARN con un rendimiento elevado (de NEB). Los oligonucleótidos compuestos por ADN de cadena simple correspondientes al complemento inverso de la secuencia del ARN diana se sintetizaron a partir los IDT y se alinearon con una secuencia correspondiente a un cebador corto T7. La transcripción con T7 se realizó durante 4 horas, y luego se purificó el ARN con un conjunto de elementos Clean-Up MEGAclear™ (de Ambion).

Purificación de la proteína Cpf1

La proteína FnCpf1 se clonó en un vector de expresión para bacterias (6-His-MBP-TEV-CPF1, un vector basado en pET, provisto amablemente a los solicitantes por Doug Daniels) ("6-His" se detalla en SEQ ID NO: 130). Se inocularon dos litros de un medio de cultivo Terrific Broth que contenía 100 g/ml de ampicilina con 10 ml de un cultivo durante una noche de células Rosetta (DE3) pLyseS (de EMD Millipore) que contenían la construcción para expresar Cpf1. Se realizó un cultivo con el inoculante a 37 ºC hasta que la densidad óptica las células a 600 nm alcanzó un valor de 0,2, y luego se redujo la temperatura hasta 21 ºC. El cultivo continuó hasta que la densidad óptica de las células a 600 nm

alcanzó un valor de 0,6, momento en el cual se agregó una concentración final de IPTG de 500 mM para inducir la expresión de MBP-Cpf1. Se indujo el cultivo durante 14-18 horas, se recolectaron las células y se las congeló a -80 ºC hasta que se decidió purificarlas.

La pasta con las células se resuspendió en 200 ml de un amortiguador de lisis (Hepes 50 mM, pH 7, NaCl 2 M, MgCl2 5 mM, imidazol 20 mM) con inhibidores de proteasas (un conjunto completo de Roche libre de EDTA) y una lisozima. Una vez homogenizadas, las células se lisaron por medio de una sonicación (con un dispositivo Branson 450) y se centrifugaron a 10000 × G durante 1 hora para separar el lisado. El lisado se filtró a través de tamices de 0,22 µm (Stericup, de Millipore), se aplicó sobre una columna de níquel (HisTrap FF, 5 ml), se lavó y luego se eluyó con un gradiente de imidazol. Las fracciones que contenían la proteína del tamaño esperado se combinaron, se agregó la proteasa TEV (de Sigma) y se dializó la muestra durante la noche en un amortiguador TEV (NaCl 500 mM, Hepes 50 mM, pH 7, MgCl 5 mM, DTT 2 mM). Después de la diálisis, se confirmó el clivaje de TEV por medio de una SDS-PAGE y se concentró la muestra hasta 500 µl antes de cargarla en una columna de filtración con un gel (Superdex 200 HiLoad 16/600), en la cual se puso en práctica una FPLC (AKTA Pure). Las fracciones filtradas en el gel se analizaron por medio de una SDS-PAGE: las fracciones que contenían Cpf1 se combinaron, se concentraron hasta 200 µl y se usaron directamente en los estudios bioquímicos o se congelaron a -80 ºC para almacenarlas. Las referencias que se usaron en esta filtración en un gel se procesaron en la misma columna, una vez equilibrada con NaCl 2 M, Hepes, pH 7,0, con el propósito de calcular el tamaño aproximado para FnCpf1.

Generación de un lisado con la proteína Cpf1

Se sintetizaron proteínas Cpf1 con codones optimizados para la expresión en los seres humanos con una marca apropiada para localizarlas en el núcleo en el extremo N y se las clonó en el plásmido de expresión pcDNA3.1 con el dispositivo Genscript. Se transfectaron 2000 ng de los plásmidos de expresión con Cpf1 en placas con 6 cavidades que contenían células HEK293FT con una confluencia de 90% usando el reactivo Lipofectamine 2000 (de Life Technologies). 48 horas después, las células se cosecharon por medio de un lavado con DPBS (de Life Technologies) y se colocaron en un amortiguador de lisis (Hepes 20 mM, pH 7,5, KCl 100 mM, MgCl2 5 mM, DTT 1 mM, 5% de glicerol, 0,1% Triton X-100, con un conjunto completo de inhibidores de proteasas de Roche). El lisado se trató con ultrasonido durante 10 minutos en un sonicador Biorupter (de Diagenode) y luego se centrifugó. Sobrenadante se congeló para su uso posterior en los estudios basados en el clivaje in vitro.

Estudio del clivaje vitro

El clivaje in vitro se llevó a cabo ya sea con una proteína purificada o con un lisado proveniente de un mamífero que comprendía la proteína a 37ºC, en un amortiguador apropiado para la disociación (un amortiguador NEB 3 con DTT 5 mM) durante 20 minutos. La reacción de clivaje se realizó con 500 ng de ARNcr o ARNgs sintetizado y 200 ng del ADN diana. El ADN diana comprendió protoseparadores clonados en pUC19 o amplicones obtenidos por medio de una PCR de las regiones de los genes provenientes del ADN genómico de las células HEK293. Las reacciones se depuraron usando columnas de purificación y una PCR (de Qiagen) y se analizaron en un gel de agarosa E al 2% (de Life Technologies). Se usaron geles nativos o desnaturalizantes para analizar el clivaje en combinación con nucleasas mutadas y se realizó una depuración en un gel de poliacrilamida con TBE al 6% o en un gel de poliacrilamida con TBE y urea al 6% (de Life Technologies)

Análisis preliminar de los PAM asociados a proteínas de la familia Cpf1 in vitro

Se llevaron a cabo reacciones de clivaje in vitro con proteínas de la familia Cpf1 en un gel de agarosa E al 2% (de Life Technologies). Las bandas correspondientes al diana no clivado se extrajeron con el conjunto de elementos QIAquick (de Qiagen), y la región correspondiente al PAM diana se amplificó y se secuenció usando un conjunto de elementos MiSeq (de Illumina), con un solo extremo durante 150 ciclos. Los resultados de la secuenciación se usaron en el abordaje para buscar PAM que se ha descripto con anterioridad.

Actividad de clivaje de Cpf1 en células 293FT

Se sintetizaron proteínas Cpf1 con codones optimizados para la expresión en los seres humanos con una marca apropiada para localizarlas en el núcleo en el extremo N y se las clonó en el plásmido de expresión pcDNA3.1 con el dispositivo Genscript. Por medio de una PCR con una polimerasa Herculase II (de Agilent Technologies), se generaron amplicones que comprendían un promotor U6 y una secuencia ARNcr específica. Se transfectaron 2000 ng de los plásmidos de expresión con Cpf1 y 100 ng de los productos de la PCR que abarcaban el ARNcr en placas con 24 cavidades con células HEK293FT con una confluencia de entre 75% y 90% usando el reactivo Lipofectamine 2000 (de Life Technologies). El ADN genómico se cosechó usando el conjunto de elementos QuickExtract™ (de Epicentre).

Estudio con la nucleasa Surveyor para modificar el genoma

Se transfectaron células 293FT con 400 ng de un plásmido de expresión con Cpf1 y 100 ng de fragmentos de ARNcr con el promotor U6 usando el reactivo Lipofectamine 2000 (de Life Technologies). Las células se incubaron a 37ºC durante 72 horas después de la transfección y antes de la extracción del ADN genómico. El ADN genómico fue extraído usando el conjunto de elementos QuickExtract™ (de Epicentre), de acuerdo con el protocolo del fabricante. La región genómica que rodeaba el sitio diana en los arreglos CRISPR de cada gen se amplificó por medio de una PCR, y los

productos se purificaron con una columna de centrifugación QiaQuick (de Qiagen), de acuerdo con el protocolo del fabricante. Se mezclaron 200-500 ng de todos los productos de la PCR con 1 µl de un amortiguador para una polimerasa de ADN Taq apropiada para una PCR 10x (de Enzymatics) y agua ultrapura en un volumen final de 10 µl y se llevó a cabo un alineamiento para permitir la formación de duplas: primero se aplicó una temperatura de 95 ºC durante un período de 10 minutos, luego se incrementó la temperatura de manera gradual desde 95 ºC hasta 85 ºC, con una velocidad de -2 ºC/segundos, luego se redujo la temperatura desde 85 ºC hasta 25 ºC con una velocidad de -0,25 ºC/segundos y finalmente se aplicó una temperatura de 25 ºC durante 1 minuto. Una vez completo el alineamiento, los productos fueron tratados con una nucleasa Surveyor y con un potenciador Surveyor S (de Integrated DNA Technologies), de acuerdo con el protocolo recomendado por el fabricante, y se analizaron en geles de poliacrilamida Novex TBE al 4%-20% (de Life Technologies). Se sometieron los geles a una coloración para el ADN con oro SYBR (de Life Technologies) durante 10 minutos y se tomaron fotografías un sistema apropiado para obtener imágenes a partir de geles, que en este caso fue un sistema Gel Doc (de Bio-rad). La cuantificación se basó en las intensidades relativas de las bandas. El porcentaje de indels se determinó de acuerdo con la fórmula 100 × (1 - (1 - (b + c)/(a + b + c)) 1/2), donde a representa la intensidad integrada del producto no digerido de la PCR y b y c son las intensidades integradas de cada producto del clivaje.

Secuenciación profunda para caracterizar los patrones de indels en Cpf1 en las células 293FT

Se transfectaron células HEK293FT y se las recolectó como se describió en el contexto de la evaluación de la actividad de clivaje de Cpf1. Se amplificó la región de acompañamiento asociada a DNMT1 en dos procedimientos de PCR, de manera tal de poder agregarles los adaptadores Illumina P5 y las secuencias características de las muestras a los amplicones. Los productos de la PCR se evaluaron en un gel E al 2% (de Invitrogen) y se extrajeron mediante el uso de una columna de centrifugación QiaQuick (de Qiagen), de acuerdo con el protocolo recomendado por el fabricante. Las muestras se combinaron y se cuantificaron con un fluorómetro Qubit 2.0 (de Life Technologies). Las bibliotecas de ADNc que se prepararon se secuenciaron con un sistema MiSeq (de Illumina). Los indels se determinaron con la función Python del programa Geneious Read Mapper 6.0.3.

Análisis informático de los loci relacionados con Cpf1

Se usó el programa PSI-BLAST (Altschul y col., 1997) para identificar los homólogos Cpf1 en la base de datos NR del NCBI usando diversas secuencias de Cpf1 conocidas como referencias, con un umbral E para Cpf1 de 0,01, una separación con una complejidad baja y las estadísticas basadas en la composición desactivadas. Se usó el programa TBLASTN con un umbral E de 0,01 y una separación con una complejidad baja para realizar una búsqueda en la base de datos WGS del NCBI con del perfil de Cpf1 descripto por Makarova et al. en 2015 como referencia. Los resultados de todas las búsquedas se combinaron. Se usó el programa HHpred con los parámetros por omisión para identificar las secuencias con una similitud remota usando un subconjunto de las secuencias de Cpf1 representativas (Söding y col., 2006). Se construyeron múltiples alineamientos con las secuencias usando Muscle (Edgar, 2004), con una corrección manual basada en los alineamientos de pares que se habían obtenido con los programas HHpred y PSI-BLAST. El análisis filogenético se realizó usando el programa FastTree, de acuerdo con un modelo evolutivo WAG y un modelo discreto Gamma con 20 tipos diferentes (Price y col., 2010). La estructura secundaria de las proteínas se predijo usando el programa Jpred 4 (Drozdetskiy y col., 2015).

Las repeticiones asociadas a los arreglos CRISPR se identificaron usando los programas PILER-CR (Edgar, 2007) y CRISPRfinder (Grissa y col., 2007). Las secuencias separadoras se determinaron por medio de búsquedas en la base de datos de nucleótidos NR del NCBI con el programa MEGABLAST (Morgulis y col., 2008), con los parámetros por omisión, excepto que el tamaño de las palabras se fijó en 20 y el valor del umbral E se fijó en 0,0001.

Tabla 1. Secuencias separadoras endógenas de *F. novicida*

Número del separador	Secuencia
1	GAGAAGTCATTTAATAAGGCCACTGTTAAAA (SEQ ID NO: 131)
2	GCTACTATTCCTGTGCCTTCAGATAATTCA (SEQ ID NO: 132)
3	GTCTAGAGCCTTTTGTATTAGTAGCCG (SEQ ID NO: 133)

Tabla 2. Oligonucleótidos y cebadores compuestos por ADN de cadena simple que se usaron en el contexto de la generación de las bibliotecas de PAM

Nombre del oligonucleótido/ cebador	Secuencia
Biblioteca de los PAM 5' (+)	GGCCAGTGAATTCGAGCTCGGTACCCGGG NNNNNNNNGAGAAGTCATTTAATAAGGC CACTGTTAAAAAGCTTGGCGTAATCATGG TCATAGCTGTTT (SEQ ID NO: 134)
Biblioteca de los PAM 3' (+)	GGCCAGTGAATTCGAGCTCGGTACCCGGG GAGAAGTCATTTAATAAGGCCACTGTTAA AANNNNNNNNAGCTTGGCGTAATCATGG TCATAGCTGTTT (SEQ ID NO: 135)
Biblioteca de los PAM (-)	GCTGACATGAAGCTGTTGTGTGAGG (SEQ ID NO: 136)

Tabla 3. Cebadores que se usaron en el contexto de la secuenciación del plásmido pUC19 y el análisis con la nucleasa Surveyor

Nombre del cebador	Secuencia
NGS pUC, cebador directo	GGCCAGTGAATTCGAGCTCGG (SEQ ID NO: 137)
NGS pUC, cebador inverso	CAATTTCACACAGGAAACAGCTATGACC (SEQ ID NO: 138)
Sanger pUC, cebador directo	CGGGGCTGGCTTAACTATGCG (SEQ ID NO: 139)
Sanger pUC, cebador inverso	GCCCAATACGCAAACCGCCT (SEQ ID NO: 140)
EMX1, cebador directo	CCATCCCCTTCTGTGAATGT (SEQ ID NO: 141)
EMX1, cebador inverso	TCTCCGTGTCTCCAATCTCC (SEQ ID NO: 142)
DNMT1, cebador directo	CTGGGACTCAGGCGGGTCAC (SEQ ID NO: 143)
DNMT1, cebador inverso	GCTGACATGAAGCTGTTGTGTGAGG (SEQ ID NO: 144)

Tabla 4. Secuencias de las guías truncadas que se usaron en el contexto del análisis del clivaje *in vitro*

Número de las guías truncadas	Secuencia
1	GAGAAGTCATTTAATAAGGCCACT (SEQ ID NO: 145)
2	GAGAAGTCATTTAATAAGGCCA (SEQ ID NO: 146)
3	GAGAAGTCATTTAATAAGGC (SEQ ID NO: 147)
4	GAGAAGTCATTTAATAAG (SEQ ID NO: 148)
5	GAGAAGTCATTTAATAA (SEQ ID NO: 149)
6	GAGAAGTCATTTAATA (SEQ ID NO: 150)

Tabla 5. Secuencias de las guías con faltas de coincidencia que se usaron en el contexto del análisis del clivaje *in vitro*

Número de las guías con faltas de coincidencia	Secuencia
1	GATAAGTCATTTAATAAGGCCACT (SEQ ID NO: 151)
2	GAGAAGGCATTTAATAAGGCCACT (SEQ ID NO: 152)
3	GAGAAGTCATGTAATAAGGCCACT (SEQ ID NO: 153)
4	GAGAAGTCATTTAAGAAGGCCACT (SEQ ID NO: 154)
5	GAGAAGTCATTTAATAAGTCCACT (SEQ ID NO: 155)
6	GAGAAGTCATTTAATAAGGCCAAT (SEQ ID NO: 156)

Tabla 6. Secuencias de las guías truncadas con repeticiones directas que se usaron en el contexto del análisis del clivaje *in vitro*

Longitud de las repeticiones directas	Secuencia
+18	ATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT (SEQ ID NO: 157)
+17	TTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT (SEQ ID NO: 158)
+16	TTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT (SEQ ID NO: 159)
+15	TCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT (SEQ ID NO: 160)
+11	CTGTTGTAGATGAGAAGTCATTTAATAAGGCCACT (SEQ ID NO: 161)
+7	TGTAGATGAGAAGTCATTTAATAAGGCCACT (SEQ ID NO: 162)

Tabla 7. Mutaciones en las porciones lineales correspondientes a las repeticiones directas que se usaron en el contexto del análisis del clivaje *in vitro*

Número de las mutaciones en las porciones lineales correspondientes a las repeticiones directas	Secuencia
1	AATTTCTGCTGTTGCAGAT (SEQ ID NO: 163)
2	AATTTCCACTGTTGTGGAT (SEQ ID NO: 164)
3	AATTCCTACTGTTGTAGGT(SEQ ID NO: 165)
4	AATTTATACTGTTGTAGAT(SEQ ID NO: 166)
5	AATTTCGACTGTTGTAGATAATTTCGACTGTTGTAGAT (SEQ ID NO: 167)
6	AATTTCTAGTGTTGTAGAT (SEQ ID NO: 168)

Tabla 8. Mutaciones en las porciones circulares correspondientes a las repeticiones directas que se usaron en el contexto del análisis del clivaje *in vitro*

Número de las mutaciones en las porciones circulares correspondientes a las repeticiones directas	Secuencia
1	AATTTCTACTATTGTAGAT (SEQ ID NO: 169)
2	AATTTCTACTGCTGTAGAT (SEQ ID NO: 170)
3	AATTTCTACTTTGTAGAT (SEQ ID NO: 171)
4	AATTTCTACTTGTAGAT (SEQ ID NO: 172)
5	AATTTCTACTTTTGTAGAA (SEQ ID NO: 173)
6	AATTTCTACTTTTGTAGAC (SEQ ID NO: 174)

Tabla 9. Secuencias de las guías específicas dirigidas a los ortólogos de DMNT1 que se aplicaron sobre las células de mamíferos

Nucleasa	Nombre	Repetición directa 5'	Secuencia
AsCpf1	DNMT1, diana 1	Repetición directa 5'	Secuencia
AsCpf1	DNMT1, diana 2	TAATTTCTACTGTTGTAGAT (SEQ ID NO: 175)	CCTCACTCCTGC TCGGTGAATTT (SEQ ID NO: 176)
AsCpf1	DNMT1, diana 3	TAATTTCTACTGTTGTAGAT (SEQ ID NO: 177)	AGGAGTGTTCAG TCTCCGTGAAC (SEQ ID NO: 178)
AsCpf1	DNMT1, diana 4	TAATTTCTACTGTTGTAGAT (SEQ ID NO: 179)	CTGATGGTCCAT GTCTGTTACTC (SEQ ID NO: 180)
Lb3Cpf1	DNMT1, diana 1	TAATTTCTACTGTTGTAGAT (SEQ ID NO: 181)	TTTCCCTTCAGCT AAAATAAAGG (SEQ ID NO: 182)
Lb3Cpf1	DNMT1, diana 2	TAATTTCTACTAAGTGTAGAT (SEQ ID NO: 183)	CCTCACTCCTGC TCGGTGAATTT (SEQ ID NO: 184)
Lb3Cpf1	DNMT1, diana 3	TAATTTCTACTAAGTGTAGAT (SEQ ID NO: 185)	AGGAGTGTTCAG TCTCCGTGAAC (SEQ ID NO: 186)
Lb3Cpf1	DNMT1, diana 4	TAATTTCTACTAAGTGTAGAT (SEQ ID NO: 187)	CTGATGGTCCAT GTCTGTTACTC (SEQ ID NO: 188)
SpCas9	DNMT1, diana 1	TAATTTCTACTAAGTGTAGAT (SEQ ID NO: 189)	TTTCCCTTCAGCT AAAATAAAGG (SEQ ID NO: 190)

SpCas9	DNMT1, diana 2	na	TCACTCCTGCTC GGTGAATT (SEQ ID NO: 191)
SpCas9	DNMT1, diana 3	na	AACCCTCTGGGG ACCGTTTG (SEQ ID NO: 192)
SpCas9	DNMT1, diana 4	na	AGTACGTTAATG TTTCCTGA (SEQ ID NO: 193)

Tabla 10. Secuencias de las repeticiones directas específicas en el ARNcr dirigidas al protoseparador 1 y a DMNT1, correspondientes a la diana 3

Origen de la repetición directa	Secuencia
FnCpf1	TAATTTCTACTGTTGTAGAT (SEQ ID NO: 195)
Lb1Cpf1	AGAAATGCATGGTTCTCATGC (SEQ ID NO: 196)
BpCpf1	AAAATTACCTAGTAATTAGGT (SEQ ID NO: 197)
PeCpf1	GGATTTCTACTTTTGTAGAT (SEQ ID NO: 198)
PbCpf1	AAATTTCTACTTTTGTAGAT (SEQ ID NO: 199)
SsCpf1	CGCGCCCACGCGGGGCGCGAC (SEQ ID NO: 200)
AsCpf1	TAATTTCTACTCTTGTAGAT (SEQ ID NO: 201)
Lb2Cpf1	GAATTTCTACTATTGTAGAT (SEQ ID NO: 202)
CMtCpf1	GAATCTCTACTCTTTGTAGAT (SEQ ID NO: 203)
EeCpf1	TAATTTCTACTTTGTAGAT (SEQ ID NO: 204)
MbCpf1	AAATTTCTACTGTTTGTAGAT (SEQ ID NO: 205)
LiCpf1	GAATTTCTACTTTTGTAGAT (SEQ ID NO: 206)
Lb3Cpf1	TAATTTCTACTAAGTGTAGAT (SEQ ID NO: 207)
PcCpf1	TAATTTCTACTATTGTAGAT (SEQ ID NO: 208)
PdCpf1	TAATTTCTACTTCGGTAGAT (SEQ ID NO: 209)
PmCpf1	TAATTTCTACTATTGTAGAT (SEQ ID NO: 210)

Ejemplo 8. Clonación de Cpf1 de Francisella tularensis subespecie novicida U112 (FnCpf1)

Los solicitantes clonaron el locus Cpf1 (FnCpf1) de *Francisella tularensis* subespecie *novicida* U112 (figura 95A) en plásmidos con una cantidad baja de copias (pFnCpf1) para posibilitar una reconstitución heteróloga en *Escherichia coli*. Típicamente, en los sistemas CRISPR-Cas que se han caracterizados hasta el momento, hay dos requisitos para que ocurra la interferencia a nivel del ADN: (i) la secuencia diana debe coincidir con uno de los separadores presentes en el arreglo CRISPR respectivo y (ii) la secuencia diana complementaria del separador (protoseparador de aquí en adelante) debe estar rodeada por un motivo protoseparador adyacente (PAM) apropiado. Debido a que la funcionalidad del locus correspondiente al arreglo CRISPR en FnCpf1 no ha sido completamente caracterizada, se diseñó un estudio basado en el agotamiento de los plásmidos para determinar la actividad de Cpf1, identificar la secuencia de los PAM y establecer su ubicación en relación con el protoseparador (5' o 3') (figura 95B). Se construyeron dos bibliotecas de plásmidos que contenían un protoseparador que coincidía con el primer separador en el arreglo CRISPR de FnCpf1 en combinación con secuencias de 7 pares de bases al azar hacia el extremo 5' o 3'. Por medio de un procedimiento de transformación, cada biblioteca de plásmidos se introdujo en bacterias *E. coli* en las que se expresaba de manera heteróloga el locus FnCpf1 o en una cepa de *E. coli* de control que contenía un vector vacío. A través de este estudio,

se determinó la secuencia y la localización de los PAM mediante la identificación de los motivos de nucleótidos que se agotaban de manera preferencial en las células en las que se había expresado el locus FnCpf1 de manera heteróloga. Se descubrió que el PAM asociado a FnCpf1 se encontraba más allá del extremo 5' de la cadena desplazada del protoseparador y comprendía una secuencia 5'-TTN (figuras 95C-D y 102). La localización del PAM más allá del extremo 5' también fue observada en los arreglos CRISPR del tipo I, pero no en los arreglos del tipo II, donde Cas9 está asociado a secuencias correspondientes a PAM que se encuentran hacia el extremo 3' del protoseparador (Mojica y col., 2009; Garneau y col., 2010). Más allá de la identificación de los PAM, en función de los resultados del estudio de agotamiento, se concluyó que la expresión de los loci relacionados con Cpf1 de manera heteróloga puede resultar claramente útil para poner en práctica una reacción de interferencia con las moléculas de ADN provenientes de los plásmidos.

Con el propósito de caracterizar los PAM con mayor detalle, se analizó la actividad de interferencia sobre los plásmidos mediante la introducción de Cpf1 por medio de un procedimiento de transformación en células con plásmidos que comprendían el protoseparador 1 rodeado por un PAM con una secuencia 5'-TTN. Todos los PAM que comprendían una secuencia 5'-TTN fueron atacados de manera eficiente (figura 1E). También fueron atacadas de manera eficiente las secuencias 5'-CTA, pero no las secuencias 5'-TCA (figura 95E), en función de lo cual podría llegarse a la conclusión de que la base T en la mitad del segmento es la más crítica para el reconocimiento de los PAM que la base T que se encuentra al principio, lo cual concuerda con lo que se había determinado a partir del estudio que se basó en el agotamiento de los PAM (figura 102D): en conclusión, este PAM podría no estar limitado a un motivo 5'-TTN.

Ejemplo 9. El arreglo CRISPR en Cpf1 es procesado de manera independiente del ARNcrtrac

Se realizó una búsqueda con ARN pequeño para determinar la identidad exacta del ARNcr que se había producido en los loci relacionados con los arreglos CRISPR relacionados con Cpf1. Al secuenciar el ARN pequeño extraído de un cultivo de *Francisella tularensis* subespecie *novicida* U112, se determinó que el arreglo CRISPR era procesado a para producir un ARNcr maduro corto, con una longitud entre 42 y 44 nucleótidos. Cada ARNcr maduro comenzó en el nucleótido 19 de la repetición directa, el cual fue seguido por 23-25 nucleótidos de una secuencia separadora (figura 96A). Esta disposición del ARNcr contrasta con la de los sistemas CRISPR-Cas del tipo II, donde el ARNcr maduro comienza con 20-24 nucleótidos de una secuencia separadora, los cuales son seguidos por ~22 nucleótidos de una repetición directa (Deltcheva y col., 2011; Chylinski y col., 2013). Inesperadamente, aparte del ARNcr, no se observó ninguna expresión robusta de transcriptos cerca del locus Cpf1 de *Francisella* que pudiera corresponder al ARNcrtrac que está asociado a los sistemas basados en Cas9.

Para confirmar que no se requiriera ARN adicional para la maduración del ARNcr y la interferencia con el ADN, se construyó un plásmido de expresión usando promotores sintéticos para dirigir la expresión de Cpf1 de *Francisella* (FnCpf1) y el arreglo CRISPR (pFnCpf1_min). Por medio de otra búsqueda con ARN pequeño en *E. coli* donde se expresaba este plásmido, se observó un procesamiento robusto del arreglo CRISPR a partir del cual se obtuvo el ARNcr maduro (figura 96B), a partir de lo cual puede concluirse que FnCpf1 y su arreglo CRISPR son suficientes para posibilitar el procesamiento del ARNcr. Por otro lado, en bacterias *E. coli* donde se expresaba pFnCpf1_min y pFnCpf1_ΔCas, un plásmido con todos los genes cas removidos pero con los promotores nativos que impulsan la expresión de FnCpf1 y el arreglo CRISPR, también se observó una interferencia robusta con el ADN, con lo cual puede establecerse que FnCpf1 y ARNcr son suficientes para posibilitar la guía hacia el ADN (figura 96C). Por el contrario, en el caso de Cas9 fue necesaria la presencia tanto del ARNcr como del ARNcrtrac para posibilitar la interferencia a nivel del ADN (Deltcheva y col., 2011; Zhang y col., 2013).

Ejemplo 10. Cpf1 es una endonucleasa única que está guiada por el ARNcr

El descubrimiento de que FnCpf1 puede mediar en la interferencia a nivel del ADN con el ARNcr por sí solo fue muy sorprendente, debido a que Cas9 puede reconocer el ARNcr través de la estructura doble que se forma entre el ARNcr y el ARNcrtrac (Jinek y col., 2012; Nishimasu y col., 2014), así como por medio de la estructura secundaria 3' del ARNcrtrac (Hsu y col., 2013; Nishimasu y col., 2014). Para asegurar que el ARNcr fuera de hecho suficiente para formar un complejo activo con FnCpf1 que resultara apropiado para posibilitar el clivaje del ADN dirigido por el ARN, se evaluó el efecto de FnCpf1 en combinación con el ARNcr sobre el clivaje del ADN diana in vitro. Se evaluó la capacidad de FnCpf1 purificado (figura 103) de clivar el mismo protoseparador 1 en el plásmido con experimentos de interferencia en el ADN de bacterias (figura 97A). FnCpf1 con un transcripto madurado in vitro en presencia del ARNcr y el protoseparador 1 pudo clivar eficientemente el plásmido diana en combinación con Mg^{2+}, lo cual también tuvo lugar de una manera dependiente del ARNcr (figura 97B). Por otra parte, FnCpf1 pudo clivar un diana de ADN superenrollado y un diana de ADN lineal (figura 97C). Con estos resultados, fue posible demostrar claramente que la presencia de FnCpf1 y de ARNcr es suficiente para posibilitar el clivaje del ADN mediado por el ARN.

El punto de corte en FnCpf1 también fue localizado mediante una secuenciación de Sanger en los extremos del ADN clivado. FnCpf1 participó en el clivaje de la porción saliente (figuras 97A, 97D y 104), a 5 nucleótidos del extremo 5' del producto, un sitio diferente del que se había obtenido con Cas9 (Garneau y col., 2010; Jinek y col., 2012; Gasiunas y col., 2012). El sitio donde efectuó el clivaje en FnCpf1 está distante del PAM: el clivaje se produjo después de la base 18a en la cadena (+) no específica y después de la base 23a en la cadena (-) que fue el diana (figuras 97A, 97D, y 104). El uso de sustratos en forma de oligonucleótidos de cadena doble que comprendían PAM con diversas secuencias también fue útil para comprobar que FnCpf1 puede clivar el ADN diana cuando el PAM 5'-TTN toma la

forma de una dupla (figura 97E), en contraste con los PAM de Cas9 (Sternberg y col., 2014).

Ejemplo 11. El dominio similar a RuvC de Cpf1 media en el clivaje del ADN operado por el ARN

El dominio similar a RuvC de Cpf1 retiene todos los residuos catalíticos propios de las endonucleasas de esta familia (figuras 98A y 105), y por lo tanto se prevé que sea una nucleasa activa. Se generaron tres mutantes, FnCpf1 (D917A), FnCpf1 (E1006A) y FnCpf1 (D1225A) (figura 98A), de modo tal de determinar si los residuos catalíticos conservados eran esenciales para la actividad de nucleasa de FnCpf1. Las mutaciones D917A y E1006A inactivaron por completo la actividad de clivaje del ADN de FnCpf1, mientras que D1255A dio como resultado una disminución significativa en la actividad nucleolítica (figura 98B). Estos resultados contrastaron con los resultados que se habían obtenido a través de la mutagénesis de Cas9 en *Streptococcus pyogenes* (SpCas9), donde la mutación en RuvC (D10A) y en HNH (N863A), que son dominios de nucleasas, habían convertido SpCas9 en una enzima apropiada para cortar el ADN (es decir, se había producido la inactivación de cada uno de los dos dominios de las nucleasas y se había anulado el clivaje de una de las cadenas del ADN) (Jinek y col., 2012; Gasiunas y col., 2012) (figura 98B). Sobre la base de estos descubrimientos, podría llegarse a la conclusión de que el dominio similar a RuvC de FnCpf1 puede escindir las dos cadenas del ADN diana, tal vez en una configuración dimérica (figura 103B).

Ejemplo 12. Secuencia y estructura del ARNcr de Cpf1

En comparación con el ARN de guía para Cas9, que presenta una estructura secundaria con características que posibilitan su interacción con Cas9 (Nishimasu y col., 2014), el ARN de guía para FnCpf1 es notablemente más simple y solamente comprende un único rizo, un tallo y una repetición directa (figura 97A).

Se analizaron los requisitos de la secuencia y la estructura del ARNcr para mediar en el clivaje del ADN con FnCpf1. En particular, se examinó la longitud de la secuencia de guía. Se observó que una secuencia de guía de 16 nucleótidos era necesaria para obtener un clivaje detectable del ADN, y que una secuencia de guía de 18 nucleótidos era necesaria para obtener un clivaje eficiente del ADN in vitro (figura 99A). Estas longitudes son similares a las que se habían demostrado para SpCas9, donde una secuencia separadora de 16 o 17 nucleótidos había sido suficiente para efectuar el clivaje del ADN (Cencic y col., 2014; Fu y col., 2014). Se observó que la región de partida para el ARN de guía de FnCpf1 se halló en los primeros 6 o 7 nucleótidos desde el extremo 5' de la secuencia del separador (figura 99B).

Se investigó el efecto de las mutaciones en la región de la repetición directa sobre la actividad de clivaje del ADN operada por el ARN. La porción de la repetición directa del ARNcr maduro tiene una longitud de 19 nucleótidos (figura 96A). Mediante el truncamiento de la repetición directa, se estableció que son suficientes 16 nucleótidos de la repetición directa para efectuar el clivaje, pero que óptimamente son necesarios 17 nucleótidos. Las mutaciones en el rizo o el tallo conservado en el ARN de cadena doble no afectaron la actividad de clivaje, mientras que las mutaciones que interrumpieron la estructura de la dupla del rizo y el tallo lo anularon (figura 99D). Por último, las sustituciones de bases en la región del rizo no afectaron la actividad de la nucleasa, mientras que la sustitución de la base U que se encontraba inmediatamente hacia el extremo 5' de la secuencia del separador tuvo como consecuencia una disminución sustancial en la actividad (figura 5E). Sobre la base de estos resultados, podría llegarse a la conclusión de que FnCpf1 puede reconocer el ARNcr a través de una combinación de características específicas en la secuencia y la estructura del rizo y el tallo.

Ejemplo 13. Las proteínas de la familia Cpf1 de diversas bacterias comparten presentan ARNcr y PAM con estructuras en común

Con el propósito de investigar el uso de Cpf1 como una herramienta para modificar el genoma, se aprovechó la diversidad de las proteínas de la familia Cpf1 disponibles en las bases de datos de secuencias públicas. Por medio de una búsqueda BLAST en la base de datos WGS del NCBI, se observaron 46 proteínas de la familia Cpf1 no redundantes (figura 64). Se seleccionaron 16 sobre la base de la reconstrucción filogenética que se realizó (figura 64), que fueron representantes de la diversidad de Cpf1 (figuras 100A-100B y 106). Estas proteínas de la familia Cpf1 presentaron diversas longitudes, de entre ~1200 y ~1500 aminoácidos.

Las secuencias de las repeticiones directas de cada una de estas proteínas en la familia Cpf1 presentaron una conservación fuerte en los 19 nucleótidos hacia el extremo 3', la parte de las repeticiones que se incluye en el ARNcr procesado (figura 100C). La secuencia hacia el extremo 5' de la repetición directa fue mucho más diversa. De las 16 proteínas de la familia Cpf1 que se seleccionaron para el análisis, tres (2 - Lb3Cpf1 de la bacteria Lachnospiraceae MC2017, 3 - BpCpf1 de *Butyrivibrio proteoclasticus* y 6 - SsCpf1 de *Smithella* sp. SC_K08D17) estuvieron asociadas a secuencias de repeticiones directas que fueron notablemente divergentes con relación a la de FnCpf1 (figura 100C). En particular, estas secuencias de repeticiones directas conservaron estructuras de rizos y tallos que fueron idénticas o casi idénticas a las de las repeticiones directas de FnCpf1 (figura 100D).

Se evaluó la capacidad de las secuencias de las repeticiones directas de mantener la actividad de nucleasa de FnCpf1 in vitro. Las repeticiones directas que contenían secuencias progenitoras conservadas pudieron operar indistintamente con FnCpf1. La repetición directa del candidato 3 (BpCpf1) propició una actividad de nucleasa baja en presencia de FnCpf1 (figura 100E), posiblemente debido a la conservación de los tres residuos de U hacia el extremo 3'.

Se realizó un estudio in vitro para identificar los PAM en cada proteína de la familia Cpf1 (figura 107A) y para determinar

su secuencia. Las secuencias de los PAM fueron identificadas en 7 nuevas proteínas de la familia Cpf1 (figuras 100E y 107B-C), y con el análisis se determinó que los PAM para FnCpf1 eran PAM 5'-TTN. Las secuencias de los PAM para las proteínas de la familia Cpf1 fueron predominantemente ricas en T y variaron principalmente en la cantidad de residuos de T que constituyeron cada PAM (figura 100F y 107B-C).

Ejemplo 14. Cpf1 puede aprovecharse para facilitar la modificación del genoma en las células humanas

Las proteínas de la familia Cpf1 fueron sometidas a una optimización a nivel de los codones y fueron unidas una señal de localización nuclear en el extremo C (NLS) para expresarlas en el núcleo de las células humanas (figura 101A). Para evaluar la actividad de cada proteína de la familia Cpf1, se seleccionó un sitio diana en el ARN de guía dentro del gen DNMT1 (figura 101B). Cada una de las proteínas de la familia Cpf1, junto con el ARNcr respectivo, que había sido diseñado de manera tal que resultara útil para efectuar la guía hacia DNMT1, dio como resultado el clivaje de un amplicón que se obtuvo medio de una PCR a partir de la región genómica de DNMT1 in vitro (figura 101C). Cuando se realizó un análisis en células de riñón embrionario humano 293FT (HEK 293FT), 2 de las proteínas de la familia Cpf1 (7 - AsCpf1 y 13 - LbCpf1) dieron como resultado un nivel detectable de indels inducidos por las nucleasas bajo las condiciones que se emplearon (figura 101C y D).

Cada una de las proteínas de la familia Cpf1 fue evaluada con blancos genómicos adicionales. Consistentemente, AsCpf1 y LbCpf1 dieron como resultado una modificación robusta del genoma en las células HEK293FT (figuras 101E y 108). Cuando se las comparó con Cas9, AsCpf1 y LbCpf1 dieron como resultado un nivel comparable de formación de indels (figura 101E). Por otra parte, se evaluó el clivaje in vitro, se llevó a cabo una secuenciación de Sanger en los extremos del ADN clivados y se determinó que 7 - AsCpf1 y 13 - LbCpf1 también generaron sitios de clivaje escalonados (figuras 101D y 107E).

A continuación se proveen las secuencias de nucleótidos y de aminoácidos de las construcciones y los ortólogos de FnCpf1.

Secuencias de locus FnCpf1

pFnCpf1

Extremo 5' de la acetiltransferasa endógena de F. novicida (hacia el extremo 5' del locus de FnCpf1) (en cursiva)

FnCpf1 (con subrayado simple)

Cas4 (con subrayado individual)

Cas1 (con subrayado individual)

Cas2 (con subrayado individual)

Repeticiones directas (en negrita)

Separador (con subrayado doble)

CATCAAGGAATTGGTTCTAAGCTTATAGAAGCAATGATTAAGGAAGCCA
AAAAAAATAATATTGATGCAATATTTGTCTTAGGTCATCCAAGTTATTATCCAAAATT
TGGTTTTAAACCAGCCACAGAATATCAGATAAAATGTGAATATGATGTCCCAGCGG
ATGTTTTTATGGTACTAGATTTGTCAGCTAAACTAGCTAGTTTAAAAGGACAAACTG
*TCTACTATGCCGATGAGTTTGGCAAAATTTTTTAG*ATCTACAAAATTATAAACTA
AATAAAGATTCTTATAATAACTTTATATATAATCGAAATGTAGAGAATTTTAT
AAGGAGTCTTTATCATGTCAATTTATCAAGAATTTGTTAATAAATATAGTTTA
AGTAAAACTCTAAGATTTGAGTTAATCCCACAGGGTAAAACACTTGAAAACA
TAAAAGCAAGAGGTTTGATTTTAGATGATGAGAAAAGAGCTAAAGACTACA
AAAAGGCTAAACAAATAATTGATAAATATCATCAGTTTTTTATAGAGGAGAT
ATTAAGTTCGGTTTGTATTAGCGAAGATTTATTACAAAACTATTCTGATGTTT
ATTTTAAACTTAAAAAGAGTGATGATGATAATCTACAAAAAGATTTTAAAAG
TGCAAAAGATACGATAAAGAAACAAATATCTGAATATATAAAGGACTCAGA
GAAATTTAAGAATTTGTTTAATCAAAACCTTATCGATGCTAAAAAAGGGCAA

GAGTCAGATTTAATTCTATGGCTAAAGCAATCTAAGGATAATGGTATAGAAC
TATTTAAAGCCAATAGTGATATCACAGATATAGATGAGGCGTTAGAAATAAT
CAAATCTTTTAAAGGTTGGACAACTTATTTTAAGGGTTTTCATGAAAATAGA
AAAAATGTTTATAGTAGCAATGATATTCCTACATCTATTATTTATAGGATAGT
AGATGATAATTTGCCTAAATTTCTAGAAAATAAAGCTAAGTATGAGAGTTTA
AAAGACAAAGCTCCAGAAGCTATAAACTATGAACAAATTAAAAAAGATTTG
GCAGAAGAGCTAACCTTTGATATTGACTACAAAACATCTGAAGTTAATCAAA
GAGTTTTTTCACTTGATGAAGTTTTTGAGATAGCAAACTTTAATAATTATCTA
AATCAAAGTGGTATTACTAAATTTAATACTATTATTGGTGGTAAATTTGTAAA
TGGTGAAAATACAAAGAGAAAAGGTATAAATGAATATATAAATCTATACTC
ACAGCAAATAAATGATAAAACACTCAAAAAATATAAAATGAGTGTTTTATTT
AAGCAAATTTTAAGTGATACAGAATCTAAATCTTTTGTAATTGATAAGTTAG
AAGATGATAGTGATGTAGTTACAACGATGCAAAGTTTTTATGAGCAAATAGC
AGCTTTTAAAACAGTAGAAGAAAAATCTATTAAAGAAACACTATCTTTATTA
TTTGATGATTTAAAAGCTCAAAAACTTGATTTGAGTAAAATTTATTTTAAAAA
TGATAAATCTCTTACTGATCTATCACAACAAGTTTTTGATGATTATAGTGTTA
TTGGTACAGCGGTACTAGAATATATAACTCAACAAATAGCACCTAAAAATCT
TGATAACCCTAGTAAGAAAGAGCAAGAATTAATAGCCAAAAAAACTGAAAA
AGCAAAATACTTATCTCTAGAAACTATAAAGCTTGCCTTAGAAGAATTTAAT
AAGCATAGAGATATAGATAAACAGTGTAGGTTTGAAGAAATACTTGCAAAC
TTTGCGGCTATTCCGATGATATTTGATGAAATAGCTCAAAACAAAGACAATT
TGGCACAGATATCTATCAAATATCAAAATCAAGGTAAAAAAGACCTACTTCA
AGCTAGTGCGGAAGATGATGTTAAAGCTATCAAGGATCTTTTAGATCAAACT
AATAATCTCTTACATAAACTAAAAATATTTCATATTAGTCAGTCAGAAGATA
AGGCAAATATTTTAGACAAGGATGAGCATTTTTATCTAGTATTTGAGGAGTG
CTACTTTGAGCTAGCGAATATAGTGCCTCTTTATAACAAAATTAGAAACTAT
ATAACTCAAAAGCCATATAGTGATGAGAAATTTAAGCTCAATTTTGAGAACT
CGACTTTGGCTAATGGTTGGGATAAAAATAAAGAGCCTGACAATACGGCAAT
TTTATTTATCAAAGATGATAAATATTATCTGGGTGTGATGAATAAGAAAAAT
AACAAAATATTTGATGATAAAGCTATCAAAGAAAATAAAGGCGAGGGTTAT
AAAAAAATTGTTTATAAACTTTTACCTGGCGCAAATAAAATGTTACCTAAGG
TTTTCTTTTCTGCTAAATCTATAAAATTTTATAATCCTAGTGAAGATATACTT
AGAATAAGAAATCATTCCACACATACAAAAAATGGTAGTCCTCAAAAAGGA

TATGAAAAATTTGAGTTTAATATTGAAGATTGCCGAAAATTTATAGATTTTTA
TAAACAGTCTATAAGTAAGCATCCGGAGTGGAAAGATTTTGGATTTAGATTT
TCTGATACTCAAAGATATAATTCTATAGATGAATTTTATAGAGAAGTTGAAA
ATCAAGGCTACAAACTAACTTTTGAAAATATATCAGAGAGCTATATTGATAG
CGTAGTTAATCAGGGTAAATTGTACCTATTCCAAATCTATAATAAAGATTTTT
CAGCTTATAGCAAAGGGCGACCAAATCTACATACTTTATATTGGAAAGCGCT
GTTTGATGAGAGAAATCTTCAAGATGTGGTTTATAAGCTAAATGGTGAGGCA
GAGCTTTTTTATCGTAAACAATCAATACCTAAAAAAATCACTCACCCAGCTA
AAGAGGCAATAGCTAATAAAAACAAAGATAATCCTAAAAAAGAGAGTGTTT
TTGAATATGATTTAATCAAAGATAAACGCTTTACTGAAGATAAGTTTTTCTTT
CACTGTCCTATTACAATCAATTTTAAATCTAGTGGAGCTAATAAGTTTAATGA
TGAAATCAATTTATTGCTAAAAGAAAAAGCAAATGATGTTCATATATTAAGT
ATAGATAGAGGTGAAAGACATTTAGCTTACTATACTTTGGTAGATGGTAAAG
GCAATATCATCAAACAAGATACTTTCAACATCATTGGTAATGATAGAATGAA
AACAAACTACCATGATAAGCTTGCTGCAATAGAGAAAGATAGGGATTCAGC
TAGGAAAGACTGGAAAAAGATAAATAACATCAAAGAGATGAAAGAGGGCT
ATCTATCTCAGGTAGTTCATGAAATAGCTAAGCTAGTTATAGAGTATAATGC
TATTGTGGTTTTTGAGGATTTAAATTTTGGATTTAAAAGAGGGCGTTTCAAGG
TAGAGAAGCAGGTCTATCAAAAGTTAGAAAAAATGCTAATTGAGAAACTAA
ACTATCTAGTTTTCAAAGATAATGAGTTTGATAAAACTGGGGGAGTGCTTAG
AGCTTATCAGCTAACAGCACCTTTTGAGACTTTTAAAAAGATGGGTAAACAA
ACAGGTATTATCTACTATGTACCAGCTGGTTTTACTTCAAAAATTTGTCCTGT
AACTGGTTTTGTAAATCAGTTATATCCTAAGTATGAAAGTGTCAGCAAATCT
CAAGAGTTCTTTAGTAAGTTTGACAAGATTTGTTATAACCTTGATAAGGGCT
ATTTTGAGTTTAGTTTTGATTATAAAAACTTTGGTGACAAGGCTGCCAAAGG
CAAGTGGACTATAGCTAGCTTTGGGAGTAGATTGATTAACTTTAGAAATTCA
GATAAAAAATCATAATTGGGATACTCGAGAAGTTTATCCAACTAAAGAGTTGG
AGAAATTGCTAAAAGATTATTCTATCGAATATGGGCATGGCGAATGTATCAA
AGCAGCTATTTGCGGTGAGAGCGACAAAAAGTTTTTTGCTAAGCTAACTAGT
GTCCTAAATACTATCTTACAAATGCGTAACTCAAAAACAGGTACTGAGTTAG
ATTATCTAATTTCACCAGTAGCAGATGTAAATGGCAATTTCTTTGATTCGCGA
CAGGCGCCAAAAAATATGCCTCAAGATGCTGATGCCAATGGTGCTTATCATA
TTGGGCTAAAAGGTCTGATGCTACTAGGTAGGATCAAAAATAATCAAGAGG

GCAAAAAACTCAATTTGGTTATCAAAAATGAAGAGTATTTTGAGTTCGTGCA
GAATAGGAATAACTAATTCATTCAAGAATATATTACCCTGTCAGTTTAGCGA
CTATTACCTCTTTAATAATTTGCAGGGGAATTATTTTAGTAATAGTAATATAC
ACAAGAGTTATTGATTATATGGAAAATTATATTTAGATAACATGGTTAAATG
ATTTTATATTCTGTCCTTACTCGATATATTTGCATAATATCTATAGTAATGCCT
CAGATACTACATACTATTCATCTAGCCAAACAAAAGGGCGCGATGCTCATAA
AAGTATCGATAAAGGAATCTATAGTACCAAAAAAGATGACCTGATCGGTATC
GATGTTATTAACCATAAATATGGTTTGGTTGGTAAAATTGATGTTTTTCATAA
AGATAAGGGCTTACTTGTGGAGAGAAAAAGGCAAATCAAGACTATCTATGA
TGGCTATAAATATCAGCTTTATGCGCAATATTTTTGTCTCCAAGAGATGGGCT
ATGATGTCAAAGCCATTAAATTTTATTCGATGGTTGATAATAAATCATACCC
AATAGCTATACCAACTTCAGCTGAGTTAGAAAAGTTTGAAAAACATATTCAA
ACAATCAAGCAATATAATCCAATGGATAACTCATTTAGGCAAAATATTGAAA
AGTGTAAATTTTGTATATATGCAAACTTATGTGATAAAACGGACTTGTAGAT
TATGTTTAGTAAAAATGATATTGAATCAAAGAATATAGTTTTTGTTAATATTT
TTGATGGAGTGAAACTTAGTCTATCATTGGGGAATATAGTTATAAAAGATAA
AGAAACTGATGAGGTGAAAACTAAGCTTTCTGTTCATAAAGTTCTTGCATTG
TTTATCGTAGGTAATATGACGATGACCTCGCAACTTTTAGAGACCTGTAAGA
AAAATGCTATACAGCTAGTTTTTATGAAAAATAGCTTTAGACCATATCTATGT
TTTGGTGATATTGCTGAGGCTAATTTTTTAGCTAGATATAAGCAATATAGTGT
AGTTGAGCAAGATATAAGTTTAGCAAGGATTTTTATAACATCAAAGATACGC
AATCAACATAACTTAGTCAAAAGCCTAAGAGATAAAACTCCAGAGCAGCAA
GAGATAGTCAAAAAGAATAAACAGCTAATAGCAGAGTTAGAAAATACAACA
AGCCTAGCGGAGCTAATGGGTATAGAGGGCAATGTTGCCAAAAATTTCTTCA
AAGGATTCTATGGACATTTAGATAGTTGGCAAGGGCGCAAACCTAGAATAA
AACAGGATCCATATAATGTTGTTTTAGACTTGGGCTATAGTATGTTGTTTAAT
TTTGTAGAGTGTTTTTTGCGACTTTTTGGCTTTGATTTATACAAGGGCTTTTGT
CATCAGACTTGGTATAAGCGTAAATCCCTAGTTTGTGACTTTGTTGAGCCATT
TAGATGTATAGTGGATAACCAAGTTAGAAAATCATGGAATCTCGGGCAATTT
TCTGTAGAGGATTTTGGTTGCAAAAATGAGCAGTTTTATATAAAAAAAGATA
AAACAAAAGACTACTCAAAAATACTTTTTGCCGAGATTATCAGCTACAAGCT
AGAGATATTTGAATATGTAAGAGAATTTTATCGTGCCTTTATGCGAGGCAAA
GAAATTGCAGAGTATCCAATATTTTGTTATGAAACTAGGAGGGTGTATGTTG

ATAGTCAGTTATGATTTTAGTAATAATAAAGTACGTGCAAAGTTTGCCAAAT
TTCTAGAAAGTTATGGTGTACGTTTACAATATTCGGTATTTGAGCTCAAATAT
AGCAAGAGAATGTTAGACTTGATTTTAGCTGAGATAGAAAATAACTATGTAC
CACTATTTACAAATGCTGATAGTGTTTTAATCTTTAATGCTCCAGATAAAGAT
GTGATAAAATATGGTTATGCGATTCATAGAGAACAAGAGGTTGTTTTTATAG
ACTAAAAATTGCAAACCTTAGTCTTTATGTTAAAATAACTACTAAGTTCTTAG
AGATATTTAAAAATATGACTGTTGTTATATATCAAAATGCTAAAAAAATCAT
AGATTTTAGGTCTTTTTTTGCTGATTTAGGCAAAAACGG**GTCTAAGAACTTT**
AAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACTGTTA
AAA**GTCTAAGAACTTTAAATAATTTCTACTGTTGTAGAT**GCTACTATTCCT
GTGCCTTCAGATAATTCA**GTCTAAGAACTTTAAATAATTTCTACTGTTGTA**
GATGTCTAGAGCCTTTTGTATTAGTAGCCG**GTCTAAGAACTTTAAATAATTT**
CTACTGTTGTAGATTAGCGATTTATGAAGGTCATTTTTTTGTCT (SEQ ID NO:
211)

pFnCpf1_min

Promotor Lac (con subrayado individual)

secuencia de Shine-Dalgarno (en cursiva)

FnCpf1 (con subrayado simple)

Promotor J23119 (con subrayado punteado)

Repeticiones directas (en negrita)

Separador (con subrayado doble)

TTTACACTTTATGCTTCCGGCTCGTATGTT*AGGAGG*TCTTTATCATGTC
AATTTATCAAGAATTTGTTAATAAATATAGTTTAAGTAAAACTCTAAGATTTGAGTT
AATCCCACAGGGTAAAACACTTGAAAACATAAAAGCAAGAGGTTTGATTTTAGATG
ATGAGAAAAGAGCTAAAGACTACAAAAAGGCTAAACAAATAATTGATAAATATCAT
CAGTTTTTTATAGAGGAGATATTAAGTTCGGTTTGTATTAGCGAAGATTTATTACAA
AACTATTCTGATGTTTATTTTAAACTTAAAAAGAGTGATGATGATAATCTACAAAAA
GATTTTAAAAGTGCAAAAGATACGATAAAGAAACAAATATCTGAATATATAAAGGA
CTCAGAGAAATTTAAGAATTTGTTTAATCAAAACCTTATCGATGCTAAAAAAGGGCA
AGAGTCAGATTTAATTCTATGGCTAAAGCAATCTAAGGATAATGGTATAGAACTATT
TAAAGCCAATAGTGATATCACAGATATAGATGAGGCGTTAGAAATAATCAAATCTTT
TAAAGGTTGGACAACTTATTTTAAGGGTTTTCATGAAAATAGAAAAAATGTTTATAG
TAGCAATGATATTCCTACATCTATTATTTATAGGATAGTAGATGATAATTTGCCTAAA
TTTCTAGAAAATAAAGCTAAGTATGAGAGTTTAAAAGACAAAGCTCCAGAAGCTAT

AAACTATGAACAAATTAAAAAAGATTTGGCAGAAGAGCTAACCTTTGATATTGACT
ACAAAACATCTGAAGTTAATCAAAGAGTTTTTTCACTTGATGAAGTTTTTGAGATAG
CAAACTTTAATAATTATCTAAATCAAAGTGGTATTACTAAATTTAATACTATTATTGG
TGGTAAATTTGTAAATGGTGAAAATACAAAGAGAAAAGGTATAAATGAATATATAA
ATCTATACTCACAGCAAATAAATGATAAAACACTCAAAAAATATAAAATGAGTGTTT
TATTTAAGCAAATTTTAAGTGATACAGAATCTAAATCTTTTGTAATTGATAAGTTAG
AAGATGATAGTGATGTAGTTACAACGATGCAAAGTTTTTATGAGCAAATAGCAGCTT
TTAAAACAGTAGAAGAAAAATCTATTAAAGAAACACTATCTTTATTATTTGATGATT
TAAAAGCTCAAAAACTTGATTTGAGTAAAATTTATTTTAAAAATGATAAATCTCTTA
CTGATCTATCACAACAAGTTTTTGATGATTATAGTGTTATTGGTACAGCGGTACTAG
AATATATAACTCAACAAATAGCACCTAAAAATCTTGATAACCCTAGTAAGAAAGAG
CAAGAATTAATAGCCAAAAAAACTGAAAAAGCAAAATACTTATCTCTAGAAACTAT
AAAGCTTGCCTTAGAAGAATTTAATAAGCATAGAGATATAGATAAACAGTGTAGGT
TTGAAGAAATACTTGCAAACTTTGCGGCTATTCCGATGATATTTGATGAAATAGCTC
AAAACAAAGACAATTTGGCACAGATATCTATCAAATATCAAAATCAAGGTAAAAAA
GACCTACTTCAAGCTAGTGCGGAAGATGATGTTAAAGCTATCAAGGATCTTTTAGAT
CAAACTAATAATCTCTTACATAAACTAAAAATATTTCATATTAGTCAGTCAGAAGAT
AAGGCAAATATTTTAGACAAGGATGAGCATTTTTATCTAGTATTTGAGGAGTGCTAC
TTTGAGCTAGCGAATATAGTGCCTCTTTATAACAAAATTAGAAACTATATAACTCAA
AAGCCATATAGTGATGAGAAATTTAAGCTCAATTTTGAGAACTCGACTTTGGCTAAT
GGTTGGGATAAAAATAAAGAGCCTGACAATACGGCAATTTTATTTATCAAAGATGAT
AAATATTATCTGGGTGTGATGAATAAGAAAAATAACAAAATATTTGATGATAAAGC
TATCAAAGAAAATAAAGGCGAGGGTTATAAAAAAATTGTTTATAAACTTTTACCTGG
CGCAAATAAAATGTTACCTAAGGTTTTCTTTTCTGCTAAATCTATAAAATTTTATAAT
CCTAGTGAAGATATACTTAGAATAAGAAATCATTCCACACATACAAAAAATGGTAG
TCCTCAAAAAGGATATGAAAAATTTGAGTTTAATATTGAAGATTGCCGAAAATTTAT
AGATTTTTATAAACAGTCTATAAGTAAGCATCCGGAGTGGAAAGATTTTGGATTTAG
ATTTTCTGATACTCAAAGATATAATTCTATAGATGAATTTTATAGAGAAGTTGAAAA
TCAAGGCTACAAACTAACTTTTGAAAATATATCAGAGAGCTATATTGATAGCGTAGT
TAATCAGGGTAAATTGTACCTATTCCAAATCTATAATAAAGATTTTTCAGCTTATAGC
AAAGGGCGACCAAATCTACATACTTTATATTGGAAAGCGCTGTTTGATGAGAGAAAT
CTTCAAGATGTGGTTTATAAGCTAAATGGTGAGGCAGAGCTTTTTTATCGTAAACAA
TCAATACCTAAAAAAATCACTCACCCAGCTAAAGAGGCAATAGCTAATAAAAACAA
AGATAATCCTAAAAAAGAGAGTGTTTTTGAATATGATTTAATCAAAGATAAACGCTT
TACTGAAGATAAGTTTTTCTTTCACTGTCCTATTACAATCAATTTTAAATCTAGTGGA
GCTAATAAGTTTAATGATGAAATCAATTTATTGCTAAAAGAAAAAGCAAATGATGTT

CATATATTAAGTATAGATAGAGGTGAAAGACATTTAGCTTACTATACTTTGGTAGAT
GGTAAAGGCAATATCATCAAACAAGATACTTTCAACATCATTGGTAATGATAGAATG
AAAACAAACTACCATGATAAGCTTGCTGCAATAGAGAAAGATAGGGATTCAGCTAG
GAAAGACTGGAAAAAGATAAATAACATCAAAGAGATGAAAGAGGGCTATCTATCTC
AGGTAGTTCATGAAATAGCTAAGCTAGTTATAGAGTATAATGCTATTGTGGTTTTTG
AGGATTTAAATTTTGGATTTAAAAGAGGGCGTTTCAAGGTAGAGAAGCAGGTCTATC
AAAAGTTAGAAAAAATGCTAATTGAGAAACTAAACTATCTAGTTTTCAAAGATAAT
GAGTTTGATAAAACTGGGGGAGTGCTTAGAGCTTATCAGCTAACAGCACCTTTTGAG
ACTTTTAAAAAGATGGGTAAACAAACAGGTATTATCTACTATGTACCAGCTGGTTTT
ACTTCAAAAATTTGTCCTGTAACTGGTTTTGTAAATCAGTTATATCCTAAGTATGAAA
GTGTCAGCAAATCTCAAGAGTTCTTTAGTAAGTTTGACAAGATTTGTTATAACCTTG
ATAAGGGCTATTTTGAGTTTAGTTTTGATTATAAAAACTTTGGTGACAAGGCTGCCA
AAGGCAAGTGGACTATAGCTAGCTTTGGGAGTAGATTGATTAACTTTAGAAATTCAG
ATAAAAATCATAATTGGGATACTCGAGAAGTTTATCCAACTAAAGAGTTGGAGAAA
TTGCTAAAAGATTATTCTATCGAATATGGGCATGGCGAATGTATCAAAGCAGCTATT
TGCGGTGAGAGCGACAAAAAGTTTTTTGCTAAGCTAACTAGTGTCCTAAATACTATC
TTACAAATGCGTAACTCAAAAACAGGTACTGAGTTAGATTATCTAATTTCACCAGTA
GCAGATGTAAATGGCAATTTCTTTGATTCGCGACAGGCGCCAAAAAATATGCCTCAA
GATGCTGATGCCAATGGTGCTTATCATATTGGGCTAAAAGGTCTGATGCTACTAGGT
AGGATCAAAAATAATCAAGAGGGCAAAAAACTCAATTTGGTTATCAAAAATGAAGA
GTATTTTGAGTTCGTGCAGAATAGGAATAACTAATTGACAGCTAGCTCAGTCCTAGG
TATAATGCTAGCGCTGATTTAGGCAAAAACGGGTCTAAGAACTTTAAATAATTTCT
ACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACTGTTAAAAGTCTAAGAACTTT
AAATAATTTCTACTGTTGTAGATGCTACTATTCCTGTGCCTTCAGATAATTCAGTCT
AAGAACTTTAAATAATTTCTACTGTTGTAGA (SEQ ID NO: 212)

pFnCpf1_Δcas

Extremo 5' de la acetiltransferasa endógena de F. novicida (hacia el extremo 5' del locus de FnCpf1) (en cursiva)

FnCpf1 (con subrayado simple)

Repeticiones directas (en negrita)

Separador (con subrayado doble)

*CTGTCTACTATGCCGATGAGTTTGGCAAAATTTTTTAG*ATCTACAAAATTA
TAAACTAAATAAAGATTCTTATAATAACTTTATATATAATCGAAATGTAGAGAATTT
TATAAGGAGTCTTTATCATGTCAATTTATCAAGAATTTGTTAATAAATATAGTTTAAG
TAAAACTCTAAGATTTGAGTTAATCCCACAGGGTAAAACACTTGAAAACATAAAAG
CAAGAGGTTTGATTTTAGATGATGAGAAAAGAGCTAAAGACTACAAAAAGGCTAAA

CAAATAATTGATAAATATCATCAGTTTTTTATAGAGGAGATATTAAGTTCGGTTTGT
ATTAGCGAAGATTTATTACAAAACTATTCTGATGTTTATTTTAAACTTAAAAAGAGT
GATGATGATAATCTACAAAAAGATTTTAAAAGTGCAAAAGATACGATAAAGAAACA
AATATCTGAATATATAAAGGACTCAGAGAAATTTAAGAATTTGTTTAATCAAAACCT
TATCGATGCTAAAAAAGGGCAAGAGTCAGATTTAATTCTATGGCTAAAGCAATCTAA
GGATAATGGTATAGAACTATTTAAAGCCAATAGTGATATCACAGATATAGATGAGG
CGTTAGAAATAATCAAATCTTTTAAAGGTTGGACAACTTATTTTAAGGGTTTTCATG
AAAATAGAAAAAATGTTTATAGTAGCAATGATATTCCTACATCTATTATTTATAGGA
TAGTAGATGATAATTTGCCTAAATTTCTAGAAAATAAAGCTAAGTATGAGAGTTTAA
AAGACAAAGCTCCAGAAGCTATAAACTATGAACAAATTAAAAAAGATTTGGCAGAA
GAGCTAACCTTTGATATTGACTACAAAACATCTGAAGTTAATCAAAGAGTTTTTTCA
CTTGATGAAGTTTTTGAGATAGCAAACTTTAATAATTATCTAAATCAAAGTGGTATT
ACTAAATTTAATACTATTATTGGTGGTAAATTTGTAAATGGTGAAAATACAAAGAGA
AAAGGTATAAATGAATATATAAATCTATACTCACAGCAAATAAATGATAAAACACT
CAAAAAATATAAAATGAGTGTTTTATTTAAGCAAATTTTAAGTGATACAGAATCTAA
ATCTTTTGTAATTGATAAGTTAGAAGATGATAGTGATGTAGTTACAACGATGCAAAG
TTTTTATGAGCAAATAGCAGCTTTTAAAACAGTAGAAGAAAAATCTATTAAAGAAAC
ACTATCTTTATTATTTGATGATTTAAAAGCTCAAAAACTTGATTTGAGTAAAATTTAT
TTTAAAAATGATAAATCTCTTACTGATCTATCACAACAAGTTTTTGATGATTATAGTG
TTATTGGTACAGCGGTACTAGAATATATAACTCAACAAATAGCACCTAAAAATCTTG
ATAACCCTAGTAAGAAAGAGCAAGAATTAATAGCCAAAAAAACTGAAAAAGCAAA
ATACTTATCTCTAGAAACTATAAAGCTTGCCTTAGAAGAATTTAATAAGCATAGAGA
TATAGATAAACAGTGTAGGTTTGAAGAAATACTTGCAAACTTTGCGGCTATTCCGAT
GATATTTGATGAAATAGCTCAAAACAAAGACAATTTGGCACAGATATCTATCAAATA
TCAAAATCAAGGTAAAAAAGACCTACTTCAAGCTAGTGCGGAAGATGATGTTAAAG
CTATCAAGGATCTTTTAGATCAAACTAATAATCTCTTACATAAACTAAAAATATTTC
ATATTAGTCAGTCAGAAGATAAGGCAAATATTTTAGACAAGGATGAGCATTTTTATC
TAGTATTTGAGGAGTGCTACTTTGAGCTAGCGAATATAGTGCCTCTTTATAACAAAA
TTAGAAACTATATAACTCAAAAGCCATATAGTGATGAGAAATTTAAGCTCAATTTTG
AGAACTCGACTTTGGCTAATGGTTGGGATAAAAATAAAGAGCCTGACAATACGGCA
ATTTTATTTATCAAAGATGATAAATATTATCTGGGTGTGATGAATAAGAAAAATAAC
AAAATATTTGATGATAAAGCTATCAAAGAAAATAAAGGCGAGGGTTATAAAAAAAT
TGTTTATAAACTTTTACCTGGCGCAAATAAAATGTTACCTAAGGTTTTCTTTTCTGCT
AAATCTATAAAATTTTATAATCCTAGTGAAGATATACTTAGAATAAGAAATCATTCC
ACACATACAAAAAATGGTAGTCCTCAAAAAGGATATGAAAAATTTGAGTTTAATATT
GAAGATTGCCGAAAATTTATAGATTTTTATAAACAGTCTATAAGTAAGCATCCGGAG

TGGAAAGATTTTGGATTTAGATTTTCTGATACTCAAAGATATAATTCTATAGATGAA
TTTTATAGAGAAGTTGAAAATCAAGGCTACAAACTAACTTTTGAAAATATATCAGAG
AGCTATATTGATAGCGTAGTTAATCAGGGTAAATTGTACCTATTCCAAATCTATAAT
AAAGATTTTTCAGCTTATAGCAAAGGGCGACCAAATCTACATACTTTATATTGGAAA
GCGCTGTTTGATGAGAGAAATCTTCAAGATGTGGTTTATAAGCTAAATGGTGAGGCA
GAGCTTTTTTATCGTAAACAATCAATACCTAAAAAAATCACTCACCCAGCTAAAGAG
GCAATAGCTAATAAAAACAAAGATAATCCTAAAAAAGAGAGTGTTTTTGAATATGA
TTTAATCAAAGATAAACGCTTTACTGAAGATAAGTTTTTCTTTCACTGTCCTATTACA
ATCAATTTTAAATCTAGTGGAGCTAATAAGTTTAATGATGAAATCAATTTATTGCTA
AAAGAAAAAGCAAATGATGTTCATATATTAAGTATAGATAGAGGTGAAAGACATTT
AGCTTACTATACTTTGGTAGATGGTAAAGGCAATATCATCAAACAAGATACTTTCAA
CATCATTGGTAATGATAGAATGAAAACAAACTACCATGATAAGCTTGCTGCAATAG
AGAAAGATAGGGATTCAGCTAGGAAAGACTGGAAAAAGATAAATAACATCAAAGA
GATGAAAGAGGGCTATCTATCTCAGGTAGTTCATGAAATAGCTAAGCTAGTTATAGA
GTATAATGCTATTGTGGTTTTTGAGGATTTAAATTTTGGATTTAAAAGAGGGCGTTTC
AAGGTAGAGAAGCAGGTCTATCAAAAGTTAGAAAAAATGCTAATTGAGAAACTAAA
CTATCTAGTTTTCAAAGATAATGAGTTTGATAAAACTGGGGGAGTGCTTAGAGCTTA
TCAGCTAACAGCACCTTTTGAGACTTTTAAAAAGATGGGTAAACAAACAGGTATTAT
CTACTATGTACCAGCTGGTTTTACTTCAAAAATTTGTCCTGTAACTGGTTTTGTAAAT
CAGTTATATCCTAAGTATGAAAGTGTCAGCAAATCTCAAGAGTTCTTTAGTAAGTTT
GACAAGATTTGTTATAACCTTGATAAGGGCTATTTTGAGTTTAGTTTTGATTATAAAA
ACTTTGGTGACAAGGCTGCCAAAGGCAAGTGGACTATAGCTAGCTTTGGGAGTAGA
TTGATTAACTTTAGAAATTCAGATAAAAATCATAATTGGGATACTCGAGAAGTTTAT
CCAACTAAAGAGTTGGAGAAATTGCTAAAAGATTATTCTATCGAATATGGGCATGGC
GAATGTATCAAAGCAGCTATTTGCGGTGAGAGCGACAAAAAGTTTTTTGCTAAGCTA
ACTAGTGTCCTAAATACTATCTTACAAATGCGTAACTCAAAAACAGGTACTGAGTTA
GATTATCTAATTTCACCAGTAGCAGATGTAAATGGCAATTTCTTTGATTCGCGACAG
GCGCCAAAAAATATGCCTCAAGATGCTGATGCCAATGGTGCTTATCATATTGGGCTA
AAAGGTCTGATGCTACTAGGTAGGATCAAAAATAATCAAGAGGGCAAAAAACTCAA
TTTGGTTATCAAAAATGAAGAGTATTTTGAGTTCGTGCAGAATAGGAATAACTAATT
CATTCAAGAATATATTACCCTGTCAGTTTAGCGACTATTACCTCTTTAATAATTTGCA
GGGGAATTATTTTAGTAATAGTAATATACACAAGAGTTATTGATTATATGGAAAATT
ATATTTAGATAACATGGTTAAATGATTTTATATTCTGTCCTTACTCGATATATTTTTTA
TAGACTAAAAATTGCAAACCTTAGTCTTTATGTTAAAATAACTACTAAGTTCTTAGA
GATATTTAAAAATATGACTGTTGTTATATATCAAAATGCTAAAAAAATCATAGATTT
TAGGTCTTTTTTTGCTGATTTAGGCAAAAACGGGTCTAAGAACTTTAAATAATTTCT

ACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACTGTTAAAAGTCTAAGAACTTT
AAATAATTTCTACTGTTGTAGATGCTACTATTCCTGTGCCTTCAGATAATTCAGTCT
AAGAACTTTAAATAATTTCTACTGTTGTAGATGTCTAGAGCCTTTTGTATTAGTAG
CCGGTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATTAGCGATTTATGAAGGT
CATTTTTTTGTCT (SEQ ID NO: 213)

Secuencias de nucleótidos de los ortólogos de Cpf1 con codones optimizados para el ser humano

Señal de localización nuclear (NLS) (en cursiva)

Conector de glicina-serina (con subrayado simple)

Marca de 3x HA (en negrita)

1 - Francisella tularensis subespecie novicida U112 (FnCpf1)

ATGAGCATCTACCAGGAGTTCGTCAACAAGTATTCACTGAGTAAGAC
ACTGCGGTTCGAGCTGATCCCACAGGGCAAGACACTGGAGAACATCAAGGCCCGAG
GCCTGATTCTGGACGATGAGAAGCGGGCAAAAGACTATAAGAAAGCCAAGCAGATC
ATTGATAAATACCACCAGTTCTTTATCGAGGAAATTCTGAGCTCCGTGTGCATCAGT
GAGGATCTGCTGCAGAATTACTCAGACGTGTACTTCAAGCTGAAGAAGAGCGACGA
TGACAACCTGCAGAAGGACTTCAAGTCCGCCAAGGACACCATCAAGAAACAGATTA
GCGAGTACATCAAGGACTCCGAAAAGTTTAAAAATCTGTTCAACCAGAATCTGATCG
ATGCTAAGAAAGGCCAGGAGTCCGACCTGATCCTGTGGCTGAAACAGTCTAAGGAC
AATGGGATTGAACTGTTCAAGGCTAACTCCGATATCACTGATATTGACGAGGCACTG
GAAATCATCAAGAGCTTCAAGGGATGGACCACATACTTTAAAGGCTTCCACGAGAA
CCGCAAGAACGTGTACTCCAGCAACGACATTCCTACCTCCATCATCTACCGAATCGT
CGATGACAATCTGCCAAAGTTCCTGGAGAACAAGGCCAAATATGAATCTCTGAAGG
ACAAAGCTCCCGAGGCAATTAATTACGAACAGATCAAGAAAGATCTGGCTGAGGAA
CTGACATTCGATATCGACTATAAGACTAGCGAGGTGAACCAGAGGGTCTTTTCCCTG
GACGAGGTGTTTGAAATCGCCAATTTCAACAATTACCTGAACCAGTCCGGCATTACT
AAATTCAATACCATCATTGGCGGGAAGTTTGTGAACGGGGAGAATACCAAGCGCAA
GGGAATTAACGAATACATCAATCTGTATAGCCAGCAGATCAACGACAAAACTCTGA
AGAAATACAAGATGTCTGTGCTGTTCAAACAGATCCTGAGTGATACCGAGTCCAAGT
CTTTTGTCATTGATAAACTGGAAGATGACTCAGACGTGGTCACTACCATGCAGAGCT
TTTATGAGCAGATCGCCGCTTTCAAGACAGTGGAGGAAAAATCTATTAAGGAAACTC
TGAGTCTGCTGTTCGATGACCTGAAAGCCCAGAAGCTGGACCTGAGTAAGATCTACT
TCAAAAACGATAAGAGTCTGACAGACCTGTCACAGCAGGTGTTTGATGACTATTCCG
TGATTGGGACCGCCGTCCTGGAGTACATTACACAGCAGATCGCTCCAAAGAACCTGG
ATAATCCCTCTAAGAAAGAGCAGGAACTGATCGCTAAGAAAACCGAGAAGGCAAAA
TATCTGAGTCTGGAAACAATTAAGCTGGCACTGGAGGAGTTCAACAAGCACAGGGA
TATTGACAAACAGTGCCGCTTTGAGGAAATCCTGGCCAACTTCGCAGCCATCCCCAT

GATTTTTGATGAGATCGCCCAGAACAAAGACAATCTGGCTCAGATCAGTATTAAGTA
CCAGAACCAGGGCAAGAAAGACCTGCTGCAGGCTTCAGCAGAAGATGACGTGAAAG
CCATCAAGGATCTGCTGGACCAGACCAACAATCTGCTGCACAAGCTGAAAATCTTCC
ATATTAGTCAGTCAGAGGATAAGGCTAATATCCTGGATAAAGACGAACACTTCTACC
TGGTGTTCGAGGAATGTTACTTCGAGCTGGCAAACATTGTCCCCCTGTATAACAAGA
TTAGGAACTACATCACACAGAAGCCTTACTCTGACGAGAAGTTTAAACTGAACTTCG
AAAATAGTACCCTGGCCAACGGGTGGGATAAGAACAAGGAGCCTGACAACACAGCT
ATCCTGTTCATCAAGGATGACAAGTACTATCTGGGAGTGATGAATAAGAAAAACAA
TAAGATCTTCGATGACAAAGCCATTAAGGAGAACAAAGGGGAAGGATACAAGAAA
ATCGTGTATAAGCTGCTGCCCGGCGCAAATAAGATGCTGCCTAAGGTGTTCTTCAGC
GCCAAGAGTATCAAATTCTACAACCCATCCGAGGACATCCTGCGGATTAGAAATCAC
TCAACACATACTAAGAACGGGAGCCCCCAGAAGGGATATGAGAAATTTGAGTTCAA
CATCGAGGATTGCAGGAAGTTTATTGACTTCTACAAGCAGAGCATCTCCAAACACCC
TGAATGGAAGGATTTTGGCTTCCGGTTTTCCGACACACAGAGATATAACTCTATCGA
CGAGTTCTACCGCGAGGTGGAAAATCAGGGGTATAAGCTGACTTTTGAGAACATTTC
TGAAAGTTACATCGACAGCGTGGTCAATCAGGGAAAGCTGTACCTGTTCCAGATCTA
TAACAAAGATTTTTCAGCATACAGCAAGGGCAGACCAAACCTGCATACACTGTACTG
GAAGGCCCTGTTCGATGAGAGGAATCTGCAGGACGTGGTCTATAAACTGAACGGAG
AGGCCGAACTGTTTTACCGGAAGCAGTCTATTCCTAAGAAAATCACTCACCCAGCTA
AGGAGGCCATCGCTAACAAGAACAAGGACAATCCTAAGAAAGAGAGCGTGTTCGAA
TACGATCTGATTAAGGACAAGCGGTTCACCGAAGATAAGTTCTTTTTCCATTGTCCA
ATCACCATTAACTTCAAGTCAAGCGGCGCTAACAAGTTCAACGACGAGATCAATCTG
CTGCTGAAGGAAAAAGCAAACGATGTGCACATCCTGAGCATTGACCGAGGAGAGCG
GCATCTGGCCTACTATACCCTGGTGGATGGCAAAGGGAATATCATTAAGCAGGATAC
ATTCAACATCATTGGCAATGACCGGATGAAAACCAACTACCACGATAAACTGGCTG
CAATCGAGAAGGATAGAGACTCAGCTAGGAAGGACTGGAAGAAAATCAACAACATT
AAGGAGATGAAGGAAGGCTATCTGAGCCAGGTGGTCCATGAGATTGCAAAGCTGGT
CATCGAATACAATGCCATTGTGGTGTTCGAGGATCTGAACTTCGGCTTTAAGAGGGG
GCGCTTTAAGGTGGAAAAACAGGTCTATCAGAAGCTGGAGAAAATGCTGATCGAAA
AGCTGAATTACCTGGTGTTTAAAGATAACGAGTTCGACAAGACCGGAGGCGTCCTG
AGAGCCTACCAGCTGACAGCTCCCTTTGAAACTTTCAAGAAAATGGGAAAACAGAC
AGGCATCATCTACTATGTGCCAGCCGGATTCACTTCCAAGATCTGCCCCGTGACCGG
CTTTGTCAACCAGCTGTACCCTAAATATGAGTCAGTGAGCAAGTCCCAGGAATTTTT
CAGCAAGTTCGATAAGATCTGTTATAATCTGGACAAGGGGTACTTCGAGTTTTCCTT
CGATTACAAGAACTTCGGCGACAAGGCCGCTAAGGGGAAATGGACCATTGCCTCCT
TCGGATCTCGCCTGATCAACTTTCGAAATTCCGATAAAAACCACAATTGGGACACTA

GGGAGGTGTACCCAACCAAGGAGCTGGAAAAGCTGCTGAAAGACTACTCTATCGAG
TATGGACATGGCGAATGCATCAAGGCAGCCATCTGTGGCGAGAGTGATAAGAAATT
TTTCGCCAAGCTGACCTCAGTGCTGAATACAATCCTGCAGATGCGGAACTCAAAGAC
CGGGACAGAACTGGACTATCTGATTAGCCCCGTGGCTGATGTCAACGGAAACTTCTT
CGACAGCAGACAGGCACCCAAAAATATGCCTCAGGATGCAGACGCCAACGGGGCCT
ACCACATCGGGCTGAAGGGACTGATGCTGCTGGGCCGGATCAAGAACAATCAGGAG
GGGAAGAAGCTGAACCTGGTCATTAAGAACGAGGAATACTTCGAGTTTGTCCAGAA
TAGAAATAAC*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAA*G
GGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGAT
TATGCATACCCATATGATGTCCCCGACTATGCC (SEQ ID NO: 214)

3 - La bacteria Lachnospiraceae MC2017 (Lb3Cpf1)

ATGGATTACGGCAACGGCCAGTTTGAGCGGAGAGCCCCCCTGACCAA
GACAATCACCCTGCGCCTGAAGCCTATCGGCGAGACACGGGAGACAATCCGCGAGC
AGAAGCTGCTGGAGCAGGACGCCGCCTTCAGAAAGCTGGTGGAGACAGTGACCCCT
ATCGTGGACGATTGTATCAGGAAGATCGCCGATAACGCCCTGTGCCACTTTGGCACC
GAGTATGACTTCAGCTGTCTGGGCAACGCCATCTCTAAGAATGACAGCAAGGCCATC
AAGAAGGAGACAGAGAAGGTGGAGAAGCTGCTGGCCAAGGTGCTGACCGAGAATC
TGCCAGATGGCCTGCGCAAGGTGAACGACATCAATTCCGCCGCCTTTATCCAGGATA
CACTGACCTCTTTCGTGCAGGACGATGCCGACAAGCGGGTGCTGATCCAGGAGCTGA
AGGGCAAGACCGTGCTGATGCAGCGGTTCCTGACCACACGGATCACAGCCCTGACC
GTGTGGCTGCCCGACAGAGTGTTCGAGAACTTTAATATCTTCATCGAGAACGCCGAG
AAGATGAGAATCCTGCTGGACTCCCCTCTGAATGAGAAGATCATGAAGTTTGACCCA
GATGCCGAGCAGTACGCCTCTCTGGAGTTCTATGGCCAGTGCCTGTCTCAGAAGGAC
ATCGATAGCTACAACCTGATCATCTCCGGCATCTATGCCGACGATGAGGTGAAGAAC
CCTGGCATCAATGAGATCGTGAAGGAGTACAATCAGCAGATCCGGGGCGACAAGGA
TGAGTCCCCACTGCCCAAGCTGAAGAAGCTGCACAAGCAGATCCTGATGCCAGTGG
AGAAGGCCTTCTTTGTGCGCGTGCTGTCTAACGACAGCGATGCCCGGAGCATCCTGG
AGAAGATCCTGAAGGACACAGAGATGCTGCCCTCCAAGATCATCGAGGCCATGAAG
GAGGCAGATGCAGGCGACATCGCCGTGTACGGCAGCCGGCTGCACGAGCTGAGCCA
CGTGATCTACGGCGATCACGGCAAGCTGTCCCAGATCATCTATGACAAGGAGTCCAA
GAGGATCTCTGAGCTGATGGAGACACTGTCTCCAAAGGAGCGCAAGGAGAGCAAGA
AGCGGCTGGAGGGCCTGGAGGAGCACATCAGAAAGTCTACATACACCTTCGACGAG
CTGAACAGGTATGCCGAGAAGAATGTGATGGCAGCATACATCGCAGCAGTGGAGGA
GTCTTGTGCCGAGATCATGAGAAAGGAGAAGGATCTGAGGACCCTGCTGAGCAAGG
AGGACGTGAAGATCCGGGGCAACAGACACAATACACTGATCGTGAAGAACTACTTT
AATGCCTGGACCGTGTTCCGGAACCTGATCAGAATCCTGAGGCGCAAGTCCGAGGC

CGAGATCGACTCTGACTTCTACGATGTGCTGGACGATTCCGTGGAGGTGCTGTCTCT
GACATACAAGGGCGAGAATCTGTGCCGCAGCTATATCACCAAGAAGATCGGCTCCG
ACCTGAAGCCCGAGATCGCCACATACGGCAGCGCCCTGAGGCCTAACAGCCGCTGG
TGGTCCCCAGGAGAGAAGTTTAATGTGAAGTTCCACACCATCGTGCGGAGAGATGG
CCGGCTGTACTATTTCATCCTGCCCAAGGGCGCCAAGCCTGTGGAGCTGGAGGACAT
GGATGGCGACATCGAGTGTCTGCAGATGAGAAAGATCCCTAACCCAACAATCTTTCT
GCCCAAGCTGGTGTTCAAGGACCCTGAGGCCTTCTTTAGGGATAATCCAGAGGCCGA
CGAGTTCGTGTTTCTGAGCGGCATGAAGGCCCCCGTGACAATCACCAGAGAGACAT
ACGAGGCCTACAGGTATAAGCTGTATACCGTGGGCAAGCTGCGCGATGGCGAGGTG
TCCGAAGAGGAGTACAAGCGGGCCCTGCTGCAGGTGCTGACCGCCTACAAGGAGTT
TCTGGAGAACAGAATGATCTATGCCGACCTGAATTTCGGCTTTAAGGATCTGGAGGA
GTATAAGGACAGCTCCGAGTTTATCAAGCAGGTGGAGACACACAACACCTTCATGT
GCTGGGCCAAGGTGTCTAGCTCCCAGCTGGACGATCTGGTGAAGTCTGGCAACGGCC
TGCTGTTCGAGATCTGGAGCGAGCGCCTGGAGTCCTACTATAAGTACGGCAATGAGA
AGGTGCTGCGGGGCTATGAGGGCGTGCTGCTGAGCATCCTGAAGGATGAGAACCTG
GTGTCCATGCGGACCCTGCTGAACAGCCGGCCCATGCTGGTGTACCGGCCAAAGGA
GTCTAGCAAGCCTATGGTGGTGCACCGGGATGGCAGCAGAGTGGTGGACAGGTTTG
ATAAGGACGGCAAGTACATCCCCCCTGAGGTGCACGACGAGCTGTATCGCTTCTTTA
ACAATCTGCTGATCAAGGAGAAGCTGGGCGAGAAGGCCCGGAAGATCCTGGACAAC
AAGAAGGTGAAGGTGAAGGTGCTGGAGAGCGAGAGAGTGAAGTGGTCCAAGTTCTA
CGATGAGCAGTTTGCCGTGACCTTCAGCGTGAAGAAGAACGCCGATTGTCTGGACAC
CACAAAGGACCTGAATGCCGAAGTGATGGAGCAGTATAGCGAGTCCAACAGACTGA
TCCTGATCAGGAATACCACAGATATCCTGTACTATCTGGTGCTGGACAAGAATGGCA
AGGTGCTGAAGCAGAGATCCCTGAACATCATCAATGACGGCGCCAGGGATGTGGAC
TGGAAGGAGAGGTTCCGCCAGGTGACAAAGGATAGAAACGAGGGCTACAATGAGT
GGGATTATTCCAGGACCTCTAACGACCTGAAGGAGGTGTACCTGAATTATGCCCTGA
AGGAGATCGCCGAGGCCGTGATCGAGTACAACGCCATCCTGATCATCGAGAAGATG
TCTAATGCCTTTAAGGACAAGTATAGCTTCCTGGACGACGTGACCTTCAAGGGCTTC
GAGACAAAGCTGCTGGCCAAGCTGAGCGATCTGCACTTTAGGGGCATCAAGGACGG
CGAGCCATGTTCCTTCACAAACCCCCTGCAGCTGTGCCAGAACGATTCTAATAAGAT
CCTGCAGGACGGCGTGATCTTTATGGTGCCAAATTCTATGACACGGAGCCTGGACCC
CGACACCGGCTTCATCTTTGCCATCAACGACCACAATATCAGGACCAAGAAGGCCA
AGCTGAACTTTCTGAGCAAGTTCGATCAGCTGAAGGTGTCCTCTGAGGGCTGCCTGA
TCATGAAGTACAGCGGCGATTCCCTGCCTACACACAACACCGACAATCGCGTGTGGA
ACTGCTGTTGCAATCACCCAATCACAAACTATGACCGGGAGACAAAGAAGGTGGAG
TTCATCGAGGAGCCCGTGGAGGAGCTGTCCCGCGTGCTGGAGGAGAATGGCATCGA

GACAGACACCGAGCTGAACAAGCTGAATGAGCGGGAGAACGTGCCTGGCAAGGTG
GTGGATGCCATCTACTCTCTGGTGCTGAATTATCTGCGCGGCACAGTGAGCGGAGTG
GCAGGACAGAGGGCCGTGTACTATAGCCCTGTGACCGGCAAGAAGTACGATATCTC
CTTTATCCAGGCCATGAACCTGAATAGGAAGTGTGACTACTATAGGATCGGCTCCAA
GGAGAGGGGAGAGTGGACCGATTTCGTGGCCCAGCTGATCAAC*AAAAGGCCGGCGG*
*CCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAA*GGATCC**TACCCATACGATGTTCC**
AGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCC
CGACTATGCC (SEQ ID NO: 215)

4 - Butyrivibrio proteoclasticus (BpCpf1)

ATGAGCATCTACCAGGAGTTCGTCAACAAGTATTCACTGAGTAAGAC
ACTGCGGTTCGAGCTGATCCCACAGGGCAAGACACTGGAGAACATCAAGGCCCGAG
GCCTGATTCTGGACGATGAGAAGCGGGCAAAAGACTATAAGAAAGCCAAGCAGATC
ATTGATAAATACCACCAGTTCTTTATCGAGGAAATTCTGAGCTCCGTGTGCATCAGT
GAGGATCTGCTGCAGAATTACTCAGACGTGTACTTCAAGCTGAAGAAGAGCGACGA
TGACAACCTGCAGAAGGACTTCAAGTCCGCCAAGGACACCATCAAGAAACAGATTA
GCGAGTACATCAAGGACTCCGAAAAGTTTAAAAATCTGTTCAACCAGAATCTGATCG
ATGCTAAGAAAGGCCAGGAGTCCGACCTGATCCTGTGGCTGAAACAGTCTAAGGAC
AATGGGATTGAACTGTTCAAGGCTAACTCCGATATCACTGATATTGACGAGGCACTG
GAAATCATCAAGAGCTTCAAGGGATGGACCACATACTTTAAAGGCTTCCACGAGAA
CCGCAAGAACGTGTACTCCAGCAACGACATTCCTACCTCCATCATCTACCGAATCGT
CGATGACAATCTGCCAAAGTTCCTGGAGAACAAGGCCAAATATGAATCTCTGAAGG
ACAAAGCTCCCGAGGCAATTAATTACGAACAGATCAAGAAAGATCTGGCTGAGGAA
CTGACATTCGATATCGACTATAAGACTAGCGAGGTGAACCAGAGGGTCTTTTCCCTG
GACGAGGTGTTTGAAATCGCCAATTTCAACAATTACCTGAACCAGTCCGGCATTACT
AAATTCAATACCATCATTGGCGGGAAGTTTGTGAACGGGGAGAATACCAAGCGCAA
GGGAATTAACGAATACATCAATCTGTATAGCCAGCAGATCAACGACAAAACTCTGA
AGAAATACAAGATGTCTGTGCTGTTCAAACAGATCCTGAGTGATACCGAGTCCAAGT
CTTTTGTCATTGATAAACTGGAAGATGACTCAGACGTGGTCACTACCATGCAGAGCT
TTTATGAGCAGATCGCCGCTTTCAAGACAGTGGAGGAAAAATCTATTAAGGAAACTC
TGAGTCTGCTGTTCGATGACCTGAAAGCCCAGAAGCTGGACCTGAGTAAGATCTACT
TCAAAAACGATAAGAGTCTGACAGACCTGTCACAGCAGGTGTTTGATGACTATTCCG
TGATTGGGACCGCCGTCCTGGAGTACATTACACAGCAGATCGCTCCAAAGAACCTGG
ATAATCCCTCTAAGAAAGAGCAGGAACTGATCGCTAAGAAAACCGAGAAGGCAAAA
TATCTGAGTCTGGAAACAATTAAGCTGGCACTGGAGGAGTTCAACAAGCACAGGGA
TATTGACAAACAGTGCCGCTTTGAGGAAATCCTGGCCAACTTCGCAGCCATCCCCAT
GATTTTTGATGAGATCGCCCAGAACAAAGACAATCTGGCTCAGATCAGTATTAAGTA

CCAGAACCAGGGCAAGAAAGACCTGCTGCAGGCTTCAGCAGAAGATGACGTGAAAG
CCATCAAGGATCTGCTGGACCAGACCAACAATCTGCTGCACAAGCTGAAAATCTTCC
ATATTAGTCAGTCAGAGGATAAGGCTAATATCCTGGATAAAGACGAACACTTCTACC
TGGTGTTCGAGGAATGTTACTTCGAGCTGGCAAACATTGTCCCCCTGTATAACAAGA
TTAGGAACTACATCACACAGAAGCCTTACTCTGACGAGAAGTTTAAACTGAACTTCG
AAAATAGTACCCTGGCCAACGGGTGGGATAAGAACAAGGAGCCTGACAACACAGCT
ATCCTGTTCATCAAGGATGACAAGTACTATCTGGGAGTGATGAATAAGAAAAACAA
TAAGATCTTCGATGACAAAGCCATTAAGGAGAACAAAGGGGAAGGATACAAGAAA
ATCGTGTATAAGCTGCTGCCCGGCGCAAATAAGATGCTGCCTAAGGTGTTCTTCAGC
GCCAAGAGTATCAAATTCTACAACCCATCCGAGGACATCCTGCGGATTAGAAATCAC
TCAACACATACTAAGAACGGGAGCCCCCAGAAGGGATATGAGAAATTTGAGTTCAA
CATCGAGGATTGCAGGAAGTTTATTGACTTCTACAAGCAGAGCATCTCCAAACACCC
TGAATGGAAGGATTTTGGCTTCCGGTTTTCCGACACACAGAGATATAACTCTATCGA
CGAGTTCTACCGCGAGGTGGAAAATCAGGGGTATAAGCTGACTTTTGAGAACATTTC
TGAAAGTTACATCGACAGCGTGGTCAATCAGGGAAAGCTGTACCTGTTCCAGATCTA
TAACAAAGATTTTTCAGCATACAGCAAGGGCAGACCAAACCTGCATACACTGTACTG
GAAGGCCCTGTTCGATGAGAGGAATCTGCAGGACGTGGTCTATAAACTGAACGGAG
AGGCCGAACTGTTTTACCGGAAGCAGTCTATTCCTAAGAAAATCACTCACCCAGCTA
AGGAGGCCATCGCTAACAAGAACAAGGACAATCCTAAGAAAGAGAGCGTGTTCGAA
TACGATCTGATTAAGGACAAGCGGTTCACCGAAGATAAGTTCTTTTTCCATTGTCCA
ATCACCATTAACTTCAAGTCAAGCGGCGCTAACAAGTTCAACGACGAGATCAATCTG
CTGCTGAAGGAAAAAGCAAACGATGTGCACATCCTGAGCATTGACCGAGGAGAGCG
GCATCTGGCCTACTATACCCTGGTGGATGGCAAAGGGAATATCATTAAGCAGGATAC
ATTCAACATCATTGGCAATGACCGGATGAAAACCAACTACCACGATAAACTGGCTG
CAATCGAGAAGGATAGAGACTCAGCTAGGAAGGACTGGAAGAAAATCAACAACATT
AAGGAGATGAAGGAAGGCTATCTGAGCCAGGTGGTCCATGAGATTGCAAAGCTGGT
CATCGAATACAATGCCATTGTGGTGTTCGAGGATCTGAACTTCGGCTTTAAGAGGGG
GCGCTTTAAGGTGGAAAAACAGGTCTATCAGAAGCTGGAGAAAATGCTGATCGAAA
AGCTGAATTACCTGGTGTTTAAAGATAACGAGTTCGACAAGACCGGAGGCGTCCTG
AGAGCCTACCAGCTGACAGCTCCCTTTGAAACTTTCAAGAAAATGGGAAAACAGAC
AGGCATCATCTACTATGTGCCAGCCGGATTCACTTCCAAGATCTGCCCCGTGACCGG
CTTTGTCAACCAGCTGTACCCTAAATATGAGTCAGTGAGCAAGTCCCAGGAATTTTT
CAGCAAGTTCGATAAGATCTGTTATAATCTGGACAAGGGGTACTTCGAGTTTTCCTT
CGATTACAAGAACTTCGGCGACAAGGCCGCTAAGGGGAAATGGACCATTGCCTCCT
TCGGATCTCGCCTGATCAACTTTCGAAATTCCGATAAAAACCACAATTGGGACACTA
GGGAGGTGTACCCAACCAAGGAGCTGGAAAAGCTGCTGAAAGACTACTCTATCGAG

TATGGACATGGCGAATGCATCAAGGCAGCCATCTGTGGCGAGAGTGATAAGAAATT
TTTCGCCAAGCTGACCTCAGTGCTGAATACAATCCTGCAGATGCGGAACTCAAAGAC
CGGGACAGAACTGGACTATCTGATTAGCCCCGTGGCTGATGTCAACGGAAACTTCTT
CGACAGCAGACAGGCACCCAAAAATATGCCTCAGGATGCAGACGCCAACGGGGCCT
ACCACATCGGGCTGAAGGGACTGATGCTGCTGGGCCGGATCAAGAACAATCAGGAG
GGGAAGAAGCTGAACCTGGTCATTAAGAACGAGGAATACTTCGAGTTTGTCCAGAA
TAGAAATAA*CAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*
GGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGAT**
TATGCATACCCATATGATGTCCCCGACTATGCC (SEQ ID NO: 216)

5 - La bacteria Peregrinibacteria GW2011_GWA_33_10 (PeCpf1)

ATGTCCAACTTCTTTAAGAATTTCACCAACCTGTATGAGCTGTCCAAG
ACACTGAGGTTTGAGCTGAAGCCCGTGGGCGACACCCTGACAAACATGAAGGACCA
CCTGGAGTACGATGAGAAGCTGCAGACCTTCCTGAAGGATCAGAATATCGACGATG
CCTATCAGGCCCTGAAGCCTCAGTTCGACGAGATCCACGAGGAGTTTATCACAGATT
CTCTGGAGAGCAAGAAGGCCAAGGAGATCGACTTCTCCGAGTACCTGGATCTGTTTC
AGGAGAAGAAGGAGCTGAACGACTCTGAGAAGAAGCTGCGCAACAAGATCGGCGA
GACATTCAACAAGGCCGGCGAGAAGTGGAAGAAGGAGAAGTACCCTCAGTATGAGT
GGAAGAAGGGCTCCAAGATCGCCAATGGCGCCGACATCCTGTCTTGCCAGGATATG
CTGCAGTTTATCAAGTATAAGAACCCAGAGGATGAGAAGATCAAGAATTACATCGA
CGATACACTGAAGGGCTTCTTTACCTATTTCGGCGGCTTTAATCAGAACAGGGCCAA
CTACTATGAGACAAAGAAGGAGGCCTCCACCGCAGTGGCAACAAGGATCGTGCACG
AGAACCTGCCAAAGTTCTGTGACAATGTGATCCAGTTTAAGCACATCATCAAGCGGA
AGAAGGATGGCACCGTGGAGAAAACCGAGAGAAAGACCGAGTACCTGAACGCCTA
CCAGTATCTGAAGAACAATAACAAGATCACACAGATCAAGGACGCCGAGACAGAGA
AGATGATCGAGTCTACACCCATCGCCGAGAAGATCTTCGACGTGTACTACTTCAGCA
GCTGCCTGAGCCAGAAGCAGATCGAGGAGTACAACCGGATCATCGGCCACTATAAT
CTGCTGATCAACCTGTATAACCAGGCCAAGAGATCTGAGGGCAAGCACCTGAGCGC
CAACGAGAAGAAGTATAAGGACCTGCCTAAGTTCAAGACCCTGTATAAGCAGATCG
GCTGCGGCAAGAAGAAGGACCTGTTTTACACAATCAAGTGTGATACCGAGGAGGAG
GCCAATAAGTCCCGGAACGAGGGCAAGGAGTCCCACTCTGTGGAGGAGATCATCAA
CAAGGCCCAGGAGGCCATCAATAAGTACTTCAAGTCTAATAACGACTGTGAGAATA
TCAACACCGTGCCCGACTTCATCAACTATATCCTGACAAAGGAGAATTACGAGGGCG
TGTATTGGAGCAAGGCCGCCATGAACACCATCTCCGACAAGTACTTCGCCAATTATC
ACGACCTGCAGGATAGACTGAAGGAGGCCAAGGTGTTTCAGAAGGCCGATAAGAAG
TCCGAGGACGATATCAAGATCCCAGAGGCCATCGAGCTGTCTGGCCTGTTCGGCGTG
CTGGACAGCCTGGCCGATTGGCAGACCACACTGTTTAAGTCTAGCATCCTGAGCAAC

GAGGACAAGCTGAAGATCATCACAGATTCCCAGACCCCCTCTGAGGCCCTGCTGAA
GATGATCTTCAATGACATCGAGAAGAACATGGAGTCCTTTCTGAAGGAGACAAACG
ATATCATCACCCTGAAGAAGTATAAGGGCAATAAGGAGGGCACCGAGAAGATCAAG
CAGTGGTTCGACTATACACTGGCCATCAACCGGATGCTGAAGTACTTTCTGGTGAAG
GAGAATAAGATCAAGGGCAACTCCCTGGATACCAATATCTCTGAGGCCCTGAAAAC
CCTGATCTACAGCGACGATGCCGAGTGGTTCAAGTGGTACGACGCCCTGAGAAACT
ATCTGACCCAGAAGCCTCAGGATGAGGCCAAGGAGAATAAGCTGAAGCTGAATTTC
GACAACCCATCTCTGGCCGGCGGCTGGGATGTGAACAAGGAGTGCAGCAATTTTTGC
GTGATCCTGAAGGACAAGAACGAGAAGAAGTACCTGGCCATCATGAAGAAGGGCG
AGAATACCCTGTTCCAGAAGGAGTGGACAGAGGGCCGGGGCAAGAACCTGACAAA
GAAGTCTAATCCACTGTTCGAGATCAATAACTGCGAGATCCTGAGCAAGATGGAGT
ATGACTTTTGGGCCGACGTGAGCAAGATGATCCCCAAGTGTAGCACCCAGCTGAAG
GCCGTGGTGAACCACTTCAAGCAGTCCGACAATGAGTTCATCTTTCCTATCGGCTAC
AAGGTGACAAGCGGCGAGAAGTTTAGGGAGGAGTGCAAGATCTCCAAGCAGGACTT
CGAGCTGAATAACAAGGTGTTTAATAAGAACGAGCTGAGCGTGACCGCCATGCGCT
ACGATCTGTCCTCTACACAGGAGAAGCAGTATATCAAGGCCTTCCAGAAGGAGTACT
GGGAGCTGCTGTTTAAGCAGGAGAAGCGGGACACCAAGCTGACAAATAACGAGATC
TTCAACGAGTGGATCAATTTTTGCAACAAGAAGTATAGCGAGCTGCTGTCCTGGGAG
AGAAAGTACAAGGATGCCCTGACCAATTGGATCAACTTCTGTAAGTACTTTCTGAGC
AAGTATCCCAAGACCACACTGTTCAACTACTCTTTTAAGGAGAGCGAGAATTATAAC
TCCCTGGACGAGTTCTACCGGGACGTGGATATCTGTTCTTACAAGCTGAATATCAAC
ACCACAATCAATAAGAGCATCCTGGATAGACTGGTGGAGGAGGGCAAGCTGTACCT
GTTTGAGATCAAGAATCAGGACAGCAACGATGGCAAGTCCATCGGCCACAAGAATA
ACCTGCACACCATCTACTGGAACGCCATCTTCGAGAATTTTGACAACAGGCCTAAGC
TGAATGGCGAGGCCGAGATCTTCTATCGCAAGGCCATCTCCAAGGATAAGCTGGGC
ATCGTGAAGGGCAAGAAAACCAAGAACGGCACCGAGATCATCAAGAATTACAGATT
CAGCAAGGAGAAGTTTATCCTGCACGTGCCAATCACCCTGAACTTCTGCTCCAATAA
CGAGTATGTGAATGACATCGTGAACACAAAGTTCTACAATTTTTCCAACCTGCACTT
TCTGGGCATCGATAGGGGCGAGAAGCACCTGGCCTACTATTCTCTGGTGAATAAGAA
CGGCGAGATCGTGGACCAGGGCACACTGAACCTGCCTTTCACCGACAAGGATGGCA
ATCAGCGCAGCATCAAGAAGGAGAAGTACTTTTATAACAAGCAGGAGGACAAGTGG
GAGGCCAAGGAGGTGGATTGTTGGAATTATAACGACCTGCTGGATGCCATGGCCTCT
AACCGGGACATGGCCAGAAAGAATTGGCAGAGGATCGGCACCATCAAGGAGGCCA
AGAACGGCTACGTGAGCCTGGTCATCAGGAAGATCGCCGATCTGGCCGTGAATAAC
GAGCGCCCCGCCTTCATCGTGCTGGAGGACCTGAATACAGGCTTTAAGCGGTCCAGA
CAGAAGATCGATAAGAGCGTGTACCAGAAGTTCGAGCTGGCCCTGGCCAAGAAGCT

GAACTTTCTGGTGGACAAGAATGCCAAGCGCGATGAGATCGGCTCCCCTACAAAGG
CCCTGCAGCTGACCCCCCCTGTGAATAACTACGGCGACATTGAGAACAAGAAGCAG
GCCGGCATCATGCTGTATACCCGGGCCAATTATACCTCTCAGACAGATCCAGCCACA
GGCTGGAGAAAGACCATCTATCTGAAGGCCGGCCCCGAGGAGACAACATACAAGAA
GGACGGCAAGATCAAGAACAAGAGCGTGAAGGACCAGATCATCGAGACATTCACCG
ATATCGGCTTTGACGGCAAGGATTACTATTTCGAGTACGACAAGGGCGAGTTTGTGG
ATGAGAAAACCGGCGAGATCAAGCCCAAGAAGTGGCGGCTGTACTCCGGCGAGAAT
GGCAAGTCCCTGGACAGGTTCCGCGGAGAGAGGGAGAAGGATAAGTATGAGTGGA
AGATCGACAAGATCGATATCGTGAAGATCCTGGACGATCTGTTCGTGAATTTTGACA
AGAACATCAGCCTGCTGAAGCAGCTGAAGGAGGGCGTGGAGCTGACCCGGAATAAC
GAGCACGGCACAGGCGAGTCCCTGAGATTCGCCATCAACCTGATCCAGCAGATCCG
GAATACCGGCAATAACGAGAGAGACAACGATTTCATCCTGTCCCCAGTGAGGGACG
AGAATGGCAAGCACTTTGACTCTCGCGAGTACTGGGATAAGGAGACAAAGGGCGAG
AAGATCAGCATGCCCAGCTCCGGCGATGCCAATGGCGCCTTCAACATCGCCCGGAA
GGGCATCATCATGAACGCCCACATCCTGGCCAATAGCGACTCCAAGGATCTGTCCCT
GTTCGTGTCTGACGAGGAGTGGGATCTGCACCTGAATAACAAGACCGAGTGGAAGA
AGCAGCTGAACATCTTTTCTAGCAGGAAGGCCATGGCCAAGCGCAAGAAG*AAAAGG*
*CCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*GGATCCTACCCATACG
ATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATG
ATGTCCCCGACTATGCCTAAGAATTC (SEQ ID NO: 217)

6 - La bacteria Parcubacteria GWC2011_GWC2_44_17 (PbCpf1)

ATGGAGAACATCTTCGACCAGTTTATCGGCAAGTACAGCCTGTCCAA
GACCCTGAGATTCGAGCTGAAGCCCGTGGGCAAGACAGAGGACTTCCTGAAGATCA
ACAAGGTGTTTGAGAAGGATCAGACCATCGACGATAGCTACAATCAGGCCAAGTTC
TATTTTGATTCCCTGCACCAGAAGTTTATCGACGCCGCCCTGGCCTCCGATAAGACA
TCCGAGCTGTCTTTCCAGAACTTTGCCGACGTGCTGGAGAAGCAGAATAAGATCATC
CTGGATAAGAAGAGAGAGATGGGCGCCCTGAGGAAGCGCGACAAGAACGCCGTGG
GCATCGATAGGCTGCAGAAGGAGATCAATGACGCCGAGGATATCATCCAGAAGGAG
AAGGAGAAGATCTACAAGGACGTGCGCACCCTGTTCGATAACGAGGCCGAGTCTTG
GAAAACCTACTATCAGGAGCGGGAGGTGGACGGCAAGAAGATCACCTTCAGCAAGG
CCGACCTGAAGCAGAAGGGCGCCGATTTTCTGACAGCCGCCGGCATCCTGAAGGTG
CTGAAGTATGAGTTCCCCGAGGAGAAGGAGAAGGAGTTTCAGGCCAAGAACCAGCC
CTCCCTGTTCGTGGAGGAGAAGGAGAATCCTGGCCAGAAGAGGTACATCTTCGACTC
TTTTGATAAGTTCGCCGGCTATCTGACCAAGTTTCAGCAGACAAAGAAGAATCTGTA
CGCAGCAGACGGCACCAGCACAGCAGTGGCCACCCGCATCGCCGATAACTTTATCA
TCTTCCACCAGAATACCAAGGTGTTCCGGGACAAGTACAAGAACAATCACACAGAC

CTGGGCTTCGATGAGGAGAACATCTTTGAGATCGAGAGGTATAAGAATTGCCTGCTG
CAGCGCGAGATCGAGCACATCAAGAATGAGAATAGCTACAACAAGATCATCGGCCG
GATCAATAAGAAGATCAAGGAGTATCGGGACCAGAAGGCCAAGGATACCAAGCTG
ACAAAGTCCGACTTCCCTTTCTTTAAGAACCTGGATAAGCAGATCCTGGGCGAGGTG
GAGAAGGAGAAGCAGCTGATCGAGAAAACCCGGGAGAAAACCGAGGAGGACGTGC
TGATCGAGCGGTTCAAGGAGTTCATCGAGAACAATGAGGAGAGGTTCACCGCCGCC
AAGAAGCTGATGAATGCCTTCTGTAACGGCGAGTTTGAGTCCGAGTACGAGGGCAT
CTATCTGAAGAATAAGGCCATCAACACAATCTCCCGGAGATGGTTCGTGTCTGACAG
AGATTTTGAGCTGAAGCTGCCTCAGCAGAAGTCCAAGAACAAGTCTGAGAAGAATG
AGCCAAAGGTGAAGAAGTTCATCTCCATCGCCGAGATCAAGAACGCCGTGGAGGAG
CTGGACGGCGATATCTTTAAGGCCGTGTTCTACGACAAGAAGATCATCGCCCAGGGC
GGCTCTAAGCTGGAGCAGTTCCTGGTCATCTGGAAGTACGAGTTTGAGTATCTGTTC
CGGGACATCGAGAGAGAGAACGGCGAGAAGCTGCTGGGCTATGATAGCTGCCTGAA
GATCGCCAAGCAGCTGGGCATCTTCCCACAGGAGAAGGAGGCCCGCGAGAAGGCAA
CCGCCGTGATCAAGAATTACGCCGACGCCGGCCTGGGCATCTTCCAGATGATGAAGT
ATTTTTCTCTGGACGATAAGGATCGGAAGAACACCCCCGGCCAGCTGAGCACAAATT
TCTACGCCGAGTATGACGGCTACTACAAGGATTTCGAGTTTATCAAGTACTACAACG
AGTTTAGGAACTTCATCACCAAGAAGCCTTTCGACGAGGATAAGATCAAGCTGAACT
TTGAGAATGGCGCCCTGCTGAAGGGCTGGGACGAGAACAAGGAGTACGATTTCATG
GGCGTGATCCTGAAGAAGGAGGGCCGCCTGTATCTGGGCATCATGCACAAGAACCA
CCGGAAGCTGTTTCAGTCCATGGGCAATGCCAAGGGCGACAACGCCAATAGATACC
AGAAGATGATCTATAAGCAGATCGCCGACGCCTCTAAGGATGTGCCCAGGCTGCTG
CTGACCAGCAAGAAGGCCATGGAGAAGTTCAAGCCTTCCCAGGAGATCCTGAGAAT
CAAGAAGGAGAAAACCTTCAAGCGGGAGAGCAAGAACTTTTCCCTGAGAGATCTGC
ACGCCCTGATCGAGTACTATAGGAACTGCATCCCTCAGTACAGCAATTGGTCCTTTT
ATGACTTCCAGTTTCAGGATACCGGCAAGTACCAGAATATCAAGGAGTTCACAGAC
GATGTGCAGAAGTACGGCTATAAGATCTCCTTTCGCGACATCGACGATGAGTATATC
AATCAGGCCCTGAACGAGGGCAAGATGTACCTGTTCGAGGTGGTGAACAAGGATAT
CTATAACACCAAGAATGGCTCCAAGAATCTGCACACACTGTACTTTGAGCACATCCT
GTCTGCCGAGAACCTGAATGACCCAGTGTTCAAGCTGTCTGGCATGGCCGAGATCTT
TCAGCGGCAGCCCAGCGTGAACGAAAGAGAGAAGATCACCACACAGAAGAATCAG
TGTATCCTGGACAAGGGCGATAGAGCCTACAAGTATAGGCGCTACACCGAGAAGAA
GATCATGTTCCACATGAGCCTGGTGCTGAACACAGGCAAGGGCGAGATCAAGCAGG
TGCAGTTTAATAAGATCATCAACCAGAGGATCAGCTCCTCTGACAACGAGATGAGG
GTGAATGTGATCGGCATCGATCGCGGCGAGAAGAACCTGCTGTACTATAGCGTGGT
GAAGCAGAATGGCGAGATCATCGAGCAGGCCTCCCTGAACGAGATCAATGGCGTGA

ACTACCGGGACAAGCTGATCGAGAGGGAGAAGGAGCGCCTGAAGAACCGGCAGAG
CTGGAAGCCTGTGGTGAAGATCAAGGATCTGAAGAAGGGCTACATCTCCCACGTGA
TCCACAAGATCTGCCAGCTGATCGAGAAGTATTCTGCCATCGTGGTGCTGGAGGACC
TGAATATGAGATTCAAGCAGATCAGGGGAGGAATCGAGCGGAGCGTGTACCAGCAG
TTCGAGAAGGCCCTGATCGATAAGCTGGGCTATCTGGTGTTTAAGGACAACAGGGAT
CTGAGGGCACCAGGAGGCGTGCTGAATGGCTACCAGCTGTCTGCCCCCTTTGTGAGC
TTCGAGAAGATGCGCAAGCAGACCGGCATCCTGTTCTACACACAGGCCGAGTATAC
CAGCAAGACAGACCCAATCACCGGCTTTCGGAAGAACGTGTATATCTCTAATAGCGC
CTCCCTGGATAAGATCAAGGAGGCCGTGAAGAAGTTCGACGCCATCGGCTGGGATG
GCAAGGAGCAGTCTTACTTCTTTAAGTACAACCCTTACAACCTGGCCGACGAGAAGT
ATAAGAACTCTACCGTGAGCAAGGAGTGGGCCATCTTTGCCAGCGCCCCAAGAATC
CGGAGACAGAAGGGCGAGGACGGCTACTGGAAGTATGATAGGGTGAAAGTGAATG
AGGAGTTCGAGAAGCTGCTGAAGGTCTGGAATTTTGTGAACCCAAAGGCCACAGAT
ATCAAGCAGGAGATCATCAAGAAGGAGAAGGCAGGCGACCTGCAGGGAGAGAAGG
AGCTGGATGGCCGGCTGAGAAACTTTTGGCACTCTTTCATCTACCTGTTTAACCTGGT
GCTGGAGCTGCGCAATTCTTTCAGCCTGCAGATCAAGATCAAGGCAGGAGAAGTGA
TCGCAGTGGACGAGGGCGTGGACTTCATCGCCAGCCCAGTGAAGCCCTTCTTTACCA
CACCCAACCCTTACATCCCCTCCAACCTGTGCTGGCTGGCCGTGGAGAATGCAGACG
CAAACGGAGCCTATAATATCGCCAGGAAGGGCGTGATGATCCTGAAGAAGATCCGC
GAGCACGCCAAGAAGGACCCCGAGTTCAAGAAGCTGCCAAACCTGTTTATCAGCAA
TGCAGAGTGGGACGAGGCAGCCCGGGATTGGGGCAAGTACGCAGGCACCACAGCCC
TGAACCTGGACCAC*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAA*
*AAAG*GGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCT**
GATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC (SEQ ID NO: 218)

7 - Smithella sp. SC_K08D17 (SsCpf1)

ATGCAGACCCTGTTTGAGAACTTCACAAATCAGTACCCAGTGTCCAA
GACCCTGCGCTTTGAGCTGATCCCCCAGGGCAAGACAAAGGACTTCATCGAGCAGA
AGGGCCTGCTGAAGAAGGATGAGGACCGGGCCGAGAAGTATAAGAAGGTGAAGAA
CATCATCGATGAGTACCACAAGGACTTCATCGAGAAGTCTCTGAATGGCCTGAAGCT
GGACGGCCTGGAGAAGTACAAGACCCTGTATCTGAAGCAGGAGAAGGACGATAAG
GATAAGAAGGCCTTTGACAAGGAGAAGGAGAACCTGCGCAAGCAGATCGCCAATGC
CTTCCGGAACAATGAGAAGTTTAAGACACTGTTCGCCAAGGAGCTGATCAAGAACG
ATCTGATGTCTTTCGCCTGCGAGGAGGACAAGAAGAATGTGAAGGAGTTTGAGGCC
TTCACCACATACTTCACCGGCTTCCACCAGAACCGCGCCAATATGTACGTGGCCGAT
GAGAAGAGAACAGCCATCGCCAGCAGGCTGATCCACGAGAACCTGCCAAAGTTTAT

CGACAATATCAAGATCTTCGAGAAGATGAAGAAGGAGGCCCCCGAGCTGCTGTCTC
CTTTCAACCAGACCCTGAAGGATATGAAGGACGTGATCAAGGGCACCACACTGGAG
GAGATCTTTAGCCTGGATTATTTCAACAAGACCCTGACACAGAGCGGCATCGACATC
TACAATTCCGTGATCGGCGGCAGAACCCCTGAGGAGGGCAAGACAAAGATCAAGGG
CCTGAACGAGTACATCAATACCGACTTCAACCAGAAGCAGACAGACAAGAAGAAGC
GGCAGCCAAAGTTCAAGCAGCTGTATAAGCAGATCCTGAGCGATAGGCAGAGCCTG
TCCTTTATCGCCGAGGCCTTCAAGAACGACACCGAGATCCTGGAGGCCATCGAGAA
GTTTTACGTGAATGAGCTGCTGCACTTCAGCAATGAGGGCAAGTCCACAAACGTGCT
GGACGCCATCAAGAATGCCGTGTCTAACCTGGAGAGCTTTAACCTGACCAAGATGTA
TTTCCGCTCCGGCGCCTCTCTGACAGACGTGAGCCGGAAGGTGTTTGGCGAGTGGAG
CATCATCAATAGAGCCCTGGACAACTACTATGCCACCACATATCCAATCAAGCCCAG
AGAGAAGTCTGAGAAGTACGAGGAGAGGAAGGAGAAGTGGCTGAAGCAGGACTTC
AACGTGAGCCTGATCCAGACCGCCATCGATGAGTACGACAACGAGACAGTGAAGGG
CAAGAACAGCGGCAAAGTGATCGCCGATTATTTTGCCAAGTTCTGCGACGATAAGG
AGACAGACCTGATCCAGAAGGTGAACGAGGGCTACATCGCCGTGAAGGATCTGCTG
AATACACCCTGTCCTGAGAACGAGAAGCTGGGCAGCAATAAGGACCAGGTGAAGCA
GATCAAGGCCTTTATGGATTCTATCATGGACATCATGCACTTCGTGCGCCCCCTGAG
CCTGAAGGATACCGACAAGGAGAAGGATGAGACATTCTACTCCCTGTTCACACCTCT
GTACGACCACCTGACCCAGACAATCGCCCTGTATAACAAGGTGCGGAACTATCTGAC
CCAGAAGCCTTACAGCACAGAGAAGATCAAGCTGAACTTCGAGAACAGCACCCTGC
TGGGCGGCTGGGATCTGAATAAGGAGACAGACAACACAGCCATCATCCTGAGGAAG
GATAACCTGTACTATCTGGGCATCATGGACAAGAGGCACAATCGCATCTTTCGGAAC
GTGCCCAAGGCCGATAAGAAGGACTTCTGCTACGAGAAGATGGTGTATAAGCTGCT
GCCTGGCGCCAACAAGATGCTGCCAAAGGTGTTCTTTTCTCAGAGCAGAATCCAGGA
GTTTACCCCTTCCGCCAAGCTGCTGGAGAACTACGCCAATGAGACACACAAGAAGG
GCGATAATTTCAACCTGAATCACTGTCACAAGCTGATCGATTTCTTTAAGGACTCTAT
CAACAAGCACGAGGATTGGAAGAATTTCGACTTTAGGTTCAGCGCCACCTCCACCTA
CGCCGACCTGAGCGGCTTTTACCACGAGGTGGAGCACCAGGGCTACAAGATCTCTTT
TCAGAGCGTGGCCGATTCCTTCATCGACGATCTGGTGAACGAGGGCAAGCTGTACCT
GTTCCAGATCTATAATAAGGACTTTTCCCCATTCTCTAAGGGCAAGCCCAACCTGCA
CACCCTGTACTGGAAGATGCTGTTTGATGAGAACAATCTGAAGGACGTGGTGTATAA
GCTGAATGGCGAGGCCGAGGTGTTCTACCGCAAGAAGAGCATTGCCGAGAAGAACA
CCACAATCCACAAGGCCAATGAGTCCATCATCAACAAGAATCCTGATAACCCAAAG
GCCACCAGCACCTTCAACTATGATATCGTGAAGGACAAGAGATACACCATCGACAA
GTTTCAGTTCCACATCCCAATCACAATGAACTTTAAGGCCGAGGGCATCTTCAACAT
GAATCAGAGGGTGAATCAGTTCCTGAAGGCCAATCCCGATATCAACATCATCGGCAT

CGACAGAGGCGAGAGGCACCTGCTGTACTATGCCCTGATCAACCAGAAGGGCAAGA
TCCTGAAGCAGGATACCCTGAATGTGATCGCCAACGAGAAGCAGAAGGTGGACTAC
CACAATCTGCTGGATAAGAAGGAGGGCGACCGCGCAACCGCAAGGCAGGAGTGGG
GCGTGATCGAGACAATCAAGGAGCTGAAGGAGGGCTATCTGTCCCAGGTCATCCAC
AAGCTGACCGATCTGATGATCGAGAACAATGCCATCATCGTGATGGAGGACCTGAA
CTTTGGCTTCAAGCGGGGCAGACAGAAGGTGGAGAAGCAGGTGTATCAGAAGTTTG
AGAAGATGCTGATCGATAAGCTGAATTACCTGGTGGACAAGAATAAGAAGGCAAAC
GAGCTGGGAGGCCTGCTGAACGCATTCCAGCTGGCCAATAAGTTTGAGTCCTTCCAG
AAGATGGGCAAGCAGAACGGCTTTATCTTCTACGTGCCCGCCTGGAATACCTCTAAG
ACAGATCCTGCCACCGGCTTTATCGACTTCCTGAAGCCCCGCTATGAGAACCTGAAT
CAGGCCAAGGATTTCTTTGAGAAGTTTGACTCTATCCGGCTGAACAGCAAGGCCGAT
TACTTTGAGTTCGCCTTTGACTTCAAGAATTTCACCGAGAAGGCCGATGGCGGCAGA
ACCAAGTGGACAGTGTGCACCACAAACGAGGACAGATATGCCTGGAATAGGGCCCT
GAACAATAACAGGGGCAGCCAGGAGAAGTACGACATCACAGCCGAGCTGAAGTCCC
TGTTCGATGGCAAGGTGGACTATAAGTCTGGCAAGGATCTGAAGCAGCAGATCGCC
AGCCAGGAGTCCGCCGACTTCTTTAAGGCCCTGATGAAGAACCTGTCCATCACCCTG
TCTCTGAGACACAATAACGGCGAGAAGGGCGATAATGAGCAGGACTACATCCTGTC
CCCTGTGGCCGATTCTAAGGGCCGCTTCTTTGACTCCCGGAAGGCCGACGATGACAT
GCCAAAGAATGCCGACGCCAACGGCGCCTATCACATCGCCCTGAAGGGCCTGTGGT
GTCTGGAGCAGATCAGCAAGACCGATGACCTGAAGAAGGTGAAGCTGGCCATCTCC
AACAAGGAGTGGCTGGAGTTCGTGCAGACACTGAAGGG*CAAAAGGCCCGGCGGCCAC*
*GAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*GGATCC**TACCCATACGATGTTCCAGA**
TTACGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGA
CTATGCCTAAGAATTC (SEQ ID NO: 219)

8 - Acidaminococcus sp. BV3L6 (AsCpf1)

ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCA
AGACACTGCGGTTTGAGCTGATCCCACAGGGCAAGACCCTGAAGCACATCCA
GGAGCAGGGCTTCATCGAGGAGGACAAGGCCCGCAATGATCACTACAAGGA
GCTGAAGCCCATCATCGATCGGATCTACAAGACCTATGCCGACCAGTGCCTG
CAGCTGGTGCAGCTGGATTGGGAGAACCTGAGCGCCGCCATCGACTCCTATA
GAAAGGAGAAAACCGAGGAGACAAGGAACGCCCTGATCGAGGAGCAGGCC
ACATATCGCAATGCCATCCACGACTACTTCATCGGCCGGACAGACAACCTGA
CCGATGCCATCAATAAGAGACACGCCGAGATCTACAAGGGCCTGTTCAAGG
CCGAGCTGTTTAATGGCAAGGTGCTGAAGCAGCTGGGCACCGTGACCACAAC
CGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTTACAACCTACTTC

TCCGGCTTTTATGAGAACAGGAAGAACGTGTTCAGCGCCGAGGATATCAGCA
CAGCCATCCCACACCGCATCGTGCAGGACAACTTCCCCAAGTTTAAGGAGAA
TTGTCACATCTTCACACGCCTGATCACCGCCGTGCCCAGCCTGCGGGAGCAC
TTTGAGAACGTGAAGAAGGCCATCGGCATCTTCGTGAGCACCTCCATCGAGG
AGGTGTTTTCCTTCCCTTTTTATAACCAGCTGCTGACACAGACCCAGATCGAC
CTGTATAACCAGCTGCTGGGAGGAATCTCTCGGGAGGCAGGCACCGAGAAG
ATCAAGGGCCTGAACGAGGTGCTGAATCTGGCCATCCAGAAGAATGATGAG
ACAGCCCACATCATCGCCTCCCTGCCACACAGATTCATCCCCCTGTTTAAGCA
GATCCTGTCCGATAGGAACACCCTGTCTTTCATCCTGGAGGAGTTTAAGAGC
GACGAGGAAGTGATCCAGTCCTTCTGCAAGTACAAGACACTGCTGAGAAAC
GAGAACGTGCTGGAGACAGCCGAGGCCCTGTTTAACGAGCTGAACAGCATC
GACCTGACACACATCTTCATCAGCCACAAGAAGCTGGAGACAATCAGCAGC
GCCCTGTGCGACCACTGGGATACACTGAGGAATGCCCTGTATGAGCGGAGA
ATCTCCGAGCTGACAGGCAAGATCACCAAGTCTGCCAAGGAGAAGGTGCAG
CGCAGCCTGAAGCACGAGGATATCAACCTGCAGGAGATCATCTCTGCCGCAG
GCAAGGAGCTGAGCGAGGCCTTCAAGCAGAAAACCAGCGAGATCCTGTCCC
ACGCACACGCCGCCCTGGATCAGCCACTGCCTACAACCCTGAAGAAGCAGG
AGGAGAAGGAGATCCTGAAGTCTCAGCTGGACAGCCTGCTGGGCCTGTACC
ACCTGCTGGACTGGTTTGCCGTGGATGAGTCCAACGAGGTGGACCCCGAGTT
CTCTGCCCGGCTGACCGGCATCAAGCTGGAGATGGAGCCTTCTCTGAGCTTC
TACAACAAGGCCAGAAATTATGCCACCAAGAAGCCCTACTCCGTGGAGAAG
TTCAAGCTGAACTTTCAGATGCCTACACTGGCCTCTGGCTGGGACGTGAATA
AGGAGAAGAACAATGGCGCCATCCTGTTTGTGAAGAACGGCCTGTACTATCT
GGGCATCATGCCAAAGCAGAAGGGCAGGTATAAGGCCCTGAGCTTCGAGCC
CACAGAGAAAACCAGCGAGGGCTTTGATAAGATGTACTATGACTACTTCCCT
GATGCCGCCAAGATGATCCCAAAGTGCAGCACCCAGCTGAAGGCCGTGACA
GCCCACTTTCAGACCCACACAACCCCCATCCTGCTGTCCAACAATTTCATCGA
GCCTCTGGAGATCACAAAGGAGATCTACGACCTGAACAATCCTGAGAAGGA
GCCAAAGAAGTTTCAGACAGCCTACGCCAAGAAAACCGGCGACCAGAAGGG
CTACAGAGAGGCCCTGTGCAAGTGGATCGACTTCACAAGGGATTTTCTGTCC
AAGTATACCAAGACAACCTCTATCGATCTGTCTAGCCTGCGGCCATCCTCTC
AGTATAAGGACCTGGGCGAGTACTATGCCGAGCTGAATCCCCTGCTGTACCA
CATCAGCTTCCAGAGAATCGCCGAGAAGGAGATCATGGATGCCGTGGAGAC

AGGCAAGCTGTACCTGTTCCAGATCTATAACAAGGACTTTGCCAAGGGCCAC
CACGGCAAGCCTAATCTGCACACACTGTATTGGACCGGCCTGTTTTCTCCAG
AGAACCTGGCCAAGACAAGCATCAAGCTGAATGGCCAGGCCGAGCTGTTCT
ACCGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAGA
TGCTGAACAAGAAGCTGAAGGATCAGAAAACCCCAATCCCCGACACCCTGT
ACCAGGAGCTGTACGACTATGTGAATCACAGACTGTCCCACGACCTGTCTGA
TGAGGCCAGGGCCCTGCTGCCCAACGTGATCACCAAGGAGGTGTCTCACGAG
ATCATCAAGGATAGGCGCTTTACCAGCGACAAGTTCTTTTTCCACGTGCCTAT
CACACTGAACTATCAGGCCGCCAATTCCCCATCTAAGTTCAACCAGAGGGTG
AATGCCTACCTGAAGGAGCACCCCGAGACACCTATCATCGGCATCGATCGGG
GCGAGAGAAACCTGATCTATATCACAGTGATCGACTCCACCGGCAAGATCCT
GGAGCAGCGGAGCCTGAACACCATCCAGCAGTTTGATTACCAGAAGAAGCT
GGACAACAGGGAGAAGGAGAGGGTGGCAGCAAGGCAGGCCTGGTCTGTGGT
GGGCACAATCAAGGATCTGAAGCAGGGCTATCTGAGCCAGGTCATCCACGA
GATCGTGGACCTGATGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTG
AATTTCGGCTTTAAGAGCAAGAGGACCGGCATCGCCGAGAAGGCCGTGTAC
CAGCAGTTCGAGAAGATGCTGATCGATAAGCTGAATTGCCTGGTGCTGAAGG
ACTATCCAGCAGAGAAAGTGGGAGGCGTGCTGAACCCATACCAGCTGACAG
ACCAGTTCACCTCCTTTGCCAAGATGGGCACCCAGTCTGGCTTCCTGTTTTAC
GTGCCTGCCCCATATACATCTAAGATCGATCCCCTGACCGGCTTCGTGGACC
CCTTCGTGTGGAAAACCATCAAGAATCACGAGAGCCGCAAGCACTTCCTGGA
GGGCTTCGACTTTCTGCACTACGACGTGAAAACCGGCGACTTCATCCTGCAC
TTTAAGATGAACAGAAATCTGTCCTTCCAGAGGGGCCTGCCCGGCTTTATGC
CTGCATGGGATATCGTGTTCGAGAAGAACGAGACACAGTTTGACGCCAAGG
GCACCCCTTTCATCGCCGGCAAGAGAATCGTGCCAGTGATCGAGAATCACAG
ATTCACCGGCAGATACCGGGACCTGTATCCTGCCAACGAGCTGATCGCCCTG
CTGGAGGAGAAGGGCATCGTGTTCAGGGATGGCTCCAACATCCTGCCAAAG
CTGCTGGAGAATGACGATTCTCACGCCATCGACACCATGGTGGCCCTGATCC
GCAGCGTGCTGCAGATGCGGAACTCCAATGCCGCCACAGGCGAGGACTATA
TCAACAGCCCCGTGCGCGATCTGAATGGCGTGTGCTTCGACTCCCGGTTTCA
GAACCCAGAGTGGCCCATGGACGCCGATGCCAATGGCGCCTACCACATCGCC
CTGAAGGGCCAGCTGCTGCTGAATCACCTGAAGGAGAGCAAGGATCTGAAG
CTGCAGAACGGCATCTCCAATCAGGACTGGCTGGCCTACATCCAGGAGCTGC

GCAAC*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*
GGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCC**
TGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC (SEQ
ID NO: 220)

9 - La bacteria Lachnospiraceae MA2020 (Lb2Cpf1)

ATGTACTATGAGTCCCTGACCAAGCAGTACCCCGTGTCTAAGACAAT
CCGGAATGAGCTGATCCCTATCGGCAAGACACTGGATAACATCCGCCAGAACAATA
TCCTGGAGAGCGACGTGAAGCGGAAGCAGAACTACGAGCACGTGAAGGGCATCCTG
GATGAGTATCACAAGCAGCTGATCAACGAGGCCCTGGACAATTGCACCCTGCCATCC
CTGAAGATCGCCGCCGAGATCTACCTGAAGAATCAGAAGGAGGTGTCTGACAGAGA
GGATTTCAACAAGACACAGGACCTGCTGAGGAAGGAGGTGGTGGAGAAGCTGAAG
GCCCACGAGAACTTTACCAAGATCGGCAAGAAGGACATCCTGGATCTGCTGGAGAA
GCTGCCTTCCATCTCTGAGGACGATTACAATGCCCTGGAGAGCTTCCGCAACTTTTA
CACCTATTTCACATCCTACAACAAGGTGCGGGAGAATCTGTATTCTGATAAGGAGAA
GAGCTCCACAGTGGCCTACAGACTGATCAACGAGAATTTCCCAAAGTTTCTGGACAA
TGTGAAGAGCTATAGGTTTGTGAAAACCGCAGGCATCCTGGCAGATGGCCTGGGAG
AGGAGGAGCAGGACTCCCTGTTCATCGTGGAGACATTCAACAAGACCCTGACACAG
GACGGCATCGATACCTACAATTCTCAAGTGGGCAAGATCAACTCTAGCATCAATCTG
TATAACCAGAAGAATCAGAAGGCCAATGGCTTCAGAAAGATCCCCAAGATGAAGAT
GCTGTATAAGCAGATCCTGTCCGATAGGGAGGAGTCTTTCATCGACGAGTTTCAGAG
CGATGAGGTGCTGATCGACAACGTGGAGTCTTATGGCAGCGTGCTGATCGAGTCTCT
GAAGTCCTCTAAGGTGAGCGCCTTCTTTGATGCCCTGAGAGAGTCTAAGGGCAAGAA
CGTGTACGTGAAGAATGACCTGGCCAAGACAGCCATGAGCAACATCGTGTTCGAGA
ATTGGAGGACCTTTGACGATCTGCTGAACCAGGAGTACGACCTGGCCAACGAGAAC
AAGAAGAAGGACGATAAGTATTTCGAGAAGCGCCAGAAGGAGCTGAAGAAGAATA
AGAGCTACTCCCTGGAGCACCTGTGCAACCTGTCCGAGGATTCTTGTAACCTGATCG
AGAATTATATCCACCAGATCTCCGACGATATCGAGAATATCATCATCAACAATGAGA
CATTCCTGCGCATCGTGATCAATGAGCACGACAGGTCCCGCAAGCTGGCCAAGAAC
CGGAAGGCCGTGAAGGCCATCAAGGACTTTCTGGATTCTATCAAGGTGCTGGAGCG
GGAGCTGAAGCTGATCAACAGCTCCGGCCAGGAGCTGGAGAAGGATCTGATCGTGT
ACTCTGCCCACGAGGAGCTGCTGGTGGAGCTGAAGCAGGTGGACAGCCTGTATAAC
ATGACCAGAAATTATCTGACAAAGAAGCCTTTCTCTACCGAGAAGGTGAAGCTGAA
CTTTAATCGCAGCACACTGCTGAACGGCTGGGATCGGAATAAGGAGACAGACAACC
TGGGCGTGCTGCTGCTGAAGGACGGCAAGTACTATCTGGGCATCATGAACACAAGC
GCCAATAAGGCCTTCGTGAATCCCCCTGTGGCCAAGACCGAGAAGGTGTTTAAGAA
GGTGGATTACAAGCTGCTGCCAGTGCCCAACCAGATGCTGCCAAAGGTGTTCTTTGC

CAAGAGCAATATCGACTTCTATAACCCCTCTAGCGAGATCTACTCCAATTATAAGAA
GGGCACCCACAAGAAGGGCAATATGTTTTCCCTGGAGGATTGTCACAACCTGATCGA
CTTCTTTAAGGAGTCTATCAGCAAGCACGAGGACTGGAGCAAGTTCGGCTTTAAGTT
CAGCGATACAGCCTCCTACAACGACATCTCCGAGTTCTATCGCGAGGTGGAGAAGC
AGGGCTACAAGCTGACCTATACAGACATCGATGAGACATACATCAATGATCTGATC
GAGCGGAACGAGCTGTACCTGTTCCAGATCTATAATAAGGACTTTAGCATGTACTCC
AAGGGCAAGCTGAACCTGCACACACTGTATTTCATGATGCTGTTTGATCAGCGCAAT
ATCGACGACGTGGTGTATAAGCTGAACGGAGAGGCAGAGGTGTTCTATAGGCCAGC
CTCCATCTCTGAGGACGAGCTGATCATCCACAAGGCCGGCGAGGAGATCAAGAACA
AGAATCCTAACCGGGCCAGAACCAAGGAGACAAGCACCTTCAGCTACGACATCGTG
AAGGATAAGCGGTATAGCAAGGATAAGTTTACCCTGCACATCCCCATCACAATGAA
CTTCGGCGTGGATGAGGTGAAGCGGTTCAACGACGCCGTGAACAGCGCCATCCGGA
TCGATGAGAATGTGAACGTGATCGGCATCGACCGGGGCGAGAGAAATCTGCTGTAC
GTGGTGGTCATCGACTCTAAGGGCAACATCCTGGAGCAGATCTCCCTGAACTCTATC
ATCAATAAGGAGTACGACATCGAGACAGATTATCACGCACTGCTGGATGAGAGGGA
GGGCGGCAGAGATAAGGCCCGGAAGGACTGGAACACCGTGGAGAATATCAGGGAC
CTGAAGGCCGGCTACCTGAGCCAGGTGGTGAACGTGGTGGCCAAGCTGGTGCTGAA
GTATAATGCCATCATCTGCCTGGAGGACCTGAACTTTGGCTTCAAGAGGGGCCGCCA
GAAGGTGGAGAAGCAGGTGTACCAGAAGTTCGAGAAGATGCTGATCGATAAGCTGA
ATTACCTGGTCATCGACAAGAGCCGCGAGCAGACATCCCCTAAGGAGCTGGGAGGC
GCCCTGAACGCACTGCAGCTGACCTCTAAGTTCAAGAGCTTTAAGGAGCTGGGCAA
GCAGTCCGGCGTGATCTACTATGTGCCTGCCTACCTGACCTCTAAGATCGATCCAAC
CACAGGCTTCGCCAATCTGTTTTATATGAAGTGTGAGAACGTGGAGAAGTCCAAGAG
ATTCTTTGACGGCTTTGATTTCATCAGGTTCAACGCCCTGGAGAACGTGTTCGAGTTC
GGCTTTGACTACCGGAGCTTCACCCAGAGGGCCTGCGGCATCAATTCCAAGTGGACC
GTGTGCACCAACGGCGAGCGCATCATCAAGTATCGGAATCCAGATAAGAACAATAT
GTTCGACGAGAAGGTGGTGGTGGTGACCGATGAGATGAAGAACCTGTTTGAGCAGT
ACAAGATCCCCTATGAGGATGGCAGAAATGTGAAGGACATGATCATCAGCAACGAG
GAGGCCGAGTTCTACCGGAGACTGTATAGGCTGCTGCAGCAGACCCTGCAGATGAG
AAACAGCACCTCCGACGGCACAAGGGATTACATCATCTCCCCTGTGAAGAATAAGA
GAGAGGCCTACTTCAACAGCGAGCTGTCCGACGGCTCTGTGCCAAAGGACGCCGAT
GCCAACGGCGCCTACAATATCGCCAGAAAGGGCCTGTGGGTGCTGGAGCAGATCAG
GCAGAAGAGCGAGGGCGAGAAGATCAATCTGGCCATGACCAACGCCGAGTGGCTGG
AGTATGCCCAGACACACCTGCTG*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGG*
*CAAAAAAGAAAAAG*GGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCTAC

GACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC

(SEQ ID NO: 221)

10 - La bacteria Methanoplasma termitum candidata (CMtCpf1)

ATGAACAATTACGACGAGTTCACCAAGCTGTATCCTATCCAGAAAAC
CATCCGGTTTGAGCTGAAGCCACAGGGCAGAACCATGGAGCACCTGGAGACATTCA
ACTTCTTTGAGGAGGACCGGGATAGAGCCGAGAAGTATAAGATCCTGAAGGAGGCC
ATCGACGAGTACCACAAGAAGTTTATCGATGAGCACCTGACCAATATGTCCCTGGAT
TGGAACTCTCTGAAGCAGATCAGCGAGAAGTACTATAAGAGCAGGGAGGAGAAGG
ACAAGAAGGTGTTCCTGTCCGAGCAGAAGAGGATGCGCCAGGAGATCGTGTCTGAG
TTTAAGAAGGACGATCGCTTCAAGGACCTGTTTTCCAAGAAGCTGTTCTCTGAGCTG
CTGAAGGAGGAGATCTACAAGAAGGGCAACCACCAGGAGATCGACGCCCTGAAGA
GCTTCGATAAGTTTTCCGGCTATTTCATCGGCCTGCACGAGAATAGGAAGAACATGT
ACTCCGACGGCGATGAGATCACCGCCATCTCCAATCGCATCGTGAATGAGAACTTCC
CCAAGTTTCTGGATAACCTGCAGAAGTACCAGGAGGCCAGGAAGAAGTATCCTGAG
TGGATCATCAAGGCCGAGAGCGCCCTGGTGGCCCACAATATCAAGATGGACGAGGT
GTTCTCCCTGGAGTACTTTAATAAGGTGCTGAACCAGGAGGGCATCCAGCGGTACAA
CCTGGCCCTGGGCGGCTATGTGACCAAGAGCGGCGAGAAGATGATGGGCCTGAATG
ATGCCCTGAACCTGGCCCACCAGTCCGAGAAGAGCTCCAAGGGCAGAATCCACATG
ACCCCCCTGTTCAAGCAGATCCTGTCCGAGAAGGAGTCCTTCTCTTACATCCCCGAC
GTGTTTACAGAGGATTCTCAGCTGCTGCCTAGCATCGGCGGCTTCTTTGCCCAGATC
GAGAATGACAAGGATGGCAACATCTTCGACCGGGCCCTGGAGCTGATCTCTAGCTA
CGCCGAGTATGATACCGAGCGGATCTATATCAGACAGGCCGACATCAATAGAGTGT
CCAACGTGATCTTTGGAGAGTGGGGCACCCTGGGAGGCCTGATGAGGGAGTACAAG
GCCGACTCTATCAATGATATCAACCTGGAGCGCACATGCAAGAAGGTGGACAAGTG
GCTGGATTCTAAGGAGTTTGCCCTGAGCGATGTGCTGGAGGCCATCAAGAGGACCG
GCAACAATGACGCCTTCAACGAGTATATCTCCAAGATGCGGACAGCCAGAGAGAAG
ATCGATGCCGCCCGCAAGGAGATGAAGTTCATCAGCGAGAAGATCTCCGGCGATGA
GGAGTCTATCCACATCATCAAGACCCTGCTGGACAGCGTGCAGCAGTTCCTGCACTT
CTTTAATCTGTTTAAGGCAAGGCAGGACATCCCACTGGATGGAGCCTTCTACGCCGA
GTTTGACGAGGTGCACAGCAAGCTGTTTGCCATCGTGCCCCTGTATAACAAGGTGCG
GAACTATCTGACCAAGAACAATCTGAACACAAAGAAGATCAAGCTGAATTTCAAGA
ACCCTACACTGGCCAATGGCTGGGACCAGAACAAGGTGTACGATTATGCCTCCCTGA
TCTTTCTGCGGGACGGCAATTACTATCTGGGCATCATCAATCCTAAGAGAAAGAAGA
ACATCAAGTTCGAGCAGGGCTCTGGCAACGGCCCCTTCTACCGGAAGATGGTGTATA
AGCAGATCCCCGGCCCTAATAAGAACCTGCCAAGAGTGTTCCTGACCTCCACAAAG
GGCAAGAAGGAGTATAAGCCCTCTAAGGAGATCATCGAGGGCTACGAGGCCGACAA

GCACATCAGGGGCGATAAGTTCGACCTGGATTTTTGTCACAAGCTGATCGATTTCTT
TAAGGAGTCCATCGAGAAGCACAAGGACTGGTCTAAGTTCAACTTCTACTTCAGCCC
AACCGAGAGCTATGGCGACATCTCTGAGTTCTACCTGGATGTGGAGAAGCAGGGCT
ATCGCATGCACTTTGAGAATATCAGCGCCGAGACAATCGACGAGTATGTGGAGAAG
GGCGATCTGTTTCTGTTCCAGATCTACAACAAGGATTTTGTGAAGGCCGCCACCGGC
AAGAAGGACATGCACACAATCTACTGGAATGCCGCCTTCAGCCCCGAGAACCTGCA
GGACGTGGTGGTGAAGCTGAACGGCGAGGCCGAGCTGTTTTATAGGGACAAGTCCG
ATATCAAGGAGATCGTGCACCGCGAGGGCGAGATCCTGGTGAATAGGACCTACAAC
GGCCGCACACCAGTGCCCGACAAGATCCACAAGAAGCTGACCGATTATCACAATGG
CCGGACAAAGGACCTGGGCGAGGCCAAGGAGTACCTGGATAAGGTGAGATACTTCA
AGGCCCACTATGACATCACCAAGGATCGGAGATACCTGAACGACAAGATCTATTTCC
ACGTGCCTCTGACCCTGAACTTCAAGGCCAACGGCAAGAAGAATCTGAACAAGATG
GTCATCGAGAAGTTCCTGTCCGATGAGAAGGCCCACATCATCGGCATCGACAGGGG
CGAGCGCAATCTGCTGTACTATTCCATCATCGACAGGTCTGGCAAGATCATCGATCA
GCAGAGCCTGAATGTGATCGACGGCTTTGATTATCGGGAGAAGCTGAACCAGAGAG
AGATCGAGATGAAGGATGCCCGCCAGTCTTGGAACGCCATCGGCAAGATCAAGGAC
CTGAAGGAGGGCTACCTGAGCAAGGCCGTGCACGAGATCACCAAGATGGCCATCCA
GTATAATGCCATCGTGGTCATGGAGGAGCTGAACTACGGCTTCAAGCGGGGCCGGTT
CAAGGTGGAGAAGCAGATCTATCAGAAGTTCGAGAATATGCTGATCGATAAGATGA
ACTACCTGGTGTTTAAGGACGCACCTGATGAGTCCCCAGGAGGCGTGCTGAATGCCT
ACCAGCTGACAAACCCACTGGAGTCTTTCGCCAAGCTGGGCAAGCAGACCGGCATC
CTGTTTTACGTGCCAGCCGCCTATACATCCAAGATCGACCCCACCACAGGCTTCGTG
AATCTGTTTAACACCTCCTCTAAGACAAACGCCCAGGAGCGGAAGGAGTTCCTGCAG
AAGTTTGAGAGCATCTCCTATTCTGCCAAGGATGGCGGCATCTTTGCCTTCGCCTTTG
ACTACAGAAAGTTCGGCACCAGCAAGACAGATCACAAGAACGTGTGGACCGCCTAT
ACAAACGGCGAGAGGATGCGCTACATCAAGGAGAAGAAGCGGAATGAGCTGTTTGA
CCCTTCTAAGGAGATCAAGGAGGCCCTGACCAGCTCCGGCATCAAGTACGATGGCG
GCCAGAACATCCTGCCAGACATCCTGAGGAGCAACAATAACGGCCTGATCTACACA
ATGTATTCTAGCTTCATCGCCGCCATCCAGATGCGCGTGTACGACGGCAAGGAGGAT
TATATCATCAGCCCCATCAAGAACTCCAAGGGCGAGTTCTTTAGGACCGACCCCAAG
AGGCGCGAGCTGCCTATCGACGCCGATGCCAATGGCGCCTACAACATCGCCCTGAG
GGGAGAGCTGACAATGAGGGCAATCGCAGAGAAGTTCGACCCTGATAGCGAGAAG
ATGGCCAAGCTGGAGCTGAAGCACAAGGATTGGTTCGAGTTTATGCAGACCAGAGG
CGAC*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*GGATCC
TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGATTATGCA
TACCCATATGATGTCCCCGACTATGCCTAAGAATTC (SEQ ID NO: 222)

11 - La eubacteria seleccionada (EeCpf1)

ATGAACGGCAATAGGTCCATCGTGTACCGCGAGTTCGTGGGCGTGAT
CCCCGTGGCCAAGACCCTGAGGAATGAGCTGCGCCCTGTGGGCCACACACAGGAGC
ACATCATCCAGAACGGCCTGATCCAGGAGGACGAGCTGCGGCAGGAGAAGAGCACC
GAGCTGAAGAACATCATGGACGATTACTATAGAGAGTACATCGATAAGTCTCTGAG
CGGCGTGACCGACCTGGACTTCACCCTGCTGTTCGAGCTGATGAACCTGGTGCAGAG
CTCCCCCTCCAAGGACAATAAGAAGGCCCTGGAGAAGGAGCAGTCTAAGATGAGGG
AGCAGATCTGCACCCACCTGCAGTCCGACTCTAACTACAAGAATATCTTTAACGCCA
AGCTGCTGAAGGAGATCCTGCCTGATTTCATCAAGAACTACAATCAGTATGACGTGA
AGGATAAGGCCGGCAAGCTGGAGACACTGGCCCTGTTTAATGGCTTCAGCACATACT
TTACCGACTTCTTTGAGAAGAGGAAGAACGTGTTCACCAAGGAGGCCGTGAGCACA
TCCATCGCCTACCGCATCGTGCACGAGAACTCCCTGATCTTCCTGGCCAATATGACC
TCTTATAAGAAGATCAGCGAGAAGGCCCTGGATGAGATCGAAGTGATCGAGAAGAA
CAATCAGGACAAGATGGGCGATTGGGAGCTGAATCAGATCTTTAACCCTGACTTCTA
CAATATGGTGCTGATCCAGTCCGGCATCGACTTCTACAACGAGATCTGCGGCGTGGT
GAATGCCCACATGAACCTGTACTGTCAGCAGACCAAGAACAATTATAACCTGTTCAA
GATGCGGAAGCTGCACAAGCAGATCCTGGCCTACACCAGCACCAGCTTCGAGGTGC
CCAAGATGTTCGAGGACGATATGAGCGTGTATAACGCCGTGAACGCCTTCATCGACG
AGACAGAGAAGGGCAACATCATCGGCAAGCTGAAGGATATCGTGAATAAGTACGAC
GAGCTGGATGAGAAGAGAATCTATATCAGCAAGGACTTTTACGAGACACTGAGCTG
CTTCATGTCCGGCAACTGGAATCTGATCACAGGCTGCGTGGAGAACTTCTACGATGA
GAACATCCACGCCAAGGGCAAGTCCAAGGAGGAGAAGGTGAAGAAGGCCGTGAAG
GAGGACAAGTACAAGTCTATCAATGACGTGAACGATCTGGTGGAGAAGTATATCGA
TGAGAAGGAGAGGAATGAGTTCAAGAACAGCAATGCCAAGCAGTACATCCGCGAG
ATCTCCAACATCATCACCGACACAGAGACAGCCCACCTGGAGTATGACGATCACATC
TCTCTGATCGAGAGCGAGGAGAAGGCCGACGAGATGAAGAAGCGGCTGGATATGTA
TATGAACATGTACCACTGGGCCAAGGCCTTTATCGTGGACGAGGTGCTGGACAGAG
ATGAGATGTTCTACAGCGATATCGACGATATCTATAATATCCTGGAGAACATCGTGC
CACTGTATAATCGGGTGAGAAACTACGTGACCCAGAAGCCCTACAACTCTAAGAAG
ATCAAGCTGAATTTCCAGAGCCCTACACTGGCCAATGGCTGGTCCCAGTCTAAGGAG
TTCGACAACAATGCCATCATCCTGATCAGAGATAACAAGTACTATCTGGCCATCTTC
AATGCCAAGAACAAGCCAGACAAGAAGATCATCCAGGGCAACTCCGATAAGAAGA
ACGACAACGATTACAAGAAGATGGTGTATAACCTGCTGCCAGGCGCCAACAAGATG
CTGCCCAAGGTGTTTCTGTCTAAGAAGGGCATCGAGACATTCAAGCCCTCCGACTAT
ATCATCTCTGGCTACAACGCCCACAAGCACATCAAGACAAGCGAGAATTTTGATATC
TCCTTCTGTCGGGACCTGATCGATTACTTCAAGAACAGCATCGAGAAGCACGCCGAG

TGGAGAAAGTATGAGTTCAAGTTTTCCGCCACCGACAGCTACTCCGATATCTCTGAG
TTCTATCGGGAGGTGGAGATGCAGGGCTACAGAATCGACTGGACATATATCAGCGA
GGCCGACATCAACAAGCTGGATGAGGAGGGCAAGATCTATCTGTTTCAGATCTACA
ATAAGGATTTCGCCGAGAACAGCACCGGCAAGGAGAATCTGCACACAATGTACTTT
AAGAACATCTTCTCCGAGGAGAATCTGAAGGACATCATCATCAAGCTGAACGGCCA
GGCCGAGCTGTTTTATCGGAGAGCCTCTGTGAAGAATCCCGTGAAGCACAAGAAGG
ATAGCGTGCTGGTGAACAAGACCTACAAGAATCAGCTGGACAACGGCGACGTGGTG
AGAATCCCCATCCCTGACGATATCTATAACGAGATCTACAAGATGTATAATGGCTAC
ATCAAGGAGTCCGACCTGTCTGAGGCCGCCAAGGAGTACCTGGATAAGGTGGAGGT
GAGGACCGCCCAGAAGGACATCGTGAAGGATTACCGCTATACAGTGGACAAGTACT
TCATCCACACACCTATCACCATCAACTATAAGGTGACCGCCCGCAACAATGTGAATG
ATATGGTGGTGAAGTACATCGCCCAGAACGACGATATCCACGTGATCGGCATCGAC
CGGGGCGAGAGAAACCTGATCTACATCTCCGTGATCGATTCTCACGGCAACATCGTG
AAGCAGAAATCCTACAACATCCTGAACAACTACGACTACAAGAAGAAGCTGGTGGA
GAAGGAGAAAACCCGGGAGTACGCCAGAAAGAACTGGAAGAGCATCGGCAATATC
AAGGAGCTGAAGGAGGGCTATATCTCCGGCGTGGTGCACGAGATCGCCATGCTGAT
CGTGGAGTACAACGCCATCATCGCCATGGAGGACCTGAATTATGGCTTTAAGAGGG
GCCGCTTCAAGGTGGAGCGGCAGGTGTACCAGAAGTTTGAGAGCATGCTGATCAAT
AAGCTGAACTATTTCGCCAGCAAGGAGAAGTCCGTGGACGAGCCAGGAGGCCTGCT
GAAGGGCTATCAGCTGACCTACGTGCCCGATAATATCAAGAACCTGGGCAAGCAGT
GCGGCGTGATCTTTTACGTGCCTGCCGCCTTCACCAGCAAGATCGACCCATCCACAG
GCTTTATCTCTGCCTTCAACTTTAAGTCTATCAGCACAAATGCCTCTCGGAAGCAGTT
CTTTATGCAGTTTGACGAGATCAGATACTGTGCCGAGAAGGATATGTTCAGCTTTGG
CTTCGACTACAACAACTTCGATACCTACAACATCACAATGGGCAAGACACAGTGGA
CCGTGTATACAAACGGCGAGAGACTGCAGTCTGAGTTCAACAATGCCAGGCGCACC
GGCAAGACAAAGAGCATCAATCTGACAGAGACAATCAAGCTGCTGCTGGAGGACAA
TGAGATCAACTACGCCGACGGCCACGATATCAGGATCGATATGGAGAAGATGGACG
AGGATAAGAAGAGCGAGTTCTTTGCCCAGCTGCTGAGCCTGTATAAGCTGACCGTGC
AGATGCGCAATTCCTATACAGAGGCCGAGGAGCAGGAGAACGGCATCTCTTACGAC
AAGATCATCAGCCCTGTGATCAATGATGAGGGCGAGTTCTTTGACTCCGATAACTAT
AAGGAGTCTGACGATAAGGAGTGCAAGATGCCAAAGGACGCCGATGCCAACGGCGC
CTACTGTATCGCCCTGAAGGGCCTGTATGAGGTGCTGAAGATCAAGAGCGAGTGGA
CCGAGGACGGCTTTGATAGGAATTGCCTGAAGCTGCCACACGCAGAGTGGCTGGAC
TTCATCCAGAACAAGCGGTACGAG*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAG*
*GCAAAAAAGAAAAAG*GGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCTA

CGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC (SEQ ID NO: 223)

12 - Moraxella bovoculi 237 (MbCpf1)

ATGCTGTTCCAGGACTTTACCCACCTGTATCCACTGTCCAAGACAGTG
AGATTTGAGCTGAAGCCCATCGATAGGACCCTGGAGCACATCCACGCCAAGAACTT
CCTGTCTCAGGACGAGACAATGGCCGATATGCACCAGAAGGTGAAAGTGATCCTGG
ACGATTACCACCGCGACTTCATCGCCGATATGATGGGCGAGGTGAAGCTGACCAAG
CTGGCCGAGTTCTATGACGTGTACCTGAAGTTTCGGAAGAACCCAAAGGACGATGA
GCTGCAGAAGCAGCTGAAGGATCTGCAGGCCGTGCTGAGAAAGGAGATCGTGAAGC
CCATCGGCAATGGCGGCAAGTATAAGGCCGGCTACGACAGGCTGTTCGGCGCCAAG
CTGTTTAAGGACGGCAAGGAGCTGGGCGATCTGGCCAAGTTCGTGATCGCACAGGA
GGGAGAGAGCTCCCCAAAGCTGGCCCACCTGGCCCACTTCGAGAAGTTTTCCACCTA
TTTCACAGGCTTTCACGATAACCGGAAGAATATGTATTCTGACGAGGATAAGCACAC
CGCCATCGCCTACCGCCTGATCCACGAGAACCTGCCCCGGTTTATCGACAATCTGCA
GATCCTGACCACAATCAAGCAGAAGCACTCTGCCCTGTACGATCAGATCATCAACGA
GCTGACCGCCAGCGGCCTGGACGTGTCTCTGGCCAGCCACCTGGATGGCTATCACAA
GCTGCTGACACAGGAGGGCATCACCGCCTACAATACACTGCTGGGAGGAATCTCCG
GAGAGGCAGGCTCTCCTAAGATCCAGGGCATCAACGAGCTGATCAATTCTCACCAC
AACCAGCACTGCCACAAGAGCGAGAGAATCGCCAAGCTGAGGCCACTGCACAAGCA
GATCCTGTCCGACGGCATGAGCGTGTCCTTCCTGCCCTCTAAGTTTGCCGACGATAG
CGAGATGTGCCAGGCCGTGAACGAGTTCTATCGCCACTACGCCGACGTGTTCGCCAA
GGTGCAGAGCCTGTTCGACGGCTTTGACGATCACCAGAAGGATGGCATCTACGTGG
AGCACAAGAACCTGAATGAGCTGTCCAAGCAGGCCTTCGGCGACTTTGCACTGCTGG
GACGCGTGCTGGACGGATACTATGTGGATGTGGTGAATCCAGAGTTCAACGAGCGG
TTTGCCAAGGCCAAGACCGACAATGCCAAGGCCAAGCTGACAAAGGAGAAGGATAA
GTTCATCAAGGGCGTGCACTCCCTGGCCTCTCTGGAGCAGGCCATCGAGCACTATAC
CGCAAGGCACGACGATGAGAGCGTGCAGGCAGGCAAGCTGGGACAGTACTTCAAGC
ACGGCCTGGCCGGAGTGGACAACCCCATCCAGAAGATCCACAACAATCACAGCACC
ATCAAGGGCTTTCTGGAGAGGGAGCGCCCTGCAGGAGAGAGAGCCCTGCCAAAGAT
CAAGTCCGGCAAGAATCCTGAGATGACACAGCTGAGGCAGCTGAAGGAGCTGCTGG
ATAACGCCCTGAATGTGGCCCACTTCGCCAAGCTGCTGACCACAAAGACCACACTGG
ACAATCAGGATGGCAACTTCTATGGCGAGTTTGGCGTGCTGTACGACGAGCTGGCCA
AGATCCCCACCCTGTATAACAAGGTGAGAGATTACCTGAGCCAGAAGCCTTTCTCCA
CCGAGAAGTACAAGCTGAACTTTGGCAATCCAACACTGCTGAATGGCTGGGACCTG
AACAAGGAGAAGGATAATTTCGGCGTGATCCTGCAGAAGGACGGCTGCTACTATCT
GGCCCTGCTGGACAAGGCCCACAAGAAGGTGTTTGATAACGCCCCTAATACAGGCA

AGAGCATCTATCAGAAGATGATCTATAAGTACCTGGAGGTGAGGAAGCAGTTCCCC
AAGGTGTTCTTTTCCAAGGAGGCCATCGCCATCAACTACCACCCTTCTAAGGAGCTG
GTGGAGATCAAGGACAAGGGCCGGCAGAGATCCGACGATGAGCGCCTGAAGCTGTA
TCGGTTTATCCTGGAGTGTCTGAAGATCCACCCTAAGTACGATAAGAAGTTCGAGGG
CGCCATCGGCGACATCCAGCTGTTTAAGAAGGATAAGAAGGGCAGAGAGGTGCCAA
TCAGCGAGAAGGACCTGTTCGATAAGATCAACGGCATCTTTTCTAGCAAGCCTAAGC
TGGAGATGGAGGACTTCTTTATCGGCGAGTTCAAGAGGTATAACCCAAGCCAGGAC
CTGGTGGATCAGTATAATATCTACAAGAAGATCGACTCCAACGATAATCGCAAGAA
GGAGAATTTCTACAACAATCACCCCAAGTTTAAGAAGGATCTGGTGCGGTACTATTA
CGAGTCTATGTGCAAGCACGAGGAGTGGGAGGAGAGCTTCGAGTTTTCCAAGAAGC
TGCAGGACATCGGCTGTTACGTGGATGTGAACGAGCTGTTTACCGAGATCGAGACAC
GGAGACTGAATTATAAGATCTCCTTCTGCAACATCAATGCCGACTACATCGATGAGC
TGGTGGAGCAGGGCCAGCTGTATCTGTTCCAGATCTACAACAAGGACTTTTCCCCAA
AGGCCCACGGCAAGCCCAATCTGCACACCCTGTACTTCAAGGCCCTGTTTTCTGAGG
ACAACCTGGCCGATCCTATCTATAAGCTGAATGGCGAGGCCCAGATCTTCTACAGAA
AGGCCTCCCTGGACATGAACGAGACAACAATCCACAGGGCCGGCGAGGTGCTGGAG
AACAAGAATCCCGATAATCCTAAGAAGAGACAGTTCGTGTACGACATCATCAAGGA
TAAGAGGTACACACAGGACAAGTTCATGCTGCACGTGCCAATCACCATGAACTTTGG
CGTGCAGGGCATGACAATCAAGGAGTTCAATAAGAAGGTGAACCAGTCTATCCAGC
AGTATGACGAGGTGAACGTGATCGGCATCGATCGGGGCGAGAGACACCTGCTGTAC
CTGACCGTGATCAATAGCAAGGGCGAGATCCTGGAGCAGTGTTCCCTGAACGACAT
CACCACAGCCTCTGCCAATGGCACACAGATGACCACACCTTACCACAAGATCCTGGA
TAAGAGGGAGATCGAGCGCCTGAACGCCCGGGTGGGATGGGGCGAGATCGAGACA
ATCAAGGAGCTGAAGTCTGGCTATCTGAGCCACGTGGTGCACCAGATCAGCCAGCT
GATGCTGAAGTACAACGCCATCGTGGTGCTGGAGGACCTGAATTTCGGCTTTAAGAG
GGGCCGCTTTAAGGTGGAGAAGCAGATCTATCAGAACTTCGAGAATGCCCTGATCA
AGAAGCTGAACCACCTGGTGCTGAAGGACAAGGCCGACGATGAGATCGGCTCTTAC
AAGAATGCCCTGCAGCTGACCAACAATTTCACAGATCTGAAGAGCATCGGCAAGCA
GACCGGCTTCCTGTTTTATGTGCCCGCCTGGAACACCTCTAAGATCGACCCTGAGAC
AGGCTTTGTGGATCTGCTGAAGCCAAGATACGAGAACATCGCCCAGAGCCAGGCCT
TCTTTGGCAAGTTCGACAAGATCTGCTATAATGCCGACAAGGATTACTTCGAGTTTC
ACATCGACTACGCCAAGTTTACCGATAAGGCCAAGAATAGCCGCCAGATCTGGACA
ATCTGTTCCCACGGCGACAAGCGGTACGTGTACGATAAGACAGCCAACCAGAATAA
GGGCGCCGCCAAGGGCATCAACGTGAATGATGAGCTGAAGTCCCTGTTCGCCCGCC
ACCACATCAACGAGAAGCAGCCCAACCTGGTCATGGACATCTGCCAGAACAATGAT
AAGGAGTTTCACAAGTCTCTGATGTACCTGCTGAAAACCCTGCTGGCCCTGCGGTAC

AGCAACGCCTCCTCTGACGAGGATTTCATCCTGTCCCCCGTGGCAAACGACGAGGGC
GTGTTCTTTAATAGCGCCCTGGCCGACGATACACAGCCTCAGAATGCCGATGCCAAC
GGCGCCTACCACATCGCCCTGAAGGGCCTGTGGCTGCTGAATGAGCTGAAGAACTCC
GACGATCTGAACAAGGTGAAGCTGGCCATCGACAATCAGACCTGGCTGAATTTCGC
CCAGAACAGG*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG*
GGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCTGAT**
TATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC (SEQ ID NO: 224)

13 - Leptospira inadai (LiCpf1)

ATGGAGGACTATTCCGGCTTTGTGAACATCTACTCTATCCAGAAAAC
CCTGAGGTTCGAGCTGAAGCCAGTGGGCAAGACACTGGAGCACATCGAGAAGAAGG
GCTTCCTGAAGAAGGACAAGATCCGGGCCGAGGATTACAAGGCCGTGAAGAAGATC
ATCGATAAGTACCACAGAGCCTATATCGAGGAGGTGTTTGATTCCGTGCTGCACCAG
AAGAAGAAGAAGGACAAGACCCGCTTTTCTACACAGTTCATCAAGGAGATCAAGGA
GTTCAGCGAGCTGTACTATAAGACCGAGAAGAACATCCCCGACAAGGAGAGGCTGG
AGGCCCTGAGCGAGAAGCTGCGCAAGATGCTGGTGGGCGCCTTTAAGGGCGAGTTC
TCCGAGGAGGTGGCCGAGAAGTATAAGAACCTGTTTTCTAAGGAGCTGATCAGGAA
TGAGATCGAGAAGTTCTGCGAGACAGACGAGGAGCGCAAGCAGGTGTCTAACTTCA
AGAGCTTCACCACATACTTTACCGGCTTCCACTCCAACAGGCAGAATATCTATTCCG
ACGAGAAGAAGTCTACAGCCATCGGCTACCGCATCATCCACCAGAACCTGCCTAAG
TTCCTGGATAATCTGAAGATCATCGAGTCCATCCAGCGGCGGTTCAAGGACTTCCCA
TGGTCTGATCTGAAGAAGAACCTGAAGAAGATCGATAAGAATATCAAGCTGACCGA
GTACTTCAGCATCGACGGCTTCGTGAACGTGCTGAATCAGAAGGGCATCGATGCCTA
CAACACAATCCTGGGCGGCAAGTCCGAGGAGTCTGGCGAGAAGATCCAGGGCCTGA
ACGAGTACATCAATCTGTATCGGCAGAAGAACAATATCGACAGAAAGAACCTGCCC
AATGTGAAGATCCTGTTTAAGCAGATCCTGGGCGATAGGGAGACAAAGAGCTTTAT
CCCTGAGGCCTTCCCAGACGATCAGTCCGTGCTGAACTCTATCACAGAGTTCGCCAA
GTACCTGAAGCTGGATAAGAAGAAGAAGAGCATCATCGCCGAGCTGAAGAAGTTTC
TGAGCTCCTTCAATCGCTACGAGCTGGACGGCATCTATCTGGCCAACGATAATAGCC
TGGCCTCTATCAGCACCTTCCTGTTTGACGATTGGTCCTTTATCAAGAAGTCCGTGTC
TTTCAAGTATGACGAGTCCGTGGGCGACCCCAAGAAGAAGATCAAGTCTCCCCTGA
AGTACGAGAAGGAGAAGGAGAAGTGGCTGAAGCAGAAGTACTATACAATCTCTTTC
CTGAACGATGCCATCGAGAGCTATTCCAAGTCTCAGGACGAGAAGAGGGTGAAGAT
CCGCCTGGAGGCCTACTTTGCCGAGTTCAAGAGCAAGGACGATGCCAAGAAGCAGT
TCGACCTGCTGGAGAGGATCGAGGAGGCCTATGCCATCGTGGAGCCTCTGCTGGGA
GCAGAGTACCCAAGGGACCGCAACCTGAAGGCCGATAAGAAGGAAGTGGGCAAGA
TCAAGGACTTCCTGGATAGCATCAAGTCCCTGCAGTTCTTTCTGAAGCCTCTGCTGTC

CGCCGAGATCTTTGACGAGAAGGATCTGGGCTTCTACAATCAGCTGGAGGGCTACTA
TGAGGAGATCGATTCTATCGGCCACCTGTATAACAAGGTGCGGAATTATCTGACCGG
CAAGATCTACAGCAAGGAGAAGTTTAAGCTGAACTTCGAGAACAGCACCCTGCTGA
AGGGCTGGGACGAGAACCGGGAGGTGGCCAATCTGTGCGTGATCTTCAGAGAGGAC
CAGAAGTACTATCTGGGCGTGATGGATAAGGAGAACAATACCATCCTGTCCGACAT
CCCCAAGGTGAAGCCTAACGAGCTGTTTTACGAGAAGATGGTGTATAAGCTGATCCC
CACACCTCACATGCAGCTGCCCCGGATCATCTTCTCTAGCGACAACCTGTCTATCTAT
AATCCTAGCAAGTCCATCCTGAAGATCAGAGAGGCCAAGAGCTTTAAGGAGGGCAA
GAACTTCAAGCTGAAGGACTGTCACAAGTTTATCGATTTCTACAAGGAGTCTATCAG
CAAGAATGAGGACTGGAGCAGATTCGACTTCAAGTTCAGCAAGACCAGCAGCTACG
AGAACATCAGCGAGTTTTACCGGGAGGTGGAGAGACAGGGCTATAACCTGGACTTC
AAGAAGGTGTCTAAGTTCTACATCGACAGCCTGGTGGAGGATGGCAAGCTGTACCT
GTTCCAGATCTATAACAAGGACTTTTCTATCTTCAGCAAGGGCAAGCCCAATCTGCA
CACCATCTATTTTCGGTCCCTGTTCTCTAAGGAGAACCTGAAGGACGTGTGCCTGAA
GCTGAATGGCGAGGCCGAGATGTTCTTTCGGAAGAAGTCCATCAACTACGATGAGA
AGAAGAAGCGGGAGGGCCACCACCCCGAGCTGTTTGAGAAGCTGAAGTATCCTATC
CTGAAGGACAAGAGATACAGCGAGGATAAGTTTCAGTTCCACCTGCCCATCAGCCT
GAACTTCAAGTCCAAGGAGCGGCTGAACTTTAATCTGAAAGTGAATGAGTTCCTGAA
GAGAAACAAGGACATCAATATCATCGGCATCGATCGGGGCGAGAGAAACCTGCTGT
ACCTGGTCATGATCAATCAGAAGGGCGAGATCCTGAAGCAGACCCTGCTGGACAGC
ATGCAGTCCGGCAAGGGCCGGCCTGAGATCAACTACAAGGAGAAGCTGCAGGAGAA
GGAGATCGAGAGGGATAAGGCCCGCAAGAGCTGGGGCACAGTGGAGAATATCAAG
GAGCTGAAGGAGGGCTATCTGTCTATCGTGATCCACCAGATCAGCAAGCTGATGGTG
GAGAACAATGCCATCGTGGTGCTGGAGGACCTGAACATCGGCTTTAAGCGGGGCAG
ACAGAAGGTGGAGCGGCAGGTGTACCAGAAGTTCGAGAAGATGCTGATCGATAAGC
TGAACTTTCTGGTGTTCAAGGAGAATAAGCCAACCGAGCCAGGAGGCGTGCTGAAG
GCCTATCAGCTGACAGACGAGTTTCAGTCTTTCGAGAAGCTGAGCAAGCAGACCGG
CTTTCTGTTCTACGTGCCAAGCTGGAACACCTCCAAGATCGACCCCAGAACAGGCTT
TATCGATTTCCTGCACCCTGCCTACGAGAATATCGAGAAGGCCAAGCAGTGGATCAA
CAAGTTTGATTCCATCAGGTTCAATTCTAAGATGGACTGGTTTGAGTTCACCGCCGA
TACACGCAAGTTTTCCGAGAACCTGATGCTGGGCAAGAATCGGGTGTGGGTCATCTG
CACCACAAATGTGGAGCGGTACTTCACCAGCAAGACCGCCAACAGCTCCATCCAGT
ACAATAGCATCCAGATCACCGAGAAGCTGAAGGAGCTGTTTGTGGACATCCCTTTCA
GCAACGGCCAGGATCTGAAGCCAGAGATCCTGAGGAAGAATGACGCCGTGTTCTTT
AAGAGCCTGCTGTTTTACATCAAGACCACACTGTCCCTGCGCCAGAACAATGGCAAG
AAGGGCGAGGAGGAGAAGGACTTCATCCTGAGCCCAGTGGTGGATTCCAAGGGCCG

GTTCTTTAACTCTCTGGAGGCCAGCGACGATGAGCCCAAGGACGCCGATGCCAATGG
CGCCTACCACATCGCCCTGAAGGGCCTGATGAACCTGCTGGTGCTGAATGAGACAA
AGGAGGAGAACCTGAGCAGACCAAAGTGGAAGATCAAGAATAAGGACTGGCTGGA
GTTCGTGTGGGAGAGGAACCGC*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGC*
*AAAAAAGAAAAAG*GGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACG**
ACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC
(SEQ ID NO: 225)

14 - La bacteria Lachnospiraceae ND2006 (LbCpf1)

ATGAGCAAGCTGGAGAAGTTTACAAACTGCTACTCCCTGTCTAAGAC
CCTGAGGTTCAAGGCCATCCCTGTGGGCAAGACCCAGGAGAACATCGACAATAAGC
GGCTGCTGGTGGAGGACGAGAAGAGAGCCGAGGATTATAAGGGCGTGAAGAAGCT
GCTGGATCGCTACTATCTGTCTTTTATCAACGACGTGCTGCACAGCATCAAGCTGAA
GAATCTGAACAATTACATCAGCCTGTTCCGGAAGAAAACCAGAACCGAGAAGGAGA
ATAAGGAGCTGGAGAACCTGGAGATCAATCTGCGGAAGGAGATCGCCAAGGCCTTC
AAGGGCAACGAGGGCTACAAGTCCCTGTTTAAGAAGGATATCATCGAGACAATCCT
GCCAGAGTTCCTGGACGATAAGGACGAGATCGCCCTGGTGAACAGCTTCAATGGCTT
TACCACAGCCTTCACCGGCTTCTTTGATAACAGAGAGAATATGTTTTCCGAGGAGGC
CAAGAGCACATCCATCGCCTTCAGGTGTATCAACGAGAATCTGACCCGCTACATCTC
TAATATGGACATCTTCGAGAAGGTGGACGCCATCTTTGATAAGCACGAGGTGCAGG
AGATCAAGGAGAAGATCCTGAACAGCGACTATGATGTGGAGGATTTCTTTGAGGGC
GAGTTCTTTAACTTTGTGCTGACACAGGAGGGCATCGACGTGTATAACGCCATCATC
GGCGGCTTCGTGACCGAGAGCGGCGAGAAGATCAAGGGCCTGAACGAGTACATCAA
CCTGTATAATCAGAAAACCAAGCAGAAGCTGCCTAAGTTTAAGCCACTGTATAAGC
AGGTGCTGAGCGATCGGGAGTCTCTGAGCTTCTACGGCGAGGGCTATACATCCGATG
AGGAGGTGCTGGAGGTGTTTAGAAACACCCTGAACAAGAACAGCGAGATCTTCAGC
TCCATCAAGAAGCTGGAGAAGCTGTTCAAGAATTTTGACGAGTACTCTAGCGCCGGC
ATCTTTGTGAAGAACGGCCCCGCCATCAGCACAATCTCCAAGGATATCTTCGGCGAG
TGGAACGTGATCCGGGACAAGTGGAATGCCGAGTATGACGATATCCACCTGAAGAA
GAAGGCCGTGGTGACCGAGAAGTACGAGGACGATCGGAGAAAGTCCTTCAAGAAG
ATCGGCTCCTTTTCTCTGGAGCAGCTGCAGGAGTACGCCGACGCCGATCTGTCTGTG
GTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGATGAGATCTACAAGGTGTA
TGGCTCCTCTGAGAAGCTGTTCGACGCCGATTTTGTGCTGGAGAAGAGCCTGAAGAA
GAACGACGCCGTGGTGGCCATCATGAAGGACCTGCTGGATTCTGTGAAGAGCTTCG
AGAATTACATCAAGGCCTTCTTTGGCGAGGGCAAGGAGACAAACAGGGACGAGTCC
TTCTATGGCGATTTTGTGCTGGCCTACGACATCCTGCTGAAGGTGGACCACATCTAC
GATGCCATCCGCAATTATGTGACCCAGAAGCCCTACTCTAAGGATAAGTTCAAGCTG

TATTTTCAGAACCCTCAGTTCATGGGCGGCTGGGACAAGGATAAGGAGACAGACTA
TCGGGCCACCATCCTGAGATACGGCTCCAAGTACTATCTGGCCATCATGGATAAGAA
GTACGCCAAGTGCCTGCAGAAGATCGACAAGGACGATGTGAACGGCAATTACGAGA
AGATCAACTATAAGCTGCTGCCCGGCCCTAATAAGATGCTGCCAAAGGTGTTCTTTT
CTAAGAAGTGGATGGCCTACTATAACCCCAGCGAGGACATCCAGAAGATCTACAAG
AATGGCACATTCAAGAAGGGCGATATGTTTAACCTGAATGACTGTCACAAGCTGATC
GACTTCTTTAAGGATAGCATCTCCCGGTATCCAAAGTGGTCCAATGCCTACGATTTC
AACTTTTCTGAGACAGAGAAGTATAAGGACATCGCCGGCTTTTACAGAGAGGTGGA
GGAGCAGGGCTATAAGGTGAGCTTCGAGTCTGCCAGCAAGAAGGAGGTGGATAAGC
TGGTGGAGGAGGGCAAGCTGTATATGTTCCAGATCTATAACAAGGACTTTTCCGATA
AGTCTCACGGCACACCCAATCTGCACACCATGTACTTCAAGCTGCTGTTTGACGAGA
ACAATCACGGACAGATCAGGCTGAGCGGAGGAGCAGAGCTGTTCATGAGGCGCGCC
TCCCTGAAGAAGGAGGAGCTGGTGGTGCACCCAGCCAACTCCCCTATCGCCAACAA
GAATCCAGATAATCCCAAGAAAACCACAACCCTGTCCTACGACGTGTATAAGGATA
AGAGGTTTTCTGAGGACCAGTACGAGCTGCACATCCCAATCGCCATCAATAAGTGCC
CCAAGAACATCTTCAAGATCAATACAGAGGTGCGCGTGCTGCTGAAGCACGACGAT
AACCCCTATGTGATCGGCATCGATAGGGGCGAGCGCAATCTGCTGTATATCGTGGTG
GTGGACGGCAAGGGCAACATCGTGGAGCAGTATTCCCTGAACGAGATCATCAACAA
CTTCAACGGCATCAGGATCAAGACAGATTACCACTCTCTGCTGGACAAGAAGGAGA
AGGAGAGGTTCGAGGCCCGCCAGAACTGGACCTCCATCGAGAATATCAAGGAGCTG
AAGGCCGGCTATATCTCTCAGGTGGTGCACAAGATCTGCGAGCTGGTGGAGAAGTA
CGATGCCGTGATCGCCCTGGAGGACCTGAACTCTGGCTTTAAGAATAGCCGCGTGAA
GGTGGAGAAGCAGGTGTATCAGAAGTTCGAGAAGATGCTGATCGATAAGCTGAACT
ACATGGTGGACAAGAAGTCTAATCCTTGTGCAACAGGCGGCGCCCTGAAGGGCTAT
CAGATCACCAATAAGTTCGAGAGCTTTAAGTCCATGTCTACCCAGAACGGCTTCATC
TTTTACATCCCTGCCTGGCTGACATCCAAGATCGATCCATCTACCGGCTTTGTGAACC
TGCTGAAAACCAAGTATACCAGCATCGCCGATTCCAAGAAGTTCATCAGCTCCTTTG
ACAGGATCATGTACGTGCCCGAGGAGGATCTGTTCGAGTTTGCCCTGGACTATAAGA
ACTTCTCTCGCACAGACGCCGATTACATCAAGAAGTGGAAGCTGTACTCCTACGGCA
ACCGGATCAGAATCTTCCGGAATCCTAAGAAGAACAACGTGTTCGACTGGGAGGAG
GTGTGCCTGACCAGCGCCTATAAGGAGCTGTTCAACAAGTACGGCATCAATTATCAG
CAGGGCGATATCAGAGCCCTGCTGTGCGAGCAGTCCGACAAGGCCTTCTACTCTAGC
TTTATGGCCCTGATGAGCCTGATGCTGCAGATGCGGAACAGCATCACAGGCCGCACC
GACGTGGATTTTCTGATCAGCCCTGTGAAGAACTCCGACGGCATCTTCTACGATAGC
CGGAACTATGAGGCCCAGGAGAATGCCATCCTGCCAAAGAACGCCGACGCCAATGG
CGCCTATAACATCGCCAGAAAGGTGCTGTGGGCCATCGGCCAGTTCAAGAAGGCCG

AGGACGAGAAGCTGGATAAGGTGAAGATCGCCATCTCTAACAAGGAGTGGCTGGAG
TACGCCCAGACCAGCGTGAAGCAC*AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAG*
*GCAAAAAAGAAAAAG*GGATCCTACCCATACGATGTTCCAGATTACGCTTATCCCTA
CGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATT
C (SEQ ID NO: 226)

15 - Porphyromonas crevioricanis (PcCpf1)

ATGGACAGCCTGAAGGATTTCACCAACCTGTACCCCGTGTCCAAGAC
ACTGCGGTTTGAGCTGAAGCCTGTGGGCAAGACCCTGGAGAATATCGAGAAGGCCG
GCATCCTGAAGGAGGATGAGCACAGAGCCGAGAGCTACCGGAGAGTGAAGAAGAT
CATCGATACATATCACAAGGTGTTCATCGACAGCTCCCTGGAGAACATGGCCAAGAT
GGGCATCGAGAATGAGATCAAGGCCATGCTGCAGTCCTTTTGCGAGCTGTATAAGA
AGGACCACAGGACCGAGGGAGAGGACAAGGCCCTGGATAAGATCAGGGCCGTGCT
GAGGGGCCTGATCGTGGGAGCCTTCACCGGCGTGTGCGGCCGGCGGGAGAACACAG
TGCAGAATGAGAAGTATGAGAGCCTGTTTAAGGAGAAGCTGATCAAGGAGATCCTG
CCAGATTTCGTGCTGTCTACAGAGGCCGAGTCCCTGCCCTTTTCTGTGGAGGAGGCC
ACCAGAAGCCTGAAGGAGTTCGACTCCTTTACATCTTACTTCGCCGGCTTTTATGAG
AACCGGAAGAATATCTACTCTACCAAGCCCCAGAGCACAGCCATCGCCTATAGACT
GATCCACGAGAACCTGCCTAAGTTCATCGATAATATCCTGGTGTTTCAGAAGATCAA
GGAGCCAATCGCCAAGGAGCTGGAGCACATCAGGGCAGACTTCAGCGCCGGCGGCT
ACATCAAGAAGGATGAGCGCCTGGAGGACATCTTTTCCCTGAACTACTATATCCACG
TGCTGTCTCAGGCCGGCATCGAGAAGTACAATGCCCTGATCGGCAAGATCGTGACCG
AGGGCGATGGCGAGATGAAGGGCCTGAACGAGCACATCAACCTGTATAATCAGCAG
AGGGGCCGCGAGGACCGGCTGCCACTGTTCAGACCCCTGTATAAGCAGATCCTGTCT
GATAGGGAGCAGCTGTCCTATCTGCCAGAGTCTTTCGAGAAGGACGAGGAGCTGCT
GAGGGCCCTGAAGGAGTTTTACGATCACATCGCAGAGGACATCCTGGGAAGGACCC
AGCAGCTGATGACAAGCATCTCCGAGTACGATCTGTCCCGGATCTATGTGAGAAACG
ATAGCCAGCTGACCGACATCTCCAAGAAGATGCTGGGCGATTGGAATGCCATCTAC
ATGGCCCCGGGAGAGAGCCTATGACCACGAGCAGGCCCCCAAGCGCATCACAGCCAA
GTACGAGAGGGACCGCATCAAGGCCCTGAAGGGCGAGGAGTCTATCAGCCTGGCCA
ACCTGAACAGCTGCATCGCCTTCCTGGACAACGTGAGGGATTGTCGCGTGGACACCT
ATCTGTCTACACTGGGACAGAAGGAGGGACCTCACGGCCTGAGCAACCTGGTGGAG
AACGTGTTCGCCTCCTACCACGAGGCCGAGCAGCTGCTGTCTTTTCCCTATCCTGAG
GAGAACAATCTGATCCAGGACAAGGATAACGTGGTGCTGATCAAGAACCTGCTGGA
TAATATCAGCGACCTGCAGAGGTTCCTGAAGCCACTGTGGGGCATGGGCGATGAGC
CCGACAAGGATGAGAGGTTTTACGGCGAGTACAATTATATCAGGGGCGCCCTGGAC
CAGGTCATCCCTCTGTATAACAAGGTGCGGAATTATCTGACCCGCAAGCCATACTCC

ACACGCAAGGTGAAGCTGAACTTCGGCAATAGCCAGCTGCTGTCCGGCTGGGATAG
GAACAAGGAGAAGGACAATTCTTGCGTGATCCTGCGCAAGGGCCAGAACTTCTACC
TGGCCATCATGAACAATCGGCACAAGCGGAGCTTCGAGAATAAGATGCTGCCCGAG
TATAAGGAGGGCGAGCCTTACTTCGAGAAGATGGATTATAAGTTTCTGCCAGACCCC
AACAAGATGCTGCCCAAGGTGTTCCTGTCTAAGAAGGGCATCGAGATCTACAAGCCT
AGCCCAAAGCTGCTGGAGCAGTATGGCCACGGCACCCACAAGAAGGGCGATACCTT
CAGCATGGACGATCTGCACGAGCTGATCGACTTCTTTAAGCACTCCATCGAGGCCCA
CGAGGATTGGAAGCAGTTCGGCTTTAAGTTCAGCGACACCGCCACATACGAGAACG
TGAGCAGCTTCTACCGGGAGGTGGAGGACCAGGGCTACAAGCTGTCTTTTAGAAAG
GTGTCCGAGTCTTACGTGTATAGCCTGATCGATCAGGGCAAGCTGTACCTGTTCCAG
ATCTATAACAAGGACTTTAGCCCTTGTTCCAAGGGCACCCCAAATCTGCACACACTG
TACTGGCGGATGCTGTTCGATGAGAGAAACCTGGCCGACGTGATCTATAAGCTGGAT
GGCAAGGCCGAGATCTTCTTTCGGGAGAAGTCCCTGAAGAATGACCACCCAACCCA
CCCTGCAGGCAAGCCCATCAAGAAGAAGAGCCGGCAGAAGAAGGGCGAGGAGAGC
CTGTTCGAGTACGATCTGGTGAAGGACCGGAGATATACCATGGATAAGTTTCAGTTC
CACGTGCCAATCACAATGAACTTTAAGTGCTCTGCCGGCAGCAAGGTGAACGACAT
GGTGAATGCCCACATCAGGGAGGCCAAGGACATGCACGTGATCGGCATCGATAGGG
GCGAGCGCAATCTGCTGTATATCTGCGTGATCGACAGCCGCGGCACCATCCTGGATC
AGATCTCCCTGAACACAATCAATGACATCGATTATCACGATCTGCTGGAGTCCAGGG
ACAAGGATCGCCAGCAGGAGCACAGGAACTGGCAGACCATCGAGGGCATCAAGGA
GCTGAAGCAGGGCTACCTGTCTCAGGCCGTGCACCGCATCGCCGAGCTGATGGTGGC
CTATAAGGCCGTGGTGGCCCTGGAGGACCTGAACATGGGCTTCAAGCGGGGCAGAC
AGAAGGTGGAGAGCAGCGTGTACCAGCAGTTTGAGAAGCAGCTGATCGACAAGCTG
AATTATCTGGTGGATAAGAAGAAGCGGCCCGAGGACATCGGAGGCCTGCTGAGAGC
CTACCAGTTCACCGCCCCTTTCAAGAGCTTTAAGGAGATGGGCAAGCAGAACGGCTT
TCTGTTCTATATCCCTGCCTGGAACACATCCAATATCGACCCAACCACAGGCTTCGT
GAACCTGTTTCACGTGCAGTACGAGAATGTGGATAAGGCCAAGAGCTTCTTTCAGAA
GTTCGACAGCATCTCCTACAACCCTAAGAAGGATTGGTTTGAGTTCGCCTTTGACTA
TAAGAACTTCACCAAGAAGGCCGAGGGCTCTAGGAGCATGTGGATTCTGTGCACCC
ACGGCTCCCGGATCAAGAACTTCAGAAATTCTCAGAAGAATGGCCAGTGGGATAGC
GAGGAGTTTGCCCTGACCGAGGCCTTCAAGTCCCTGTTTGTGCGGTACGAGATCGAT
TATACCGCCGACCTGAAAACCGCCATCGTGGACGAGAAGCAGAAGGATTTCTTTGTG
GACCTGCTGAAGCTGTTCAAGCTGACCGTGCAGATGAGAAACTCCTGGAAGGAGAA
GGACCTGGATTACCTGATCTCTCCAGTGGCCGGCGCCGATGGCAGGTTCTTTGACAC
ACGCGAGGGCAATAAGAGCCTGCCCAAGGACGCAGATGCAAACGGAGCCTATAATA
TCGCCCTGAAGGGCCTGTGGGCACTGAGGCAGATCAGACAGACCTCCGAGGGCGGC

AAGCTGAAGCTGGCCATCTCTAACAAGGAGTGGCTGCAGTTTGTGCAGGAGAGATC
CTACGAGAAGGACAAAAGGCCGGCGGCCACG*AAAAAGGCCGGCCAGGCAAAAAAGAA*
*AAAG*GGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGACGTGCCT**
GATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC (SEQ ID NO: 227)

16 - Prevotella disiens (PdCpf1)

ATGGAGAACTATCAGGAGTTCACCAACCTGTTTCAGCTGAATAAGAC
ACTGAGATTCGAGCTGAAGCCCATCGGCAAGACCTGCGAGCTGCTGGAGGAGGGCA
AGATCTTCGCCAGCGGCTCCTTTCTGGAGAAGGACAAGGTGAGGGCCGATAACGTG
AGCTACGTGAAGAAGGAGATCGACAAGAAGCACAAGATCTTTATCGAGGAGACACT
GAGCTCCTTCTCTATCAGCAACGATCTGCTGAAGCAGTACTTTGACTGCTATAATGA
GCTGAAGGCCTTCAAGAAGGACTGTAAGAGCGATGAGGAGGAGGTGAAGAAAACC
GCCCTGCGCAACAAGTGTACCTCCATCCAGAGGGCCATGCGCGAGGCCATCTCTCAG
GCCTTTCTGAAGAGCCCCCAGAAGAAGCTGCTGGCCATCAAGAACCTGATCGAGAA
CGTGTTCAAGGCCGACGAGAATGTGCAGCACTTCTCCGAGTTTACCAGCTATTTCTC
CGGCTTTGAGACAAACAGAGAGAATTTCTACTCTGACGAGGAGAAGTCCACATCTAT
CGCCTATAGGCTGGTGCACGATAACCTGCCTATCTTCATCAAGAACATCTACATCTT
CGAGAAGCTGAAGGAGCAGTTCGACGCCAAGACCCTGAGCGAGATCTTCGAGAACT
ACAAGCTGTATGTGGCCGGCTCTAGCCTGGATGAGGTGTTCTCCCTGGAGTACTTTA
ACAATACCCTGACACAGAAGGGCATCGACAACTATAATGCCGTGATCGGCAAGATC
GTGAAGGAGGATAAGCAGGAGATCCAGGGCCTGAACGAGCACATCAACCTGTATAA
TCAGAAGCACAAGGACCGGAGACTGCCCTTCTTTATCTCCCTGAAGAAGCAGATCCT
GTCCGATCGGGAGGCCCTGTCTTGGCTGCCTGACATGTTCAAGAATGATTCTGAAGT
GATCAAGGCCCTGAAGGGCTTCTACATCGAGGACGGCTTTGAGAACAATGTGCTGA
CACCTCTGGCCACCCTGCTGTCCTCTCTGGATAAGTACAACCTGAATGGCATCTTTAT
CCGCAACAATGAGGCCCTGAGCTCCCTGTCCCAGAACGTGTATCGGAATTTTTCTAT
CGACGAGGCCATCGATGCCAACGCCGAGCTGCAGACCTTCAACAATTACGAGCTGA
TCGCCAATGCCCTGCGCGCCAAGATCAAGAAGGAGACAAAGCAGGGCCGGAAGTCT
TTCGAGAAGTACGAGGAGTATATCGATAAGAAGGTGAAGGCCATCGACAGCCTGTC
CATCCAGGAGATCAACGAGCTGGTGGAGAATTACGTGAGCGAGTTTAACTCTAATA
GCGGCAACATGCCAAGAAAGGTGGAGGACTACTTCAGCCTGATGAGGAAGGGCGAC
TTCGGCTCCAACGATCTGATCGAAAATATCAAGACCAAGCTGAGCGCCGCAGAGAA
GCTGCTGGGCACAAAGTACCAGGAGACAGCCAAGGACATCTTCAAGAAGGATGAGA
ACTCCAAGCTGATCAAGGAGCTGCTGGACGCCACCAAGCAGTTCCAGCACTTTATCA
AGCCACTGCTGGGCACAGGCGAGGAGGCAGATCGGGACCTGGTGTTCTACGGCGAT
TTTCTGCCCCTGTATGAGAAGTTTGAGGAGCTGACCCTGCTGTATAACAAGGTGCGG

AATAGACTGACACAGAAGCCCTATTCCAAGGACAAGATCCGCCTGTGCTTCAACAA
GCCTAAGCTGATGACAGGCTGGGTGGATTCCAAGACCGAGAAGTCTGACAACGGCA
CACAGTACGGCGGCTATCTGTTTCGGAAGAAGAATGAGATCGGCGAGTACGATTATT
TTCTGGGCATCTCTAGCAAGGCCCAGCTGTTCAGAAAGAACGAGGCCGTGATCGGC
GACTACGAGAGGCTGGATTACTATCAGCCAAAGGCCAATACCATCTACGGCTCTGCC
TATGAGGGCGAGAACAGCTACAAGGAGGACAAGAAGCGGCTGAACAAAGTGATCA
TCGCCTATATCGAGCAGATCAAGCAGACAAACATCAAGAAGTCTATCATCGAGTCC
ATCTCTAAGTATCCTAATATCAGCGACGATGACAAGGTGACCCCATCCTCTCTGCTG
GAGAAGATCAAGAAGGTGTCTATCGACAGCTACAACGGCATCCTGTCCTTCAAGTCT
TTTCAGAGCGTGAACAAGGAAGTGATCGATAACCTGCTGAAAACCATCAGCCCCCT
GAAGAACAAGGCCGAGTTTCTGGACCTGATCAATAAGGATTATCAGATCTTCACCGA
GGTGCAGGCCGTGATCGACGAGATCTGCAAGCAGAAAACCTTCATCTACTTTCCAAT
CTCCAACGTGGAGCTGGAGAAGGAGATGGGCGATAAGGACAAGCCCCTGTGCCTGT
TCCAGATCAGCAATAAGGATCTGTCCTTCGCCAAGACCTTTAGCGCCAACCTGCGGA
AGAAGAGAGGCGCCGAGAATCTGCACACAATGCTGTTTAAGGCCCTGATGGAGGGC
AACCAGGATAATCTGGACCTGGGCTCTGGCGCCATCTTCTACAGAGCCAAGAGCCTG
GACGGCAACAAGCCCACACACCCTGCCAATGAGGCCATCAAGTGTAGGAACGTGGC
CAATAAGGATAAGGTGTCCCTGTTCACCTACGACATCTATAAGAACAGGCGCTACAT
GGAGAATAAGTTCCTGTTTCACCTGAGCATCGTGCAGAACTATAAGGCCGCCAATGA
CTCCGCCCAGCTGAACAGCTCCGCCACCGAGTATATCAGAAAGGCCGATGACCTGC
ACATCATCGGCATCGATAGGGGCGAGCGCAATCTGCTGTACTATTCCGTGATCGATA
TGAAGGGCAACATCGTGGAGCAGGACTCTCTGAATATCATCAGGAACAATGACCTG
GAGACAGATTACCACGACCTGCTGGATAAGAGGGAGAAGGAGCGCAAGGCCAACC
GGCAGAATTGGGAGGCCGTGGAGGGCATCAAGGACCTGAAGAAGGGCTACCTGAGC
CAGGCCGTGCACCAGATCGCCCAGCTGATGCTGAAGTATAACGCCATCATCGCCCTG
GAGGATCTGGGCCAGATGTTTGTGACCCGCGGCCAGAAGATCGAGAAGGCCGTGTA
CCAGCAGTTCGAGAAGAGCCTGGTGGATAAGCTGTCCTACCTGGTGGACAAGAAGC
GGCCTTATAATGAGCTGGGCGGCATCCTGAAGGCCTACCAGCTGGCCTCTAGCATCA
CCAAGAACAATTCTGACAAGCAGAACGGCTTCCTGTTTTATGTGCCAGCCTGGAATA
CAAGCAAGATCGATCCCGTGACCGGCTTTACAGACCTGCTGCGGCCCAAGGCCATG
ACCATCAAGGAGGCCCAGGACTTCTTTGGCGCCTTCGATAACATCTCTTACAATGAC
AAGGGCTATTTCGAGTTTGAGACAAACTACGACAAGTTTAAGATCAGAATGAAGAG
CGCCCAGACCAGGTGGACAATCTGCACCTTCGGCAATCGGATCAAGAGAAAGAAGG
ATAAGAACTACTGGAATTATGAGGAGGTGGAGCTGACCGAGGAGTTCAAGAAGCTG
TTTAAGGACAGCAACATCGATTACGAGAACTGTAATCTGAAGGAGGAGATCCAGAA
CAAGGACAATCGCAAGTTCTTTGATGACCTGATCAAGCTGCTGCAGCTGACACTGCA

GATGCGGAACTCCGATGACAAGGGCAATGATTATATCATCTCTCCTGTGGCCAACGC
CGAGGGCCAGTTCTTTGACTCCCGCAATGGCGATAAGAAGCTGCCACTGGATGCAG
ACGCAAACGGAGCCTACAATATCGCCCGCAAGGGCCTGTGGAACATCCGGCAGATC
AAGCAGACCAAGAACGACAAGAAGCTGAATCTGAGCATCTCCTCTACAGAGTGGCT
GGATTTCGTGCGGGAGAAGCCTTACCTGAAG*AAAAGGCCGGCGGCCACGAAAAAGGC*
*CGGCCAGGCAAAAAAGAAAAA*GGATCC**TACCCATACGATGTTCCAGATTACGCTT**
ATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCT
AAGAATTC (SEQ ID NO: 228)

17 - Porphyromonas macacae (PmCpf1)

ATGAAAACCCAGCACTTCTTTGAGGACTTCACAAGCCTGTACTCTCTG
AGCAAGACCATCCGGTTTGAGCTGAAGCCAATCGGCAAGACCCTGGAGAACATCAA
GAAGAATGGCCTGATCCGGAGAGATGAGCAGAGACTGGACGATTACGAGAAGCTGA
AGAAAGTGATCGACGAGTATCACGAGGATTTCATCGCCAACATCCTGAGCTCCTTTT
CCTTCTCTGAGGAGATCCTGCAGTCCTACATCCAGAATCTGAGCGAGTCCGAGGCCA
GGGCCAAGATCGAGAAAACCATGCGCGACACACTGGCCAAGGCCTTCTCTGAGGAT
GAGAGGTACAAGAGCATCTTTAAGAAGGAGCTGGTGAAGAAGGACATCCCCGTGTG
GTGCCCTGCCTATAAGAGCCTGTGCAAGAAGTTCGATAACTTTACCACATCTCTGGT
GCCCTTCCACGAGAACAGGAAGAACCTGTATACCAGCAATGAGATCACAGCCTCTA
TCCCTTATCGCATCGTGCACGTGAACCTGCCAAAGTTTATCCAGAATATCGAGGCCC
TGTGCGAGCTGCAGAAGAAGATGGGCGCCGACCTGTACCTGGAGATGATGGAGAAC
CTGCGCAACGTGTGGCCCAGCTTCGTGAAAACCCCAGACGACCTGTGCAACCTGAA
AACCTATAATCACCTGATGGTGCAGTCTAGCATCAGCGAGTACAACAGGTTTGTGGG
CGGCTATTCCACCGAGGACGGCACAAAGCACCAGGGCATCAACGAGTGGATCAATA
TCTACAGACAGAGGAATAAGGAGATGCGCCTGCCTGGCCTGGTGTTCCTGCACAAG
CAGATCCTGGCCAAGGTGGACTCCTCTAGCTTCATCAGCGATACACTGGAGAACGAC
GATCAGGTGTTTTGCGTGCTGAGACAGTTCAGGAAGCTGTTTTGGAATACCGTGTCC
TCTAAGGAGGACGATGCCGCCTCCCTGAAGGACCTGTTCTGTGGCCTGTCTGGCTAT
GACCCTGAGGCCATCTACGTGAGCGATGCCCACCTGGCCACAATCTCCAAGAACATC
TTTGACAGATGGAATTACATCTCCGATGCCATCAGGCGCAAGACCGAGGTGCTGATG
CCACGGAAGAAGGAGAGCGTGGAGAGATATGCCGAGAAGATCTCCAAGCAGATCA
AGAAGAGACAGTCTTACAGCCTGGCCGAGCTGGACGATCTGCTGGCCCACTATAGC
GAGGAGTCCCTGCCCGCAGGCTTCTCTCTGCTGAGCTACTTTACATCTCTGGGCGGC
CAGAAGTATCTGGTGAGCGACGGCGAAGTGATCCTGTACGAGGAGGGCAGCAACAT
CTGGGACGAGGTGCTGATCGCCTTCAGGGATCTGCAGGTCATCCTGGACAAGGACTT
CACCGAGAAGAAGCTGGGCAAGGATGAGGAGGCCGTGTCTGTGATCAAGAAGGCCC
TGGACAGCGCCCTGCGCCTGCGGAAGTTCTTTGATCTGCTGTCCGGCACAGGCGCAG

AGATCAGGAGAGACAGCTCCTTCTATGCCCTGTATACCGACCGGATGGATAAGCTGA
AGGGCCTGCTGAAGATGTATGATAAGGTGAGAAACTACCTGACCAAGAAGCCTTAT
TCCATCGAGAAGTTCAAGCTGCACTTTGACAACCCATCCCTGCTGTCTGGCTGGGAT
AAGAATAAGGAGCTGAACAATCTGTCTGTGATCTTCCGGCAGAACGGCTACTATTAC
CTGGGCATCATGACACCCAAGGGCAAGAATCTGTTCAAGACCCTGCCTAAGCTGGG
CGCCGAGGAGATGTTTTATGAGAAGATGGAGTACAAGCAGATCGCCGAGCCTATGC
TGATGCTGCCAAAGGTGTTCTTTCCCAAGAAAACCAAGCCAGCCTTCGCCCCAGACC
AGAGCGTGGTGGATATCTACAACAAGAAAACCTTCAAGACAGGCCAGAAGGGCTTT
AATAAGAAGGACCTGTACCGGCTGATCGACTTCTACAAGGAGGCCCTGACAGTGCA
CGAGTGGAAGCTGTTTAACTTCTCCTTTTCTCCAACCGAGCAGTATCGGAATATCGG
CGAGTTCTTTGACGAGGTGAGAGAGCAGGCCTACAAGGTGTCCATGGTGAACGTGC
CCGCCTCTTATATCGACGAGGCCGTGGAGAACGGCAAGCTGTATCTGTTCCAGATCT
ACAATAAGGACTTCAGCCCCTACTCCAAGGGCATCCCTAACCTGCACACACTGTATT
GGAAGGCCCTGTTCAGCGAGCAGAATCAGAGCCGGGTGTATAAGCTGTGCGGAGGA
GGAGAGCTGTTTTATAGAAAGGCCAGCCTGCACATGCAGGACACCACAGTGCACCC
CAAGGGCATCTCTATCCACAAGAAGAACCTGAATAAGAAGGGCGAGACAAGCCTGT
TCAACTACGACCTGGTGAAGGATAAGAGGTTTACCGAGGACAAGTTCTTTTTCCACG
TGCCTATCTCTATCAACTACAAGAATAAGAAGATCACCAACGTGAATCAGATGGTGC
GCGATTATATCGCCCAGAACGACGATCTGCAGATCATCGGCATCGACCGCGGCGAG
CGGAATCTGCTGTATATCAGCCGGATCGATACAAGGGGCAACCTGCTGGAGCAGTTC
AGCCTGAATGTGATCGAGTCCGACAAGGGCGATCTGAGAACCGACTATCAGAAGAT
CCTGGGCGATCGCGAGCAGGAGCGGCTGAGGCGCCGGCAGGAGTGGAAGTCTATCG
AGAGCATCAAGGACCTGAAGGATGGCTACATGAGCCAGGTGGTGCACAAGATCTGT
AACATGGTGGTGGAGCACAAGGCCATCGTGGTGCTGGAGAACCTGAATCTGAGCTT
CATGAAGGGCAGGAAGAAGGTGGAGAAGTCCGTGTACGAGAAGTTTGAGCGCATGC
TGGTGGACAAGCTGAACTATCTGGTGGTGGATAAGAAGAACCTGTCCAATGAGCCA
GGAGGCCTGTATGCAGCATACCAGCTGACCAATCCACTGTTCTCTTTTGAGGAGCTG
CACAGATACCCCCAGAGCGGCATCCTGTTTTTCGTGGACCCATGGAACACCTCTCTG
ACAGATCCCAGCACAGGCTTCGTGAATCTGCTGGGCAGAATCAACTACACCAATGTG
GGCGACGCCCGCAAGTTTTTCGATCGGTTTAACGCCATCAGATATGACGGCAAGGGC
AATATCCTGTTCGACCTGGATCTGTCCAGATTTGATGTGAGGGTGGAGACACAGAGG
AAGCTGTGGACACTGACCACATTCGGCTCTCGCATCGCCAAATCCAAGAAGTCTGGC
AAGTGGATGGTGGAGCGGATCGAGAACCTGAGCCTGTGCTTTCTGGAGCTGTTCGAG
CAGTTTAATATCGGCTACAGAGTGGAGAAGGACCTGAAGAAGGCCATCCTGAGCCA
GGATAGGAAGGAGTTCTATGTGCGCCTGATCTACCTGTTTAACCTGATGATGCAGAT
CCGGAACAGCGACGGCGAGGAGGATTATATCCTGTCTCCCGCCCTGAACGAGAAGA

ATCTGCAGTTCGACAGCAGGCTGATCGAGGCCAAGGATCTGCCTGTGGACGCAGAT
GCAAACGGAGCATACAATGTGGCCCGCAAGGGCCTGATGGTGGTGCAGAGAATCAA
GAGGGGCGACCACGAGTCCATCCACAGGATCGGAAGGGCACAGTGGCTGAGATATG
TGCAGGAGGGCATCGTGGAG*AAAAGGCCCGGCGGCCACGAAAAAGGCCGGCCAGGCAA*
*AAAAGAAAAAG*GGATCC**TACCCATACGATGTTCCAGATTACGCTTATCCCTACGAC**
GTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCTAAGAATTC
(SEQ ID NO: 229)

Secuencia de aminoácidos los ortólogos de Cpf1 con codones optimizados para el ser humano

Señal de localización nuclear (NLS) (en cursiva)

Conector de glicina-serina (con subrayado simple)

Marca de 3x HA (en negrita)

1 - Franscisella tularensis subespecie novicida U112 (FnCpf1)

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKA
KQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEY
IKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKG
WTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIK
KDLAEELTFDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTK
RKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFY
EQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAV
LEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEIL
ANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLH
KLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNF
ENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIV
YKLLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRK
FIDFYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQ
GKLYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPK
KITHPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEI
NLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAI
EKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVP
AGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKAA
KGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGE
SDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDADAN
GAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN*KRPAATKKAGQAKKKK*
GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 230)**

3 - La bacteria Lachnospiraceae MC2017 (Lb3Cpf1)

MDYGNGQFERRAPLTKTITLRLKPIGETRETIREQKLLEQDAAFRK
LVETVTPIVDDCIRKIADNALCHFGTEYDFSCLGNAISKNDSKAIKKETEKVEKLL
AKVLTENLPDGLRKVNDINSAAFIQDTLTSFVQDDADKRVLIQELKGKTVLMQR
FLTTRITALTVWLPDRVFENFNIFIENAEKMRILLDSPLNEKIMKFDPDAEQYASL
EFYGQCLSQKDIDSYNLIISGIYADDEVKNPGINEIVKEYNQQIRGDKDESPLPKL
KKLHKQILMPVEKAFFVRVLSNDSDARSILEKILKDTEMLPSKIIEAMKEADAGD
IAVYGSRLHELSHVIYGDHGKLSQIIYDKESKRISELMETLSPKERKESKKRLEGL
EEHIRKSTYTFDELNRYAEKNVMAAYIAAVEESCAEIMRKEKDLRTLLSKEDVKI
RGNRHNTLIVKNYFNAWTVFRNLIRILRRKSEAEIDSDFYDVLDDSVEVLSLTYK
GENLCRSYITKKIGSDLKPEIATYGSALRPNSRWWSPGEKFNVKFHTIVRRDGRL
YYFILPKGAKPVELEDMDGDIECLQMRKIPNPTIFLPKLVFKDPEAFFRDNPEADE
FVFLSGMKAPVTITRETYEAYRYKLYTVGKLRDGEVSEEEYKRALLQVLTAYKE
FLENRMIYADLNFGFKDLEEYKDSSEFIKQVETHNTFMCWAKVSSSQLDDLVKS
GNGLLFEIWSERLESYYKYGNEKVLRGYEGVLLSILKDENLVSMRTLLNSRPML
VYRPKESSKPMVVHRDGSRVVDRFDKDGKYIPPEVHDELYRFFNNLLIKEKLGE
KARKILDNKKVKVKVLESERVKWSKFYDEQFAVTFSVKKNADCLDTTKDLNAE
VMEQYSESNRLILIRNTTDILYYLVLDKNGKVLKQRSLNIINDGARDVDWKERF
RQVTKDRNEGYNEWDYSRTSNDLKEVYLNYALKEIAEAVIEYNAILIIEKMSNA
FKDKYSFLDDVTFKGFETKLLAKLSDLHFRGIKDGEPCSFTNPLQLCQNDSNKIL
QDGVIFMVPNSMTRSLDPDTGFIFAINDHNIRTKKAKLNFLSKFDQLKVSSEGCLI
MKYSGDSLPTHNTDNRVWNCCCNHPITNYDRETKKVEFIEEPVEELSRVLEENGI
ETDTELNKLNERENVPGKVVDAIYSLVLNYLRGTVSGVAGQRAVYYSPVTGKK
YDISFIQAMNLNRKCDYYRIGSKERGEWTDFVAQLIN*KRPAATKKAGQAKKKK*GS
YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 231)

4 - Butyrivibrio proteoclasticus (BpCpf1)

MLLYENYTKRNQITKSLRLELRPQGKTLRNIKELNLLEQDKAIYALLERL
KPVIDEGIKDIARDTLKNCELSFEKLYEHFLSGDKKAYAKESERLKKEIVKTLIKNLPEGI
GKISEINSAKYLNGVLYDFIDKTHKDSEEKQNILSDILETKGYLALFSKFLTSRITTLEQSM
PKRVIENFEIYAANIPKMQDALERGAVSFAIEYESICSVDYYNQILSQEDIDSYNRLISGIM
DEDGAKEKGINQTISEKNIKIKSEHLEEKPFRILKQLHKQILEEREKAFTIDHIDSDEEVVQ
VTKEAFEQTKEQWENIKKINGFYAKDPGDITLFIVVGPNQTHVLSQLIYGEHDRIRLLLEE
YEKNTLEVLPRRTKSEKARYDKFVNAVPKKVAKESHTFDGLQKMTGDDRLFILYRDEL
ARNYMRIKEAYGTFERDILKSRRGIKGNRDVQESLVSFYDELTKFRSALRIINSGNDEKA

DPIFYNTFDGIFEKANRTYKAENLCRNYVTKSPADDARIMASCLGTPARLRTHWWNGEE
NFAINDVAMIRRGDEYYYFVLTPDVKPVDLKTKDETDAQIFVQRKGAKSFLGLPKALFK
CILEPYFESPEHKNDKNCVIEEYVSKPLTIDRRAYDIFKNGTFKKTNIGIDGLTEEKFKDD
CRYLIDVYKEFIAVYTRYSCFNMSGLKRADEYNDIGEFFSDVDTRLCTMEWIPVSFERIN
DMVDKKEGLLFLVRSMFLYNRPRKPYERTFIQLFSDSNMEHTSMLLNSRAMIQYRAASL
PRRVTHKKGSILVALRDSNGEHIPMHIREAIYKMKNNFDISSEDFIMAKAYLAEHDVAIK
KANEDIIRNRRYTEDKFFLSLSYTKNADISARTLDYINDKVEEDTQDSRMAVIVTRNLKD
LTYVAVVDEKNNVLEEKSLNEIDGVNYRELLKERTKIKYHDKTRLWQYDVSSKGLKEA
YVELAVTQISKLATKYNAVVVVESMSSTFKDKFSFLDEQIFKAFEARLCARMSDLSFNTI
KEGEAGSISNPIQVSNNNGNSYQDGVIYFLNNAYTRTLCPDTGFVDVFDKTRLITMQSKR
QFFAKMKDIRIDDGEMLFTFNLEEYPTKRLLDRKEWTVKIAGDGSYFDKDKGEYVYVN
DIVREQIIPALLEDKAVFDGNMAEKFLDKTAISGKSVELIYKWFANALYGIITKKDGEKIY
RSPITGTEIDVSKNTTYNFGKKFMFKQEYRGDGDFLDAFLNYMQAQDIAV*KRPAATKKA*
*GQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 232)**

5 - La bacteria Peregrinibacteria GW2011_GWA_33_10 (PeCpf1)

MSNFFKNFTNLYELSKTLRFELKPVGDTLTNMKDHLEYDEKLQTFLKDQ
NIDDAYQALKPQFDEIHEEFITDSLESKKAKEIDFSEYLDLFQEKKELNDSEKKLRNKIGE
TFNKAGEKWKKEKYPQYEWKKGSKIANGADILSCQDMLQFIKYKNPEDEKIKNYIDDT
LKGFFTYFGGFNQNRANYYETKKEASTAVATRIVHENLPKFCDNVIQFKHIIKRKKDGT
VEKTERKTEYLNAYQYLKNNNKITQIKDAETEKMIESTPIAEKIFDVYYFSSCLSQKQIEE
YNRIIGHYNLLINLYNQAKRSEGKHLSANEKKYKDLPKFKTLYKQIGCGKKKDLFYTIK
CDTEEEANKSRNEGKESHSVEEIINKAQEAINKYFKSNNDCENINTVPDFINYILTKENYE
GVYWSKAAMNTISDKYFANYHDLQDRLKEAKVFQKADKKSEDDIKIPEAIELSGLFGVL
DSLADWQTTLFKSSILSNEDKLKIITDSQTPSEALLKMIFNDIEKNMESFLKETNDIITLKK
YKGNKEGTEKIKQWFDYTLAINRMLKYFLVKENKIKGNSLDTNISEALKTLIYSDDAEW
FKWYDALRNYLTQKPQDEAKENKLKLNFDNPSLAGGWDVNKECSNFCVILKDKNEKK
YLAIMKKGENTLFQKEWTEGRGKNLTKKSNPLFEINNCEILSKMEYDFWADVSKMIPKC
STQLKAVVNHFKQSDNEFIFPIGYKVTSGEKFREECKISKQDFELNNKVFNKNELSVTAM
RYDLSSTQEKQYIKAFQKEYWELLFKQEKRDTKLTNNEIFNEWINFCNKKYSELLSWER
KYKDALTNWINFCKYFLSKYPKTTLFNYSFKESENYNSLDEFYRDVDICSYKLNINTTIN
KSILDRLVEEGKLYLFEIKNQDSNDGKSIGHKNNLHTIYWNAIFENFDNRPKLNGEAEIFY
RKAISKDKLGIVKGKKTKNGTEIIKNYRFSKEKFILHVPITLNFCSNNEYVNDIVNTKFYN
FSNLHFLGIDRGEKHLAYYSLVNKNGEIVDQGTLNLPFTDKDGNQRSIKKEKYFYNKQE
DKWEAKEVDCWNYNDLLDAMASNRDMARKNWQRIGTIKEAKNGYVSLVIRKIADLA
VNNERPAFIVLEDLNTGFKRSRQKIDKSVYQKFELALAKKLNFLVDKNAKRDEIGSPTK
ALQLTPPVNNYGDIENKKQAGIMLYTRANYTSQTDPATGWRKTIYLKAGPEETTYKKD

GKIKNKSVKDQIIETFTDIGFDGKDYYFEYDKGEFVDEKTGEIKPKKWRLYSGENGKSLD
RFRGEREKDKYEWKIDKIDIVKILDDLFVNFDKNISLLKQLKEGVELTRNNEHGTGESLR
FAINLIQQIRNTGNNERDNDFILSPVRDENGKHFDSREYWDKETKGEKISMPSSGDANGA
FNIARKGIIMNAHILANSDSKDLSLFVSDEEWDLHLNNKTEWKKQLNIFSSRKAMA*KRK*
*KKRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 233)**

6 - La bacteria Parcubacteria GWC2011_GWC2_44_17 (PbCpf1)

MENIFDQFIGKYSLSKTLRFELKPVGKTEDFLKINKVFEKDQTIDDSYNQ
AKFYFDSLHQKFIDAALASDKTSELSFQNFADVLEKQNKIILDKKREMGALRKRDKNAV
GIDRLQKEINDAEDIIQKEKEKIYKDVRTLFDNEAESWKTYYQEREVDGKKITFSKADLK
QKGADFLTAAGILKVLKYEFPEEKEKEFQAKNQPSLFVEEKENPGQKRYIFDSFDKFAGY
LTKFQQTKKNLYAADGTSTAVATRIADNFIIFHQNTKVFRDKYKNNHTDLGFDEENIFEI
ERYKNCLLQREIEHIKNENSYNKIIGRINKKIKEYRDQKAKDTKLTKSDFPFFKNLDKQIL
GEVEKEKQLIEKTREKTEEDVLIERFKEFIENNEERFTAAKKLMNAFCNGEFESEYEGIYL
KNKAINTISRRWFVSDRDFELKLPQQKSKNKSEKNEPKVKKFISIAEIKNAVEELDGDIFK
AVFYDKKIIAQGGSKLEQFLVIWKYEFEYLFRDIERENGEKLLGYDSCLKIAKQLGIFPQE
KEAREKATAVIKNYADAGLGIFQMMKYFSLDDKDRKNTPGQLSTNFYAEYDGYYKDFE
FIKYYNEFRNFITKKPFDEDKIKLNFENGALLKGWDENKEYDFMGVILKKEGRLYLGIM
HKNHRKLFQSMGNAKGDNANRYQKMIYKQIADASKDVPRLLLTSKKAMEKFKPSQEIL
RIKKEKTFKRESKNFSLRDLHALIEYYRNCIPQYSNWSFYDFQFQDTGKYQNIKEFTDDV
QKYGYKISFRDIDDEYINQALNEGKMYLFEVVNKDIYNTKNGSKNLHTLYFEHILSAENL
NDPVFKLSGMAEIFQRQPSVNEREKITTQKNQCILDKGDRAYKYRRYTEKKIMFHMSLV
LNTGKGEIKQVQFNKIINQRISSSDNEMRVNVIGIDRGEKNLLYYSVVKQNGEIIEQASLN
EINGVNYRDKLIEREKERLKNRQSWKPVVKIKDLKKGYISHVIHKICQLIEKYSAIVVLED
LNMRFKQIRGGIERSVYQQFEKALIDKLGYLVFKDNRDLRAPGGVLNGYQLSAPFVSFE
KMRKQTGILFYTQAEYTSKTDPITGFRKNVYISNSASLDKIKEAVKKFDAIGWDGKEQSY
FFKYNPYNLADEKYKNSTVSKEWAIFASAPRIRRQKGEDGYWKYDRVKVNEEFEKLLK
VWNFVNPKATDIKQEIIKKEKAGDLQGEKELDGRLRNFWHSFIYLFNLVLELRNSFSLQI
KIKAGEVIAVDEGVDFIASPVKPFFTTPNPYIPSNLCWLAVENADANGAYNIARKGVMIL
KKIREHAKKDPEFKKLPNLFISNAEWDEAARDWGKYAGTTALNLDH*KRPAATKKAGQA*
*KKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 234)**

7 - Smithella sp. SC_K08D17 (SsCpf1)

MQTLFENFTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKDEDRAEKYK
KVKNIIDEYHKDFIEKSLNGLKLDGLEKYKTLYLKQEKDDKDKKAFDKEKENLRKQIAN
AFRNNEKFKTLFAKELIKNDLMSFACEEDKKNVKEFEAFTTYFTGFHQNRANMYVADE
KRTAIASRLIHENLPKFIDNIKIFEKMKKEAPELLSPFNQTLKDMKDVIKGTTLEEIFSLDY

FNKTLTQSGIDIYNSVIGGRTPEEGKTKIKGLNEYINTDFNQKQTDKKKRQPKFKQLYKQ
ILSDRQSLSFIAEAFKNDTEILEAIEKFYVNELLHFSNEGKSTNVLDAIKNAVSNLESFNLT
KMYFRSGASLTDVSRKVFGEWSIINRALDNYYATTYPIKPREKSEKYEERKEKWLKQDF
NVSLIQTAIDEYDNETVKGKNSGKVIADYFAKFCDDKETDLIQKVNEGYIAVKDLLNTP
CPENEKLGSNKDQVKQIKAFMDSIMDIMHFVRPLSLKDTDKEKDETFYSLFTPLYDHLT
QTIALYNKVRNYLTQKPYSTEKIKLNFENSTLLGGWDLNKETDNTAIILRKDNLYYLGIM
DKRHNRIFRNVPKADKKDFCYEKMVYKLLPGANKMLPKVFFSQSRIQEFTPSAKLLENY
ANETHKKGDNFNLNHCHKLIDFFKDSINKHEDWKNFDFRFSATSTYADLSGFYHEVEHQ
GYKISFQSVADSFIDDLVNEGKLYLFQIYNKDFSPFSKGKPNLHTLYWKMLFDENNLKD
VVYKLNGEAEVFYRKKSIAENNTTIHKANESIINKNPDNPKATSTFNYDIVKDKRYTIDK
FQFHIPITMNFKAEGIFNMNQRVNQFLKANPDINIIGIDRGERHLLYYALINQKGKILKQD
TLNVIANEKQKVDYHNLLDKKEGDRATARQEWGVIETIKELKEGYLSQVIHKLTDLMIE
NNAIIVMEDLNFGFKRGRQKVEKQVYQKFEKMLIDKLNYLVDKNKKANELGGLLNAFQ
LANKFESFQKMGKQNGFIFYVPAWNTSKTDPATGFIDFLKPRYENLNQAKDFFEKFDSIR
LNSKADYFEFAFDFKNFTEKADGGRTKWTVCTTNEDRYAWNRALNNNRGSQEKYDIT
AELKSLFDGKVDYKSGKDLKQQIASQESADFFKALMKNLSITLSLRHNNGEKGDNEQD
YILSPVADSKGRFFDSRKADDDMPKNADANGAYHIALKGLWCLEQISKTDDLKKVKLAI
SNKEWLEFVQTLKG*KRPAATKKAGQAKKKK*GS**SYPYDVPDYAYPYDVPDYAYPYDVPD**
YA (SEQ ID NO: 235)

8 - Acidaminococcus sp. BV3L6 (AsCpf1)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKEL
KPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFI
GRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFS
GFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIF
VSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHI
IASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELN
SIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDIN
LQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHL
LDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPT
LASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDY
FPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAY
AKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLL
YHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTS
IKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSH
DLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLK
EHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWS

VVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKM
LIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLT
GFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPA
WDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDG
SNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNP
EWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN*KRPAATK*
*KAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 236)**

9 - La bacteria Lachnospiraceae MA2020 (Lb2Cpf1)

MYYESLTKQYPVSKTIRNELIPIGKTLDNIRQNNILESDVKRKQNYEHVK
GILDEYHKQLINEALDNCTLPSLKIAAEIYLKNQKEVSDREDFNKTQDLLRKEVVEKLKA
HENFTKIGKKDILDLLEKLPSISEDDYNALESFRNFYTYFTSYNKVRENLYSDKEKSSTVA
YRLINENFPKFLDNVKSYREVKTAGILADGLGEEEQDSLFIVETFNKTLTQDGIDTYNSQ
VGKINSSINLYNQKNQKANGFRKIPKMKMLYKQILSDREESFIDEFQSDEVLIDNVESYG
SVLIESLKSSKVSAFFDALRESKGKNVYVKNDLAKTAMSNIVFENWRTFDDLLNQEYDL
ANENKKKDDKYFEKRQKELKKNKSYSLEHLCNLSEDSCNLIENYIHQISDDIENIIINNET
FLRIVINEHDRSRKLAKNRKAVKAIKDFLDSIKVLERELKLINSSGQELEKDLIVYSAHEE
LLVELKQVDSLYNMTRNYLTKKPFSTEKVKLNFNRSTLLNGWDRNKETDNLGVLLLKD
GKYYLGIMNTSANKAFVNPPVAKTEKVFKKVDYKLLPVPNQMLPKVFFAKSNIDFYNPS
SEIYSNYKKGTHKKGNMFSLEDCHNLIDFFKESISKHEDWSKFGFKFSDTASYNDISEFY
REVEKQGYKLTYTDIDETYINDLIERNELYLFQIYNKDFSMYSKGKLNLHTLYFMMLFD
QRNIDDVVYKLNGEAEVFYRPASISEDELIIHKAGEEIKNKNPNRARTKETSTFSYDIVKD
KRYSKDKFTLHIPITMNFGVDEVKRFNDAVNSAIRIDENVNVIGIDRGERNLLYVVVIDS
KGNILEQISLNSIINKEYDIETDYHALLDEREGGRDKARKDWNTVENIRDLKAGYLSQVV
NVVAKLVLKYNAIICLEDLNFGFKRGRQKVEKQVYQKFEKMLIDKLNYLVIDKSREQTS
PKELGGALNALQLTSKFKSFKELGKQSGVIYYVPAYLTSKIDPTTGFANLFYMKCENVE
KSKRFFDGFDFIRFNALENVFEFGFDYRSFTQRACGINSKWTVCTNGERIIKYRNPDKNN
MFDEKVVVVTDEMKNLFEQYKIPYEDGRNVKDMIISNEEAEFYRRLYRLLQQTLQMRN
STSDGTRDYIISPVKNKREAYFNSELSDGSVPKDADANGAYNIARKGLWVLEQIRQKSE
GEKINLAMTNAEWLEYAQTHLL*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYA**
YPYDVPDYA (SEQ ID NO: 237)

10 - La bacteria Methanoplasma termitum candidata (CMtCpf1)

MNNYDEFTKLYPIQKTIRFELKPQGRTMEHLETFNFFEEDRDRAEKYKIL
KEAIDEYHKKFIDEHLTNMSLDWNSLKQISEKYYKSREEKDKKVFLSEQKRMRQEIVSE
FKKDDRFKDLFSKKLFSELLKEEIYKKGNHQEIDALKSFDKFSGYFIGLHENRKNMYSDG
DEITAISNRIVNENFPKFLDNLQKYQEARKKYPEWIIKAESALVAHNIKMDEVFSLEYFN
KVLNQEGIQRYNLALGGYVTKSGEKMMGLNDALNLAHQSEKSSKGRIHMTPLFKQILS

EKESFSYIPDVFTEDSQLLPSIGGFFAQIENDKDGNIFDRALELISSYAEYDTERIYIRQADI
NRVSNVIFGEWGTLGGLMREYKADSINDINLERTCKKVDKWLDSKEFALSDVLEAIKRT
GNNDAFNEYISKMRTAREKIDAARKEMKFISEKISGDEESIHIIKTLLDSVQQFLHFFNLFK
ARQDIPLDGAFYAEFDEVHSKLFAIVPLYNKVRNYLTKNNLNTKKIKLNFKNPTLANGW
DQNKVYDYASLIFLRDGNYYLGIINPKRKKNIKFEQGSGNGPFYRKMVYKQIPGPNKNL
PRVFLTSTKGKKEYKPSKEIIEGYEADKHIRGDKFDLDFCHKLIDFFKESIEKHKDWSKFN
FYFSPTESYGDISEFYLDVEKQGYRMHFENISAETIDEYVEKGDLFLFQIYNKDFVKAAT
GKKDMHTIYWNAAFSPENLQDVVVKLNGEAELFYRDKSDIKEIVHREGEILVNRTYNGR
TPVPDKIHKKLTDYHNGRTKDLGEAKEYLDKVRYFKAHYDITKDRRYLNDKIYFHVPLT
LNFKANGKKNLNKMVIEKFLSDEKAHIIGIDRGERNLLYYSIIDRSGKIIDQQSLNVIDGF
DYREKLNQREIEMKDARQSWNAIGKIKDLKEGYLSKAVHEITKMAIQYNAIVVMEELN
YGFKRGRFKVEKQIYQKFENMLIDKMNYLVFKDAPDESPGGVLNAYQLTNPLESFAKL
GKQTGILFYVPAAYTSKIDPTTGFVNLFNTSSKTNAQERKEFLQKFESISYSAKDGGIFAF
AFDYRKFGTSKTDHKNVWTAYTNGERMRYIKEKKRNELFDPSKEIKEALTSSGIKYDGG
QNILPDILRSNNNGLIYTMYSSFIAAIQMRVYDGKEDYIISPIKNSKGEFFRTDPKRRELPID
ADANGAYNIALRGELTMRAIAEKFDPDSEKMAKLELKHKDWFEFMQTRGD*KRPAATKK*
*AGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 238)**

11 - La eubacteria seleccionada (EeCpf1)

MNGNRSIVYREFVGVIPVAKTLRNELRPVGHTQEHIIQNGLIQEDELRQE
KSTELKNIMDDYYREYIDKSLSGVTDLDFTLLFELMNLVQSSPSKDNKKALEKEQSKMR
EQICTHLQSDSNYKNIFNAKLLKEILPDFIKNYNQYDVKDKAGKLETLALFNGFSTYFTD
FFEKRKNVFTKEAVSTSIAYRIVHENSLIFLANMTSYKKISEKALDEIEVIEKNNQDKMGD
WELNQIFNPDFYNMVLIQSGIDFYNEICGVVNAHMNLYCQQTKNNYNLFKMRKLHKQI
LAYTSTSFEVPKMFEDDMSVYNAVNAFIDETEKGNIIGKLKDIVNKYDELDEKRIYISKD
FYETLSCFMSGNWNLITGCVENFYDENIHAKGKSKEEKVKKAVKEDKYKSINDVNDLV
EKYIDEKERNEFKNSNAKQYIREISNIITDTETAHLEYDDHISLIESEEKADEMKKRLDMY
MNMYHWAKAFIVDEVLDRDEMFYSDIDDIYNILENIVPLYNRVRNYVTQKPYNSKKIKL
NFQSPTLANGWSQSKEFDNNAIILIRDNKYYLAIFNAKNKPDKKIIQGNSDKKNDNDYK
KMVYNLLPGANKMLPKVFLSKKGIETFKPSDYIISGYNAHKHIKTSENFDISFCRDLIDYF
KNSIEKHAEWRKYEFKFSATDSYSDISEFYREVEMQGYRIDWTYISEADINKLDEEGKIY
LFQIYNKDFAENSTGKENLHTMYFKNIFSEENLKDIIIKLNGQAELFYRRASVKNPVKHK
KDSVLVNKTYKNQLDNGDVVRIPIPDDIYNEIYKMYNGYIKESDLSEAAKEYLDKVEVR
TAQKDIVKDYRYTVDKYFIHTPITINYKVTARNNVNDMVVKYIAQNDDIHVIGIDRGER
NLIYISVIDSHGNIVKQKSYNILNNYDYKKKLVEKEKTREYARKNWKSIGNIKELKEGYI
SGVVHEIAMLIVEYNAIIAMEDLNYGFKRGRFKVERQVYQKFESMLINKLNYFASKEKS
VDEPGGLLKGYQLTYVPDNIKNLGKQCGVIFYVPAAFTSKIDPSTGFISAFNFKSISTNAS

RKQFFMQFDEIRYCAEKDMFSFGFDYNNFDTYNITMGKTQWTVYTNGERLQSEFNNAR
RTGKTKSINLTETIKLLLEDNEINYADGHDIRIDMEKMDEDKKSEFFAQLLSLYKLTVQM
RNSYTEAEEQENGISYDKIISPVINDEGEFFDSDNYKESDDKECKMPKDADANGAYCIAL
KGLYEVLKIKSEWTEDGFDRNCLKLPHAEWLDFIQNKRY*EKRPAATKKAGQAKKKK*GSY
PYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 239)

12 - Moraxella bovoculi 237 (MbCpf1)

MLFQDFTHLYPLSKTVRFELKPIDRTLEHIHAKNFLSQDETMADMHQKV
KVILDDYHRDFIADMMGEVKLTKLAEFYDVYLKFRKNPKDDELQKQLKDLQAVLRKEI
VKPIGNGGKYKAGYDRLFGAKLFKDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSTY
FTGFHDNRKNMYSDEDKHTAIAYRLIHENLPRFIDNLQILTTIKQKHSALYDQIINELTAS
GLDVSLASHLDGYHKLLTQEGITAYNTLLGGISGEAGSPKIQGINELINSHHNQHCHKSE
RIAKLRPLHKQILSDGMSVSFLPSKFADDSEMCQAVNEFYRHYADVFAKVQSLFDGFDD
HQKDGIYVEHKNLNELSKQAFGDFALLGRVLDGYYVDVVNPEFNERFAKAKTDNAKA
KLTKEKDKFIKGVHSLASLEQAIEHYTARHDDESVQAGKLGQYFKHGLAGVDNPIQKIH
NNHSTIKGFLERERPAGERALPKIKSGKNPEMTQLRQLKELLDNALNVAHFAKLLTTKT
TLDNQDGNFYGEFGVLYDELAKIPTLYNKVRDYLSQKPFSTEKYKLNFGNPTLLNGWD
LNKEKDNFGVILQKDGCYYLALLDKAHKKVFDNAPNTGKSIYQKMIYKYLEVRKQFPK
VFFSKEAIAINYHPSKELVEIKDKGRQRSDDERLKLYRFILECLKIHPKYDKKFEGAIGDIQ
LFKKDKKGREVPISEKDLFDKINGIFSSKPKLEMEDFFIGEFKRYNPSQDLVDQYNIYKKI
DSNDNRKKENFYNNHPKFKKDLVRYYYESMCKHEEWEESFEFSKKLQDIGCYVDVNEL
FTEIETRRLNYKISFCNINADYIDELVEQGQLYLFQIYNKDFSPKAHGKPNLHTLYFKALF
SEDNLADPIYKLNGEAQIFYRKASLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIKDK
RYTQDKFMLHVPITMNFGVQGMTIKEFNKKVNQSIQQYDEVNVIGIDRGERHLLYLTVI
NSKGEILEQCSLNDITTASANGTQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYL
SHVVHQISQLMLKYNAIVVLEDLNFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDKA
DDEIGSYKNALQLTNNFTDLKSIGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENIAQ
SQAFFGKFDKICYNADKDYFEFHIDYAKFTDKAKNSRQIWTICSHGDKRYVYDKTANQ
NKGAAKGINVNDELKSLFARHHINEKQPNLVMDICQNNDKEFHKSLMYLLKTLLALRY
SNASSDEDFILSPVANDEGVFFNSALADDTQPQNADANGAYHIALKGLWLLNELKNSDD
LNKVKLAIDNQTWLNFAQNR*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAY**
PYDVPDYA (SEQ ID NO: 240)

13 - Leptospira inadai (LiCpf1)

MEDYSGFVNIYSIQKTLRFELKPVGKTLEHIEKKGFLKKDKIRAEDYKAV
KKIIDKYHRAYIEEVFDSVLHQKKKKDKTRFSTQFIKEIKEFSELYYKTEKNIPDKERLEA
LSEKLRKMLVGAFKGEFSEEVAEKYKNLFSKELIRNEIEKFCETDEERKQVSNFKSFTTY
FTGFHSNRQNIYSDEKKSTAIGYRIIHQNLPKFLDNLKIIESIQRRFKDFPWSDLKKNLKKI

DKNIKLTEYFSIDGFVNVLNQKGIDAYNTILGGKSEESGEKIQGLNEYINLYRQKNNIDRK
NLPNVKILFKQILGDRETKSFIPEAFPDDQSVLNSITEFAKYLKLDKKKKSIIAELKKFLSSF
NRYELDGIYLANDNSLASISTFLFDDWSFIKKSVSFKYDESVGDPKKKIKSPLKYEKEKE
KWLKQKYYTISFLNDAIESYSKSQDEKRVKIRLEAYFAEFKSKDDAKKQFDLLERIEEAY
AIVEPLLGAEYPRDRNLKADKKEVGKIKDFLDSIKSLQFFLKPLLSAEIFDEKDLGFYNQL
EGYYEEIDSIGHLYNKVRNYLTGKIYSKEKFKLNFENSTLLKGWDENREVANLCVIFRED
QKYYLGVMDKENNTILSDIPKVKPNELFYEKMVYKLIPTPHMQLPRIIFSSDNLSIYNPSK
SILKIREAKSFKEGKNFKLKDCHKFIDFYKESISKNEDWSRFDFKFSKTSSYENISEFYREV
ERQGYNLDFKKVSKFYIDSLVEDGKLYLFQIYNKDFSIFSKGKPNLHTIYFRSLFSKENLK
DVCLKLNGEAEMFFRKKSINYDEKKKREGHHPELFEKLKYPILKDKRYSEDKFQFHLPIS
LNFKSKERLNFNLKVNEFLKRNKDINIIGIDRGERNLLYLVMINQKGEILKQTLLDSMQS
GKGRPEINYKEKLQEKEIERDKARKSWGTVENIKELKEGYLSIVIHQISKLMVENNAIVV
LEDLNIGFKRGRQKVERQVYQKFEKMLIDKLNFLVFKENKPTEPGGVLKAYQLTDEFQS
FEKLSKQTGFLFYVPSWNTSKIDPRTGFIDFLHPAYENIEKAKQWINKFDSIRFNSKMDW
FEFTADTRKFSENLMLGKNRVWVICTTNVERYFTSKTANSSIQYNSIQITEKLKELFVDIP
FSNGQDLKPEILRKNDAVFFKSLLFYIKTTLSLRQNNGKKGEEEKDFILSPVVDSKGRFFN
SLEASDDEPKDADANGAYHIALKGLMNLLVLNETKEENLSRPKWKIKNKDWLEFVQER
NR*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 241)**

14 - La bacteria Lachnospiraceae ND2006 (LbCpf1)

MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKG
VKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFK
GNEGYKSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIA
FRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGI
DVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGY
TSDEEVLEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWN
VIRDKWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKL
KEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFF
GEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGG
WDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNK
MLPKVFFSKKWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWS
NAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDF
SDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANK
NPDNPKKTTTLSYDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVI
GIDRGERNLLYIVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQN
WTSIENIKELKAGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEK

MLIDKLNYMVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPST
GFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSY
GNRIRIFRNPKKNNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSF
MALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYN
IARKVLWAIGQFKKAEDEKLDKVKIAISNKEWLEYAQTSVKH*KRPAATKKAGQAKKKK*G
SYPYDVPDYAYPYDVPDYAYPYDVPDYA **(SEQ ID NO: 242)**

15 - Porphyromonas crevioricanis (PcCpf1)

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRAESYRRV
KKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCELYKKDHRTEGEDKALDKIRAVLR
GLIVGAFTGVCGRRENTVQNEKYESLFKEKLIKEILPDFVLSTEAESLPFSVEEATRSLKEF
DSFTSYFAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFIDNILVFQKIKEPIAKELEHIRAD
FSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYNALIGKIVTEGDGEMKGLNEHINLYN
QQRGREDRLPLFRPLYKQILSDREQLSYLPESFEKDEELLRALKEFYDHIAEDILGRTQQL
MTSISEYDLSRIYVRNDSQLTDISKKMLGDWNAIYMARERAYDHEQAPKRITAKYERDR
IKALKGEESISLANLNSCIAFLDNVRDCRVDTYLSTLGQKEGPHGLSNLVENVFASYHEA
EQLLSFPYPEENNLIQDKDNVVLIKNLLDNISDLQRFLKPLWGMGDEPDKDERFYGEYN
YIRGALDQVIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLSGWDRNKEKDNSCVILRKG
QNFYLAIMNNRHKRSFENKMLPEYKEGEPYFEKMDYKFLPDPNKMLPKVFLSKKGIEIY
KPSPKLLEQYGHGTHKKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENV
SSFYREVEDQGYKLSFRKVSESYVYSLIDQGKLYLFQIYNKDFSPCSKGTPNLHTLYWR
MLFDERNLADVIYKLDGKAEIFFREKSLKNDHPTHPAGKPIKKKSRQKKGEESLFEYDLV
KDRRYTMDKFQFHVPITMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYIC
VIDSRGTILDQISLNTINDIDYHDLLESRDKDRQQEHRNWQTIEGIKELKQGYLSQAVHRI
AELMVAYKAVVALEDLNMGFKRGRQKVESSVYQQFEKQLIDKLNYLVDKKKRPEDIG
GLLRAYQFTAPFKSFKEMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHVQYENVDKAKSF
FQKFDSISYNPKKDWFEFAFDYKNFTKKAEGSRSMWILCTHGSRIKNFRNSQKNGQWDS
EEFALTEAFKSLFVRYEIDYTADLKTAIVDEKQKDFFVDLLKLFKLTVQMRNSWKEKDL
DYLISPVAGADGRFFDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKLA
ISNKEWLQFVQERSYEKD*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYD**
VPDYA (SEQ ID NO: 243)

16 - Prevotella disiens (PdCpf1)

MENYQEFTNLFQLNKTLRFELKPIGKTCELLEEGKIFASGSFLEKD
KVRADNVSYVKKEIDKKHKIFIEETLSSFSISNDLLKQYFDCYNELKAFKKDCKS
DEEEVKKTALRNKCTSIQRAMREAISQAFLKSPQKKLLAIKNLIENVFKADENVQ
HFSEFTSYFSGFETNRENFYSDEEKSTSIAYRLVHDNLPIFIKNIYIFEKLKEQFDA

KTLSEIFENYKLYVAGSSLDEVFSLEYFNNTLTQKGIDNYNAVIGKIVKEDKQEIQ
GLNEHINLYNQKHKDRRLPFFISLKKQILSDREALSWLPDMFKNDSEVIKALKGF
YIEDGFENNVLTPLATLLSSLDKYNLNGIFIRNNEALSSLSQNVYRNFSIDEAIDA
NAELQTFNNYELIANALRAKIKKETKQGRKSFEKYEEYIDKKVKAIDSLSIQEINE
LVENYVSEFNSNSGNMPRKVEDYFSLMRKGDFGSNDLIENIKTKLSAAEKLLGT
KYQETAKDIFKKDENSKLIKELLDATKQFQHFIKPLLGTGEEADRDLVFYGDFLP
LYEKFEELTLLYNKVRNRLTQKPYSKDKIRLCFNKPKLMTGWVDSKTEKSDNG
TQYGGYLFRKKNEIGEYDYFLGISSKAQLFRKNEAVIGDYERLDYYQPKANTIY
GSAYEGENSYKEDKKRLNKVIIAYIEQIKQTNIKKSIIESISKYPNISDDDKVTPSSL
LEKIKKVSIDSYNGILSFKSFQSVNKEVIDNLLKTISPLKNKAEFLDLINKDYQIFT
EVQAVIDEICKQKTFIYFPISNVELEKEMGDKDKPLCLFQISNKDLSFAKTFSANL
RKKRGAENLHTMLFKALMEGNQDNLDLGSGAIFYRAKSLDGNKPTHPANEAIK
CRNVANKDKVSLFTYDIYKNRRYMENKFLFHLSIVQNYKAANDSAQLNSSATE
YIRKADDLHIIGIDRGERNLLYYSVIDMKGNIVEQDSLNIIRNNDLETDYHDLLDK
REKERKANRQNWEAVEGIKDLKKGYLSQAVHQIAQLMLKYNAIIALEDLGQMF
VTRGQKIEKAVYQQFEKSLVDKLSYLVDKKRPYNELGGILKAYQLASSITKNNS
DKQNGFLFYVPAWNTSKIDPVTGFTDLLRPKAMTIKEAQDFFGAFDNISYNDKG
YFEFETNYDKFKIRMKSAQTRWTICTFGNRIKRKKDKNYWNYEEVELTEEFKKL
FKDSNIDYENCNLKEEIQNKDNRKFFDDLIKLLQLTLQMRNSDDKGNDYIISPVA
NAEGQFFDSRNGDKKLPLDADANGAYNIARKGLWNIRQIKQTKNDKKLNLSISS
TEWLDFVREKPYLK*KRPAATKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPY**
DVPDYA (SEQ ID NO: 244)

17 - Porphyromonas macacae (PmCpf1)

MKTQHFFEDFTSLYSLSKTIRFELKPIGKTLENIKKNGLIRRDEQRLDDYE
KLKKVIDEYHEDFIANILSSFSFSEEILQSYIQNLSESEARAKIEKTMRDTLAKAFSEDERY
KSIFKKELVKKDIPVWCPAYKSLCKKFDNFTTSLVPFHENRKNLYTSNEITASIPYRIVHV
NLPKFIQNIEALCELQKKMGADLYLEMMENLRNVWPSFVKTPDDLCNLKTYNHLMVQS
SISEYNRFVGGYSTEDGTKHQGINEWINIYRQRNKEMRLPGLVFLHKQILAKVDSSSFISD
TLENDDQVFCVLRQFRKLFWNTVSSKEDDAASLKDLFCGLSGYDPEAIYVSDAHLATIS
KNIFDRWNYISDAIRRKTEVLMPRKKESVERYAEKISKQIKKRQSYSLAELDDLLAHYSE
ESLPAGFSLLSYFTSLGGQKYLVSDGEVILYEEGSNIWDEVLIAFRDLQVILDKDFTEKKL
GKDEEAVSVIKKALDSALRLRKFFDLLSGTGAEIRRDSSFYALYTDRMDKLKGLLKMYD
KVRNYLTKKPYSIEKFKLHFDNPSLLSGWDKNKELNNLSVIFRQNGYYYLGIMTPKGKN
LFKTLPKLGAEEMFYEKMEYKQIAEPMLMLPKVFFPKKTKPAFAPDQSVVDIYNKKTFK

TGQKGFNKKDLYRLIDFYKEALTVHEWKLFNFSFSPTEQYRNIGEFFDEVREQAYKVSM
VNVPASYIDEAVENGKLYLFQIYNKDFSPYSKGIPNLHTLYWKALFSEQNQSRVYKLCG
GGELFYRKASLHMQDTTVHPKGISIHKKNLNKKGETSLFNYDLVKDKRFTEDKFFFHVPI
SINYKNKKITNVNQMVRDYIAQNDDLQIIGIDRGERNLLYISRIDTRGNLLEQFSLNVIES
DKGDLRTDYQKILGDREQERLRRRQEWKSIESIKDLKDGYMSQVVHKICNMVVEHKAI
VVLENLNLSFMKGRKKVEKSVYEKFERMLVDKLNYLVVDKKNLSNEPGGLYAAYQLT
NPLFSFEELHRYPQSGILFFVDPWNTSLTDPSTGFVNLLGRINYTNVGDARKFFDRFNAIR
YDGKGNILFDLDLSRFDVRVETQRKLWTLTTFGSRIAKSKKSGKWMVERIENLSLCFLEL
FEQFNIGYRVEKDLKKAILSQDRKEFYVRLIYLFNLMMQIRNSDGEEDYILSPALNEKNL
QFDSRLIEAKDLPVDADANGAYNVARKGLMVVQRIKRGDHESIHRIGRAQWLRYVQEG
IVE*KRPAATKKAGQAKKKK*<u>GS</u>**YPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO: 245)**

Ejemplo 15. Análisis informático de la estructura de Cpf1

Como resultado del análisis informático de la estructura primaria de las nucleasas Cpf1, fue posible observar tres regiones diferentes (figura 109). En primer lugar, hubo un dominio similar a RuvC en el extremo C, que fue el único dominio funcional caracterizado. En segundo lugar, hubo una región con una estructura mixta de hélice alfa y lámina beta en el extremo N, que se encontró cerca del dominio similar a RuvC.

Se predijeron diversos tramos pequeños en regiones no estructuradas dentro de la estructura primaria de Cpf1. Se previeron diversas regiones no estructuradas, que estarían expuestas al solvente, que no estarían conservadas entre los diversos ortólogos de Cpf1 y que se habrían originado como resultado de las divisiones o las inserciones de secuencias de proteínas pequeñas. Más aun, estas partes podrían ser usadas para generar proteínas quiméricas entre ortólogos de Cpf1.

Ejemplo 16. Generación de mutantes con una mayor especificidad por Cpf1

Recientemente se describió un método útil para generar ortólogos de Cas9 con una especificidad mejorada (Slaymaker et al. 2015). Esta estrategia podría usarse para mejorar la especificidad de los ortólogos de Cpf1.

Los residuos para la mutagénesis primaria fueron todos los residuos positivos dentro del dominio RuvC, ya que esta es la única estructura conocida, y aunque carece de una forma cristalina, se sabe que las mutantes con especificidad por RuvC pueden operar sobre Cas9 (véase la tabla a continuación, donde se detallan los residuos de lisina y arginina conservados dentro de RuvC).

Sin que se desee limitar la invención a una teoría, se cree que los residuos con carga positiva en esta región de Cpf1 pueden operar para estabilizar la interacción entre la enzima y el ADN, lo cual puede tener lugar como consecuencia de una interacción con el esqueleto de fosfodiéster con carga negativa de la cadena de ADN que no constituye el diana. Las interacciones con la cadena que no constituye el diana podrían ser interrumpidas como resultado de la sustitución de los residuos con carga positiva de Cpf1. Una alteración suficiente de esta interacción podría ser útil para mantener la actividad apropiada sobre los sitios diana y para disminuir simultáneamente la actividad de la enzima sobre los sitios diferentes del diana (normalmente se esperaría que hubiera interacciones más débiles con la secuencia de guía a causa de una o más faltas de coincidencia con relación a la secuencia del diana).

Otros dominios presentan características similares. Una región de interés es el dominio REC1, que puede presentar, sin limitaciones, una mutación en uno o más residuos de aminoácidos análogos a N497, R661, Q695 y Q926 de SpCas9, por ejemplo, una mutación por alanina en estas posiciones. Las mutaciones en estos residuos también podrían tener como consecuencia la ininterrupción la interacción de la enzima con la cadena principal de fosfato del ADN. Por otra parte, podrían emplearse combinaciones de mutaciones localizadas en los mismos dominios o en dominios diferentes.

Tabla. Residuos de lisina y de arginina conservados en RuvC

AsCpf1	LbCpf1
R912	R833

T923	R836
R947	K847
K949	K879
R951	K881
R955	R883
K965	R887
K968	K897
K1000	K900
R1003	K932
K1009	R935
K1017	K940
K1022	K948
K1029	K953
K1072	K960
K1086	K984
F1103	K1003
R1226	K1017
R1252	R1033
	R1138
	R1165

Otros candidatos fueron los residuos positivos conservados entre los diversos ortólogos que se detallan en la siguiente tabla.

Tabla. Residuos de lisina y de arginina conservados

Residuo	**AsCpf1**	**FnCpf1**	**LbCpf1**	**MbCpf1**
Lys	K15	K15	K15	K14
Arg	R18	R18	R18	R17
Lys/Arg	K26	K26	K26	R25
Lys/Arg	Q34	R34	K34	K33
Arg	R43	R43	R43	M42
Lys	K48	K48	K48	Q47

Lys	K51	K51	K51	K50
Lys/Arg	R56	K56	R56	D55
Lys/Arg	R84	K87	K83	K85
Lys/Arg	K85	K88	K84	N86
Lys/Arg	K87	D90	R86	K88
Arg	N93	K96	K92	K94
Lys/Arg	R103	K106	R102	R104
Lys	N104	K107	K103	K105
Lys	T118	K120	K116	K118
Lys/Arg	K123	Q125	K121	K123
Lys	K134	K143	-	K131
Arg	R176	R186	R158	R174
Lys	K177	K187	E159	K175
Arg	R192	R202	R174	R190
Lys/Arg	K200	K210	R182	R198
Lys	K226	K235	K206	I221
Lys	K273	K296	K251	K267
Lys	K275	K298	K253	Q269
Lys	T291	K314	K269	K285
Lys/Arg	R301	K320	K271	K291
Lys	K307	K326	K278	K297
Lys	K369	K397	P342	K357
Lys	S404	K444	K380	K403
Lys/Arg	V409	K449	R385	K409
Lys	K414	E454	K390	K414
Lys	K436	A483	K415	K448
Lys	K438	E491	K421	K460
Lys	K468	K527	K457	K501
Lys	D482	K541	K471	K515
Lys	K516	K581	A506	K550
Arg	R518	R583	R508	R552
Lys	K524	K589	K514	K558
Lys	K530	K595	K520	K564
Lys	K532	K597	K522	K566

Lys	K548	K613	K538	K582
Lys	K559	K624	Y548	K593
Lys	K570	K635	K560	K604
Lys/Arg	R574	K639	K564	K608
Lys	K592	K656	K580	K623
Lys	D596	K660	K584	K627
Lys	K603	K667	K591	K633
Lys	K607	K671	K595	K637
Lys	K613	K677	K601	E643
Lys	C647	K719	K634	K780
Lys/Arg	R681	K725	K640	Y787
Lys/Arg	K686	K730	R645	K792
Lys	H720	K763	K679	K830
Lys	K739	K782	K689	Q846
Lys	K748	K791	K707	K858
Lys/Arg	K757	R800	T716	K867
Lys/Arg	T766	K809	K725	K876
Lys/Arg	K780	K823	R737	K890
Arg	R790	R833	R747	R900
Lys/Arg	P791	K834	R748	K901
Lys	K796	K839	K753	M906
Lys	K809	K852	K768	K921
Lys	K815	K858	K774	K927
Lys	T816	K859	K775	K928
Lys	K860	K869	K785	K937
Lys/Arg	R862	K871	K787	K939
Arg	R863	R872	R788	R940
Lys	K868	K877	Q793	K945
Lys	K897	K905	K821	Q975
Arg	R909	R918	R833	R987
Arg	R912	R921	R836	R990
Lys	T923	K932	K847	K1001
Lys/Arg	R947	I960	K879	R1034
Lys	K949	K962	K881	I1036

Arg	R951	R964	R883	R1038
Arg	R955	R968	R887	R1042
Lys	K965	K978	K897	K1052
Lys	K968	K981	K900	K1055
Lys	K1000	K1013	K932	K1087
Arg	R1003	R1016	R935	R1090
Lys	K1009	K1021	K940	K1095
Lys	K1017	K1029	K948	N1103
Lys	K1022	K1034	K953	K1108
Lys	K1029	K1041	K960	K1115
Lys	A1053	K1065	K984	K1139
Lys	K1072	K1084	K1003	K1158
Lys/Arg	K1086	K1098	K1017	R1172
Lys/Arg	F1103	K1114	R1033	K1188
Lys	S1209	K1201	K1121	K1276
Arg	R1226	R1218	R1138	R1293
Arg	R1252	R1244	R1165	A1319
Lys	K1273	K1265	K1190	K1340
Lys	K1282	K1274	K1199	K1349
Lys	K1288	K1281	K1208	K1356

En la tabla anterior se proveen las posiciones de los residuos de lisina y de arginina conservados en un alineamiento con una nucleasa Cpf1 de *Francisella novicida* U112 (FnCpf1), de *Acidaminococcus* sp. BV3L6 (AsCpf1), de la bacteria Lachnospiraceae ND2006 (LbCpf1) o de *Moraxella bovoculi* 237 (MbCpf1). Éstos podrían ser usados para generar variantes mutadas de Cpf1 con una especificidad mejorada.

Ejemplo 17. Mejora de la especificidad de la unión a Cpf1

En este contexto, podría usarse una estrategia similar a la que se empleó para mejorar la especificidad de Cas9, por ejemplo, la especificidad de Cpf1 podría ser mejorada mediante la mutación de los residuos que estabilizan la cadena no específica de ADN. Esto podría lograrse sin una estructura cristalina, sobre la base de alineamientos de la estructura lineal que resultaran útiles para predecir (1) qué dominios de Cpf1 podrían unirse a las cadenas del ADN y (2) cuáles residuos en estos dominios podrían estar en contacto con el ADN.

Sin embargo, este abordaje podría ser limitado debido a la conservación pobre que se observa en Cpf1 en comparación con otras proteínas conocidas. Por lo tanto, podría resultar deseable evaluar la función de todos los aminoácidos que pudieran interactuar con el ADN (la lisina, la histidina y la arginina).

Los residuos con cargas positivas en el dominio RuvC están más conservados en todas las proteínas Cpf1 que los del dominio Rad50, a partir de lo cual podría concluirse que los residuos de RuvC han sido menos flexibles en el transcurso de la evolución. Sobre la base de lo anterior, podría llegarse a la conclusión de que es necesario un control rígido de la unión a los ácidos nucleicos en este dominio (con relación al dominio de Rad50). Por lo tanto, es posible que este dominio pueda clivar la cadena de ADN diana al cumplir con las exigencias que suelen observarse en el contexto de la estabilización de las duplas de ARN y ADN (precedente en Cas9). Por otro lado, hay más argininas presentes en el dominio RuvC (los residuos 904 a 1307 representan 5% de RuvC, en comparación con 3,8% en los dominios de Rad50 propuestos), sobre la base de lo cual podría llegarse a la conclusión de que RuvC está dirigido a una de las cadenas

del ADN. Las argininas participan más en la interacción con los ácidos nucleicos principales o secundarios (RoHS Nature, 2009: http://rohslab.cmb.usc.edu/Papers/Rohs_etal_Nature.pdf). Los ácidos nucleicos menos importantes solamente estarían presentes en forma de duplas (en forma de complejos compuestos por ADN y ARN), en función de lo cual puede concluirse que RuvC podría participar en el clivaje.

En las figuras 110, 111 y 112 y se proveen las estructuras cristalinas de dos dominios similares a los encontrados en Cpf1 (abarcan vacaciones con relación a RuvC u otras proteínas que participan en la reparación del ADN, tales como Rad50). Sobre la base de estas estructuras, puede deducirse que los dominios correspondientes serán similares en Cpf1 y puede inferirse que los residuos en las regiones correspondientes podrán ponerse en contacto con el ADN. En cada estructura, se han destacado los residuos que toman contacto con el ADN. En los alineamientos que se representan en la figura 113, se han destacado las regiones de AsCpf1 que corresponden a las regiones en las que tiene lugar la unión al ADN. La lista de residuos en la tabla más adelante abarca los residuos que se hallaron en los dos dominios de unión.

Tabla - Lista de residuos que podrían interactuar con el ADN

Residuos del dominio RuvC que podrían interactuar con el ADN	Residuos del dominio Rad50 que podrían interactuar con el ADN
AsCpf1	AsCpf1
R909	K324
R912	K335
R930	K337
R947	R331
K949	K369
R951	K370
R955	R386
K965	R392
K968	R393
K1000	K400
K1002	K404
R1003	K406
K1009	K408
K1017	K414
K1022	K429
K1029	K436
K1035	K438
K1054	K459
K1072	K460
K1086	K464
R1094	R670
K1095	K675
K1109	R681
K1118	K686
K1142	K689

Residuos del dominio RuvC que podrían interactuar con el ADN	Residuos del dominio Rad50 que podrían interactuar con el ADN
K1150	R699
K1158	K705
K1159	R725
R1220	K729
R1226	K739
R1242	K748
R1252	K752
	R670

A partir de estas observaciones específicas sobre AsCpf1 y los alineamientos con las secuencias, fue posible identificar diversos residuos en Cpf1 que fueron similares a los de otras especies. Se provee un ejemplo de esto en la figura 114, a través de un alineamiento entre AsCpf1 y FnCpf1 con el que fue posible realizar la identificación de los dominios de unión de Rad50 y las argininas y las lisinas en su interior.

Ejemplo 18. Formación de complejos múltiples con Cpf1 y guías en tándem

Se estudió si era posible formar complejos múltiples con la enzima Cpf1. Para este propósito, se desarrolló un ARN de guía con diversas secuencias en tándem bajo un mismo promotor y se determinó su capacidad de modificar blancos específicos en el genoma.

Se sembraron 150000 células HEK293T en 24 cavidades 24 horas antes de la transfección. Las células se transfectaron con 400 ng de un plásmido huAsCpf1 y 100 ng de un plásmido de guía en tándem que comprendía una secuencia de guía dirigida a GRIN28 y una dirigida a EMX1 colocada en tándem detrás del promotor U6 (figura 115A), mediante el uso del reactivo Lipofectamine 2000. Las células se recolectaron 72 horas después de la transfección y se evaluó la actividad de AsCpf1 mediada por las guías en tándem a través del uso de una nucleasa Surveyor.

Los resultados se representan en la figura 115B, donde puede observarse la formación de indels tanto en el gen GRIN28 como en el gen EMX1.

Por consiguiente, se determinó que en AsCpf1, y por analogía, en LbCpf1, pueden emplearse dos guías expresadas a partir de un mismo promotor U6 sin que haya una merma en la actividad. La posición dentro del tándem no tiene ninguna influencia sobre la formación de los indels. Con esto, fue posible demostrar que Cpf1 puede usarse para formar complejos múltiples con dos o más guías.

REIVINDICACIONES

1. Un método *in vitro* o *ex vivo* de direccionamiento de un polinucleótido, que comprende: poner en contacto una muestra que puede comprender el polinucleótido con un complejo de CRISPR-Cas que comprende (a) una proteína Cas de tipo V que comprende un dominio de nucleasa RuvC pero que no comprende un dominio HNH y (b) una guía manipulada que es capaz de dirigir la unión específica de secuencia del complejo a una secuencia diana del polinucleótido, dirigiéndose así o detectando dicho polinucleótido.

2. Un método de direccionamiento de un polinucleótido, que comprende: poner en contacto una muestra que puede comprender el polinucleótido con un complejo CRISPR-Cas que comprende (a) una proteína Cas de tipo V que comprende un dominio de nucleasa RuvC pero que no comprende un dominio HNH y (b) una guía manipulada que es capaz de dirigir la unión específica de secuencia del complejo a una secuencia diana del polinucleótido, dirigiéndose así al polinucleótido, en donde el método no es un método de modificación de la identidad genética de la línea germinal de seres humanos.

3. El método de cualquiera de las reivindicaciones precedentes, en donde se modifica un producto génico o se modifica la cantidad o expresión de un producto génico.

4. El método de cualquiera de las reivindicaciones precedentes, que comprende además detectar la unión del complejo a la secuencia diana.

5. El método de cualquier reivindicación precedente, en donde la proteína Cas está asociada con uno o más dominios funcionales o comprende al menos una mutación.

6. El método de cualquier reivindicación precedente, que comprende además la introducción de rotura de cadena, inactivación génica específica de sitio, edición de bases específica de sitio o edición del genoma específica de sitio.

7. El método de cualquier reivindicación precedente, en donde dicha secuencia diana es ARN.

8. El método de cualquier reivindicación precedente, en donde dicha secuencia diana es ADN.

9. El método de la reivindicación 7 u 8, en donde dicha secuencia diana está asociada a enfermedad.

10. Una composición que comprende: una proteína Cas de tipo V que comprende un dominio de nucleasa de RuvC pero que no comprende un dominio HNH o un polinucleótido que codifica la proteína, y un ARN guía manipulado que es capaz de formar un complejo CRISPR-Cas con la proteína Cas y de dirigir la unión específica de secuencia del complejo CRISPR-Cas a una secuencia diana o un polinucleótido que codifica el ARN guía.

11. La composición de la reivindicación 10, en donde la proteína Cas está asociada con uno o más dominios funcionales o al menos una mutación, y para su uso en una célula eucariota.

12. Un método *in vitro* o *ex vivo* de direccionamiento de un polinucleótido, que comprende: poner en contacto una muestra que puede comprender el polinucleótido con la composición según la reivindicación 10, en donde el ARN guía manipulado es capaz de formar un complejo CRISPR-Cas con la proteína Cas y dirigir la unión específica de secuencia del complejo CRISPR-Cas a una secuencia diana del polinucleótido.

13. Un método de direccionamiento de un polinucleótido, que comprende: poner en contacto una muestra que puede comprender el polinucleótido con la composición según la reivindicación 10, en donde el ARN guía manipulado es capaz de formar un complejo CRISPR-Cas con la proteína Cas y dirigir la unión específica de secuencia del complejo CRISPR-Cas a una secuencia diana del polinucleótido, en donde el método no es un método de modificación de la identidad genética de la línea germinal de seres humanos.

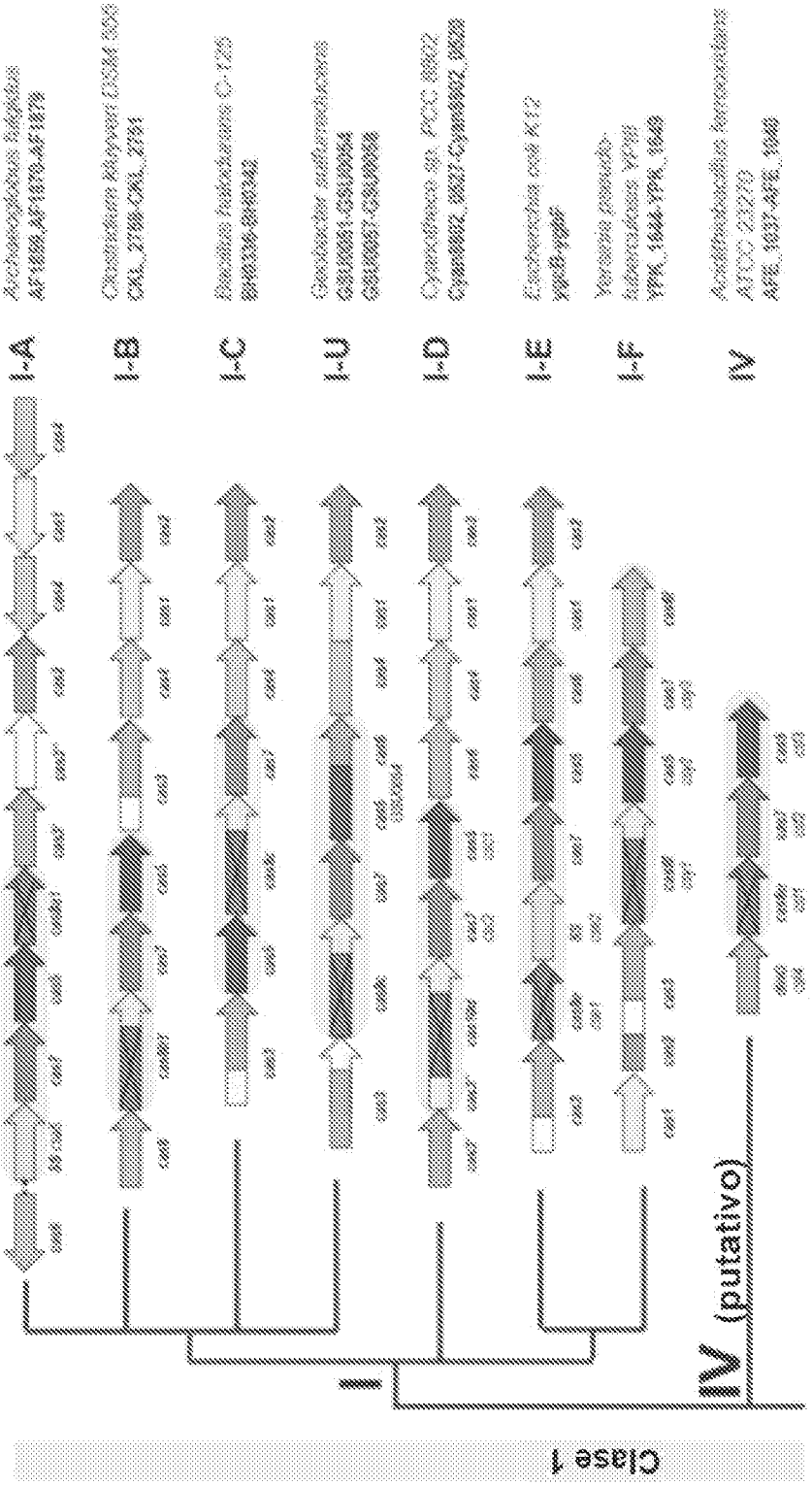


FIG. 1A

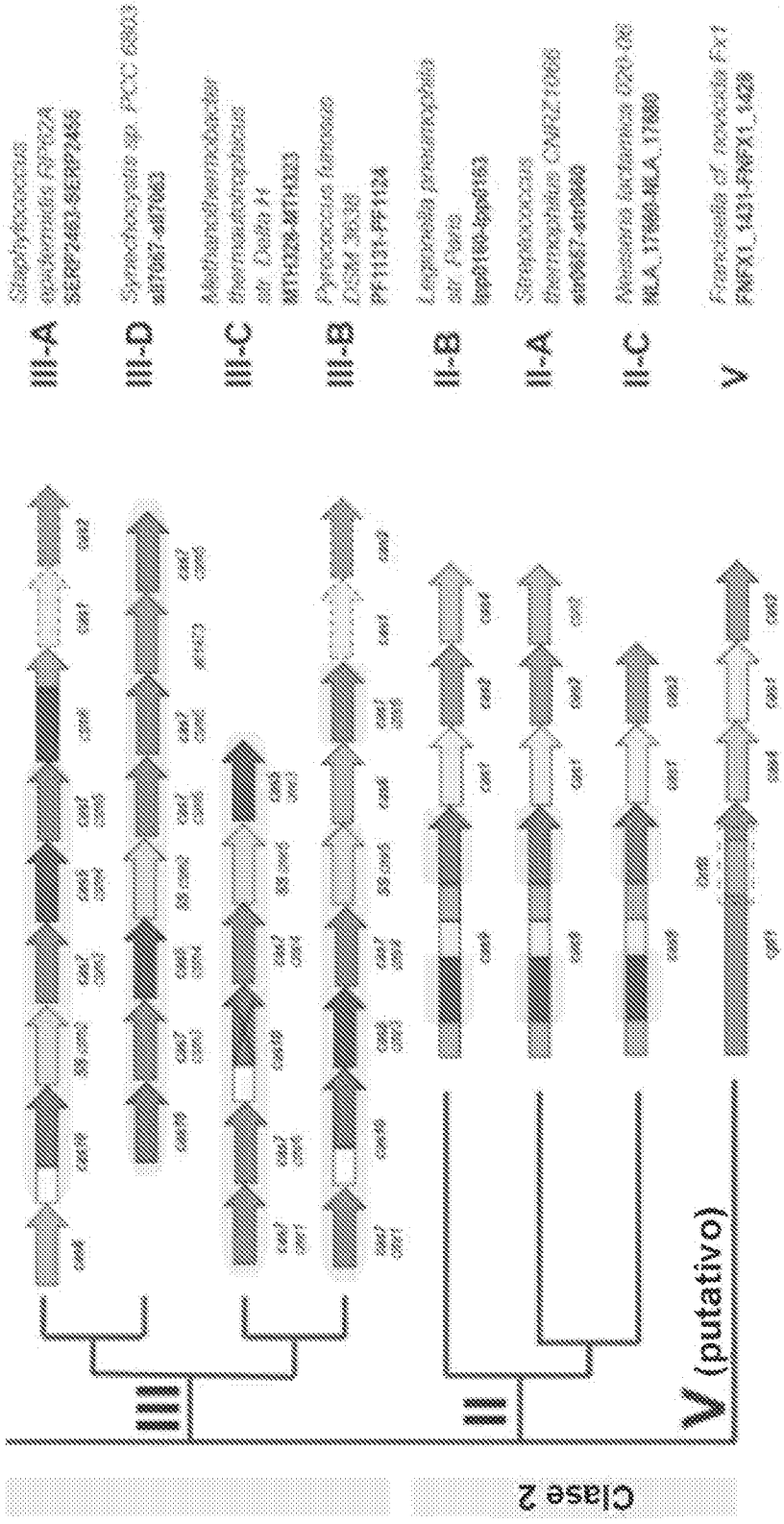


FIG. 1B

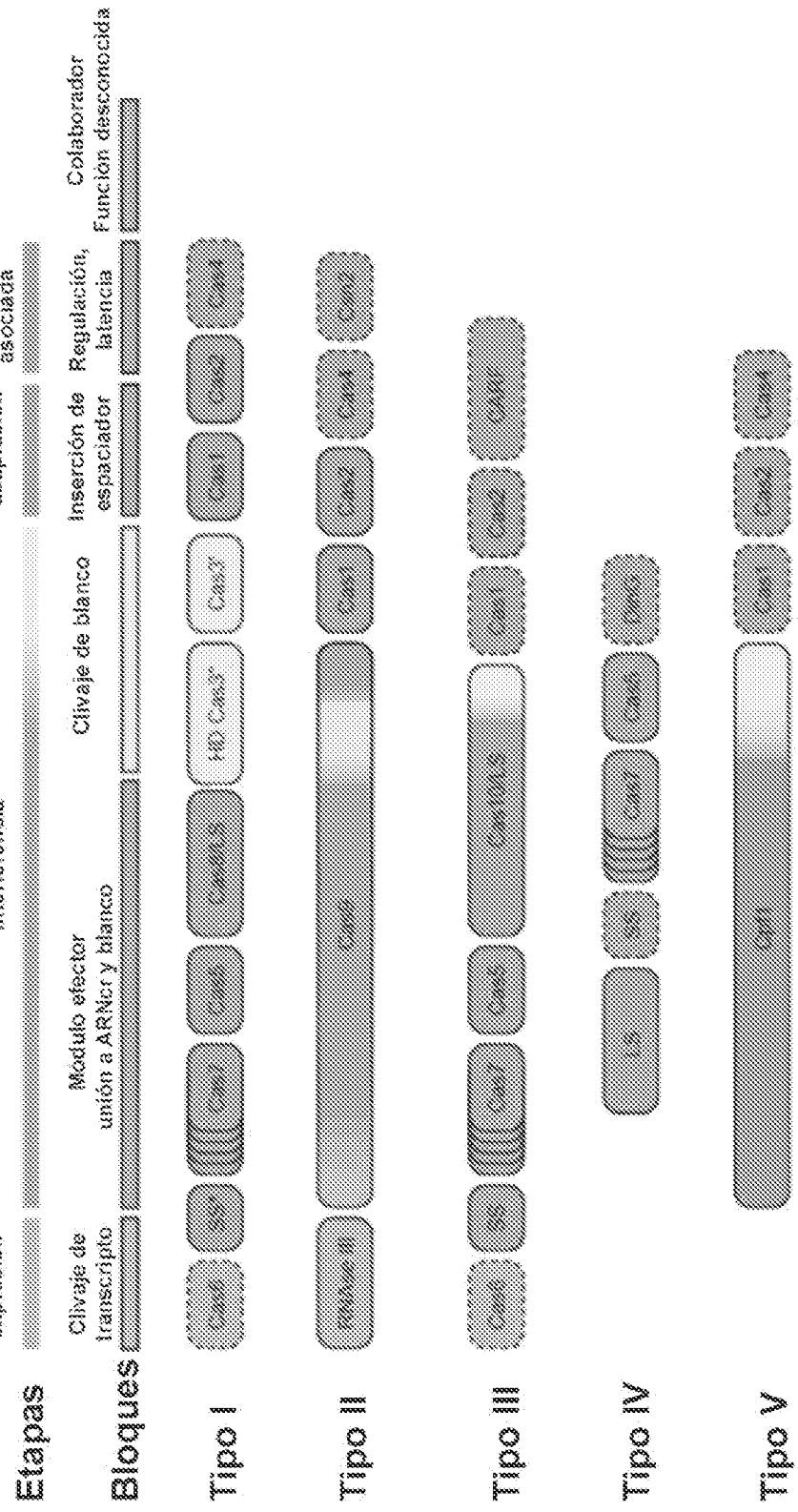

Etapas
expresión
interferencia
adaptación
inmunidad asociada
Bloques
Clivaje de transcripto
Módulo efector unión a ARNcr y blanco
Clivaje de blanco
Inserción de espaciador
Regulación, latencia
Colaborador Función desconocida
Tipo I
Tipo II
Tipo III
Tipo IV
Tipo V

FIG. 2

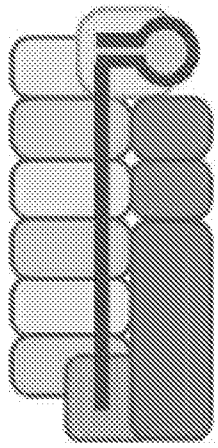


FIG. 3A

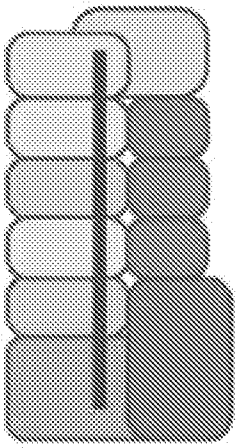


FIG. 3B

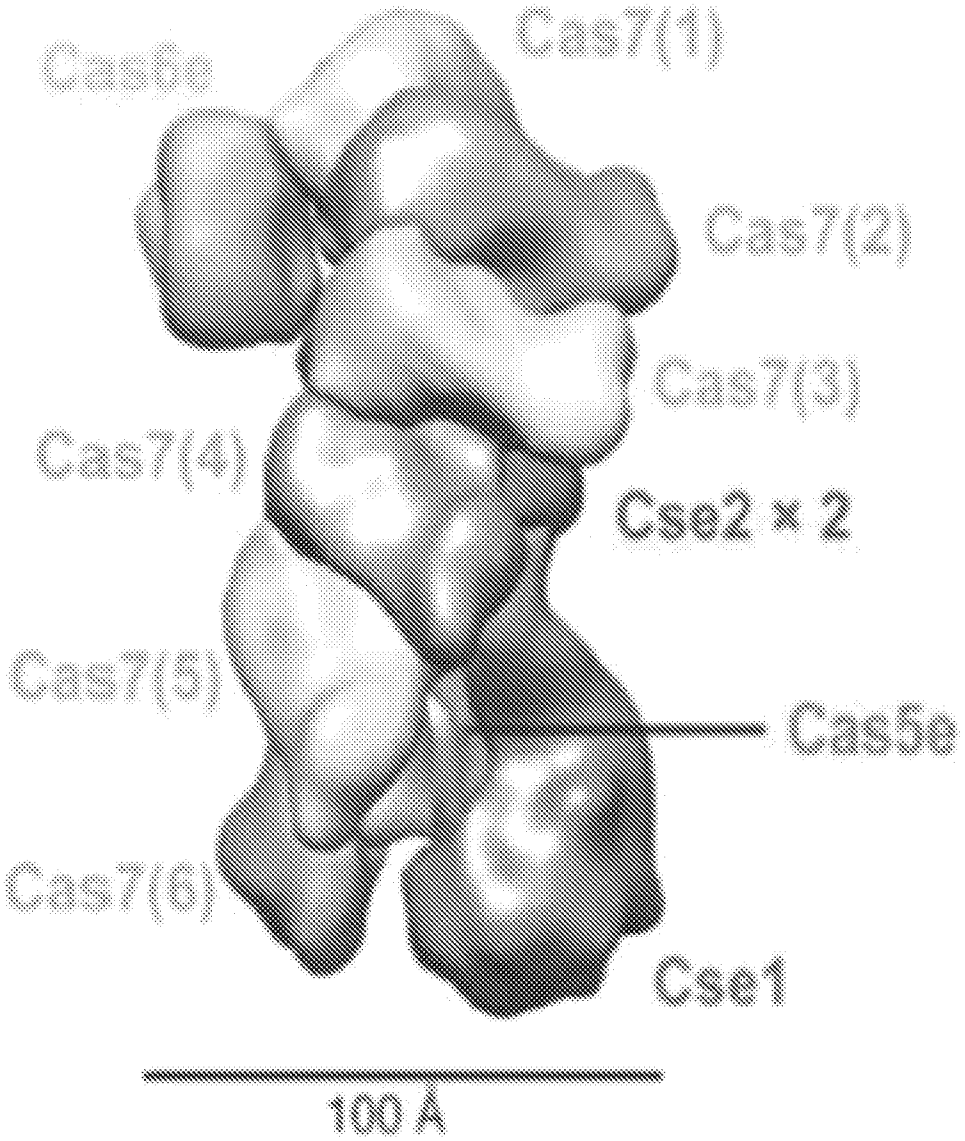

Cas6e
Cas7(1)
Cas7(2)
Cas7(3)
Cas7(4)
Cse2 × 2
Cas7(5)
Cas5e
Cas7(6)
Cse1
100 Å

FIG. 3C

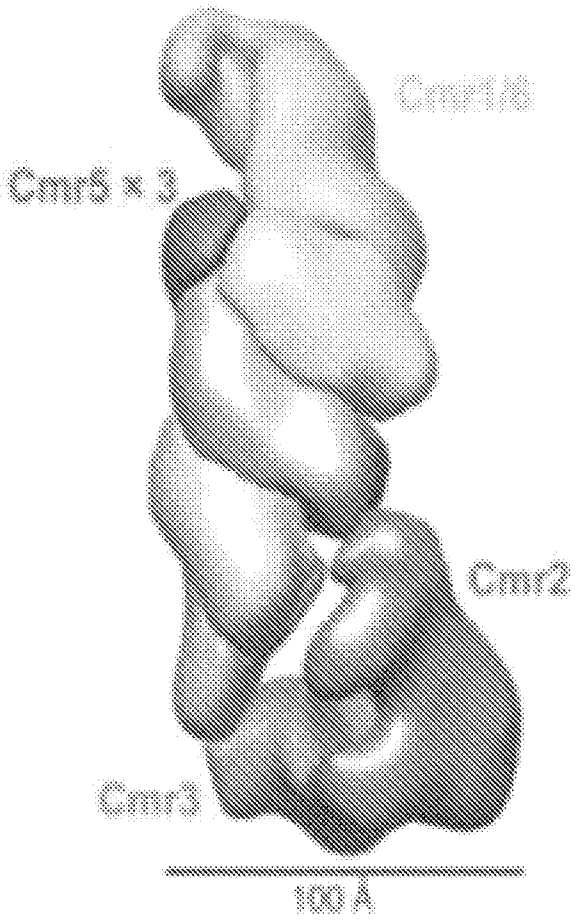
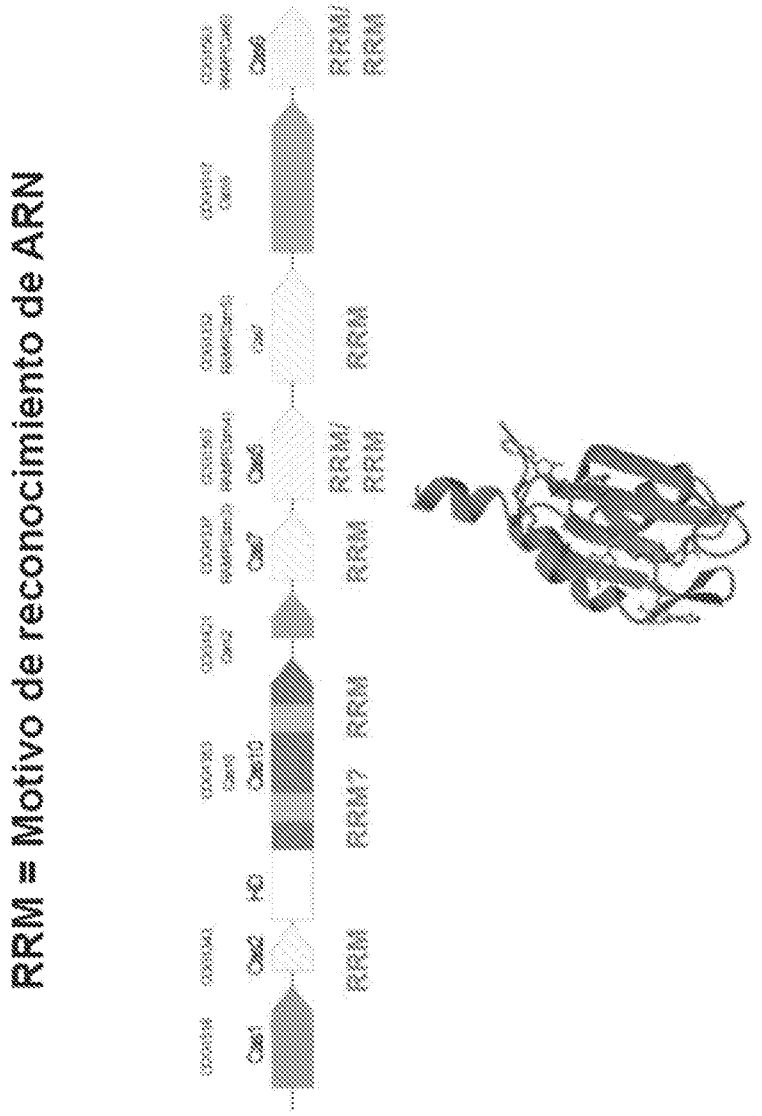


FIG. 3D


RRM = Motivo de reconocimiento de ARN
Dominio
Dominio
Dominio

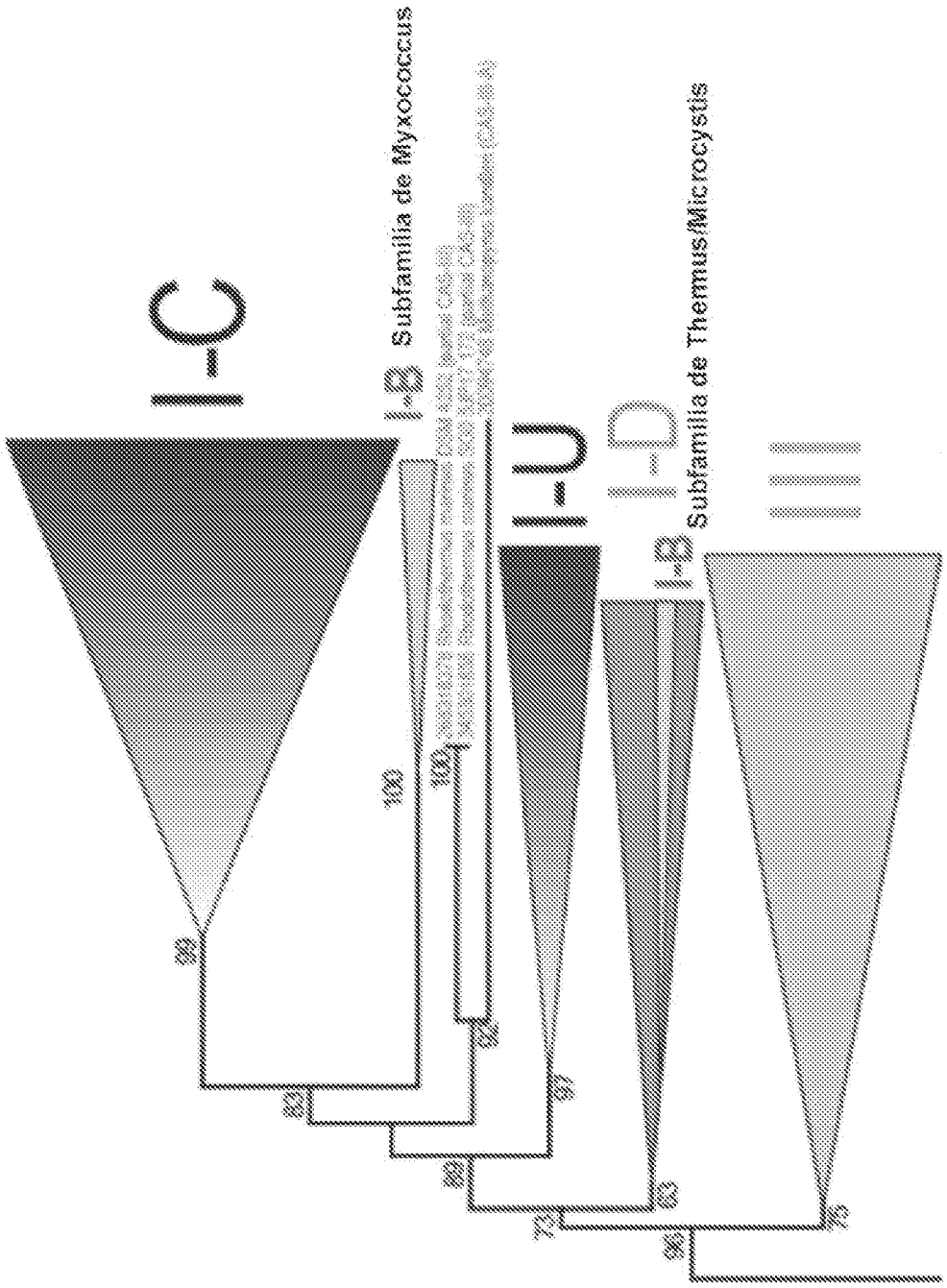

I-C
I-B
Subfamilia de Myxococcus
I-A
I-U
I-D
I-B
Subfamilia de Thermus/Microcystis
III
100
100
82
89
97
74
63
90
76

FIG. 5A

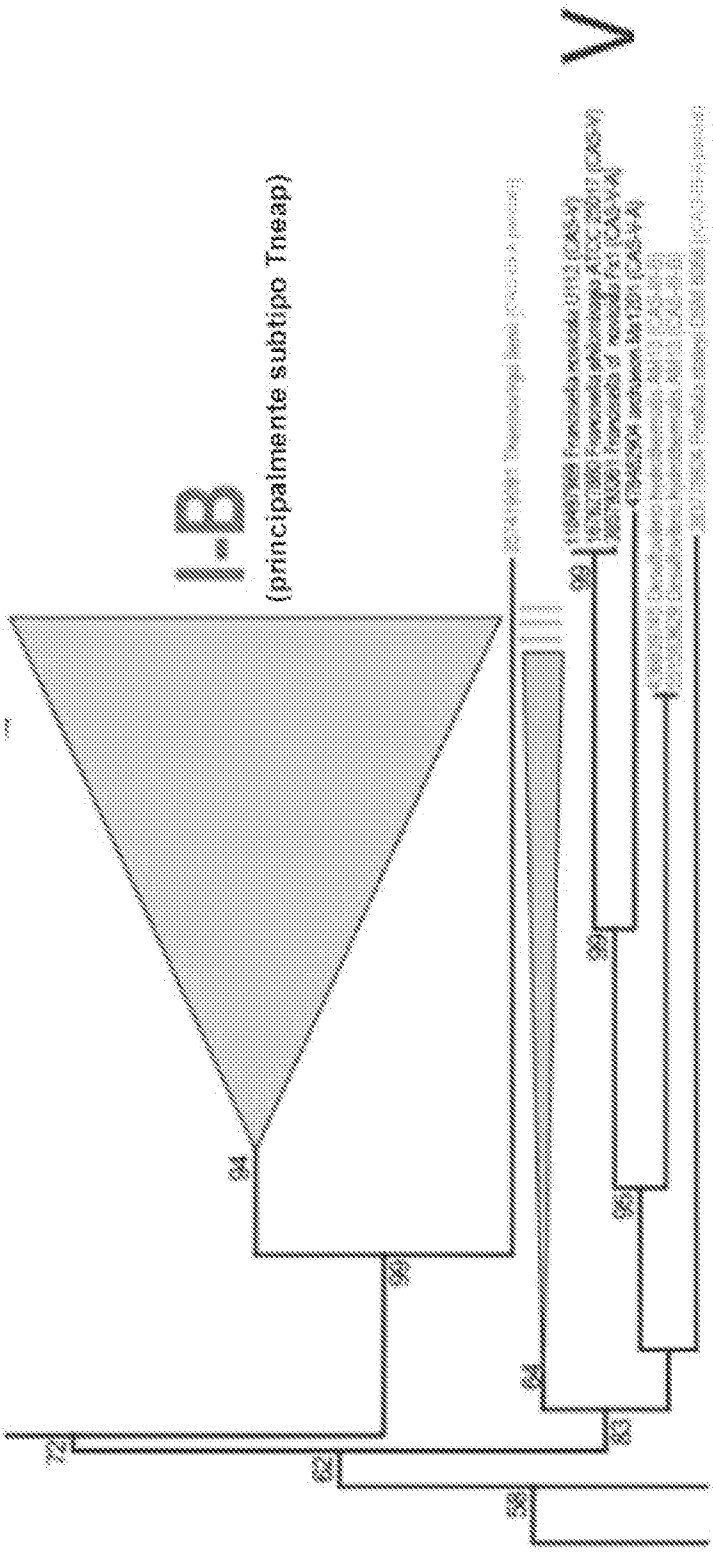


FIG. 5B

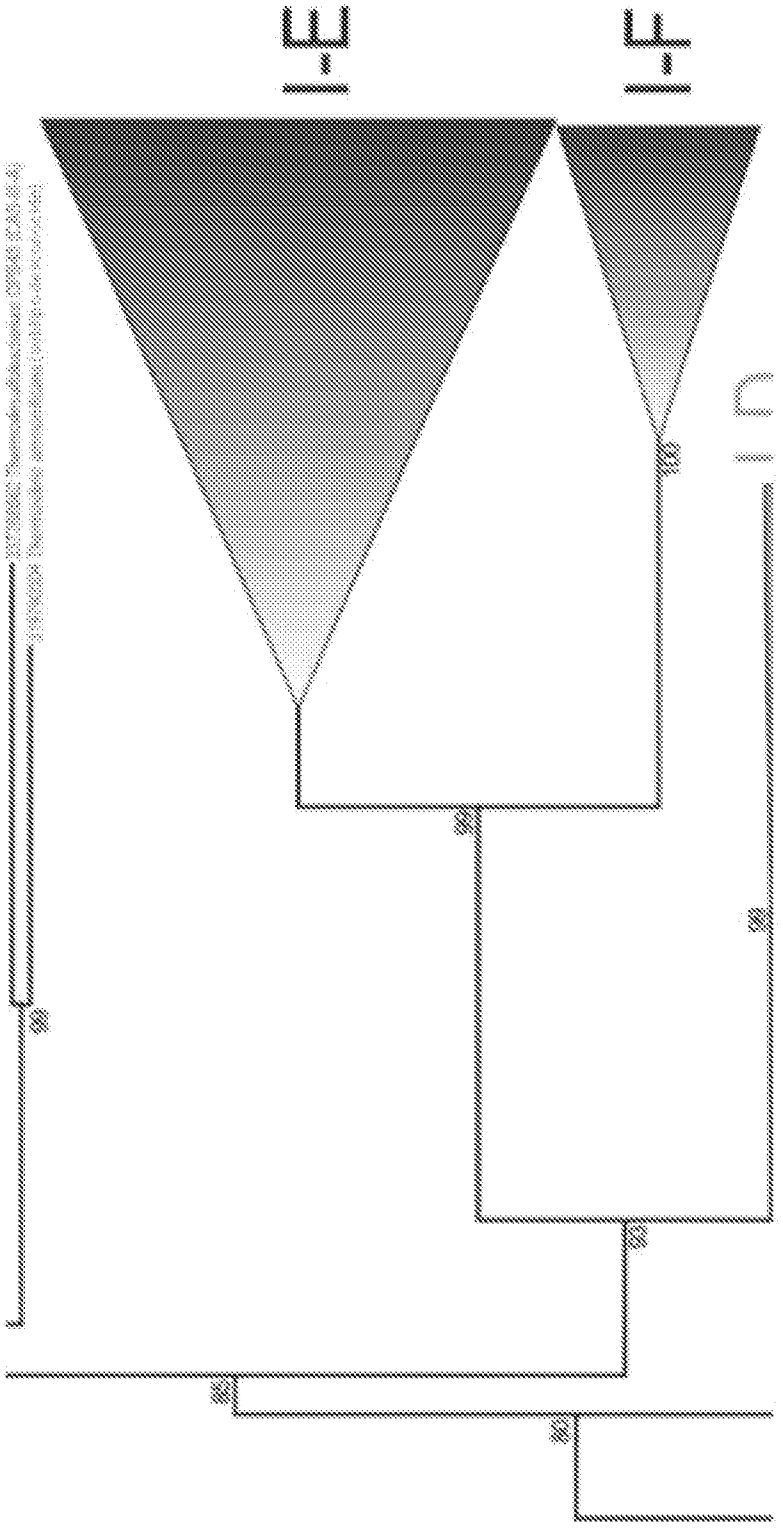


FIG. 5C

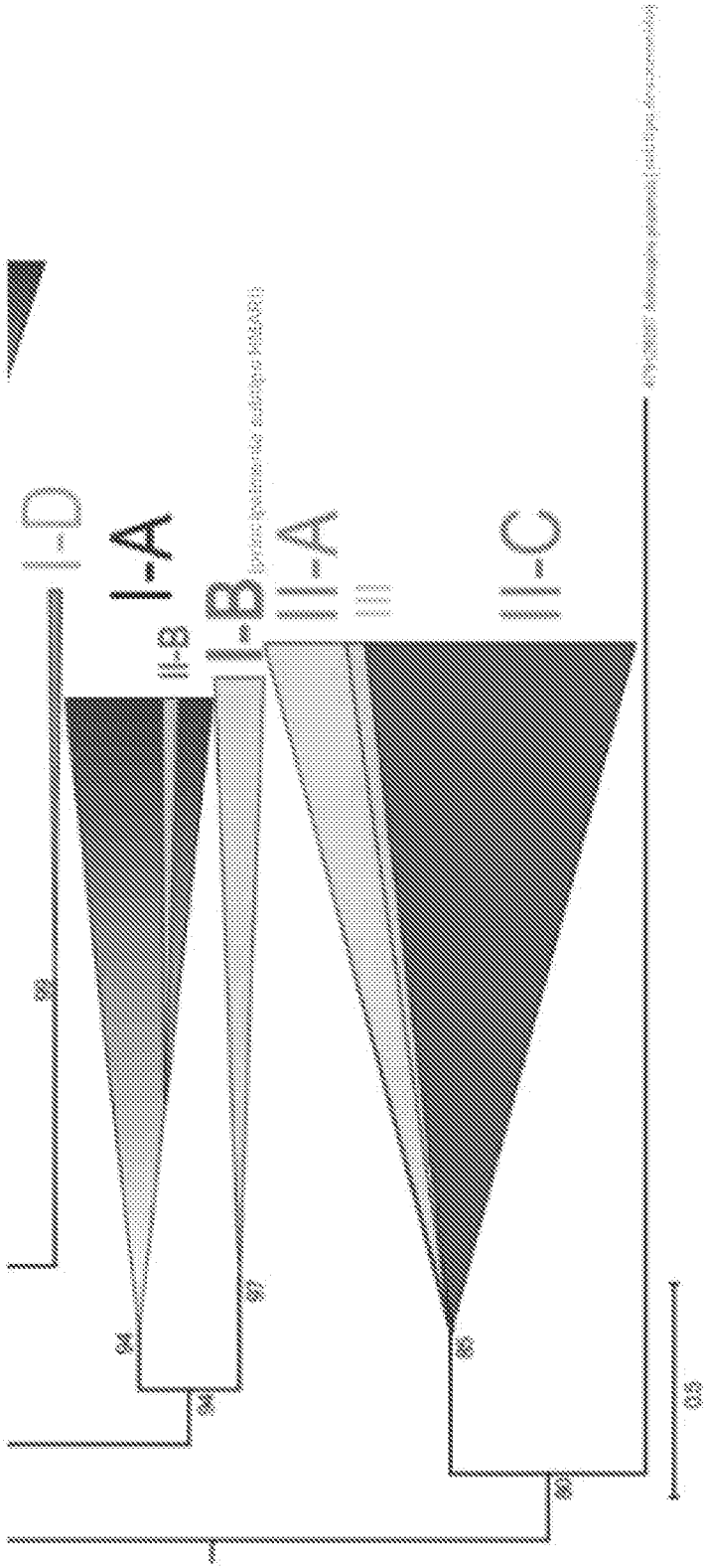
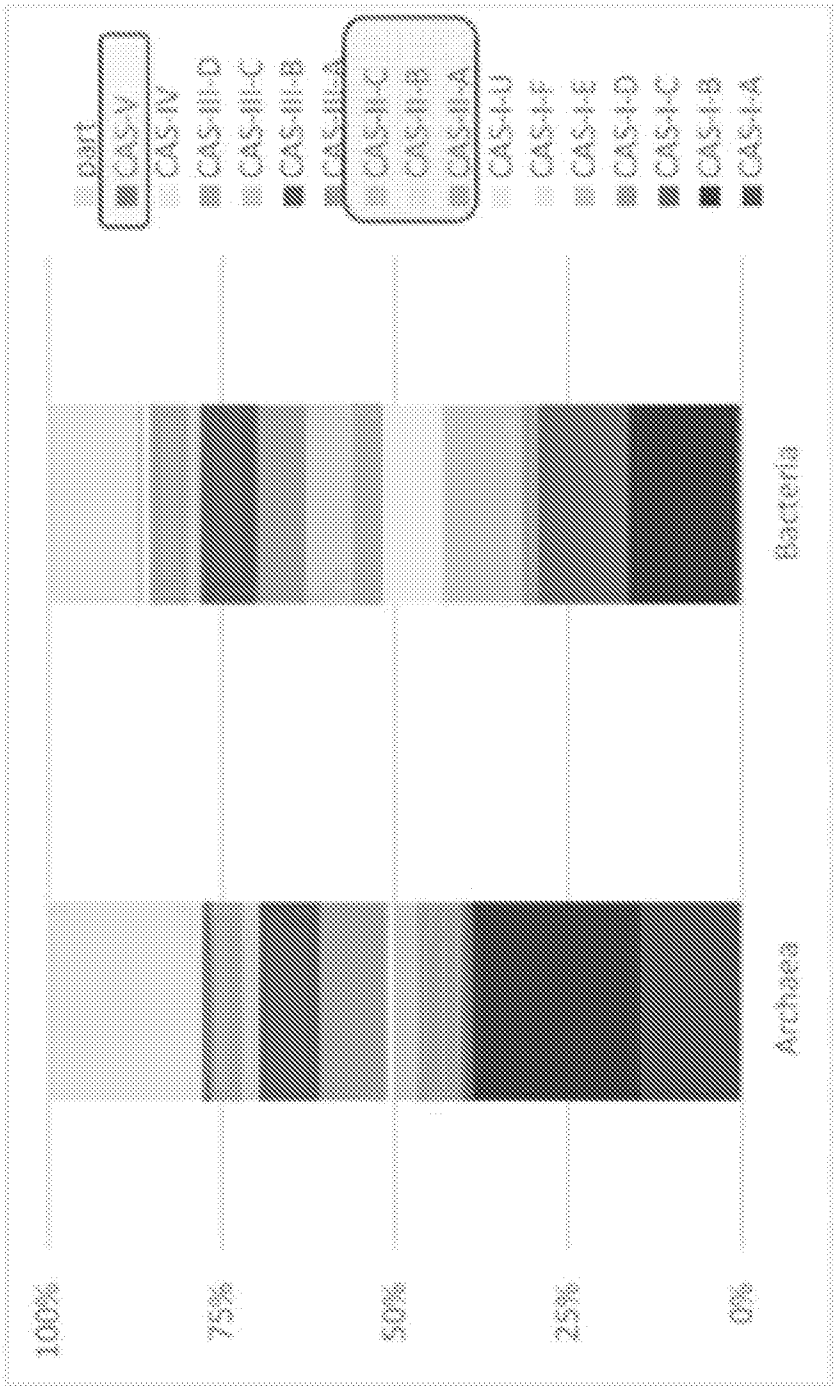


FIG. 5D



Tipo II: específico de bacterias, aproximadamente en 10% de bacterias
Tipo V: raro – varias bacterias, una arqueo

FIG. 6

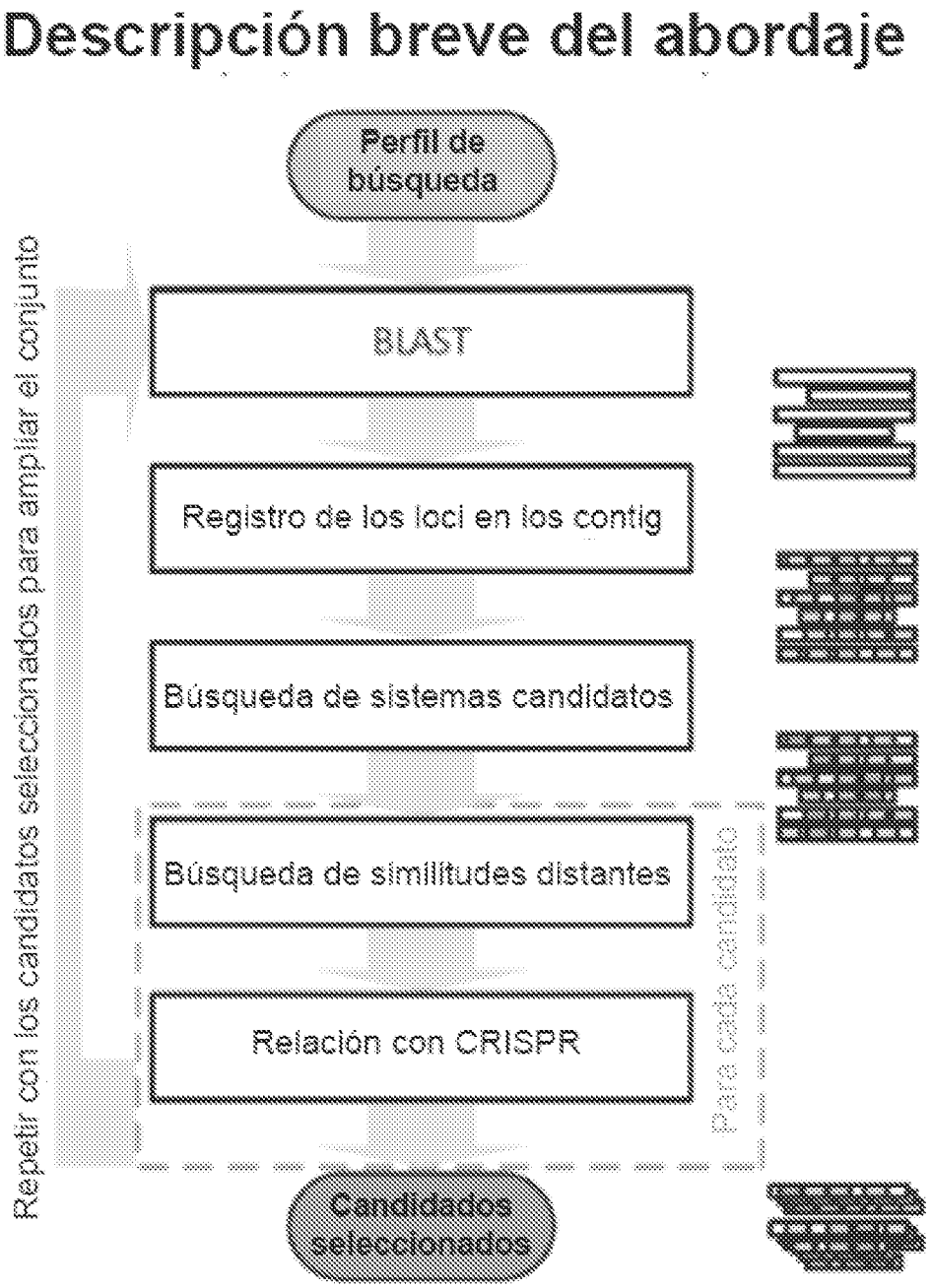

Descripción breve del abordaje
Perfil de búsqueda
BLAST
Registro de los loci en los contig
Búsqueda de sistemas candidatos
Búsqueda de similitudes distantes
Relación con CRISPR
Para cada candidato
Repetir con los candidatos seleccionados para ampliar el conjunto
Candidatos seleccionados

FIG. 7

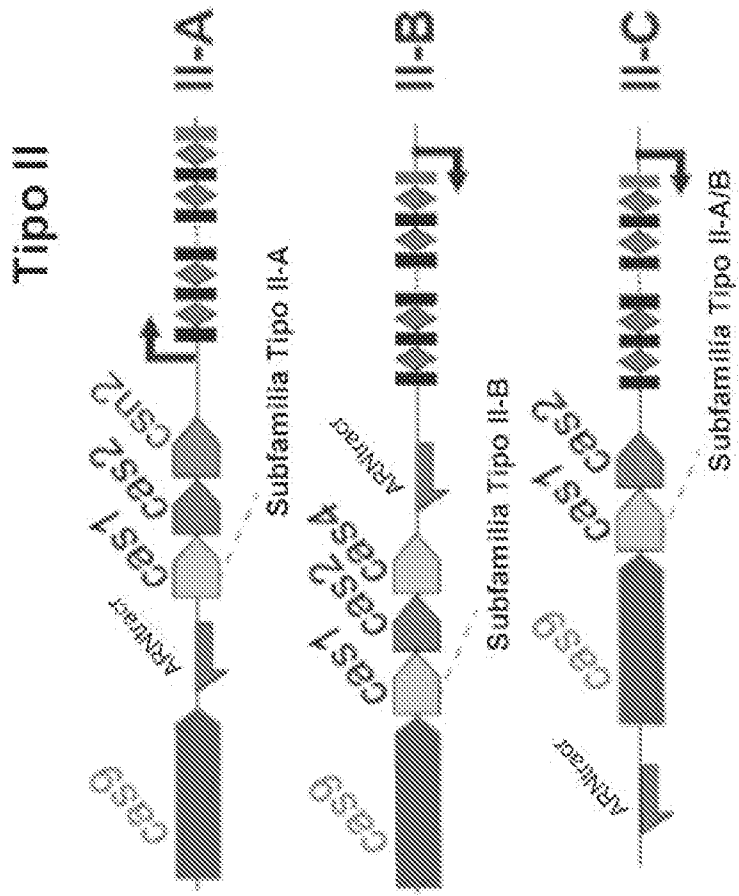

Tipo II
II-A
Subfamilia Tipo II-A
cas9
cas1
cas2
csn2
ARNtracr
II-B
Subfamilia Tipo II-B
cas9
cas1
cas2
csn2
cas4
ARNtracr
II-C
Subfamilia Tipo II-A/B
cas9
cas1
cas2
ARNtracr
Ampliamente distribuido en las bacterias, ausente en las arqueas, raro en las termófilas

FIG. 8A


PreFan/Tipo V-A? subtipo
Sin conexión con cas1
Subfamilia Tipo V
cas2
cas1
cas4
cas6
cas7
cas8
cas10
cas5
cas3
cpf1

FIG. 8B

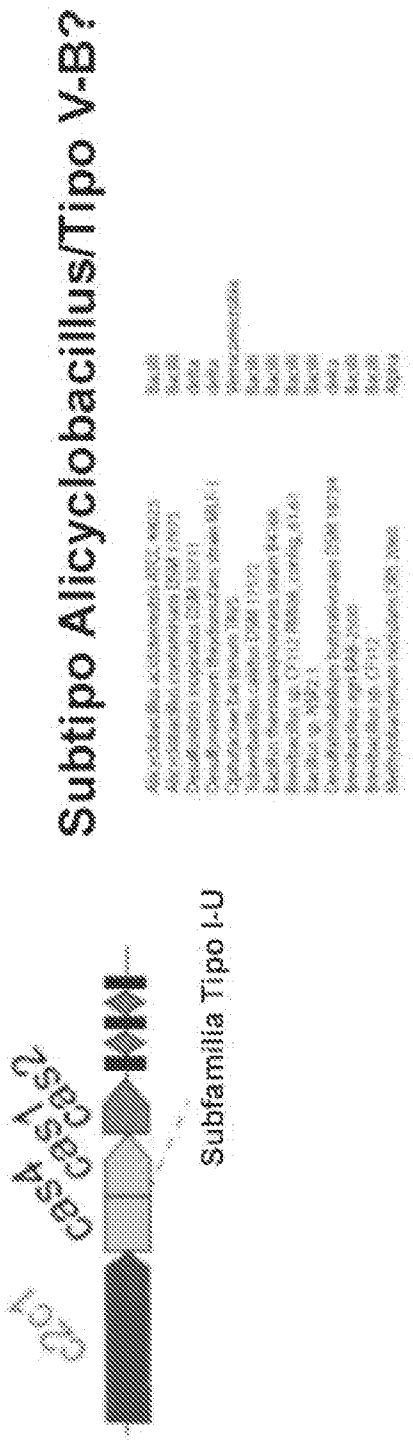


FIG. 8C

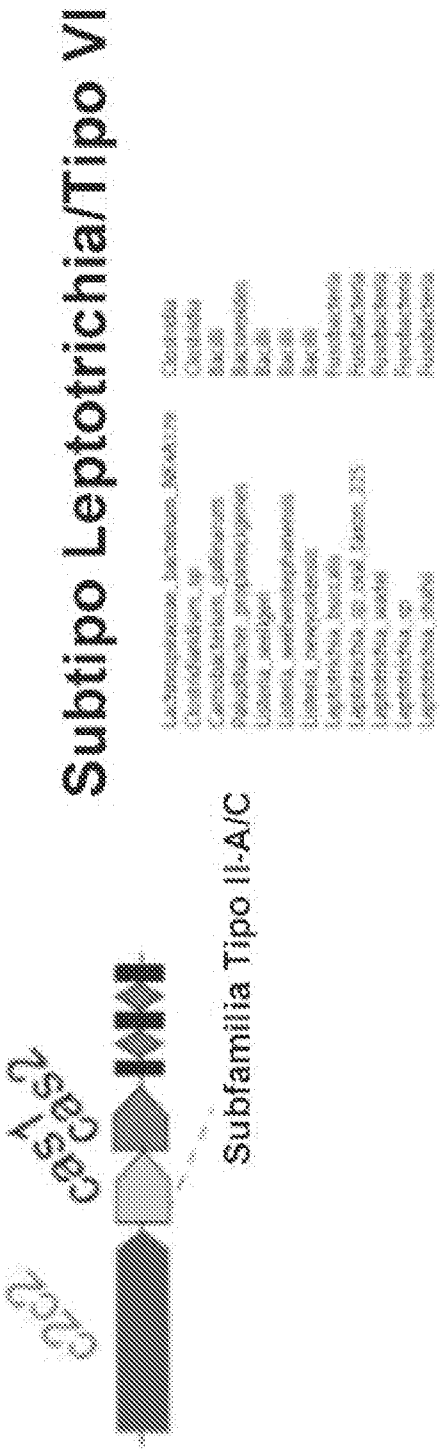


FIG. 8D

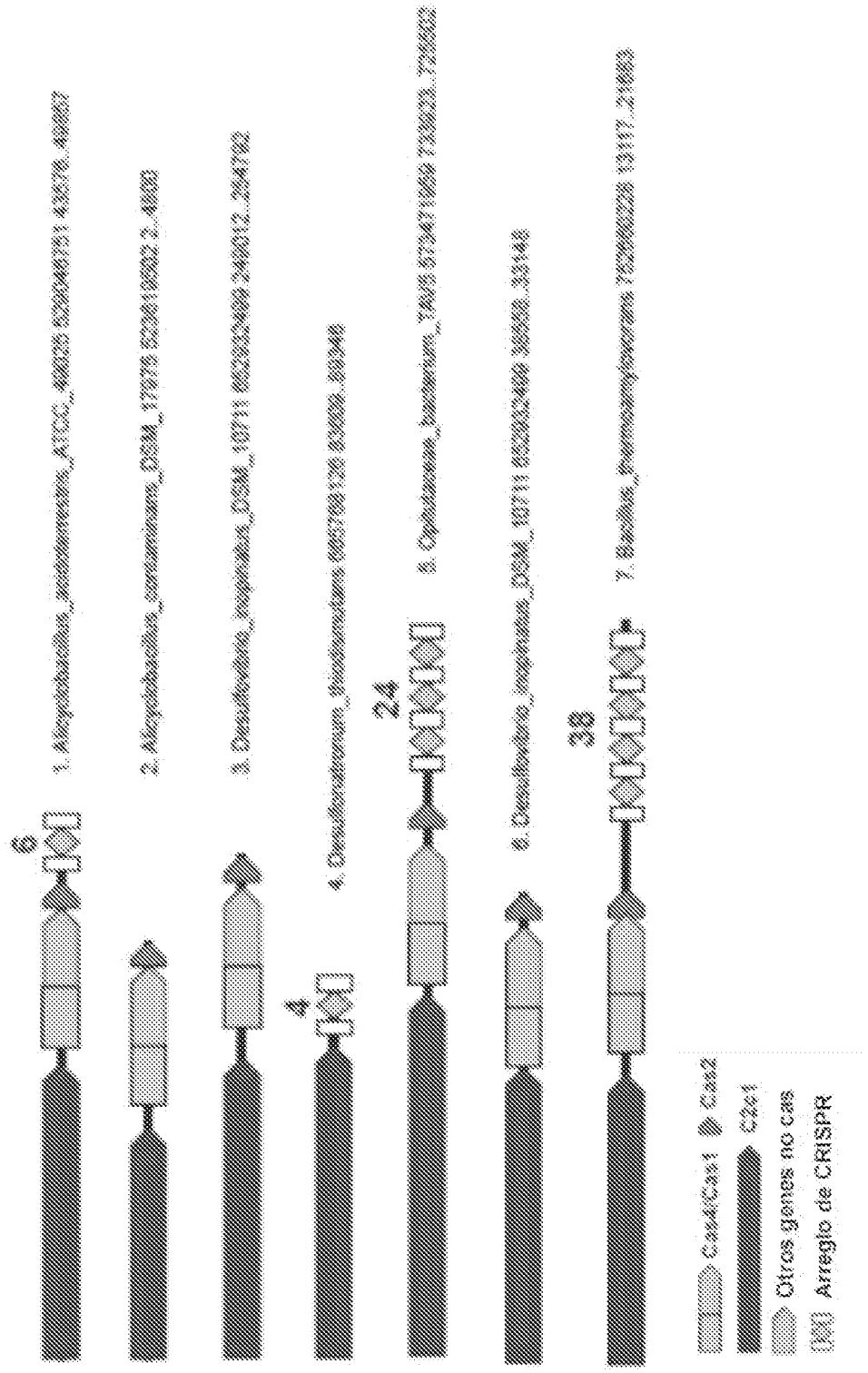

6
1. Alicyclobacillus_acidocaldarius_ATCC_ 40026 62004781 63078..64697
2. Alicyclobacillus_acidocaldarius_DSM_ 17975 62061602 1..4008
3. Desulfovibrio_hospitalis_DSM_ 10711 65303469 245012..254702
4
4. Desulfurobacterium_thermolithotrophum 66578128 63609..68046
24
5. Opitutaceae_bacterium_TAV5 67041669 732623..772622
38
6. Desulfovibrio_hospitalis_DSM_ 10711 65303469 30528..33146
7. Bacillus_thermoamylovorans 75208228 13117..21693
Cas4/Cas1
Cas2
C2c1
Otros genes no cas
Arreglo de CRISPR

FIG. 9A

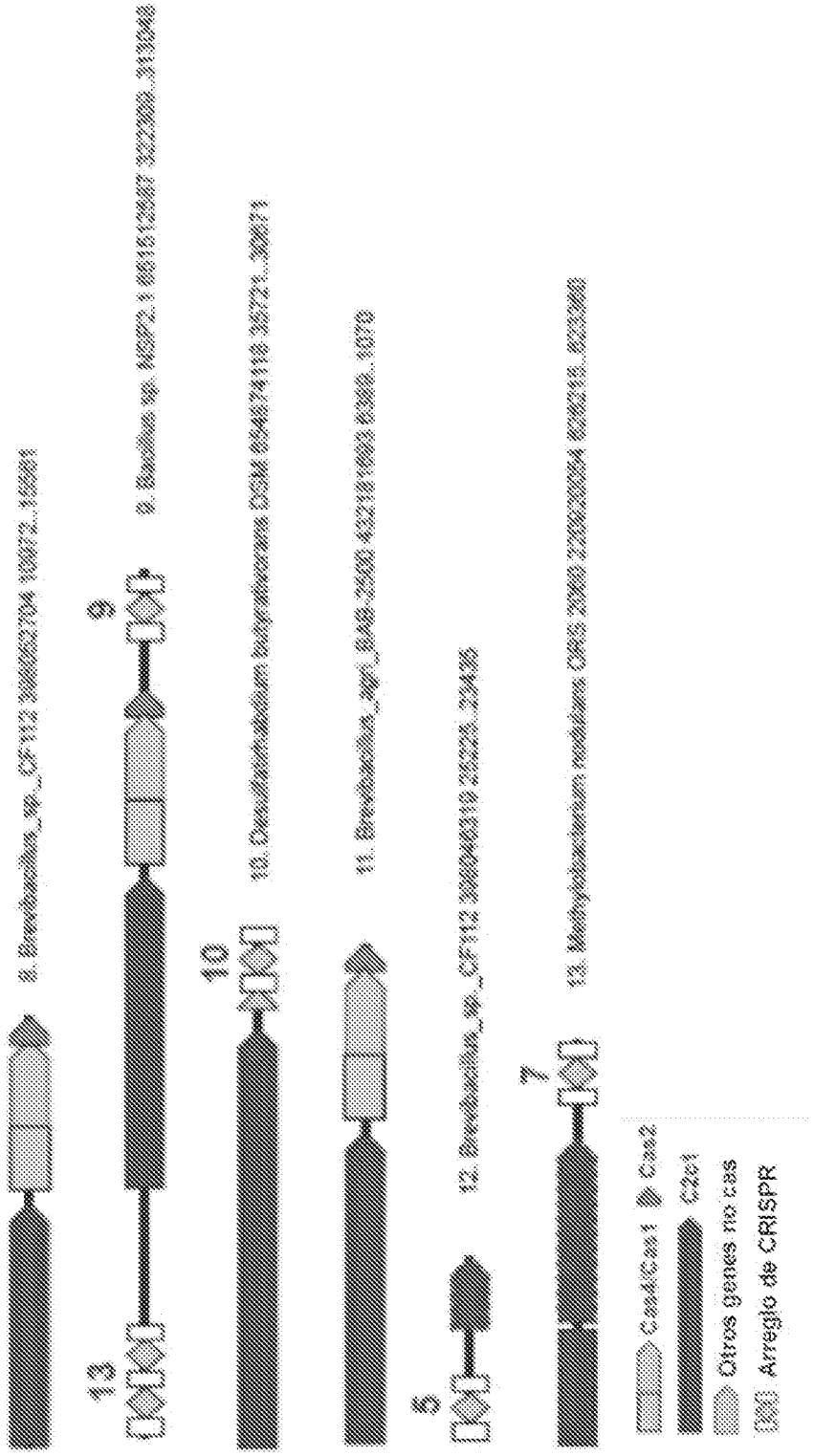

13
9
10
7
5
Cas4/Cas1
Cas2
C2c1
Otros genes no cas
Arreglo de CRISPR

FIG. 9B

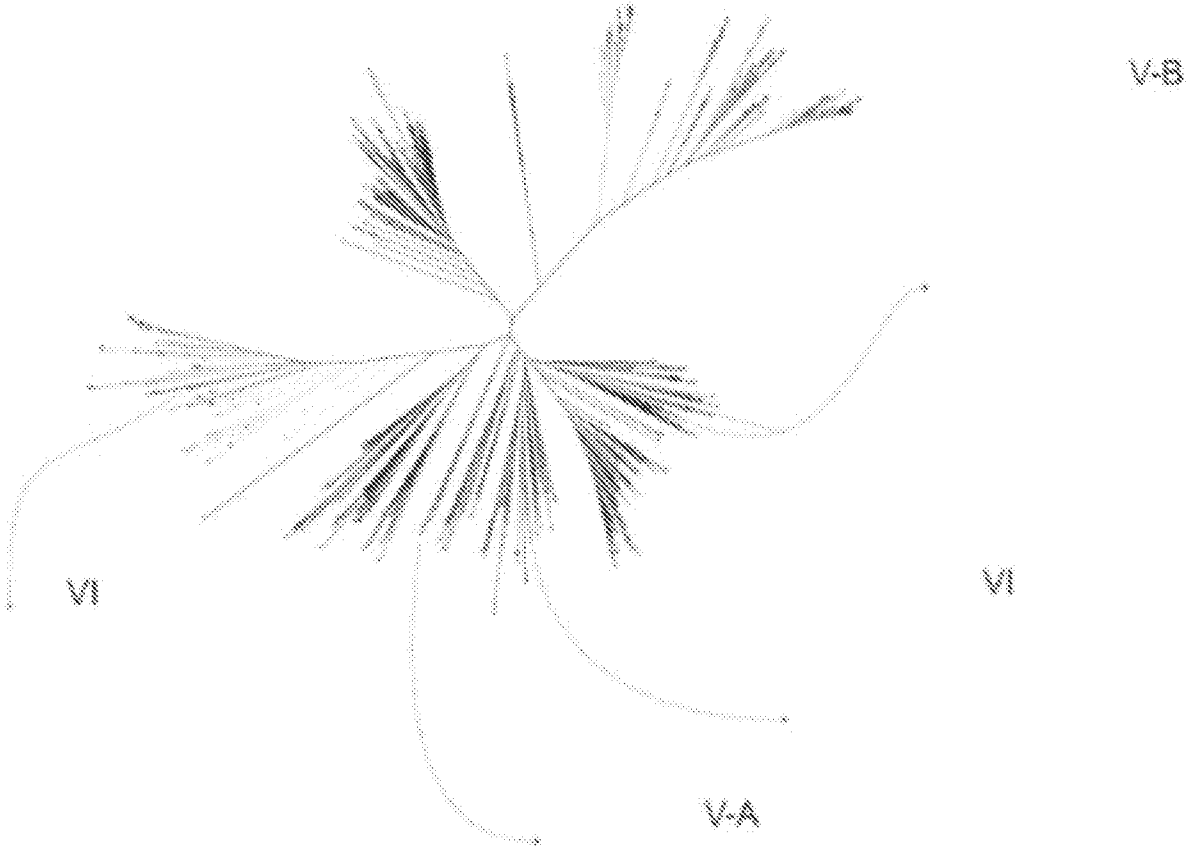


FIG. 10A

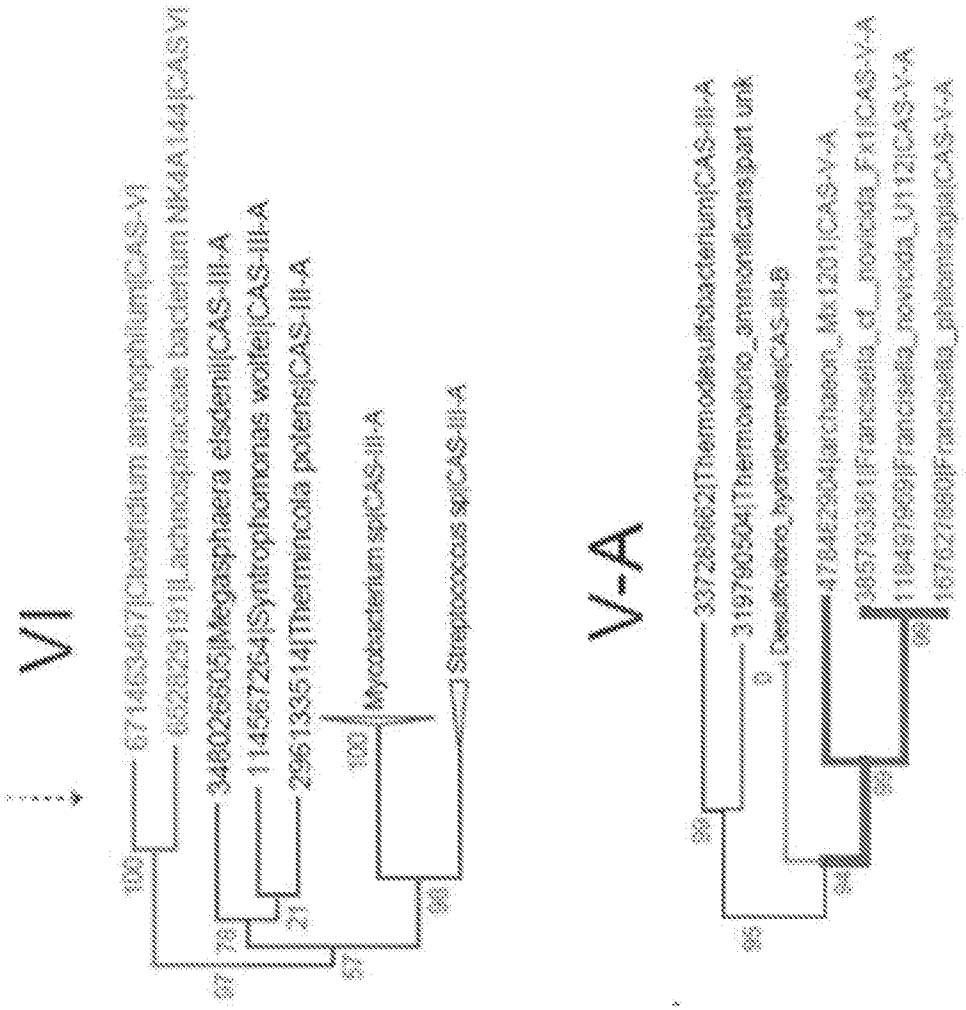

VI
100
67
71463467|Candidatus ammoniphilum|CAS-VI
60
60282910|Lachnospiraceae bacterium NK4A144|CAS-VI
78
34802660|Megasphaera elsdenii|CAS-III-A
21
114267204|Syntrophomonas wolfei|CAS-III-A
57
296133514|Thermincola potens|CAS-III-A
100
Mycobacterium sp|CAS-III-A
100
Streptococcus sp|CAS-III-A
V-A
90
337280662|Thermodesulfobacterium|CAS-III-A
86
319760004|Thermodesulfovibrio_yellowstonii_unk
0
Desulfurivibrio_hydrocarbonoxidans|CAS-III-B
94
47846220|Marichromatium_bh1|201|CAS-V-A
100
99
2067103|Francisella_cf_novicida_Fx1|CAS-V-A
1184178|Francisella_novicida_U112|CAS-V-A
100
1070718|Francisella_philomiragia|CAS-V-A

FIG. 10B

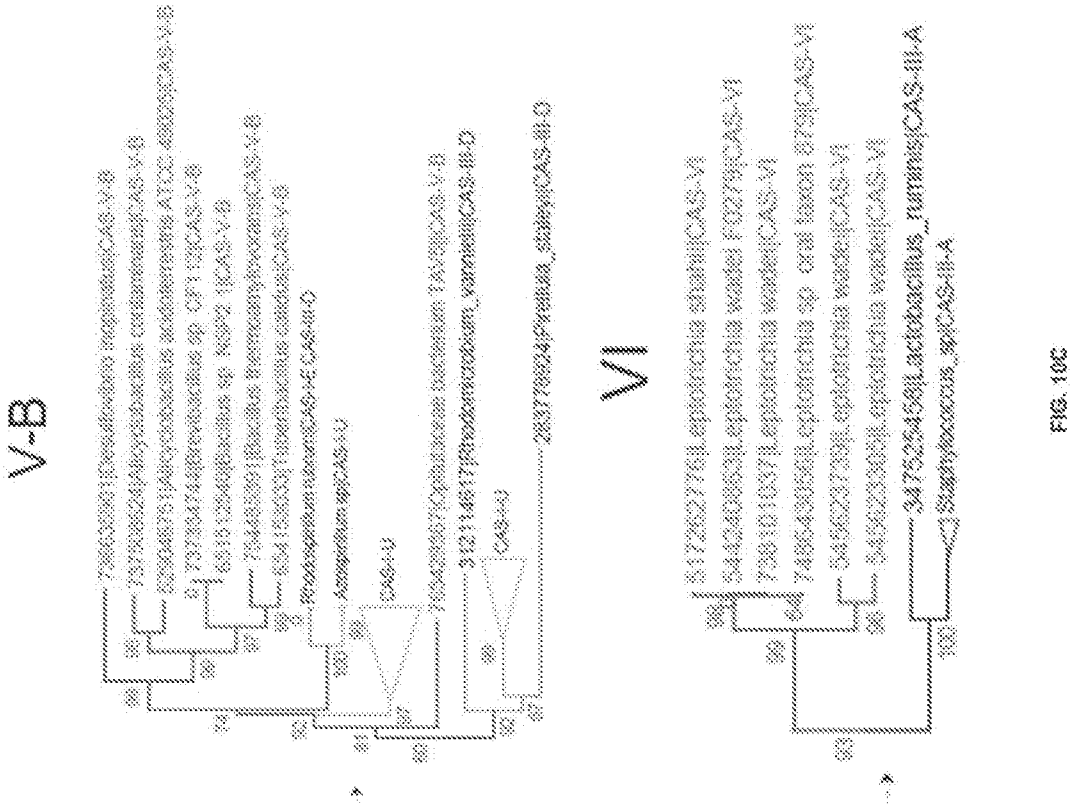


FIG. 10C

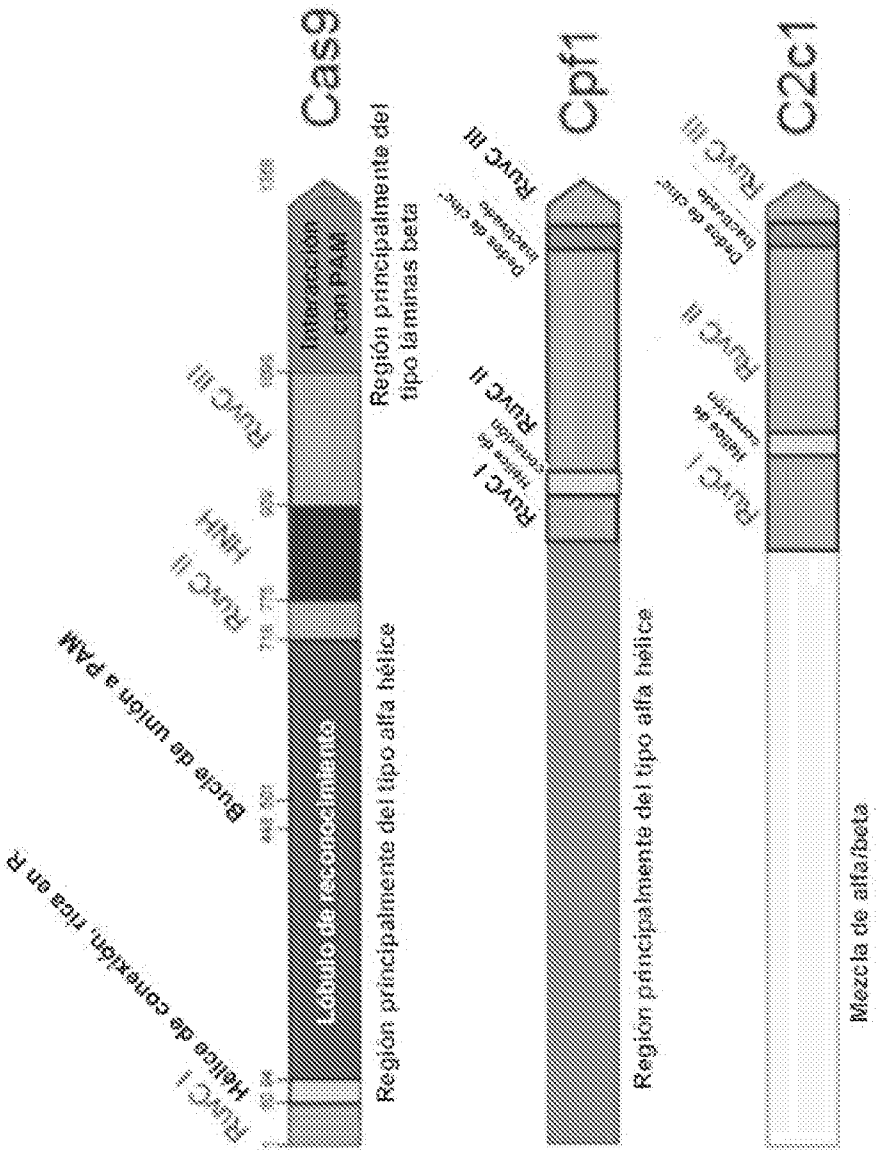
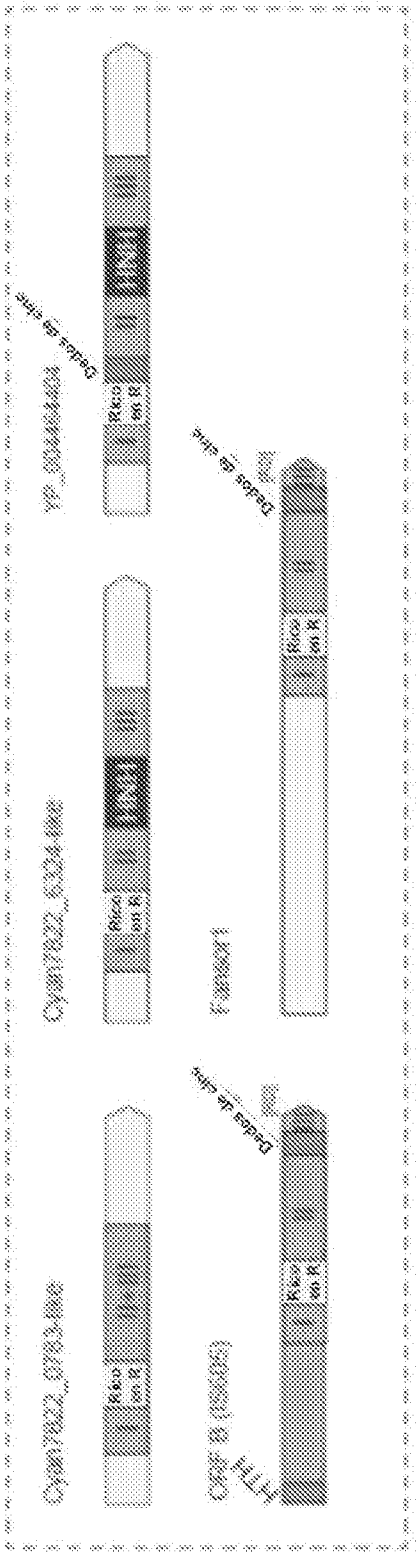

Cas9
RuvC I
RuvC II
RuvC III
HNH
Dominio de reconocimiento
Región principalmente del tipo alfa hélice
Región principalmente del tipo láminas beta
Hélice de conexión, rica en R
Bucle de unión a PAM
Cpf1
Region principalmente del tipo alfa hélice
C2c1
Mezcla de alfa/beta

FIG. 11A


Cpa1782_6784-like
Cpa1782_6784-like
Cpa1782_6784-like

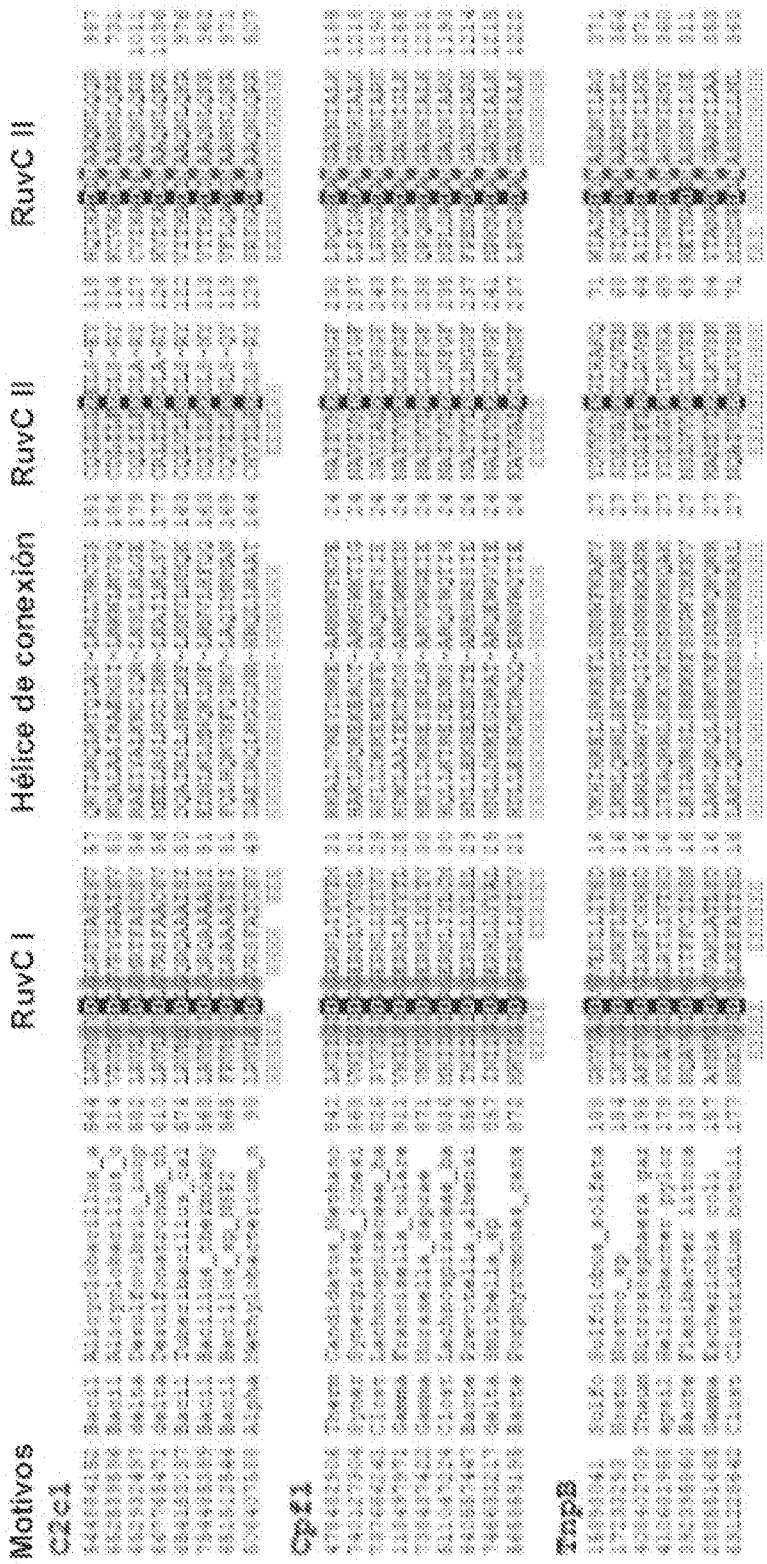

Motivos
RuvC I
Hélice de conexión
RuvC II
RuvC III
C2c1
Cpf1
TnpB

FIG. 12A

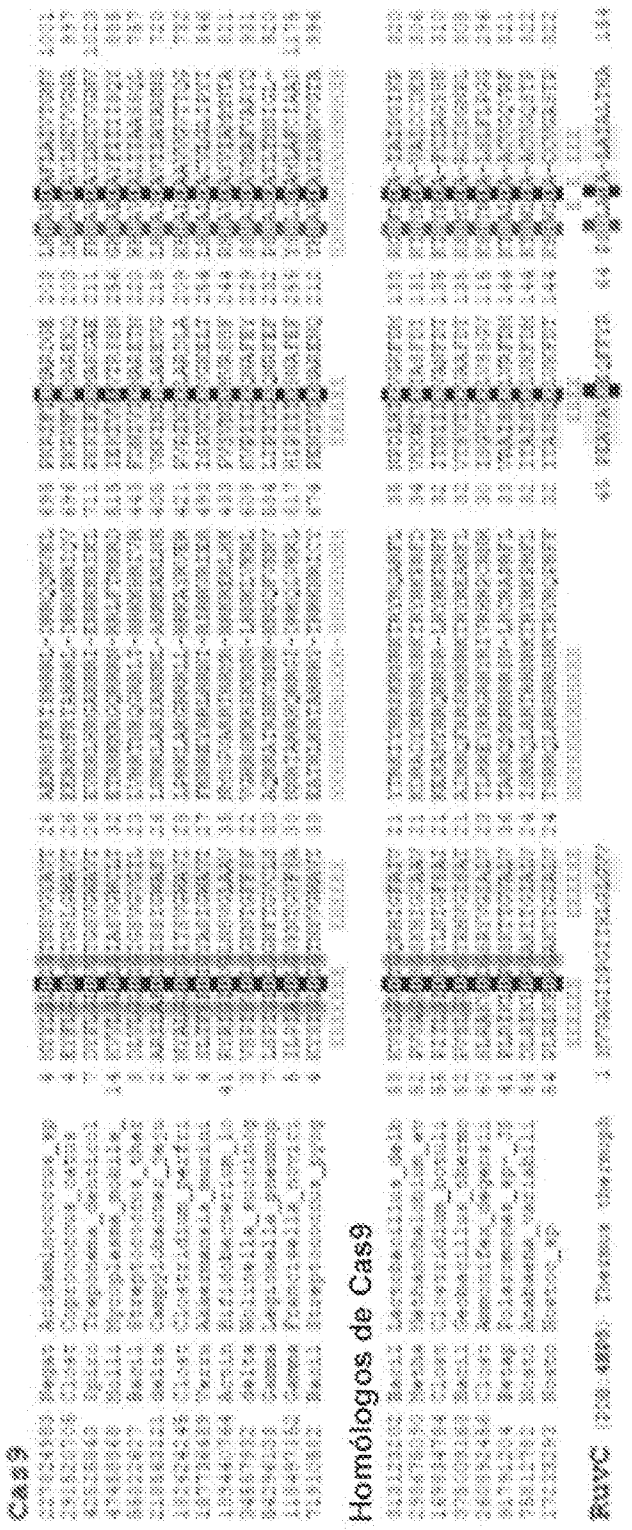


FIG. 12B

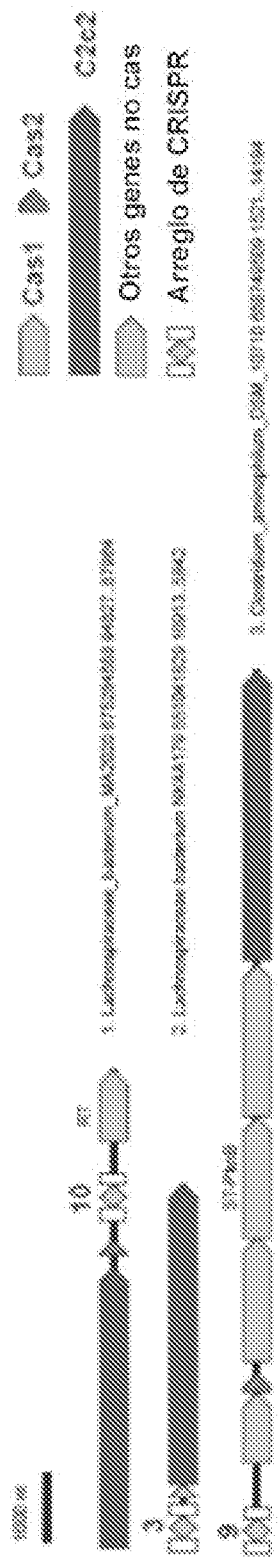


FIG. 13A

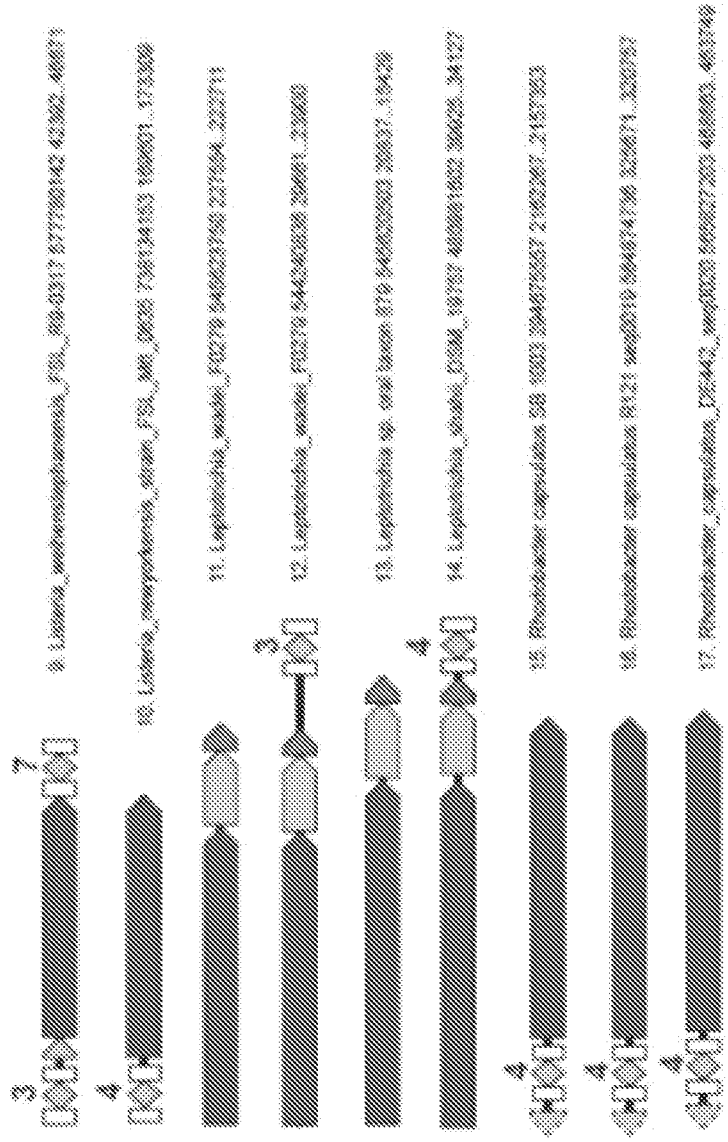


FIG. 13B

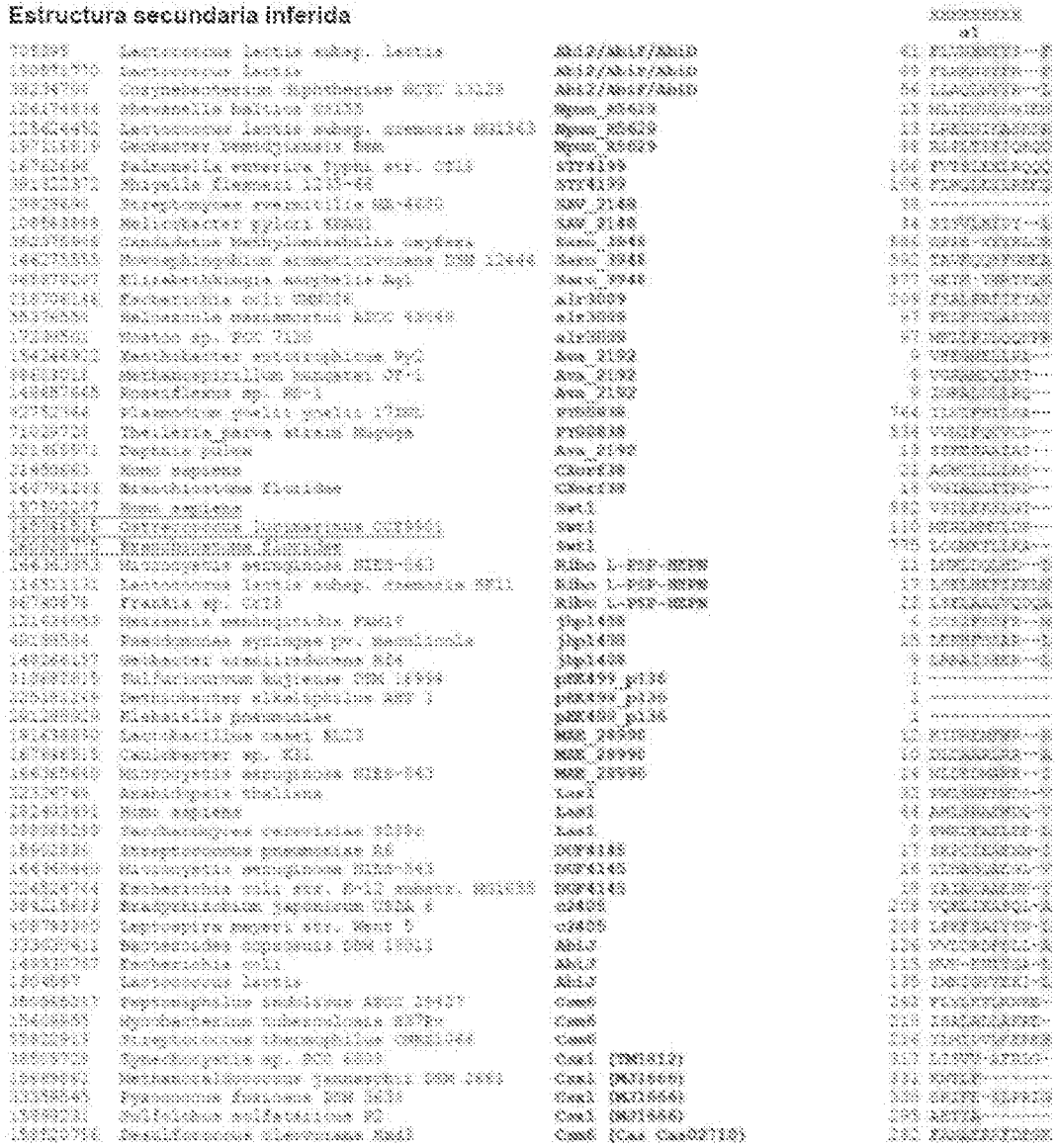


FIG. 14A

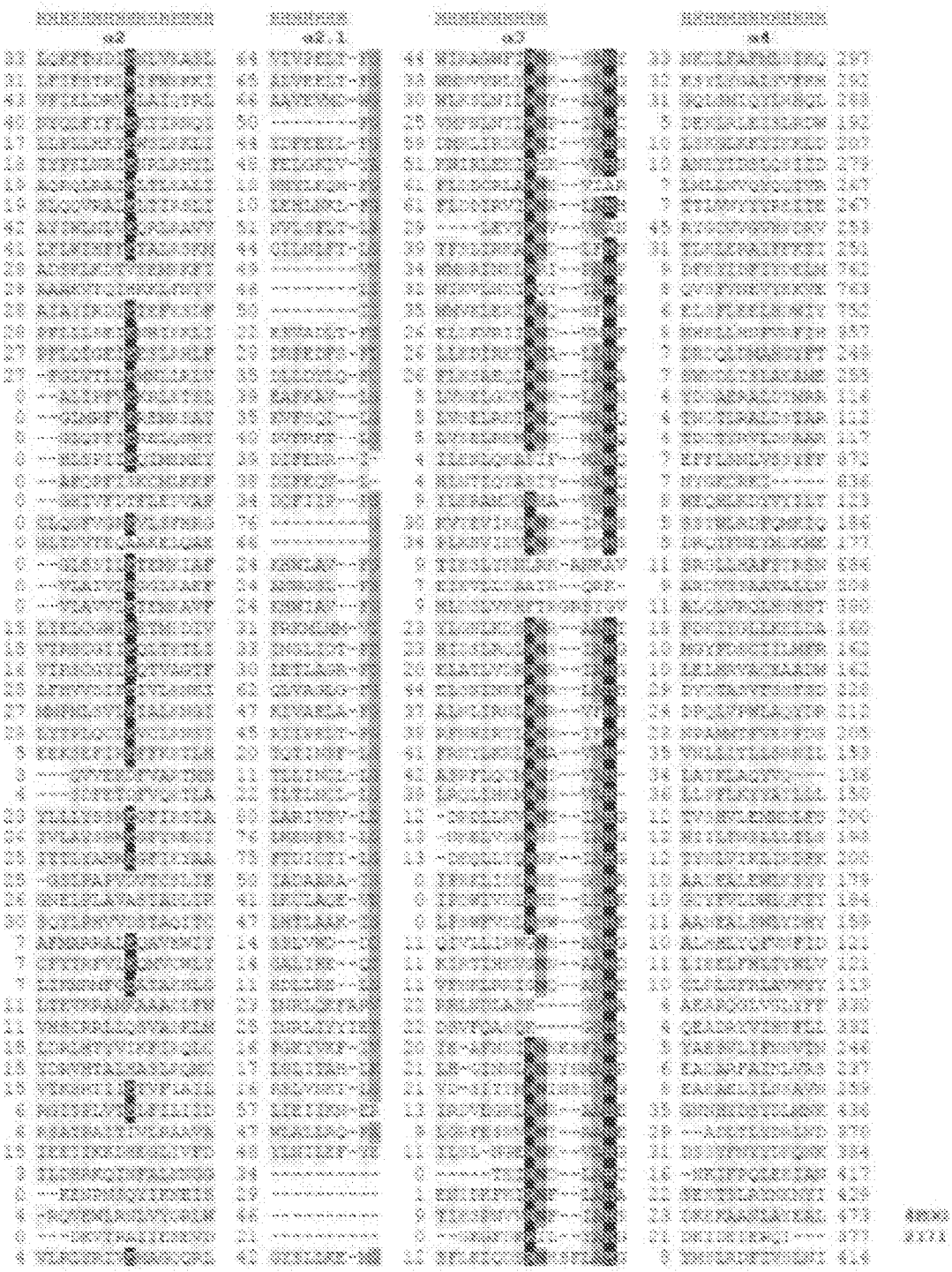


FIG. 14B

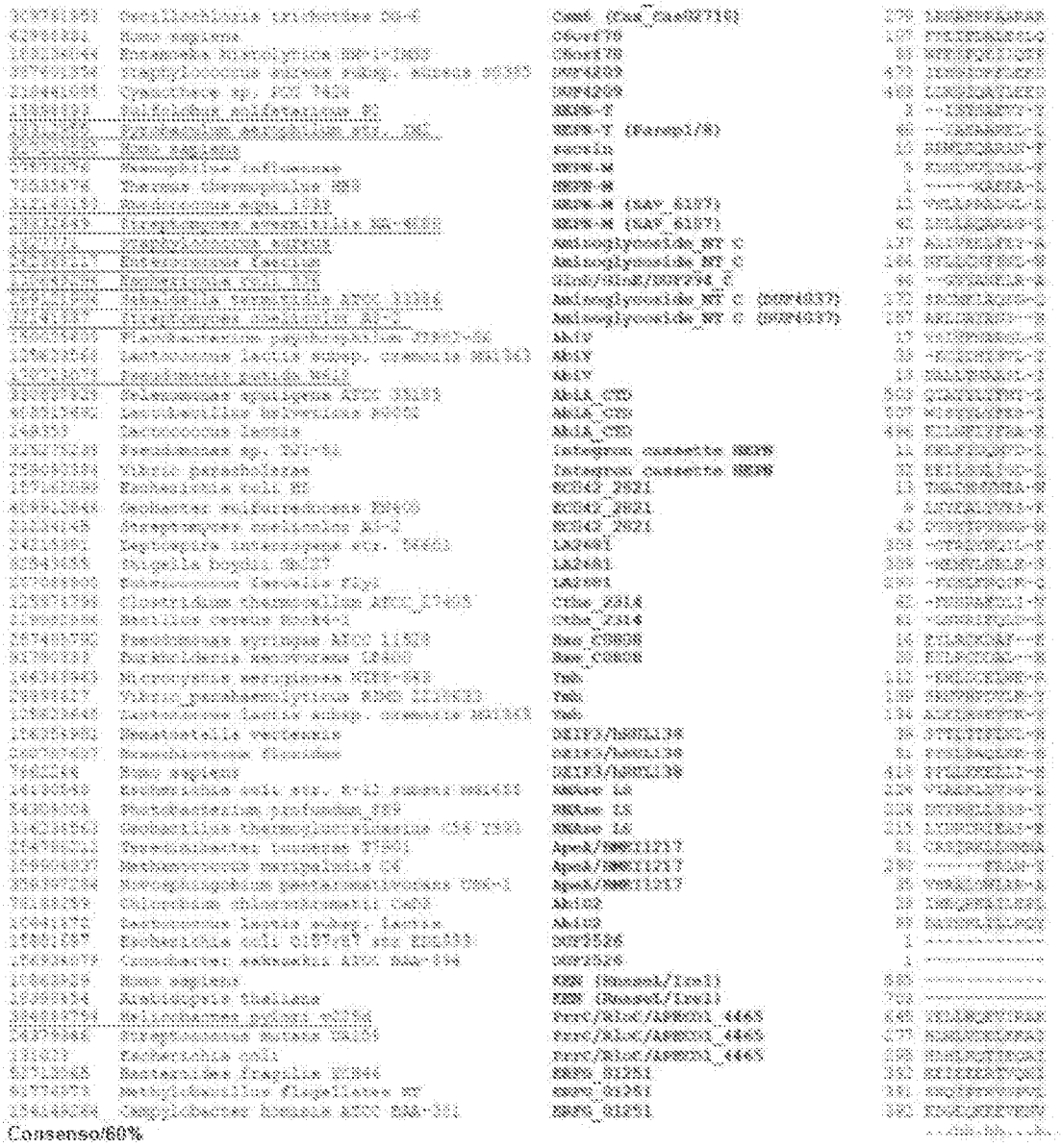


FIG. 14C

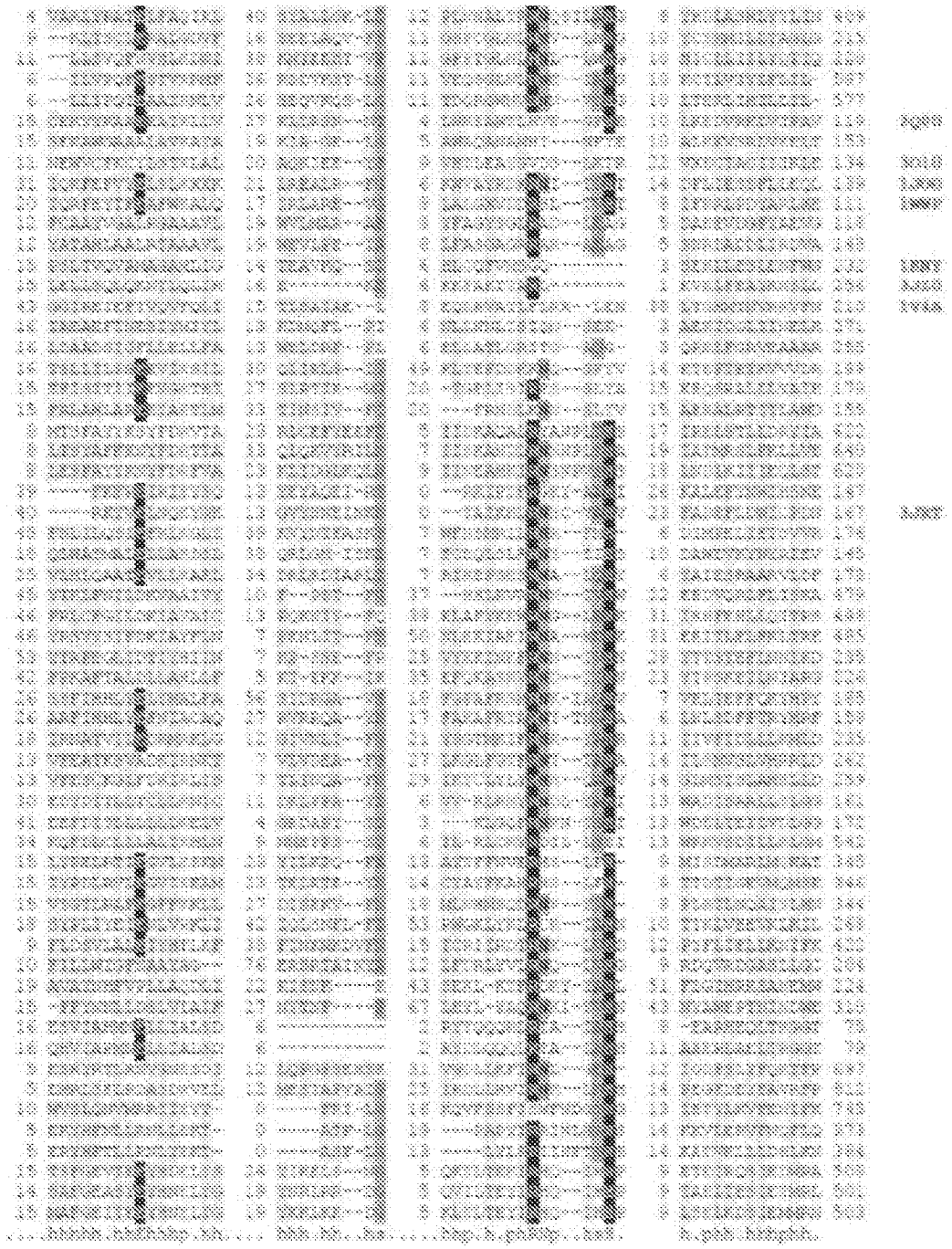


FIG. 14D

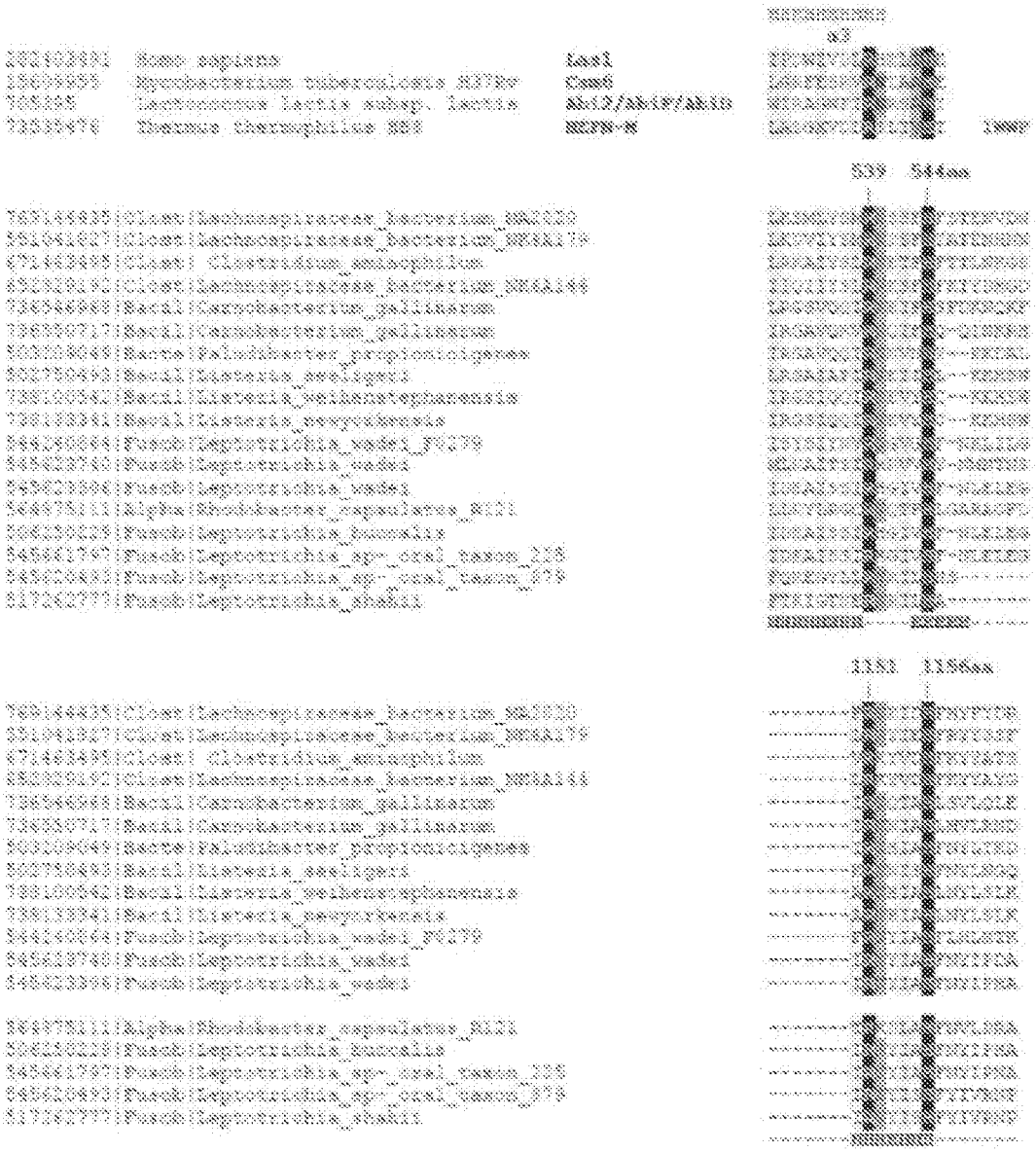
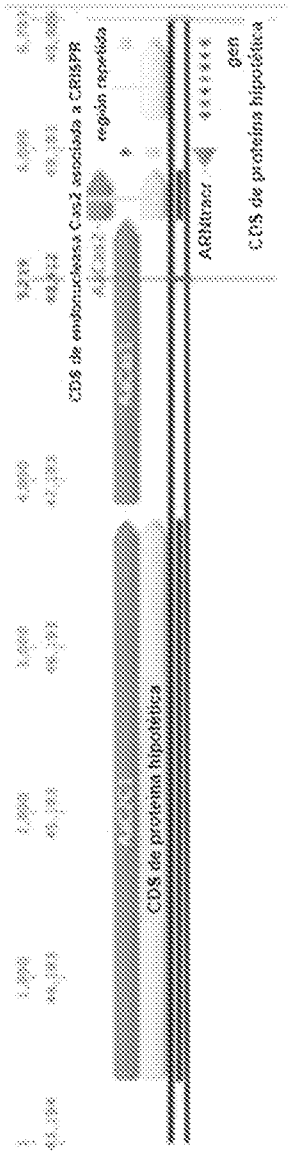


FIG. 14E



DR : GUGCCACUUCUCAGAUCGCUCGCUCAGUGAUCCGAC

ARNtracr:

GUCAGAACACUGAGCGAGCGUUCUUUUUGAGAAGCUCAACGGCCUUUGCCACCUGGAA
AGUGGCAUUGGCACACCGUUGAAAAAUUCUGUCCUCUAGACAGA

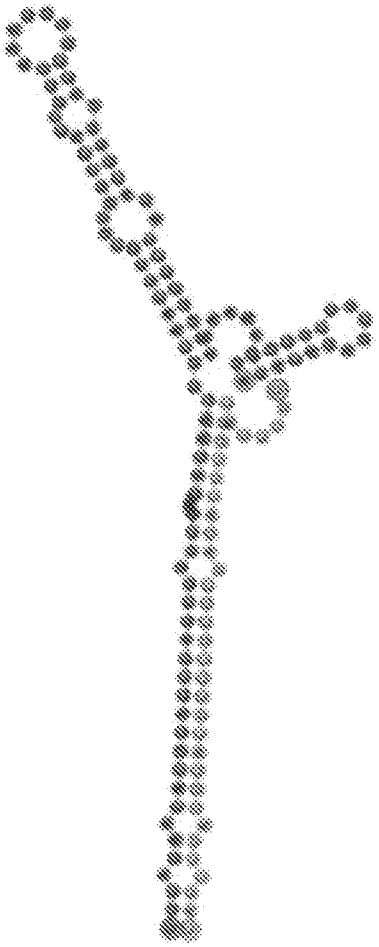


FIG. 15

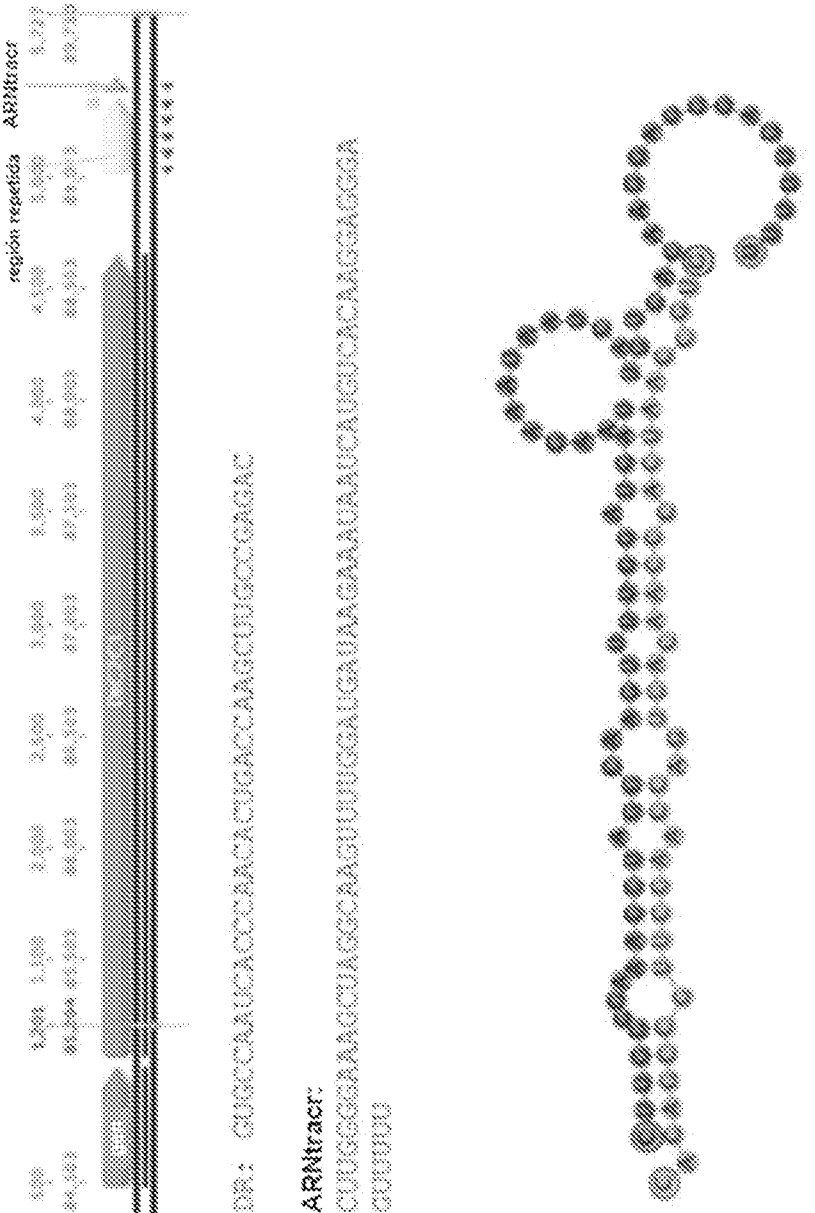

región repetida
ARNtracr
DR: GUGCCAAUGAUCCCAACACUGAACCAAGCUUGCCGAGAC
ARNtracr:
CUUGGCGAAGCUACGCAAGUUUUGCAUGAUAAGAAUAAUCAUGUCACAAGGAGCCA
GUUUUU

FIG. 16

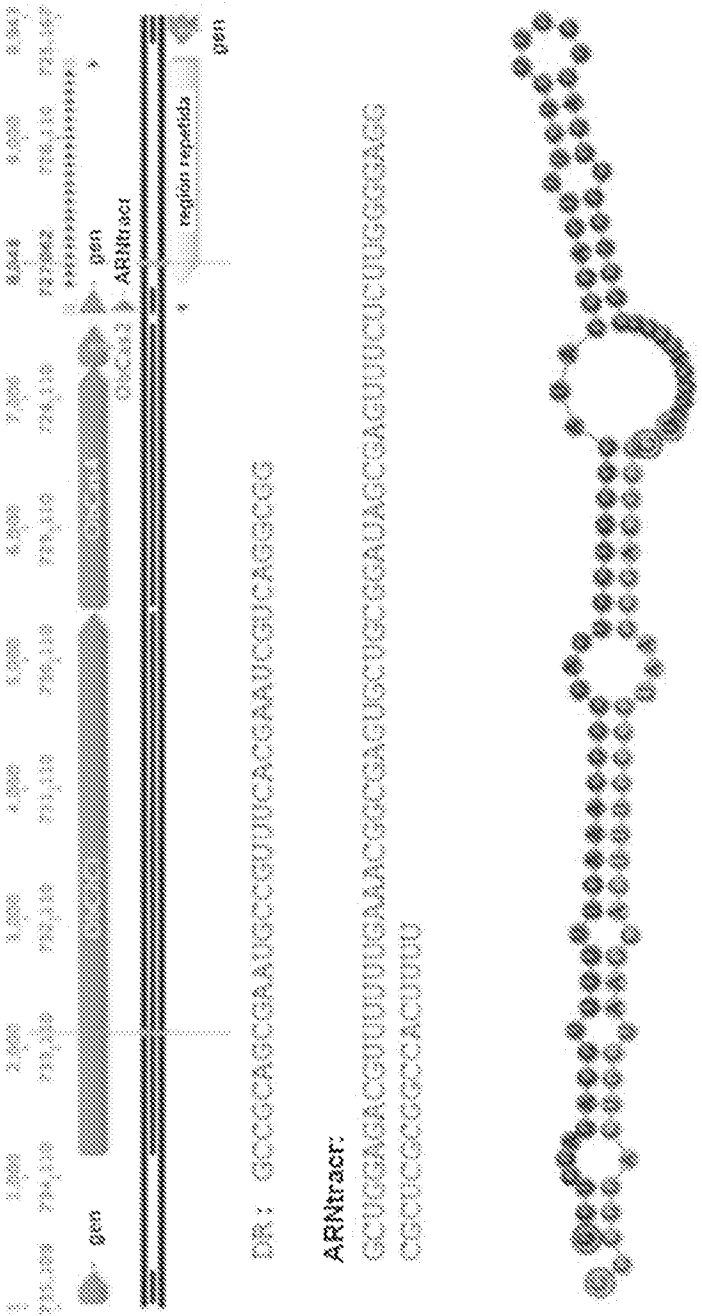
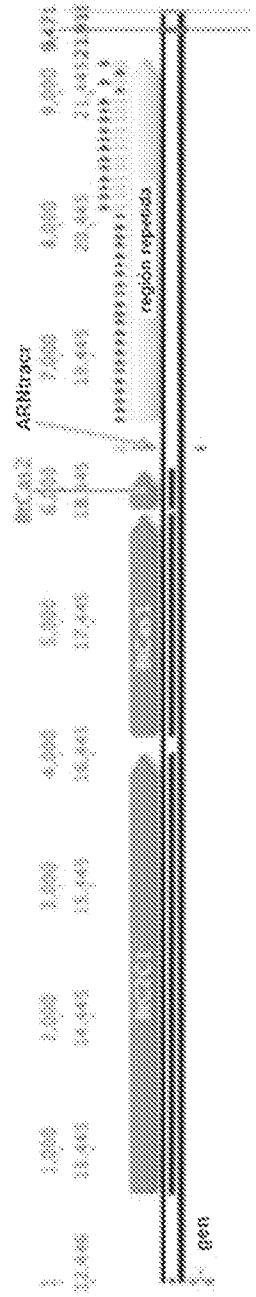

gen
ARNtracr
region repetida
gen
DR: GCCGCAGGCGAAUGCCGUUUUCACGAAUCGUCAGGCGG
ARNtracr:
GCUGCAGACGUUUUUUGAAACGCCGACUGCUGCGAUAGCGAGUUUCUCUUGCGCAGG
CGCUCGCCGCCCACUUUU

FIG. 17



ER : GUCCAAGAAAAAGAAAUGAUACGAGCAUUAGCAC

ARNtracr:

CUCAACGAUGUCUCUUUUAUUUCUUUUUCUUGCAUCUGACUACGAGCACCCACAUUG

CACAUUUCCCAUCGUCGGUCCUCCUCUACUAUACGUAAAAACAAACCUUUUU

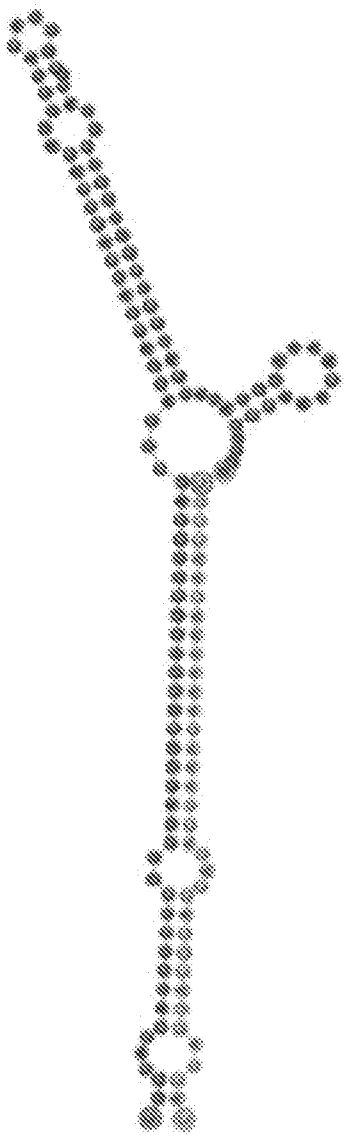
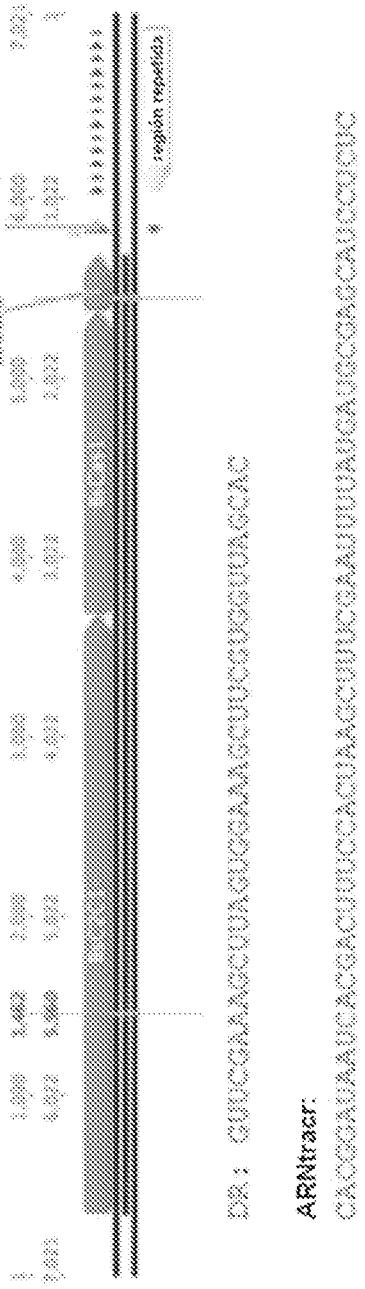


FIG. 18



DR 1 GUUCGAAAGCUUAGUCGAAAGCUUCGUUGAUUAGCAC

ARNtracr:

CACGCAUAAUCACGACGUUUCCACUAAGCUUUCGAAUUUUAUGCAUGCGAGCAUCGCUC

AGCUCAAAAAA

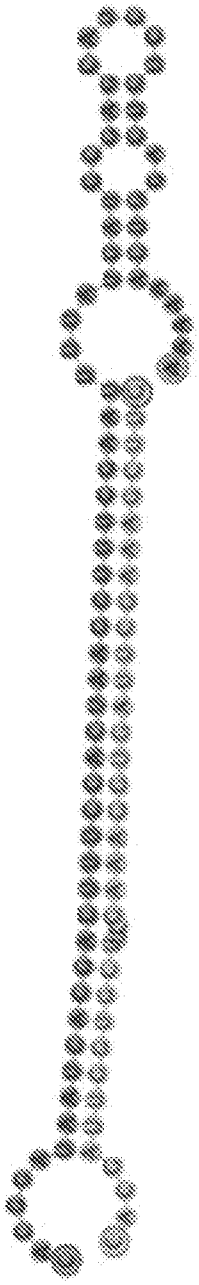


FIG. 19

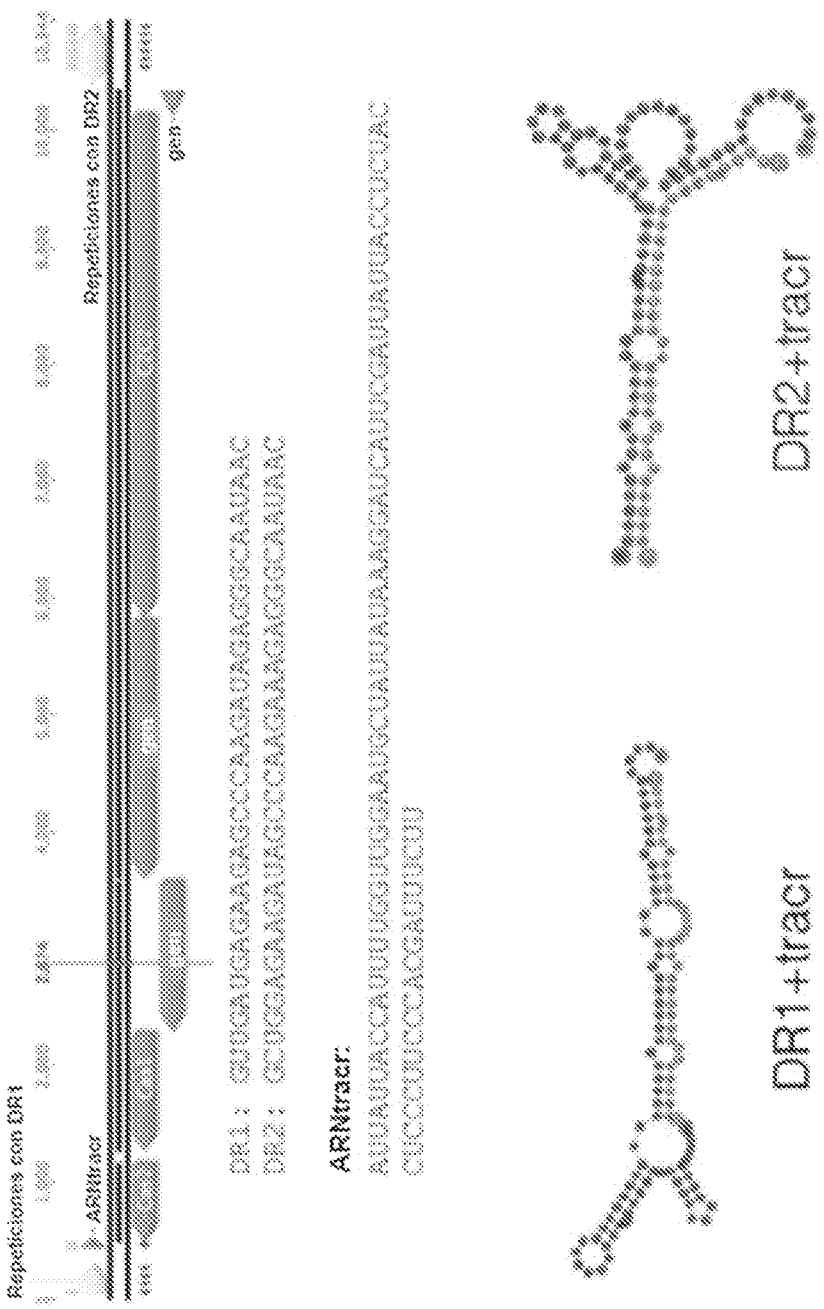
DR: GUAUUGAGAAAAGCCAGAUAUACUUUGCCAAUAGAC

ARNtracr:

AUAUUUUGAUUCCCAUUUAUGCUUAUUUACCAUAAAUGCGAAUCAACUAAAAAAUAUU
UUUU

La orientación parece extraña, pero posiblemente el sistema sea diferente de los del tipo II

FIG. 20



La orientación parece extraña, pero posiblemente el sistema sea diferente de los del tipo II

FIG. 21

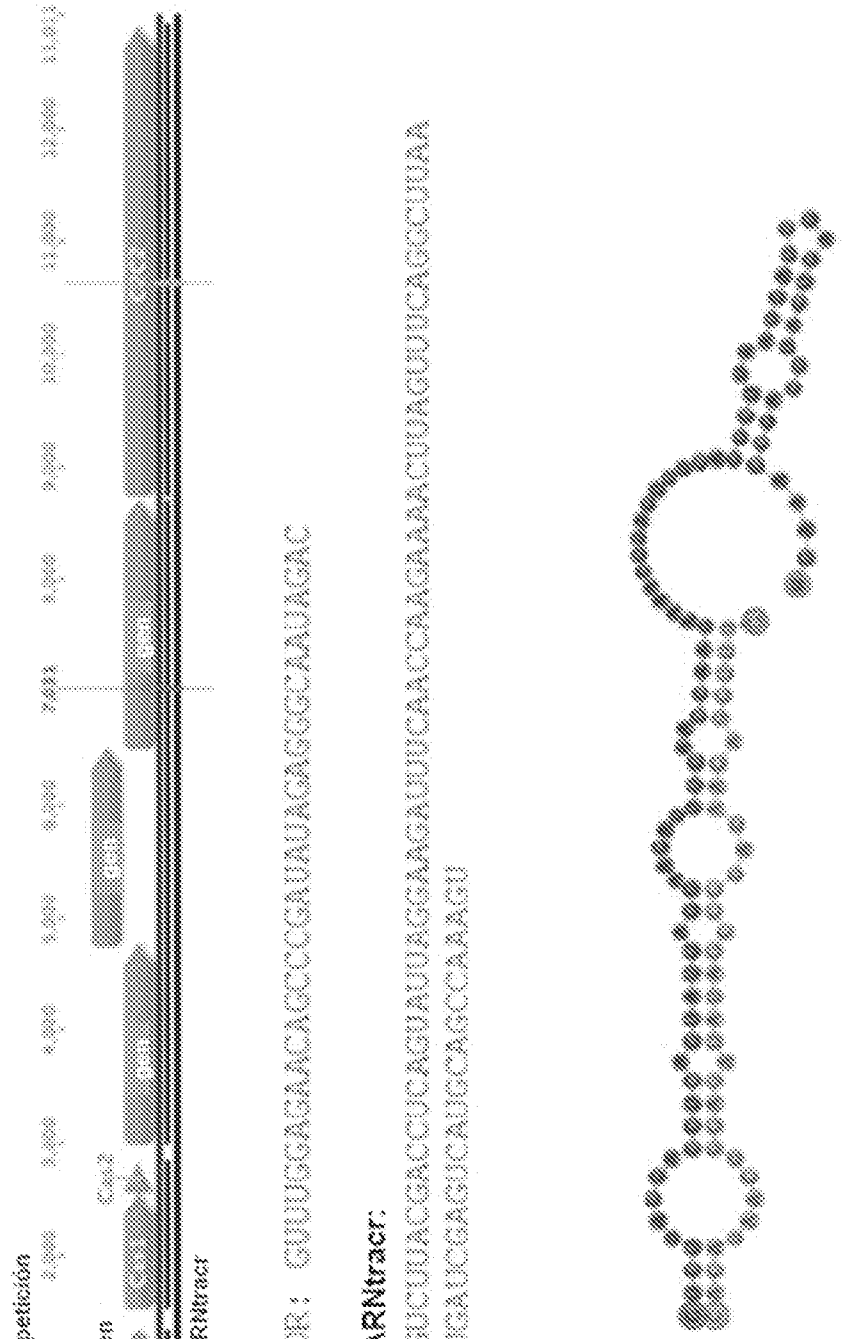
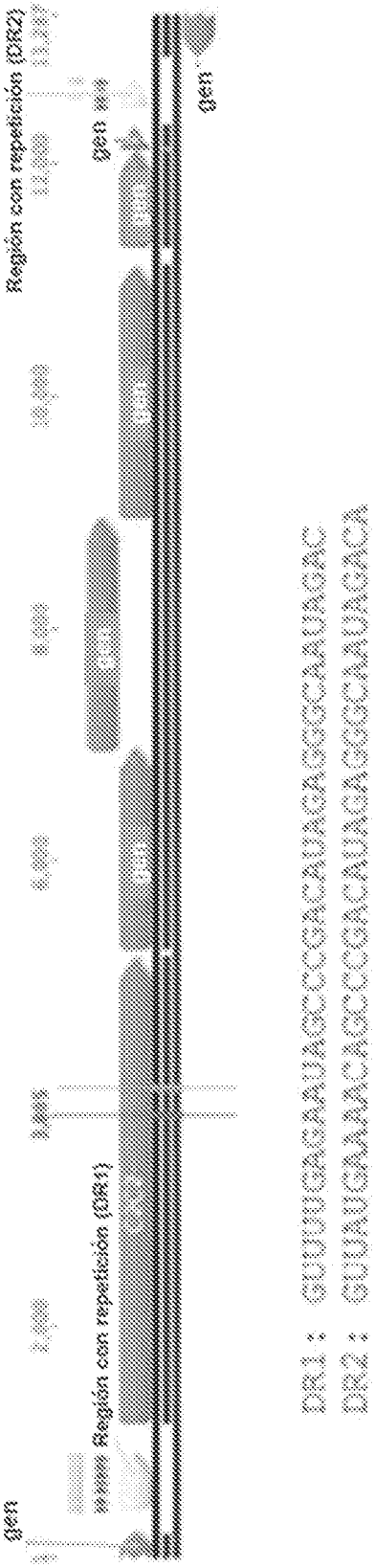

Región de repetición
cas3
gen
ARNtracr
DR: GUUGCAGAACAGCCCGAUAUAGAGGCAAUAGAC
ARNtracr:
GUCUUACGACCUCAGUAUUAGGAAGAUUUCAACCAAGAAAACUUAGUUCAGCCUUAA
UGAUCGAGUCAUGCAGCCAAAGU

FIG. 22

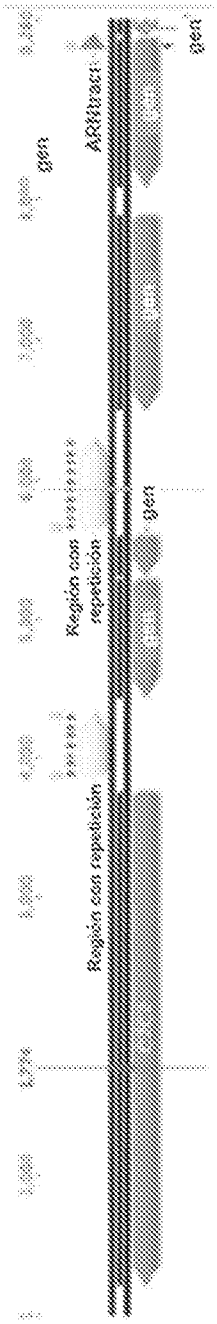


DR1 : GUUUUGAGAAUAGCCCGACAUAGAGGGCAAUAGAC

DR2 : GUUAUGAAAACAGCCCGACAUAGAGGGCAAUAGACA

ARNtracr: ???

FIG. 23



DR: GUUAUAGUCCUCUUACAUUUAGAGGUAGUCUUUAAU

ARNtracr:
UCUUAAGAACUUCUCUACCUGAAGUUGAUUAUAAAUGACUCUUGCUCUCAUAGAUAU
CCUCCUUUCAAAAUAUACACUGCCCAUUAAUUACCGUUU

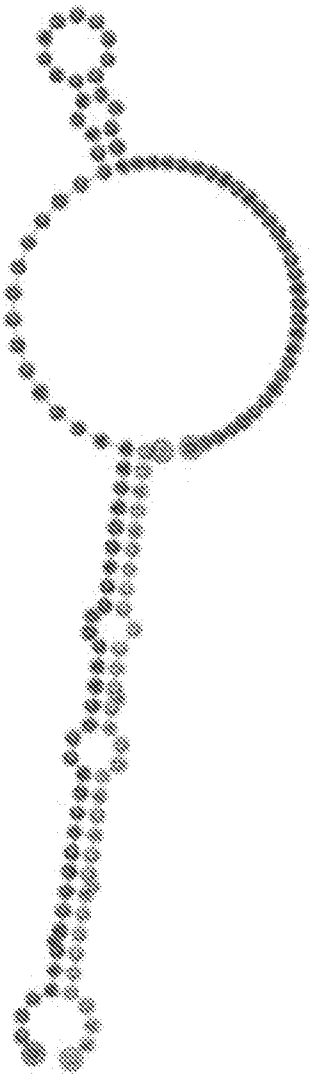


FIG. 24

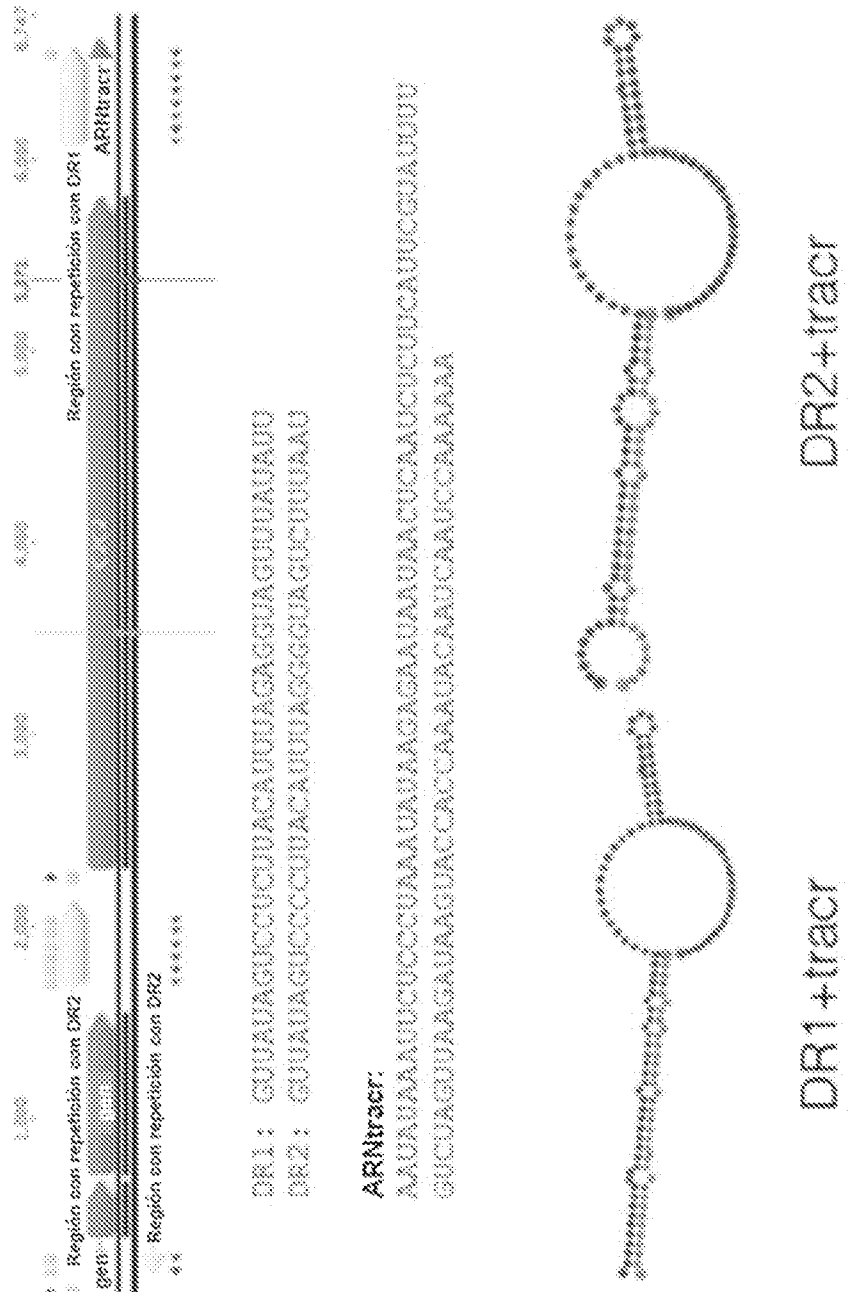
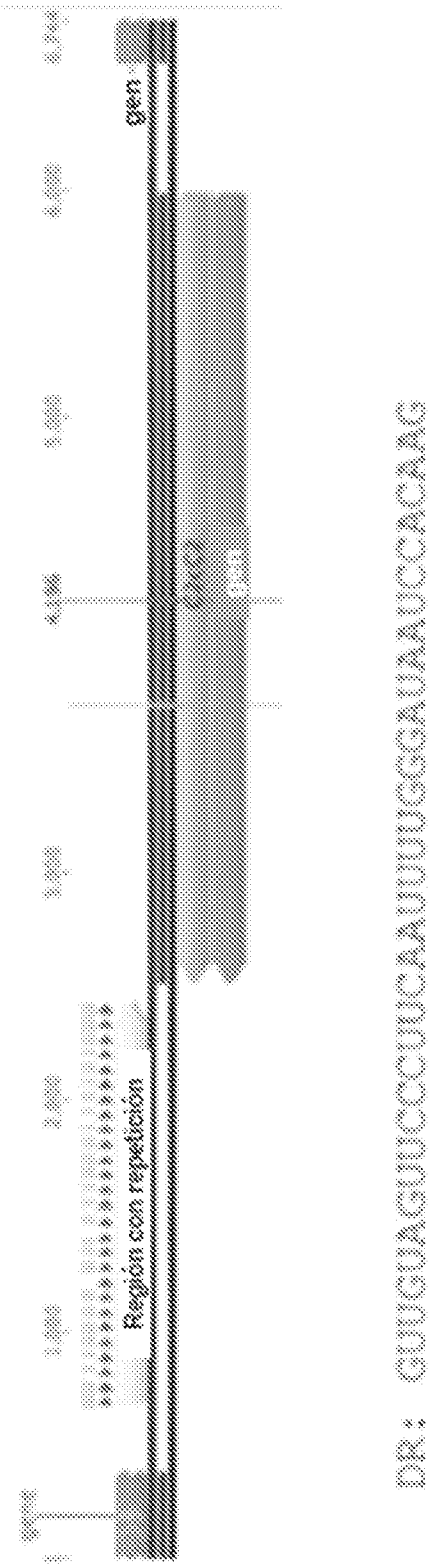

Región con repetición con DR2
Región con repetición con DR1
ARNtracr
Región con repetición con DR2
DR1 : GUUAUAGUCCUCUUACAUUUAGAGUAGUUAUAUU
DR2 : GUUAUAGUCCCUUACAUUUAGGGUAGUCUUUAAU
ARNtracr:
AAUAUAAAUUCUCCCUAAAUAUAAGACAAUAAUAACUCAAUCUCUUCAUUCCUAUUUU
GUCUACUUAAGAUAGUACCACCAAAUACAAUCAAUCCAAAAAA
DR1+tracr
DR2+tracr

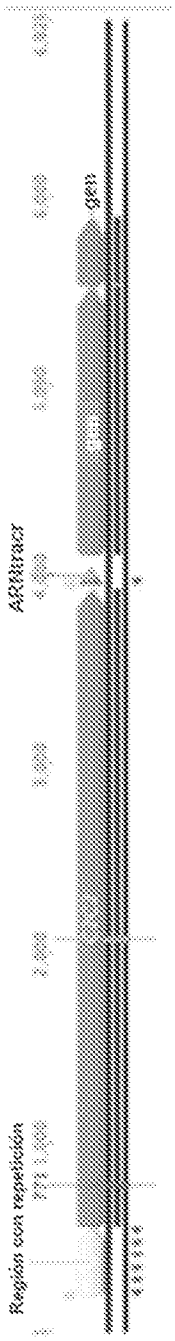
FIG. 25



DR: GUUGUAGUUCCUUCAAUUUUGGGAUAAUCCACAAG

ARNtracr: ???

FIG. 26



DR: GUUUUAGUCCCUCUUUCAUAUAGCGUAGGUCUCUUAC

ARNtracr:

AUGAAAAGAGCACUAAAACUGAAAGGACUAAAACACCAGAUCUCGAUAACUAUAUU

AGUCCCUAUUAAAAAUUCGUCGAUAUUACAGCAGCAAACUUU

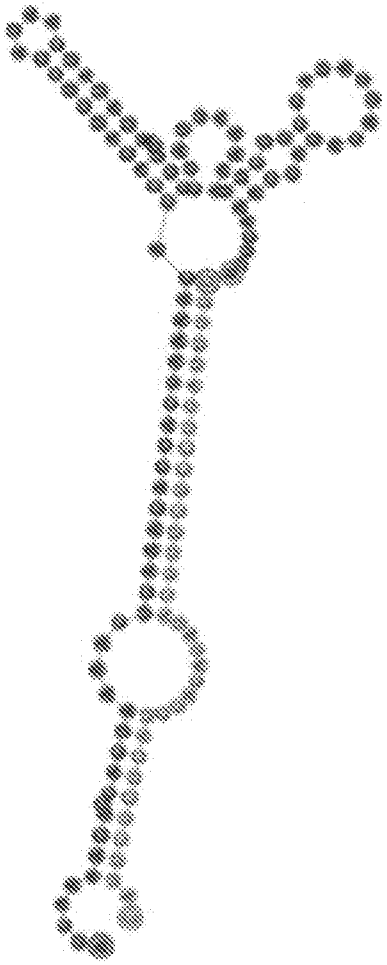


FIG. 27

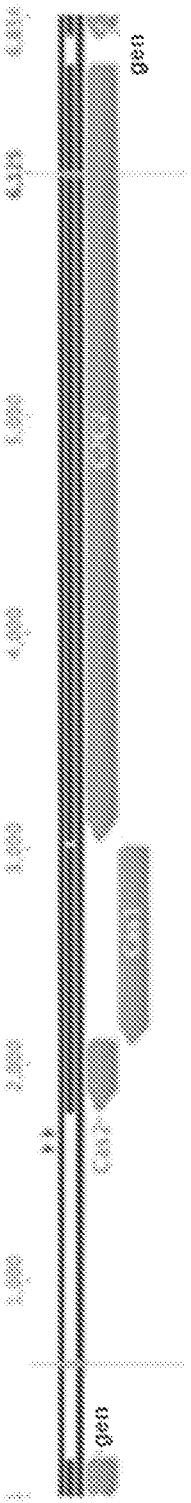
DR: GUUUUAGACCUCUUCUAUUUUGAGGUACUCUAAAUC

ARNtracr 777

FIG. 28

Región con repetición
ARNtracr
DR.: GUUUUAGUCCCUCUUUUGUUUUGAGGUACUCUAAAUC
ARNtracr:
AAGUCAGCCCACAACAAAGAAGAUGACGAACAAAAUCUCUCCCCAUCUUCUUAAAAUU
AUUUCCACACAGCCAACAUUAUAAGCGUUAAAACCAGCACCAUGACUACAUUCACC
CAACAAUCAGAAUCCCCGUUUCUCCGUUUUU

FIG. 29



DR: GUUUUAGAUCCCCUUCGUUUUGGGGUUAUCUAUAUC

ARNtracr: ???

FIG. 30

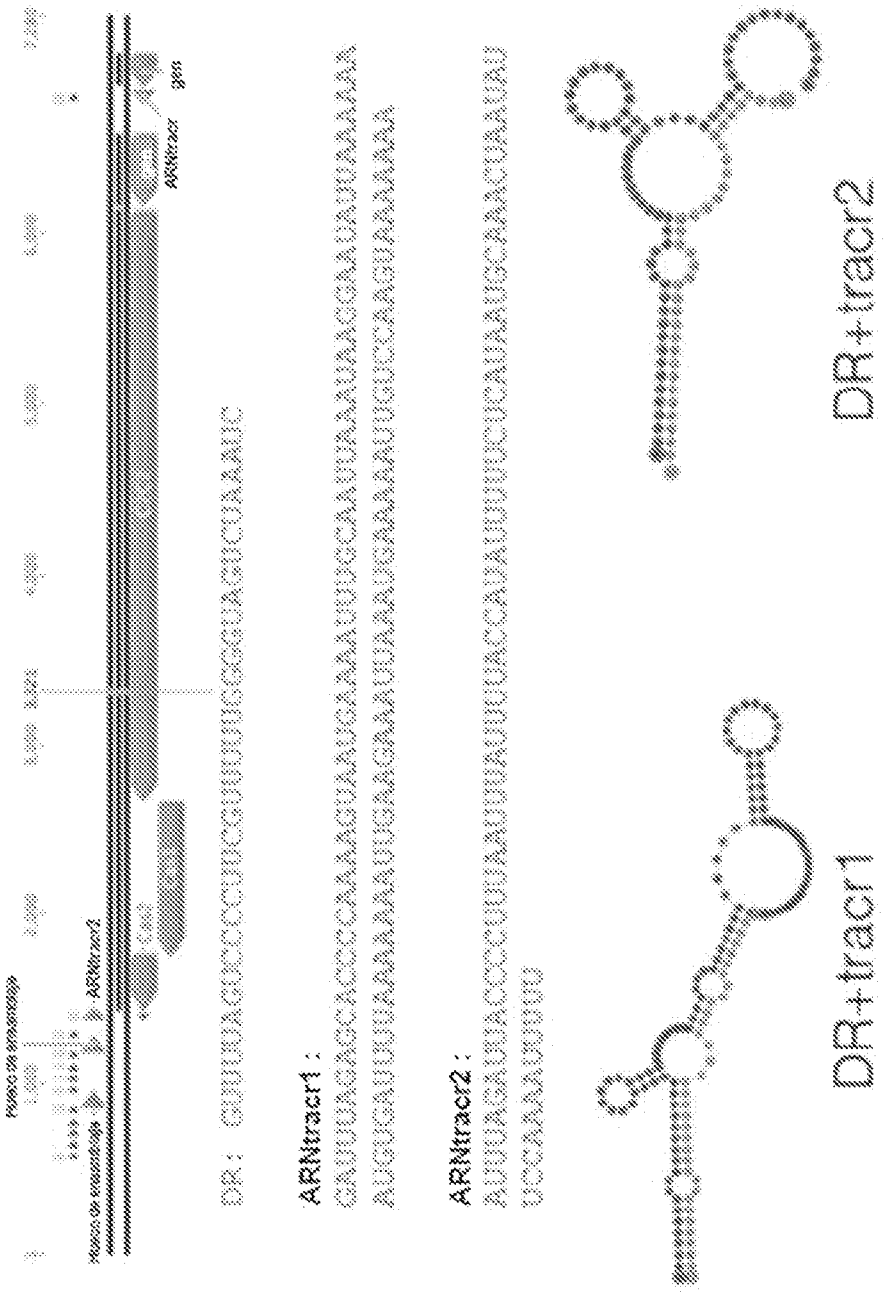


FIG. 31

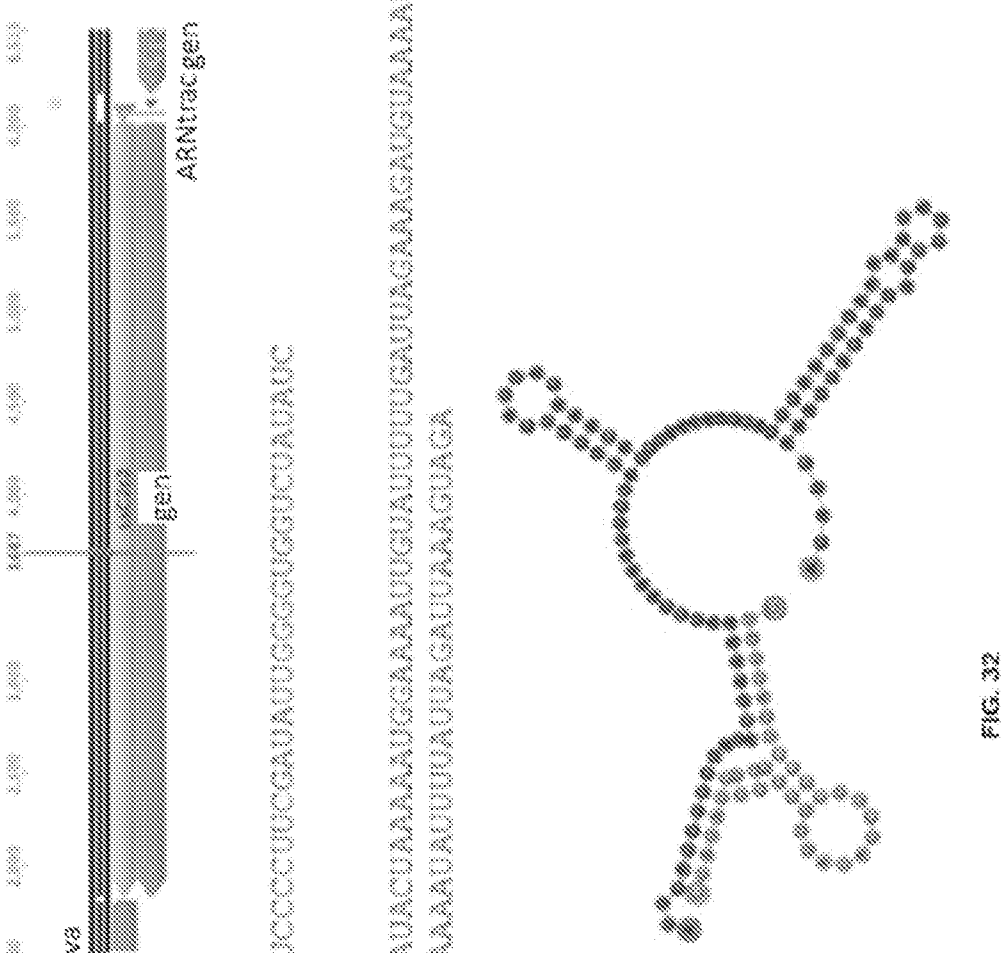

región repetitiva
gen
gen
ARNtracr gen
DR: GUUUUAGUCCCCUUCGAUAUUGGGGUCGUCUAUAUC
ARNtracr:
AUUGAUGUCGUAUACUAAAAAUGGAAAAUUGUAUUUUUGAUUAGAAACAUGUAAAAUU
GAUUUAAUUUAAAAAUAUUUUAUUUAGAUUAAAGUAGA

FIG. 32

región repetitiva
DR1: GUUCAGUCCGCCGUCGUCUUGGCGGGUGAUGUGAGGC
DR2: GUUCAGUCCGCCGUCAUUUGGCGGUGAUGUGCUCC
ARNtracr: ??

FIG. 33

DR1 : GUUCAGUCCGCCGUCGUCUUGGCGGUGAUGUGAGGC

DR2 : GUUCAGUCCGCCGUCAUUUUGCGGUGAUGUCCUCC

ARNtracr : ? ? ?

FIG. 34

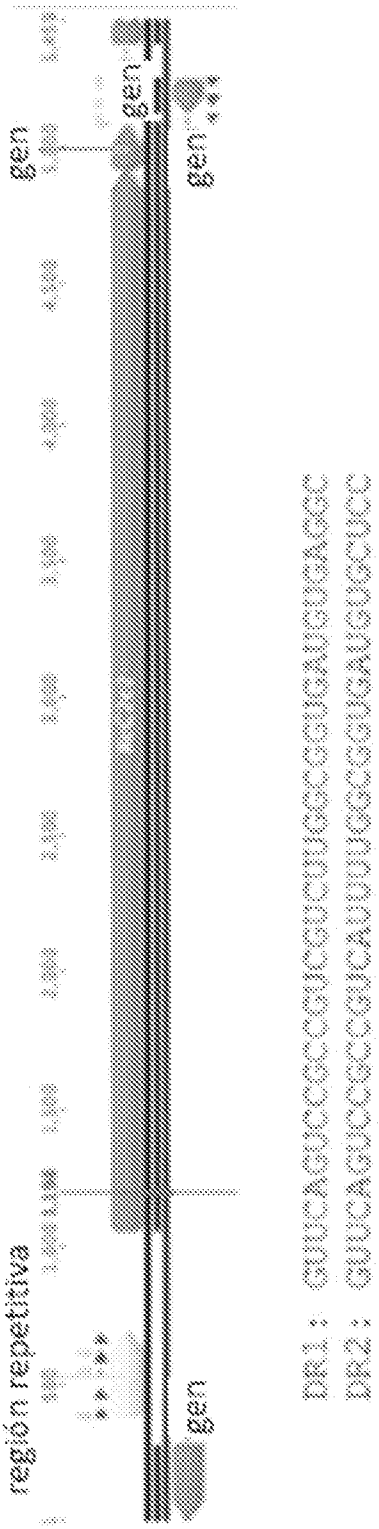

región repetitiva
gen
gen
gen
gen
DR1 : GUUCAGUCCGCCCGUCGUCUUGCGCGGUGAUGUGAGCC
DR2 : GUUCAGUCCGCCCGUCAUUUUGCGCGGUGAUGUGCUCC
ARNtracr : ???

FIG. 35

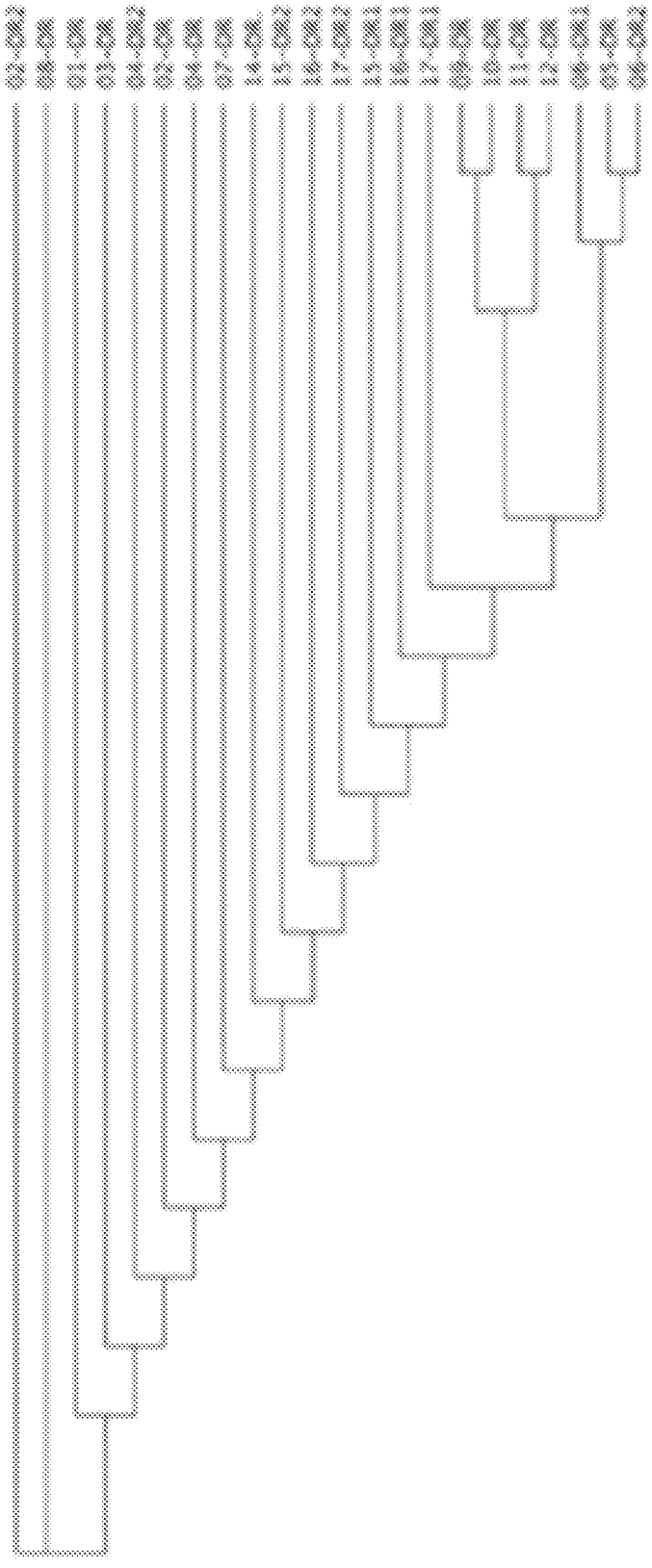


FIG. 3b

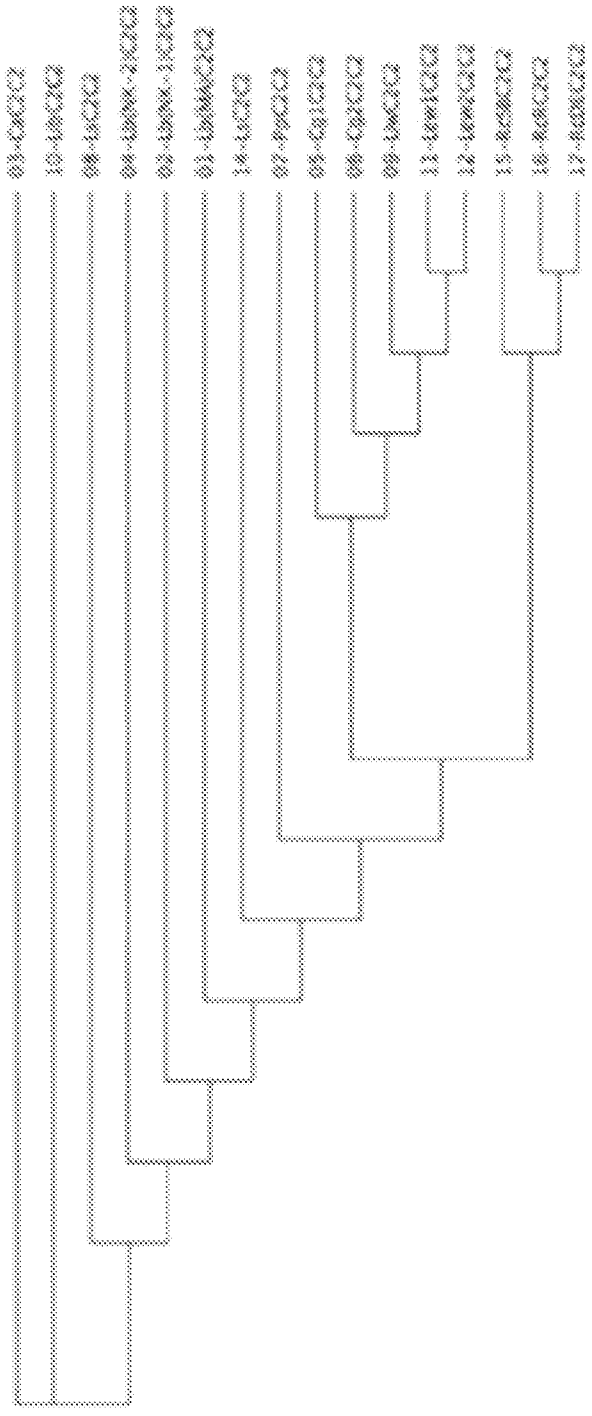


FIG. 37

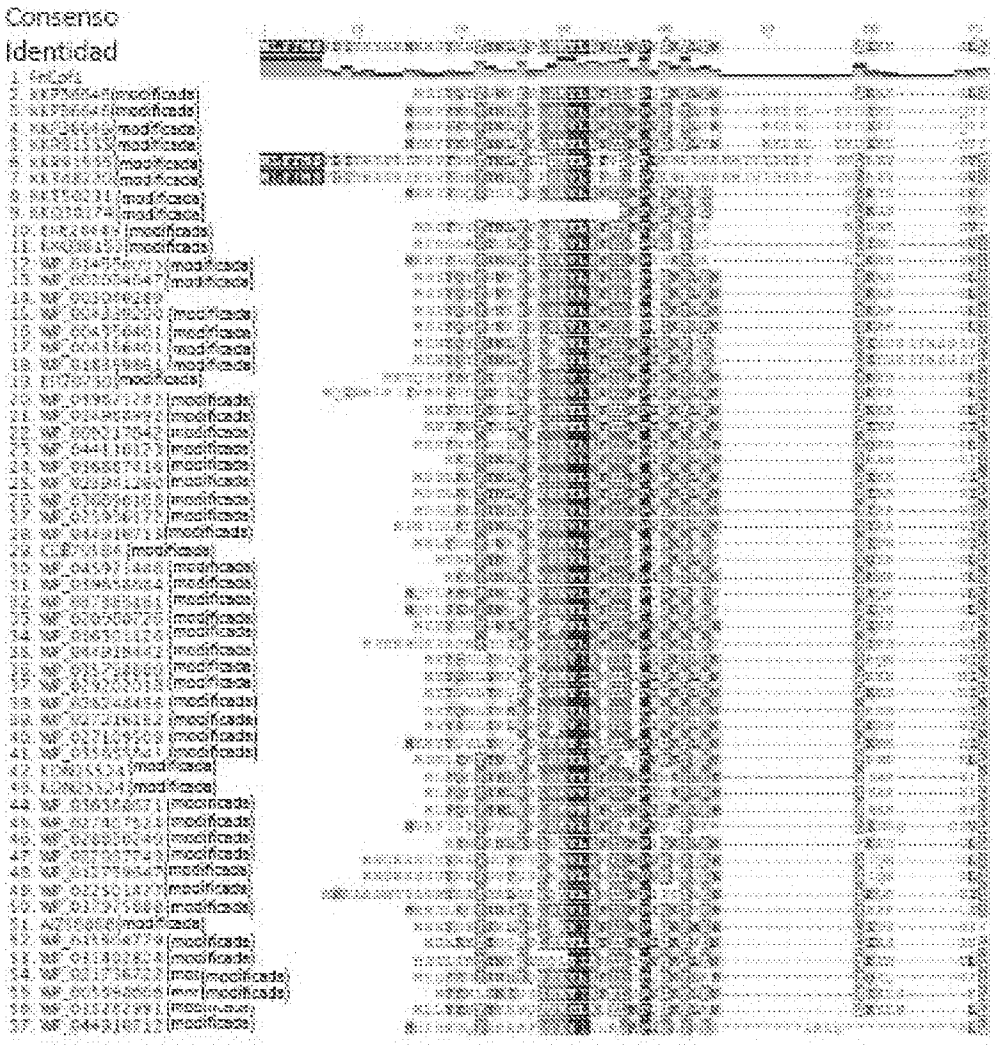


FIG. 38A

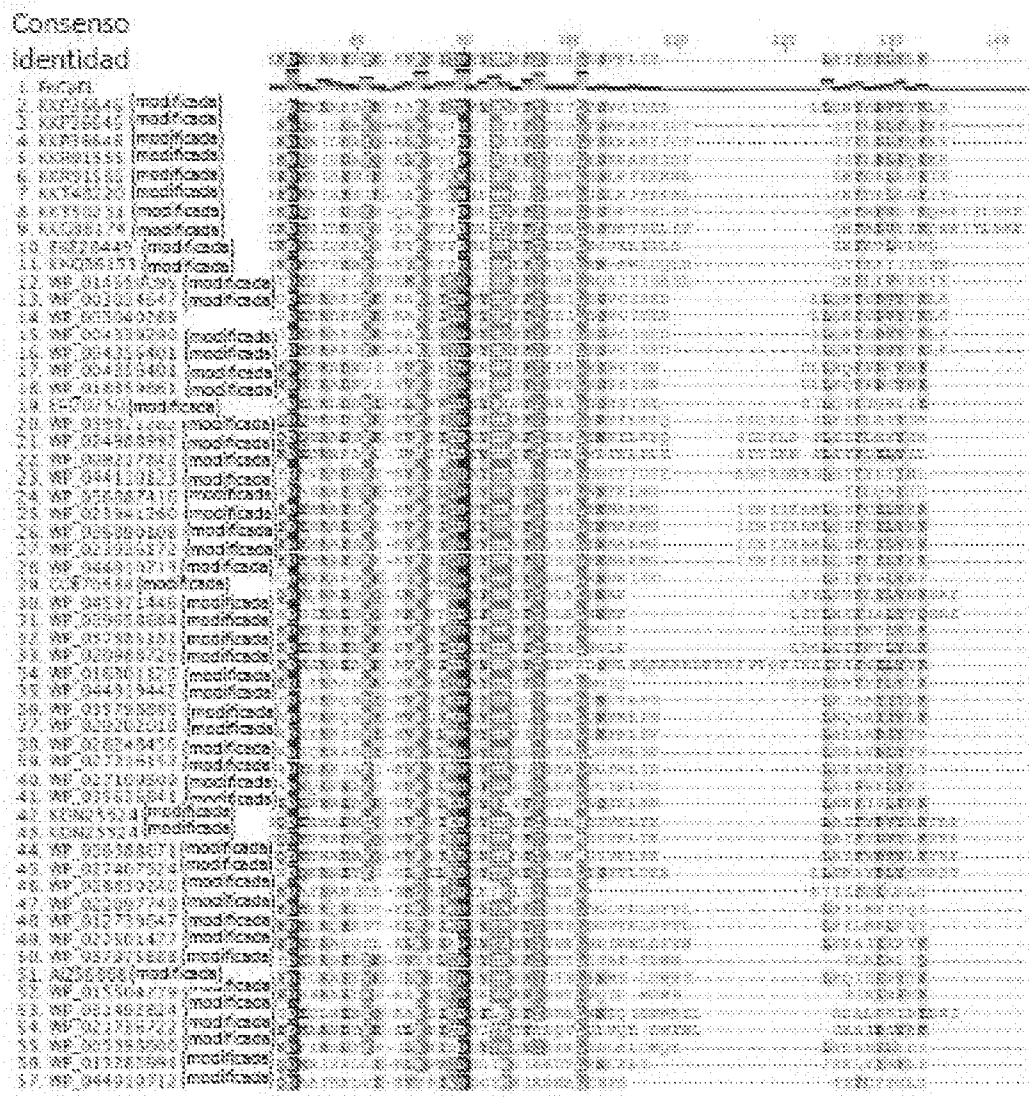


FIG. 38B

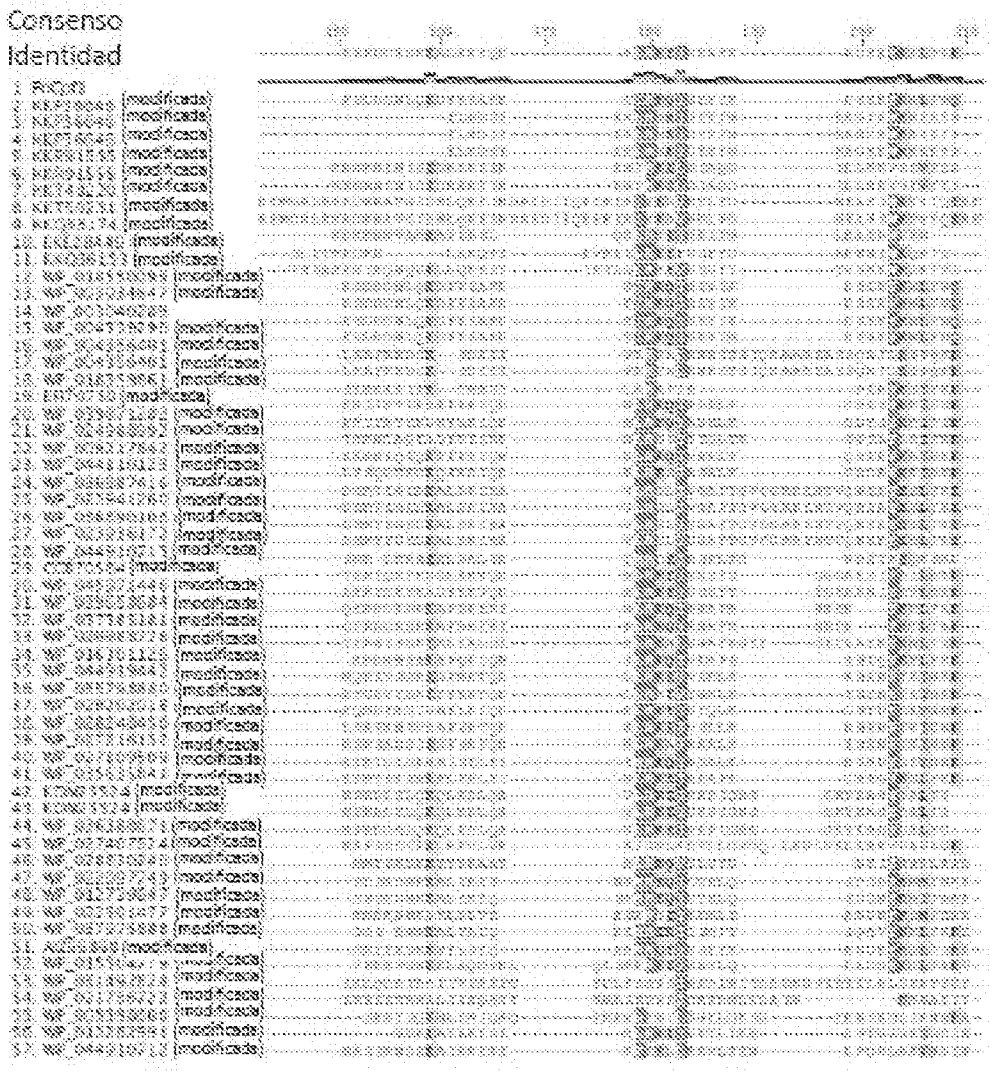


FIG. 38C

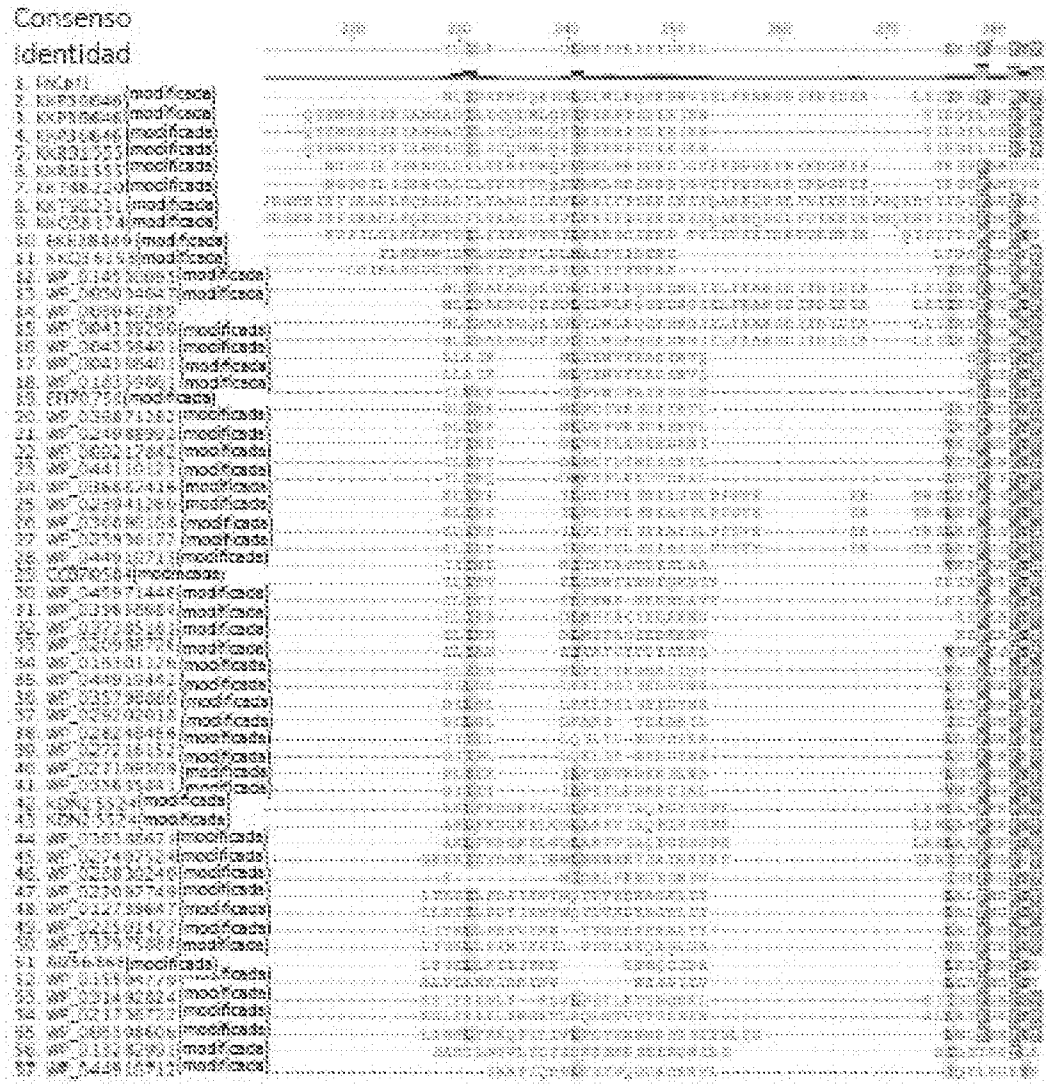


FIG. 38D

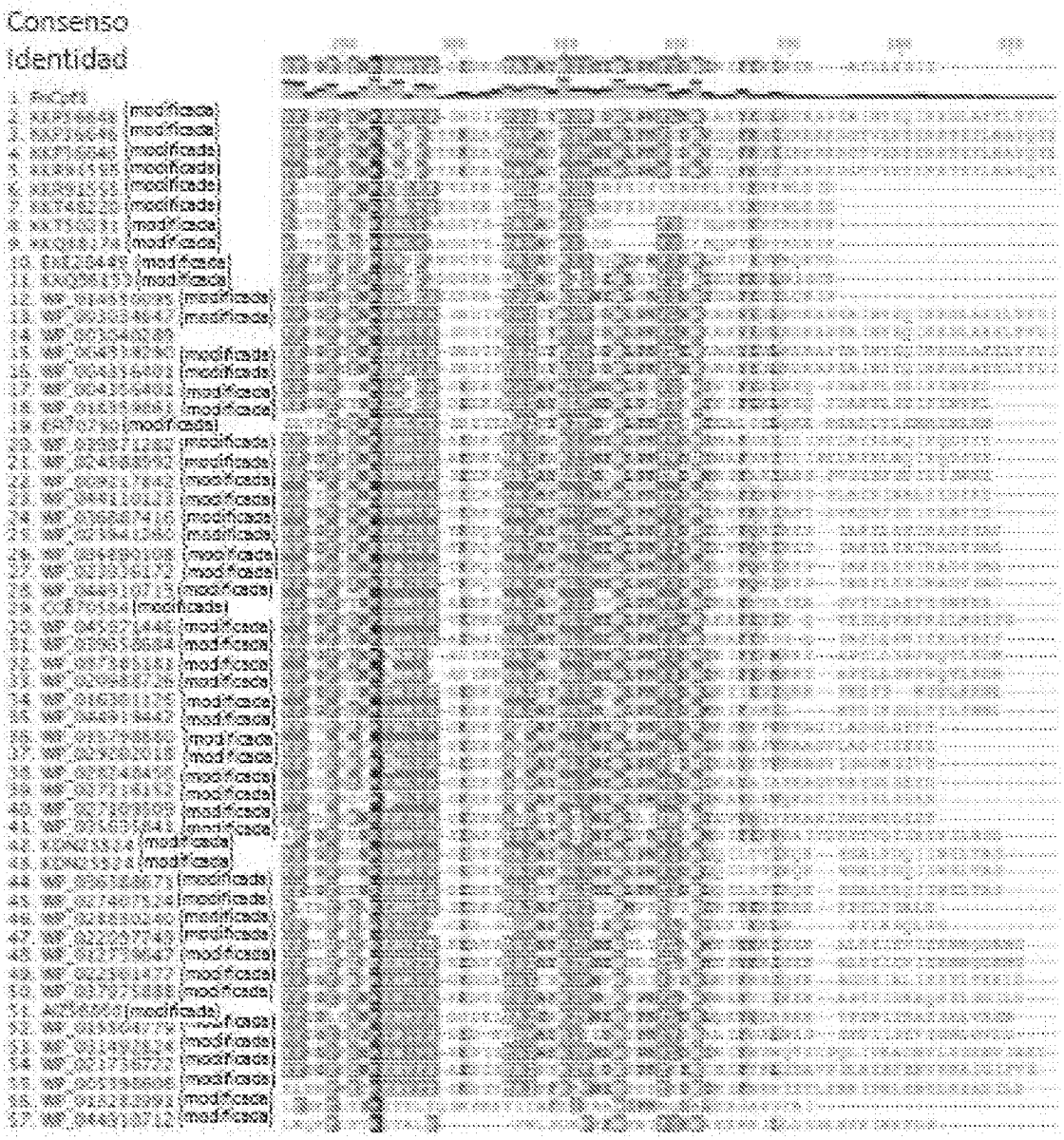


FIG. 38E

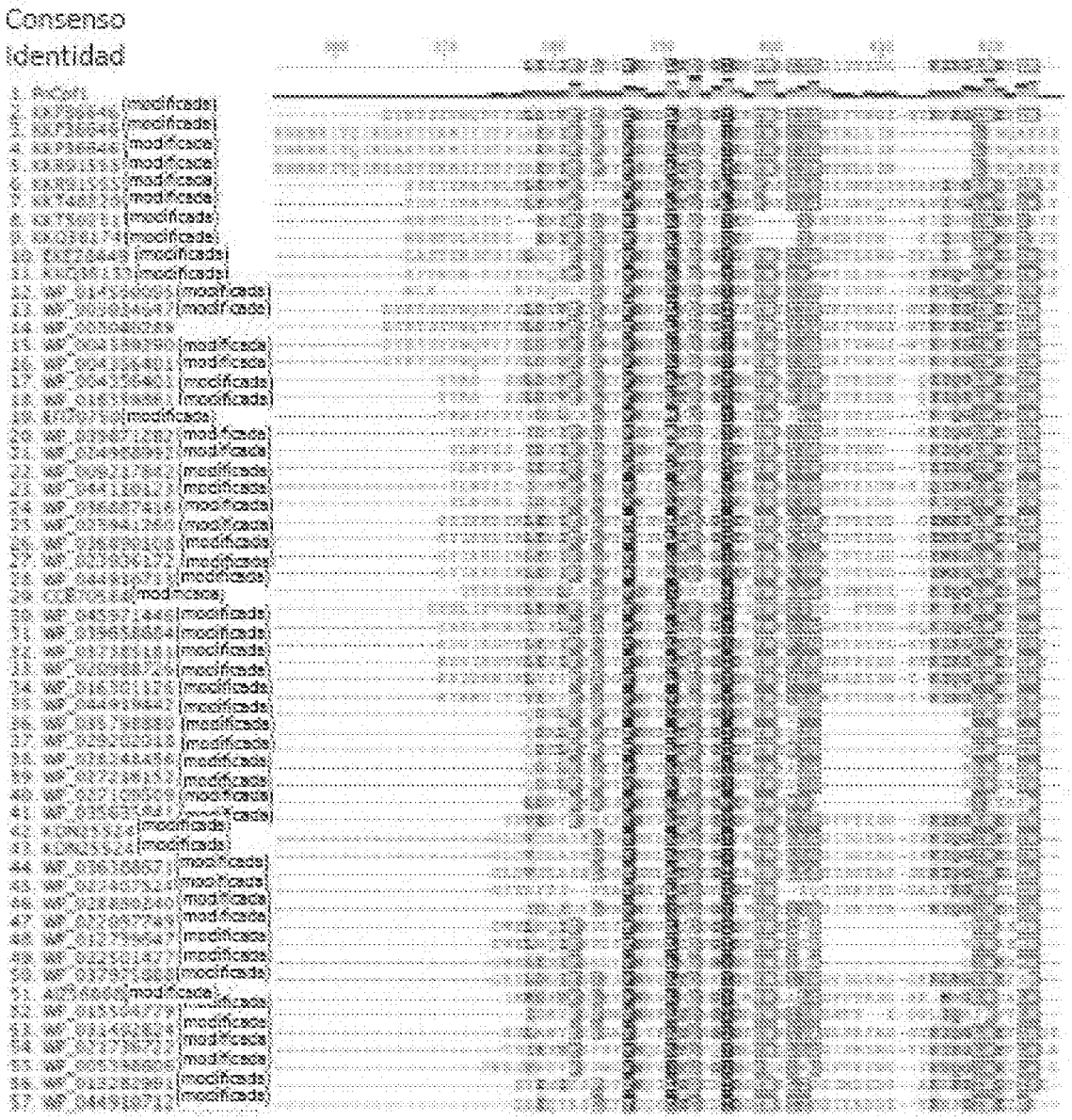


FIG. 38F

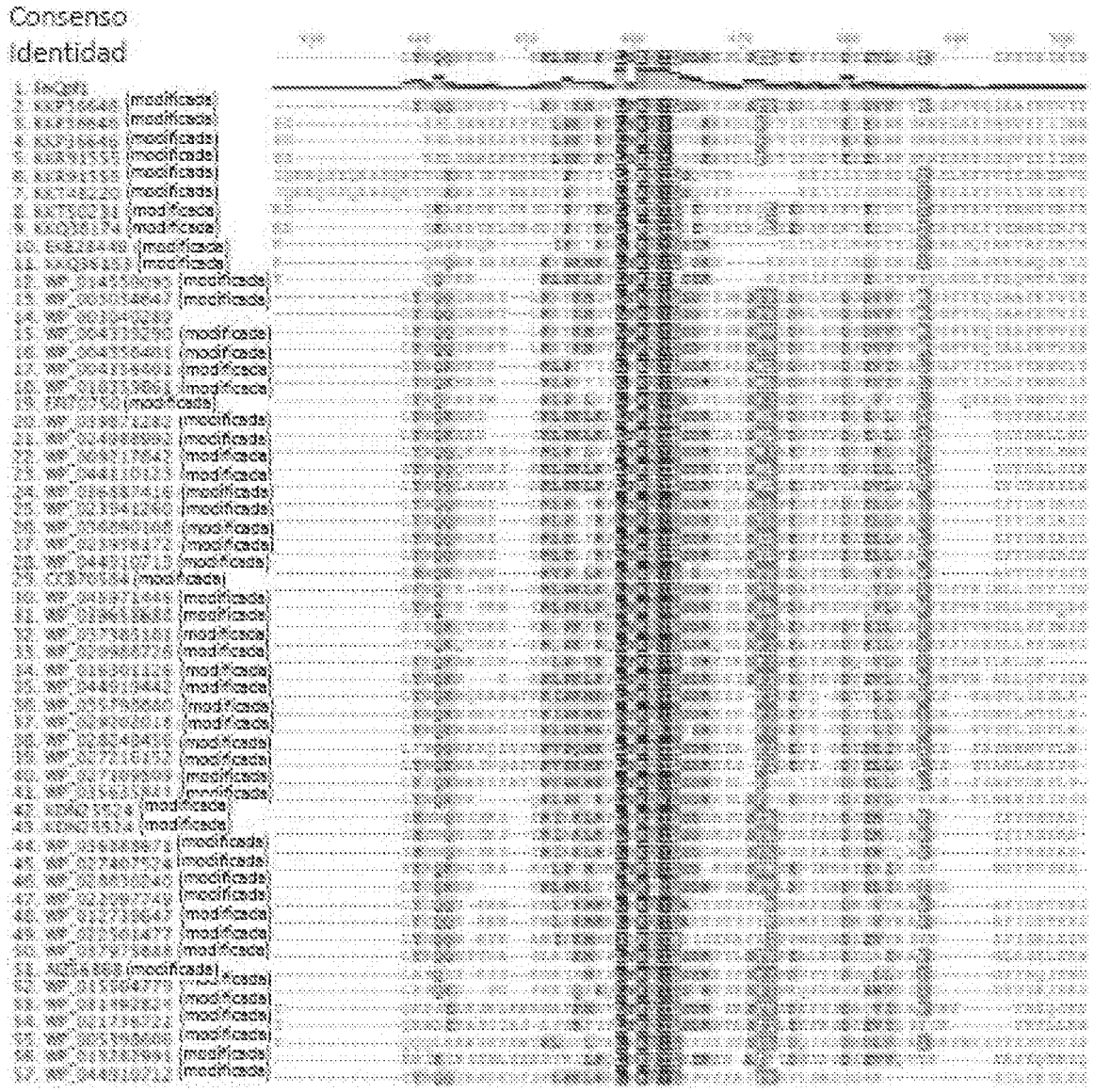


FIG. 38Q

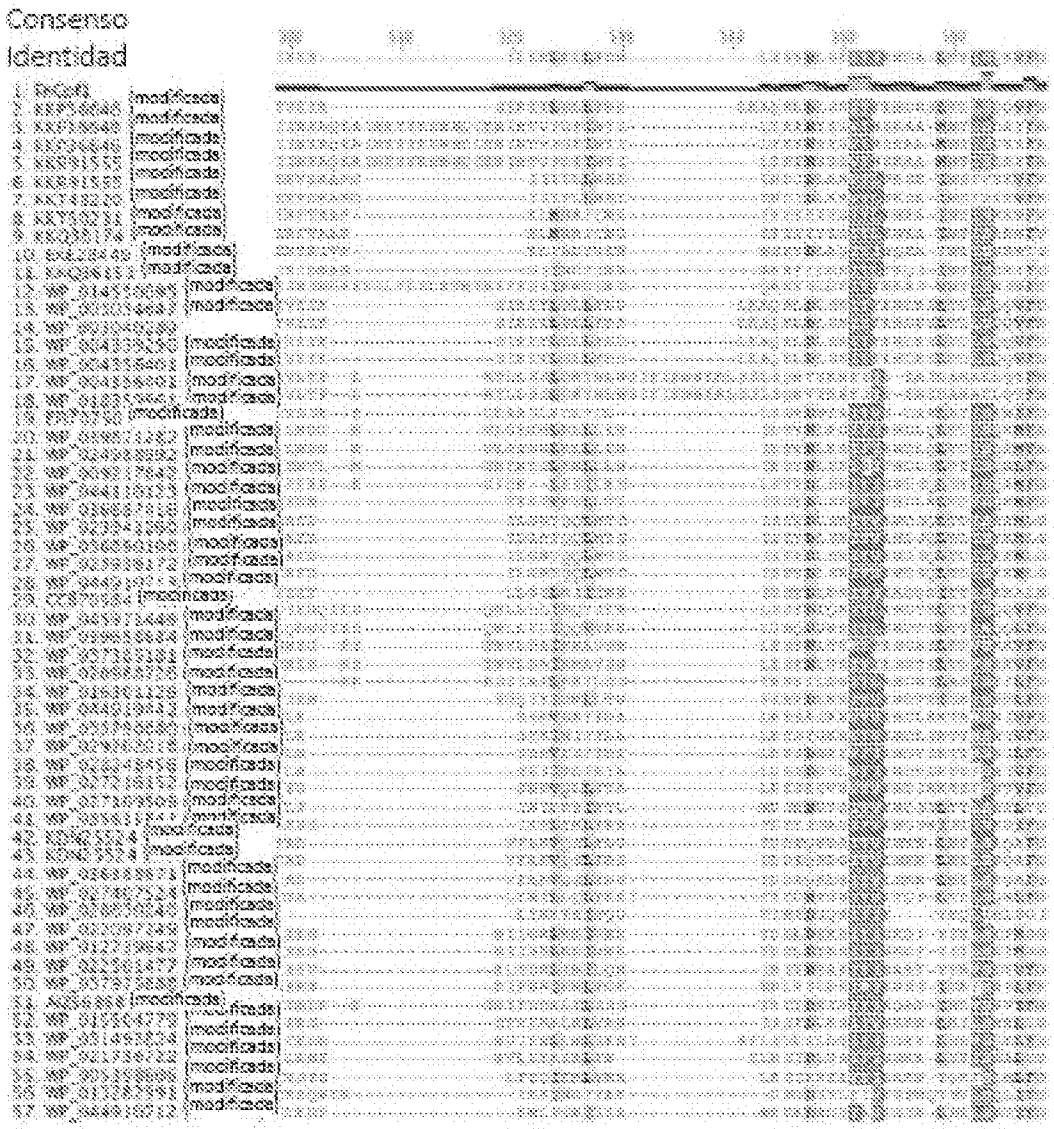


FIG. 38H

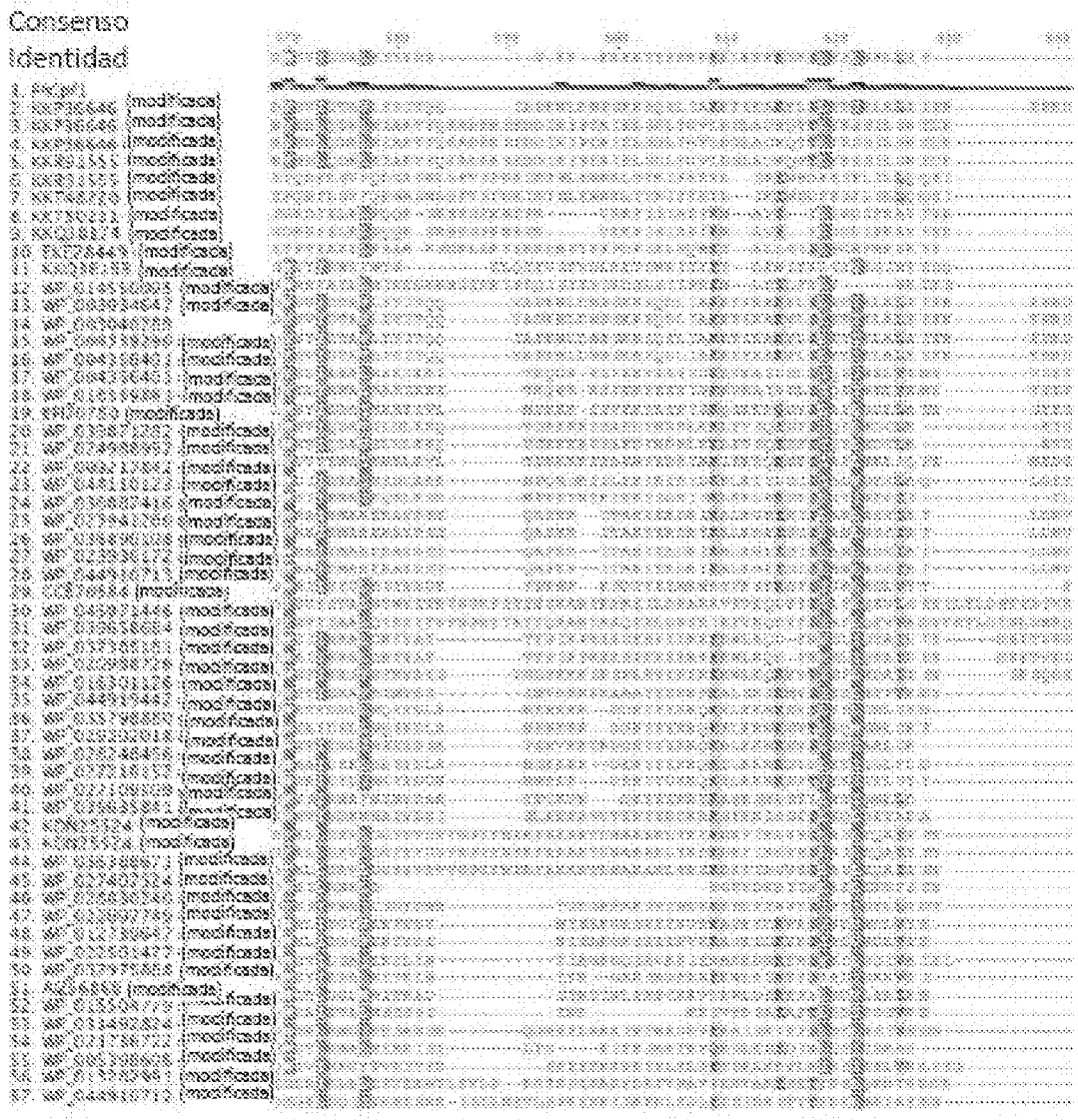


FIG. 38I

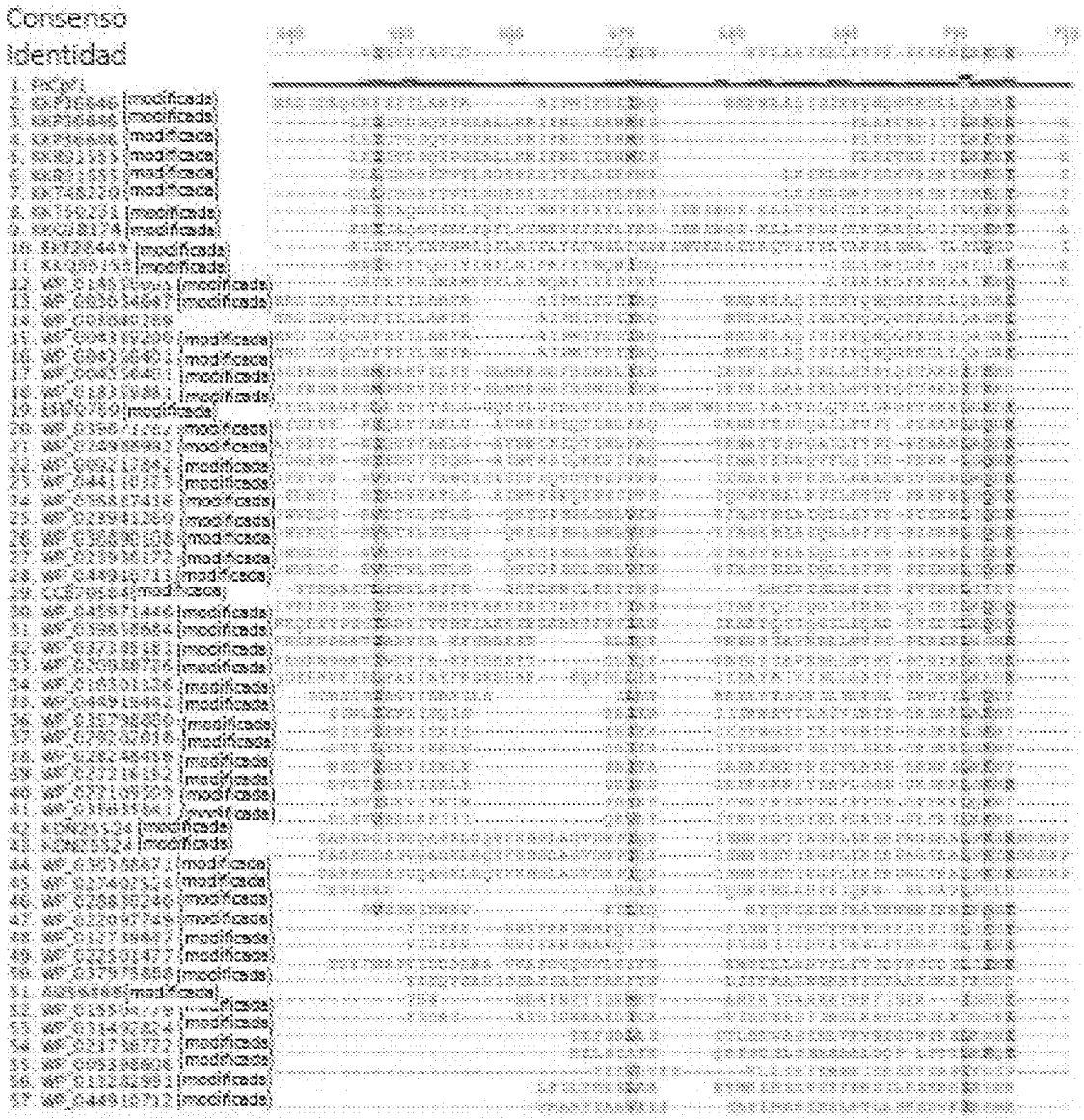


FIG. 38J

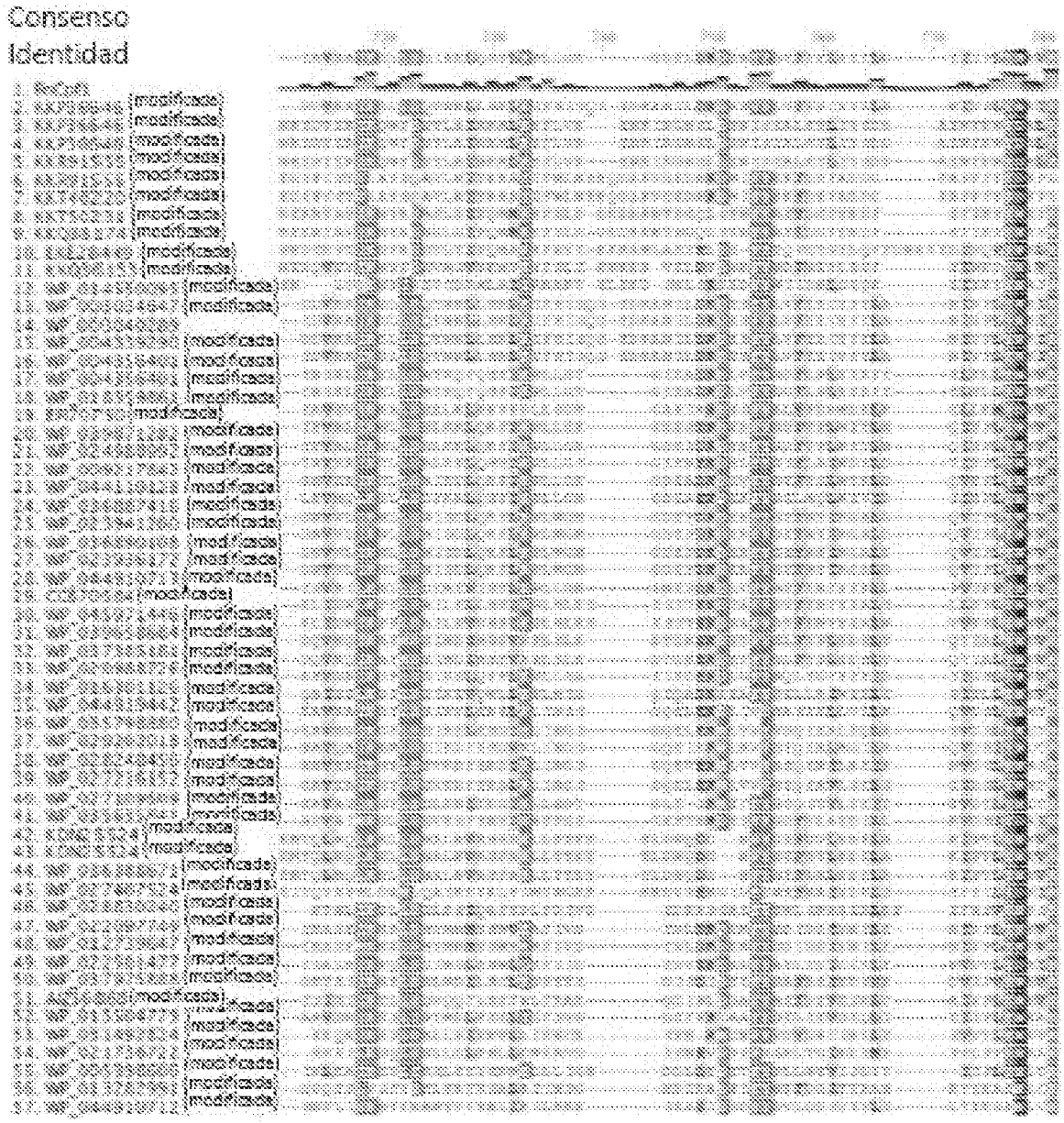


FIG. 38K

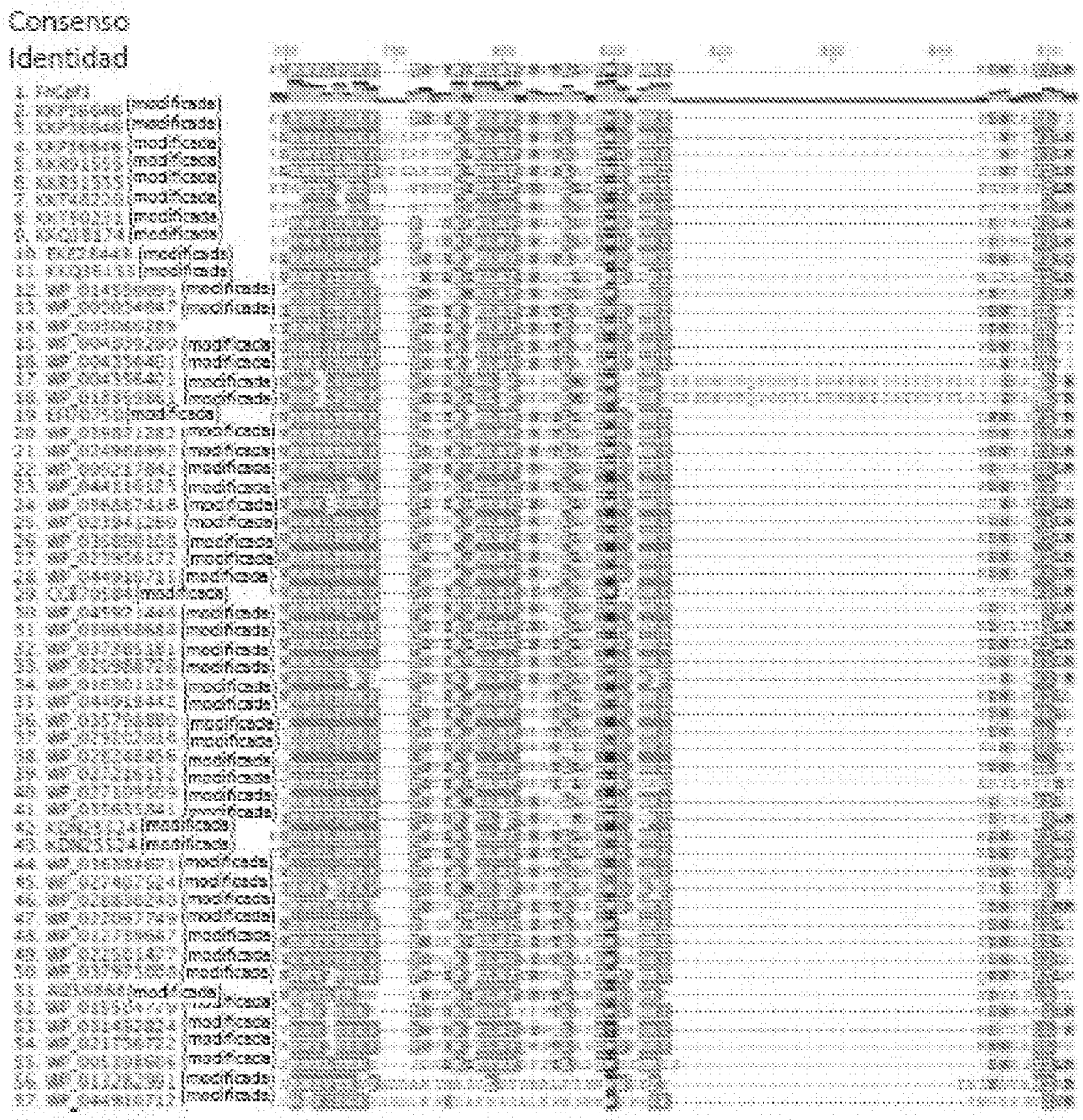


FIG. 38L

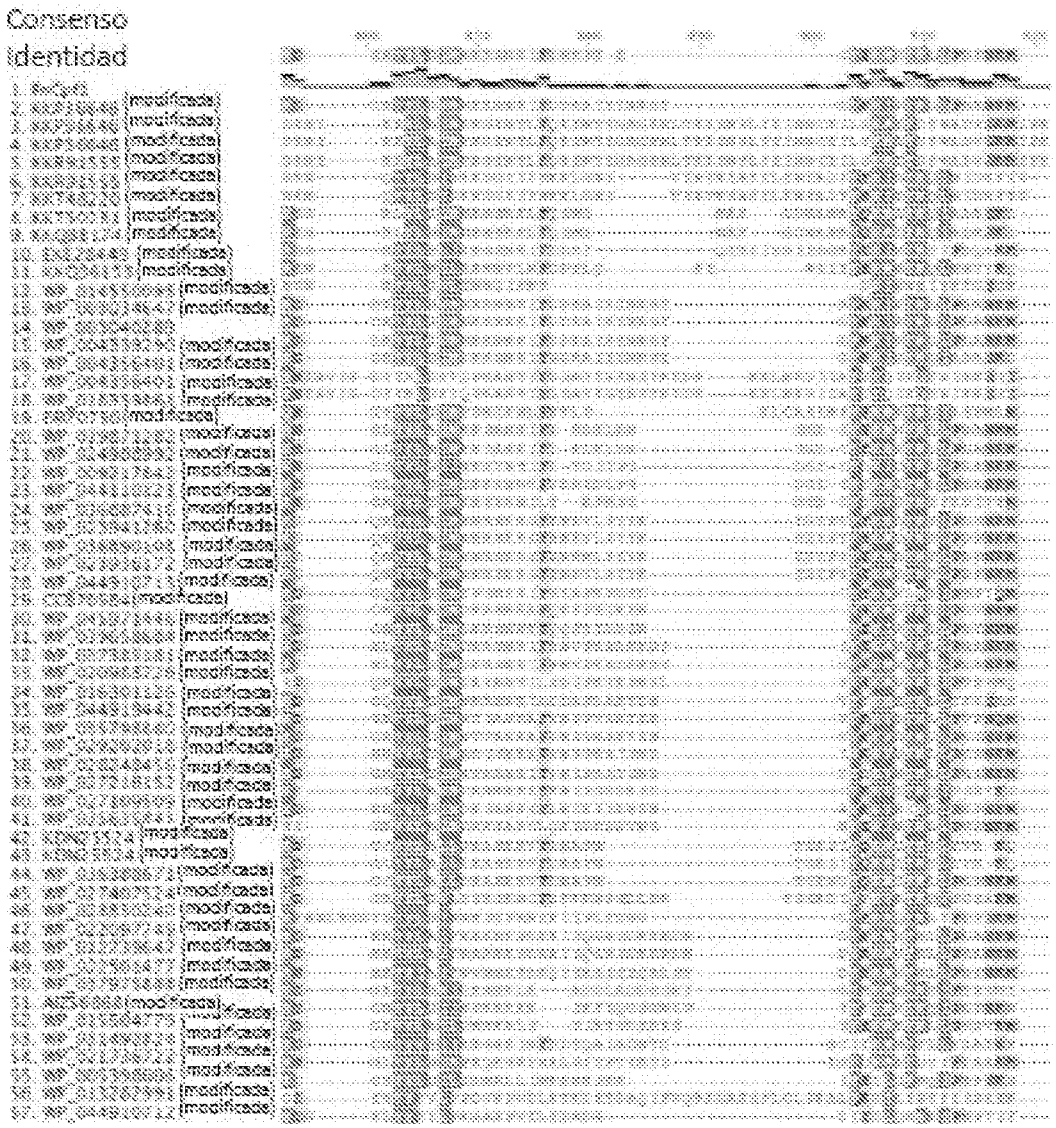


FIG. 38M

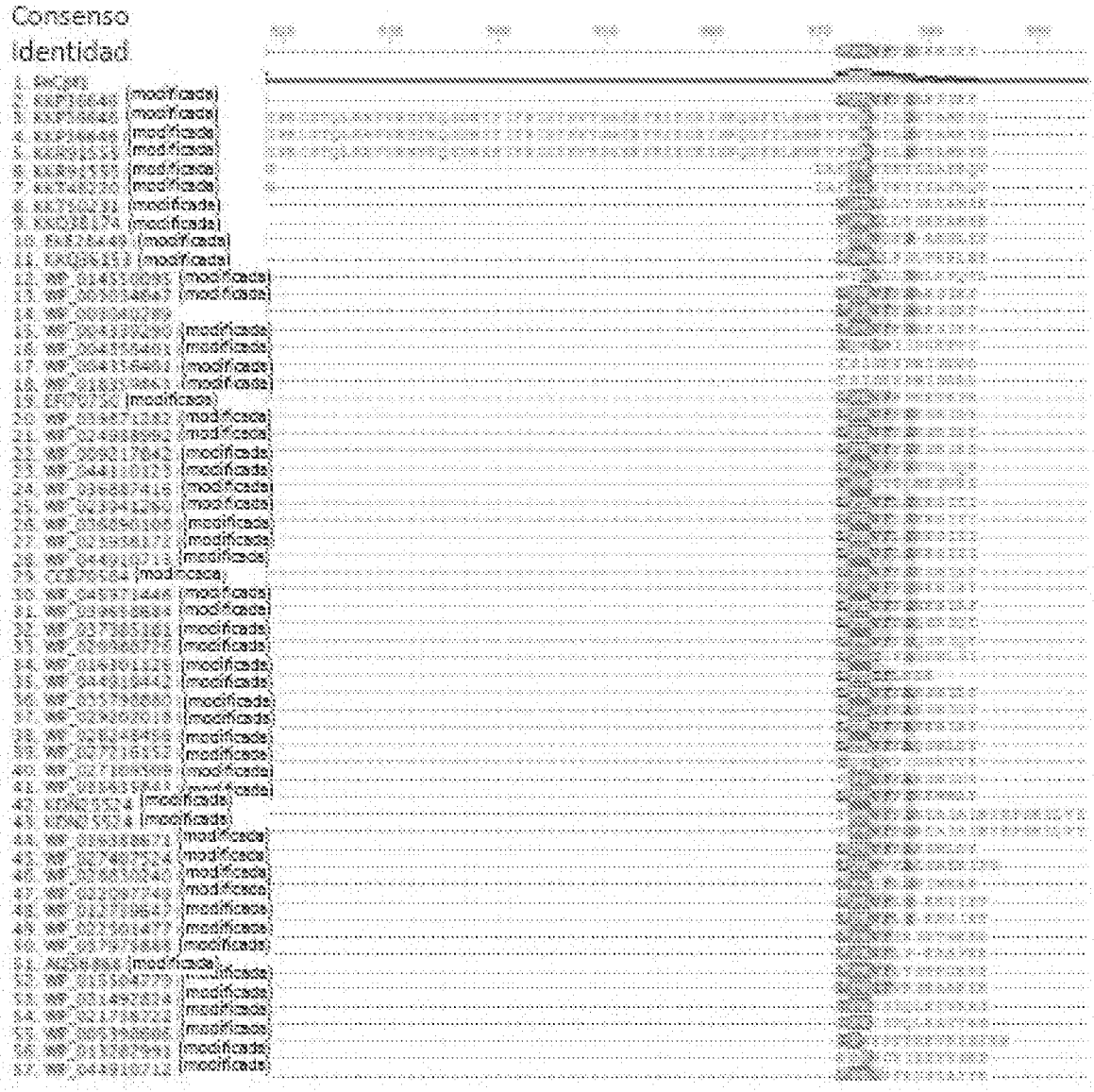


FIG. 38N

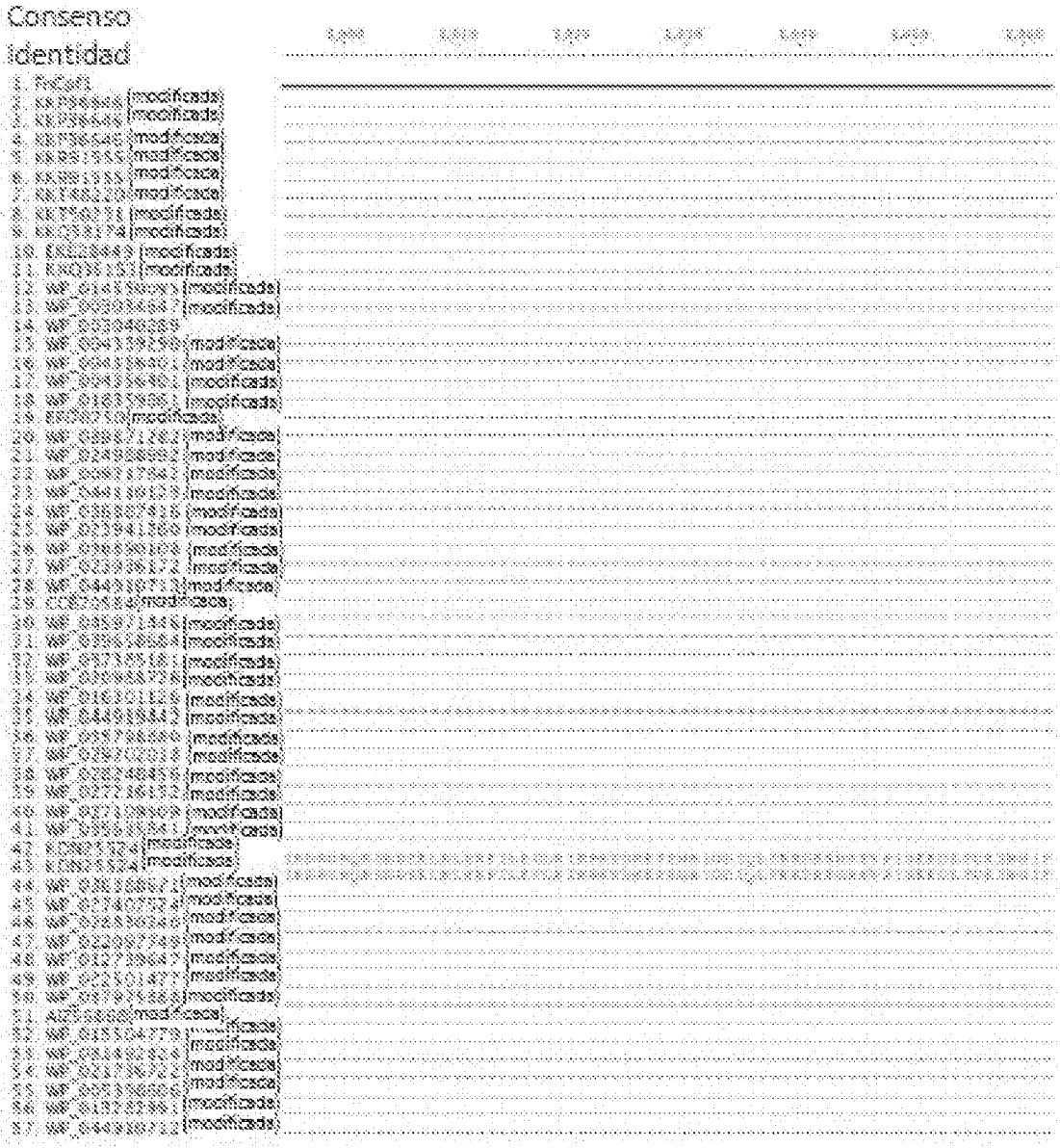


FIG. 38O

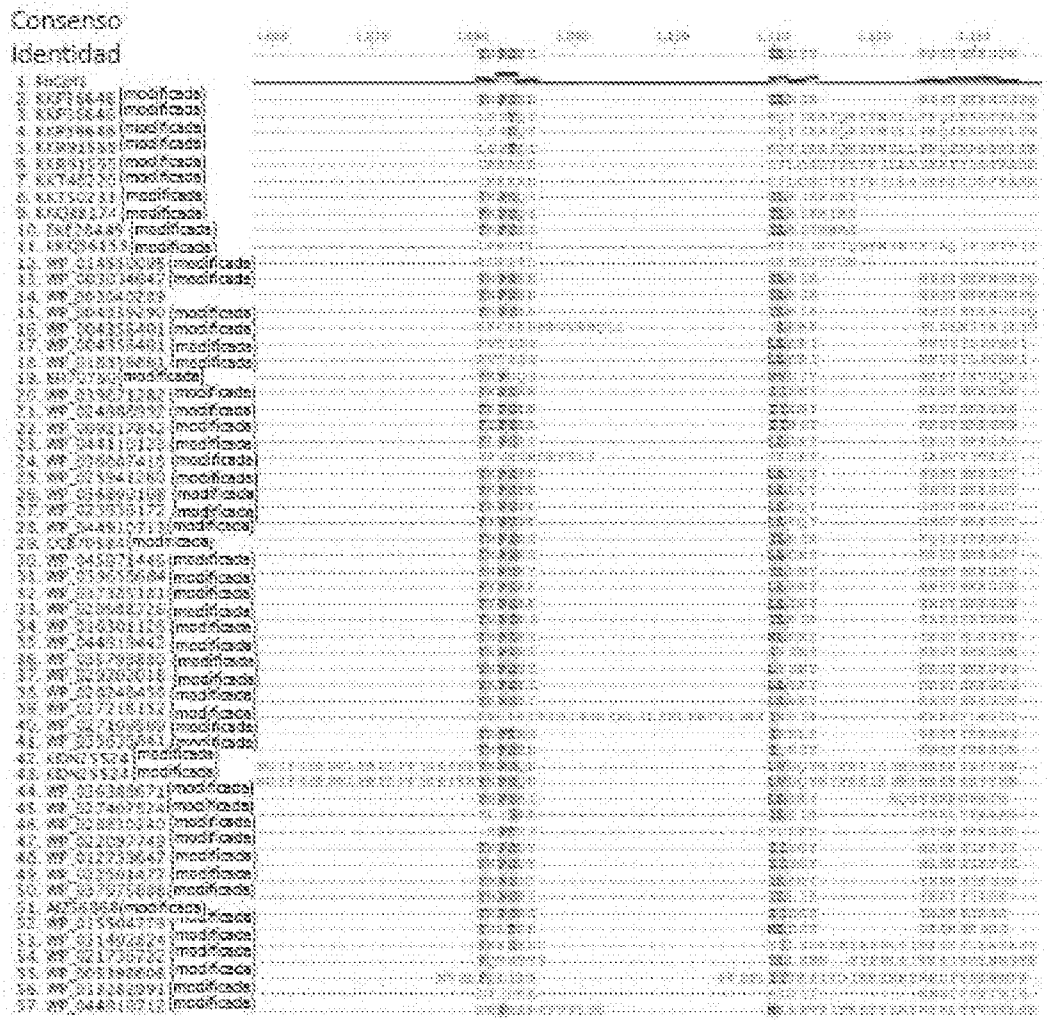


FIG. 38P

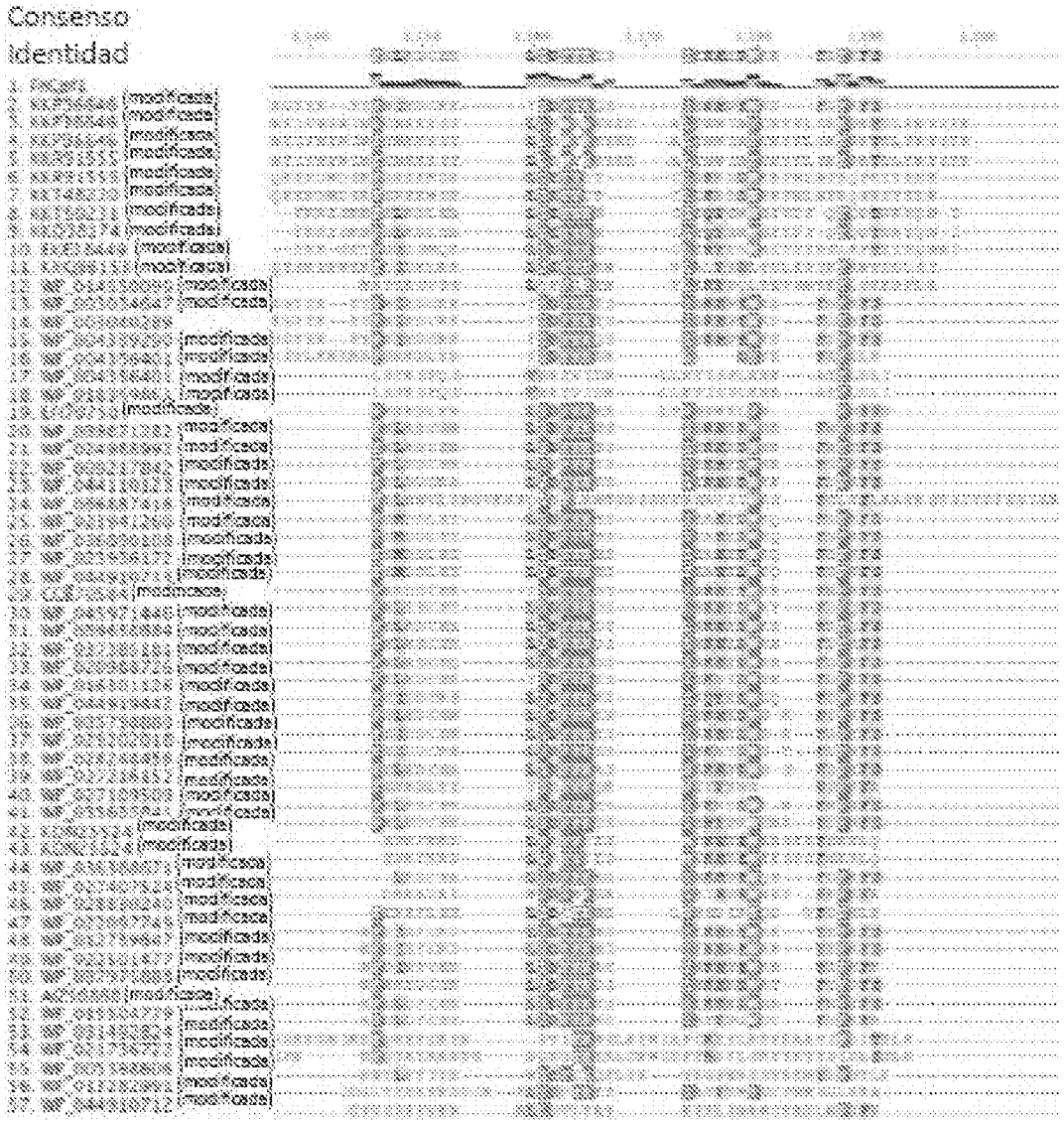


FIG. 38Q

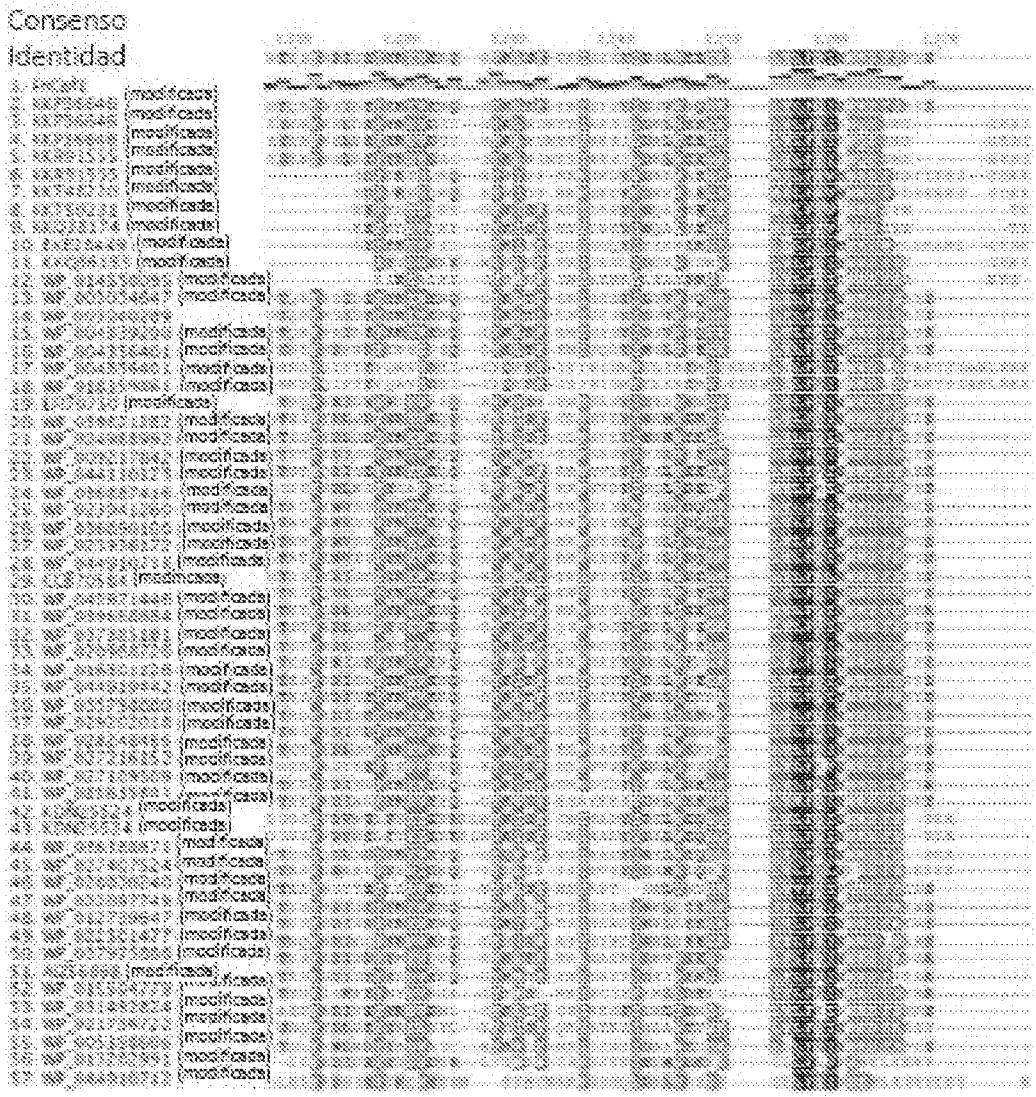


FIG. 38R

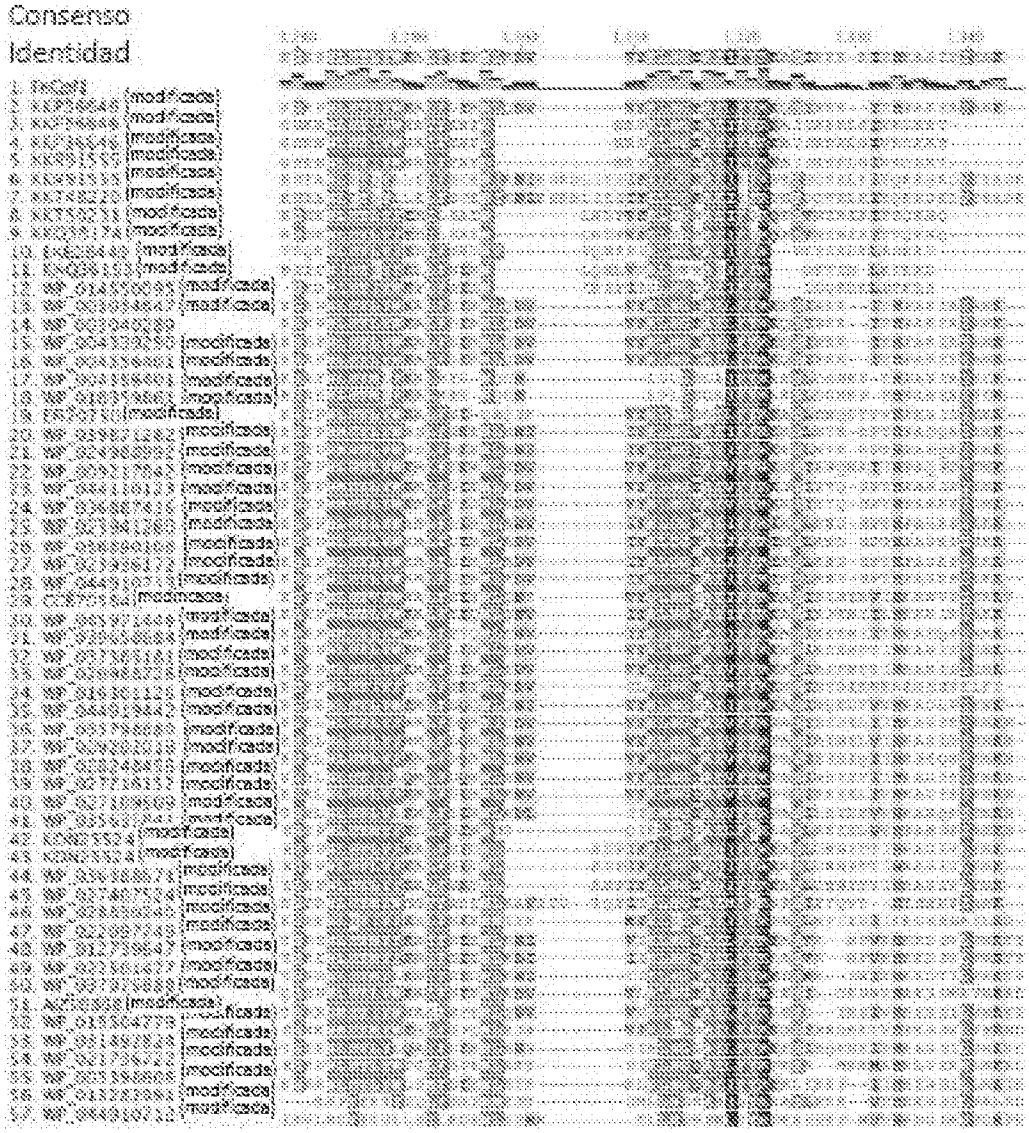


FIG. 38S

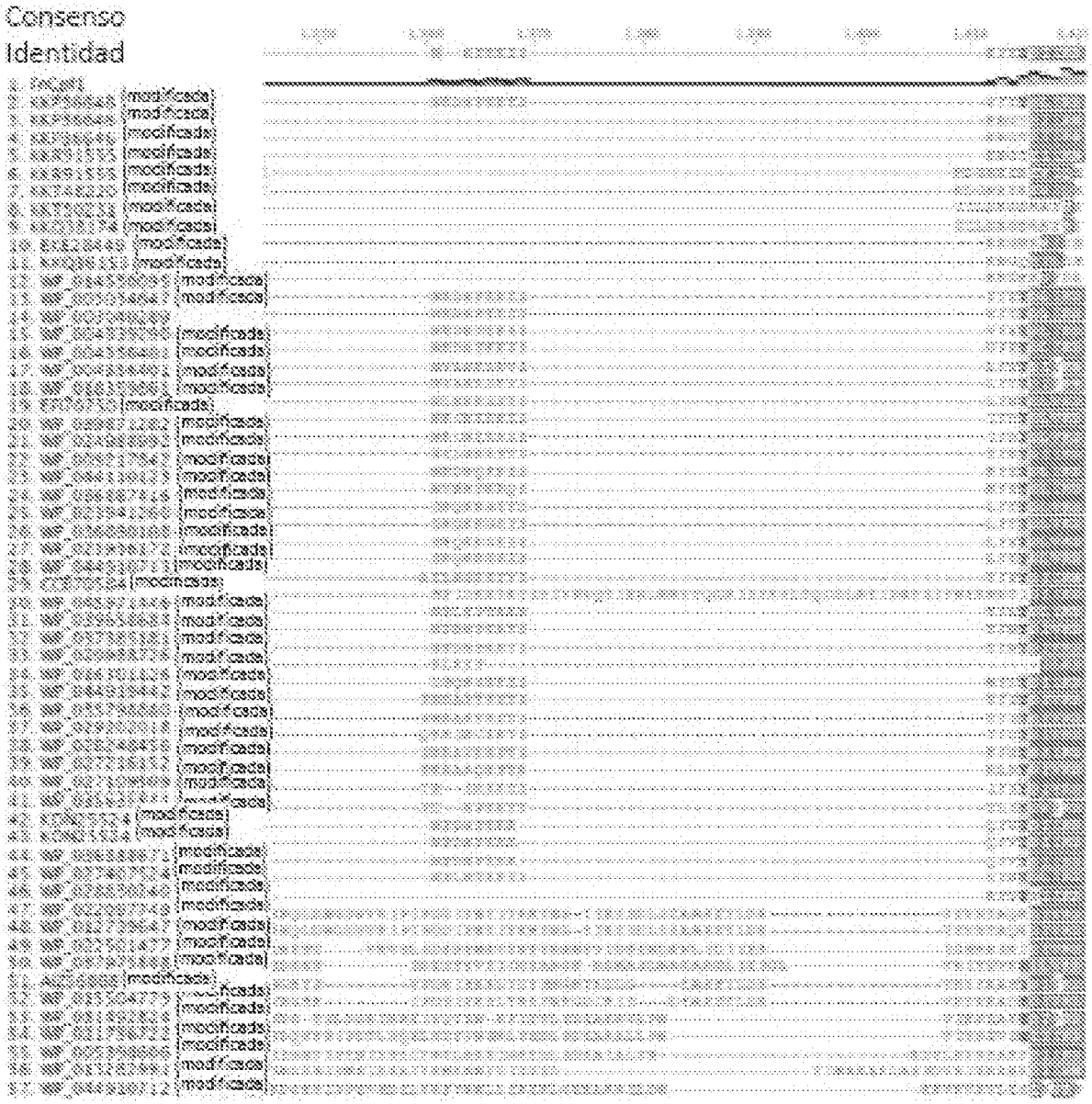


FIG. 38T

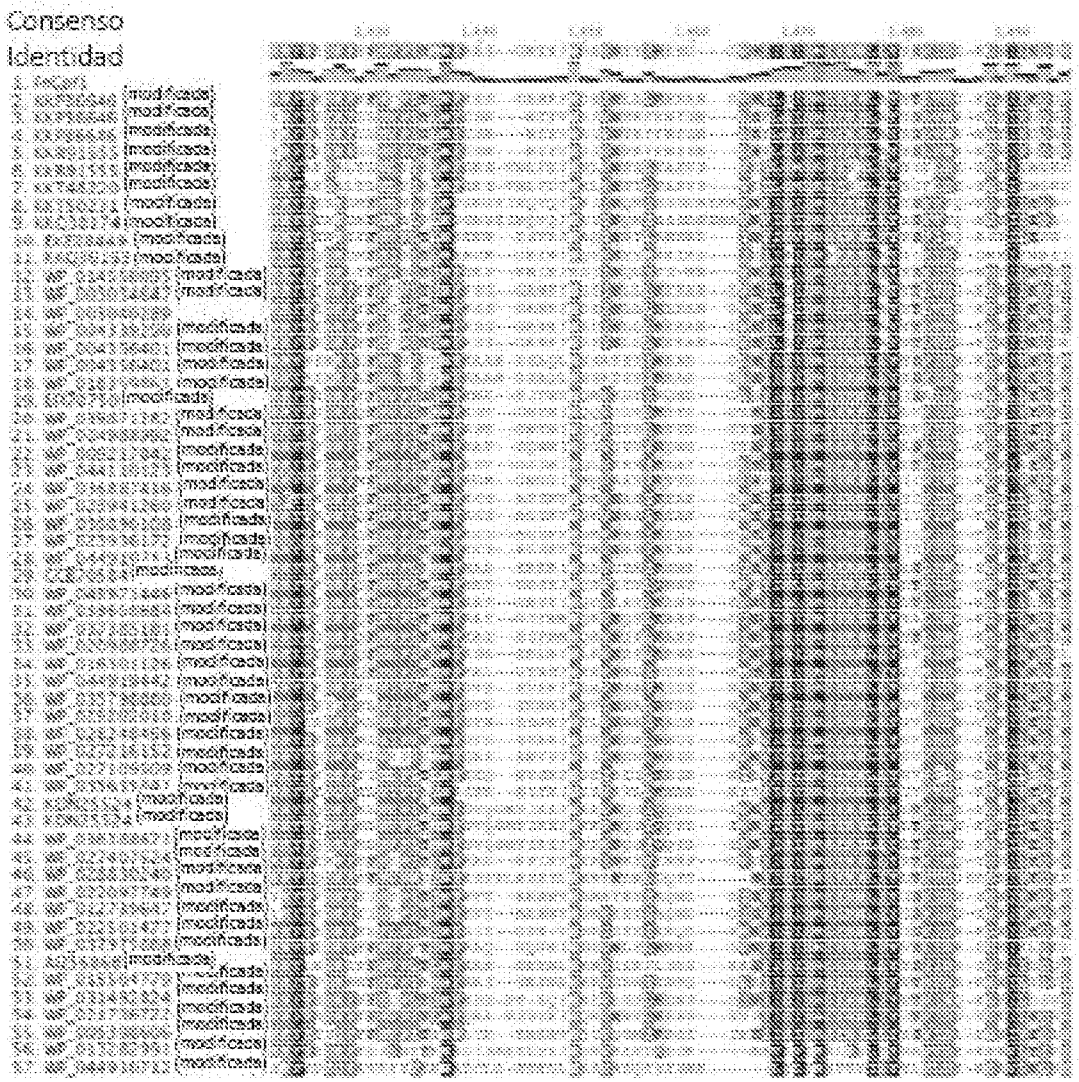


FIG. 38U

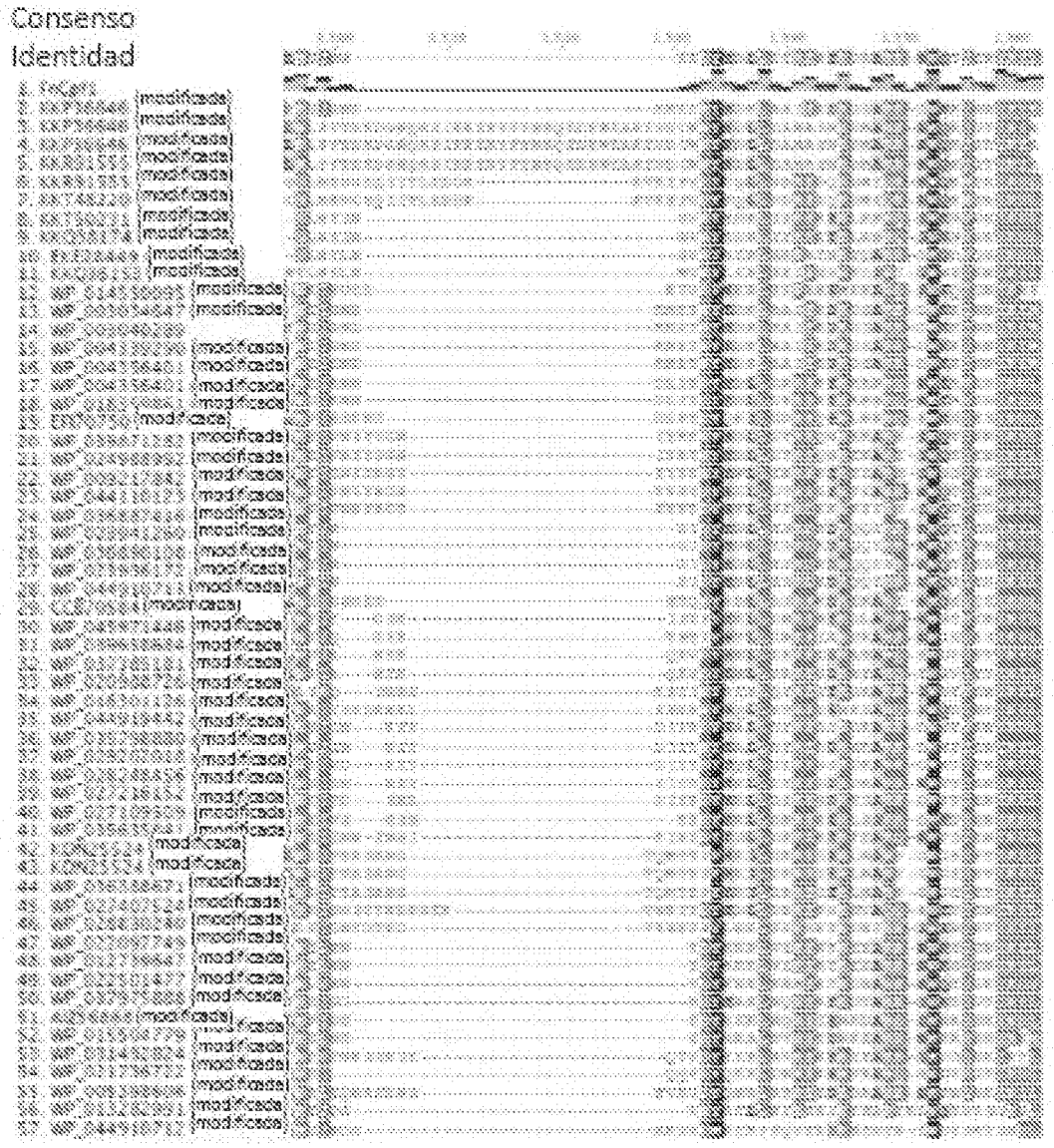


FIG. 38V

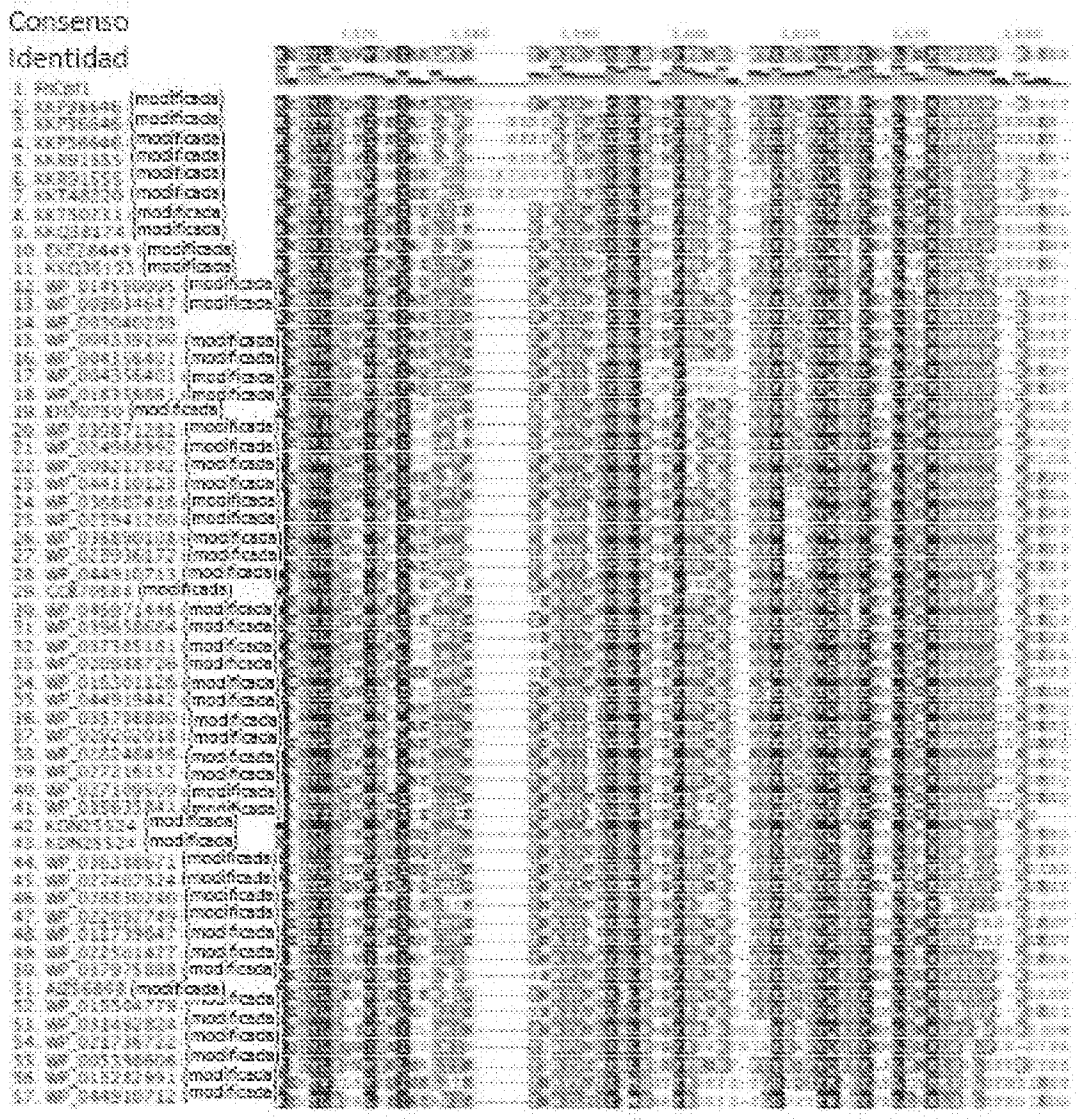


FIG. 38W

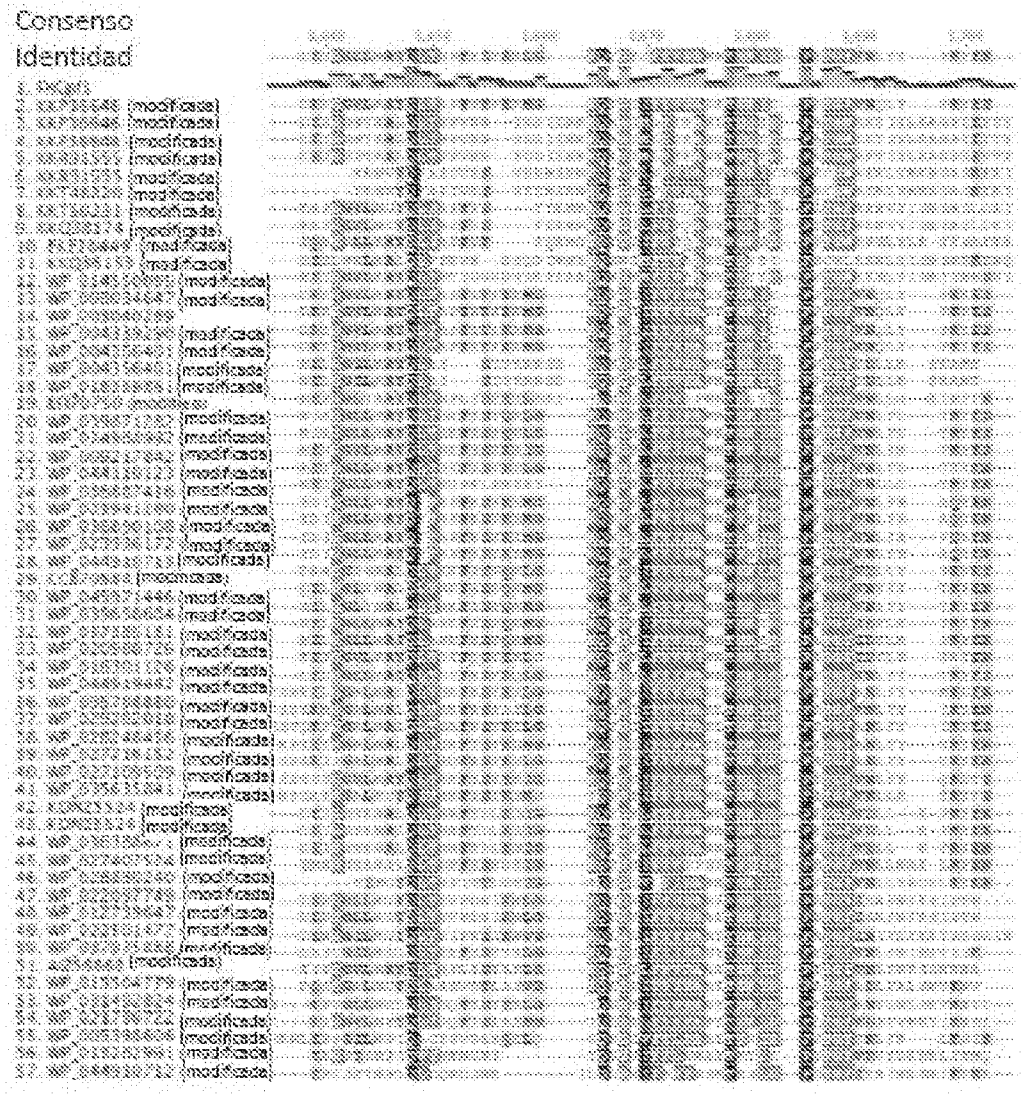


FIG. 38X

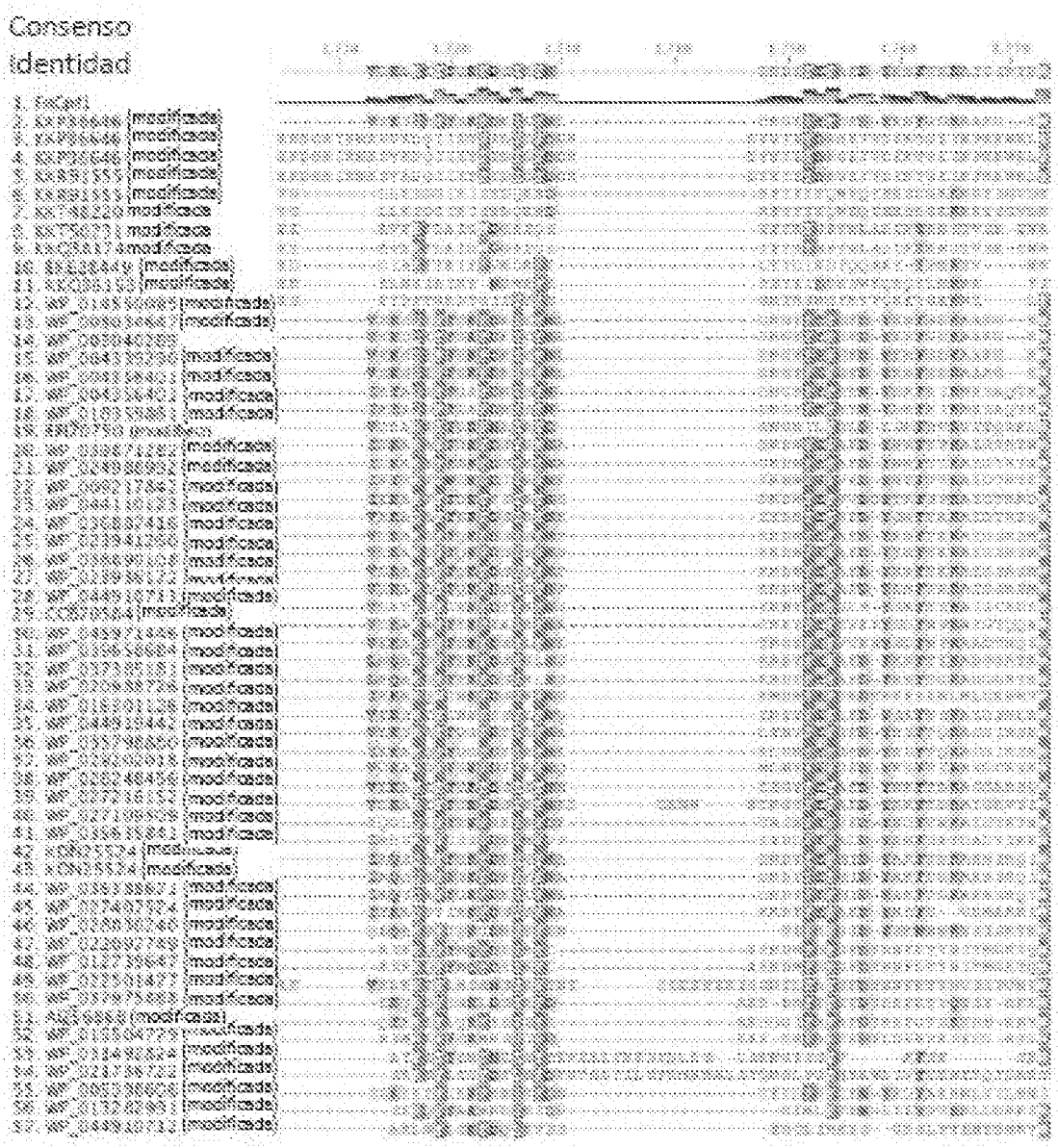


FIG. 38Y

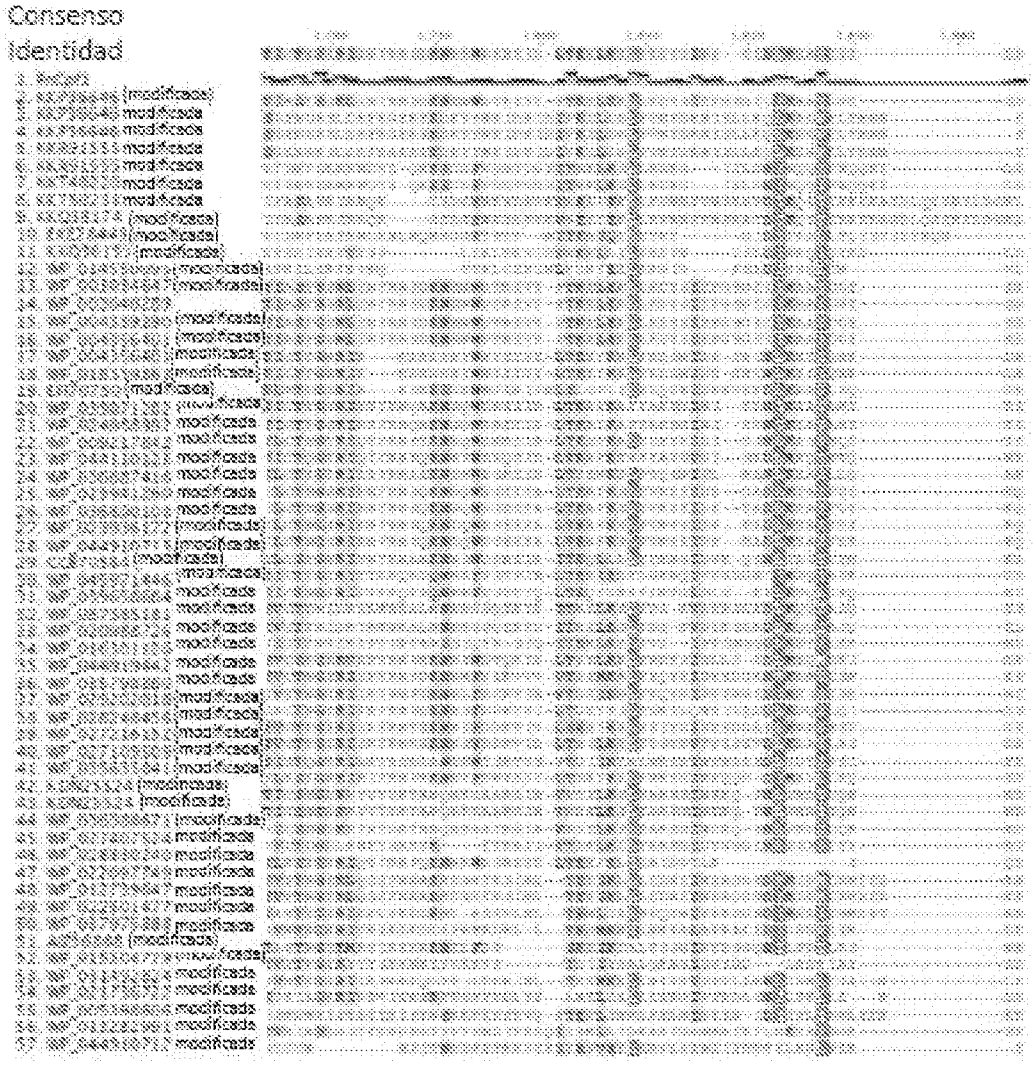


FIG. 38Z

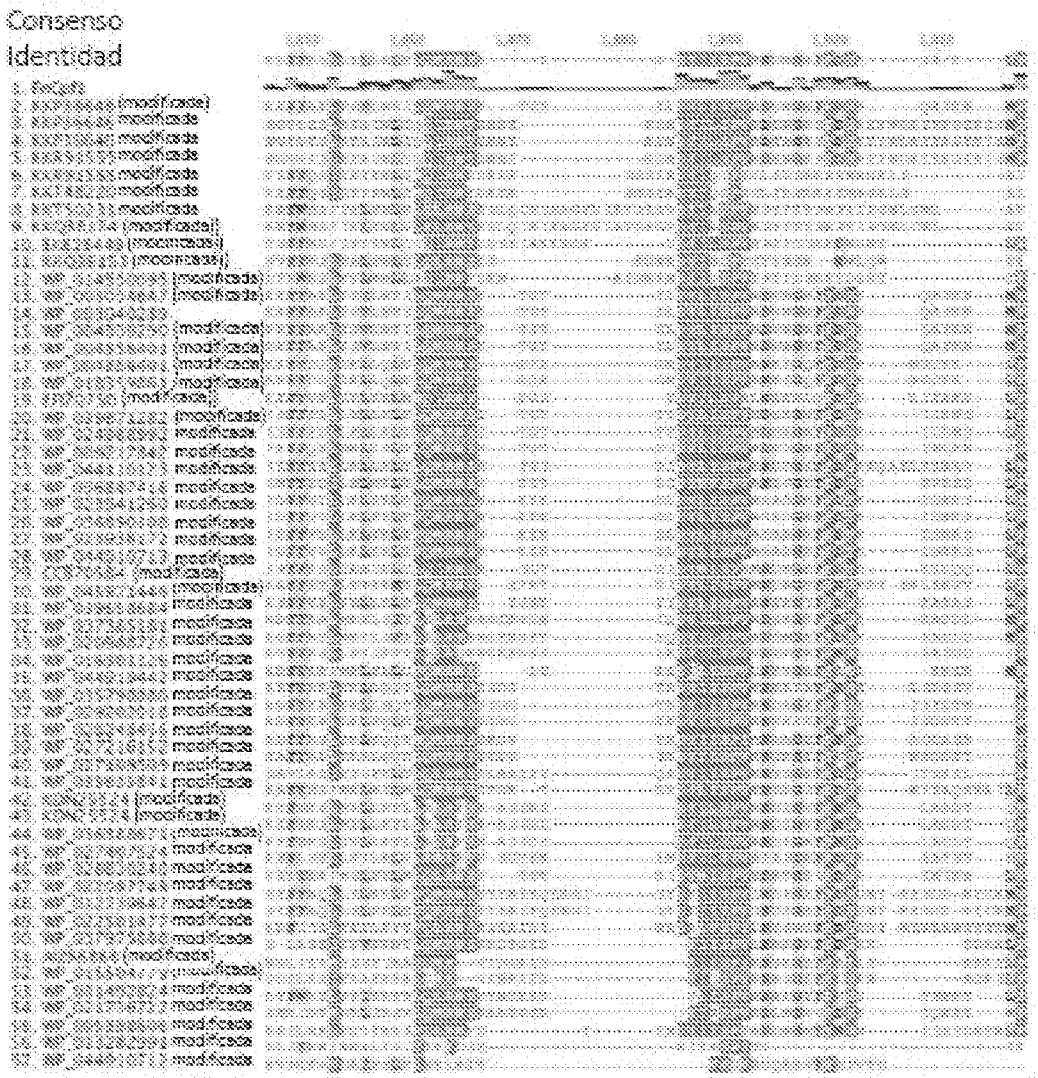


FIG. 38AA

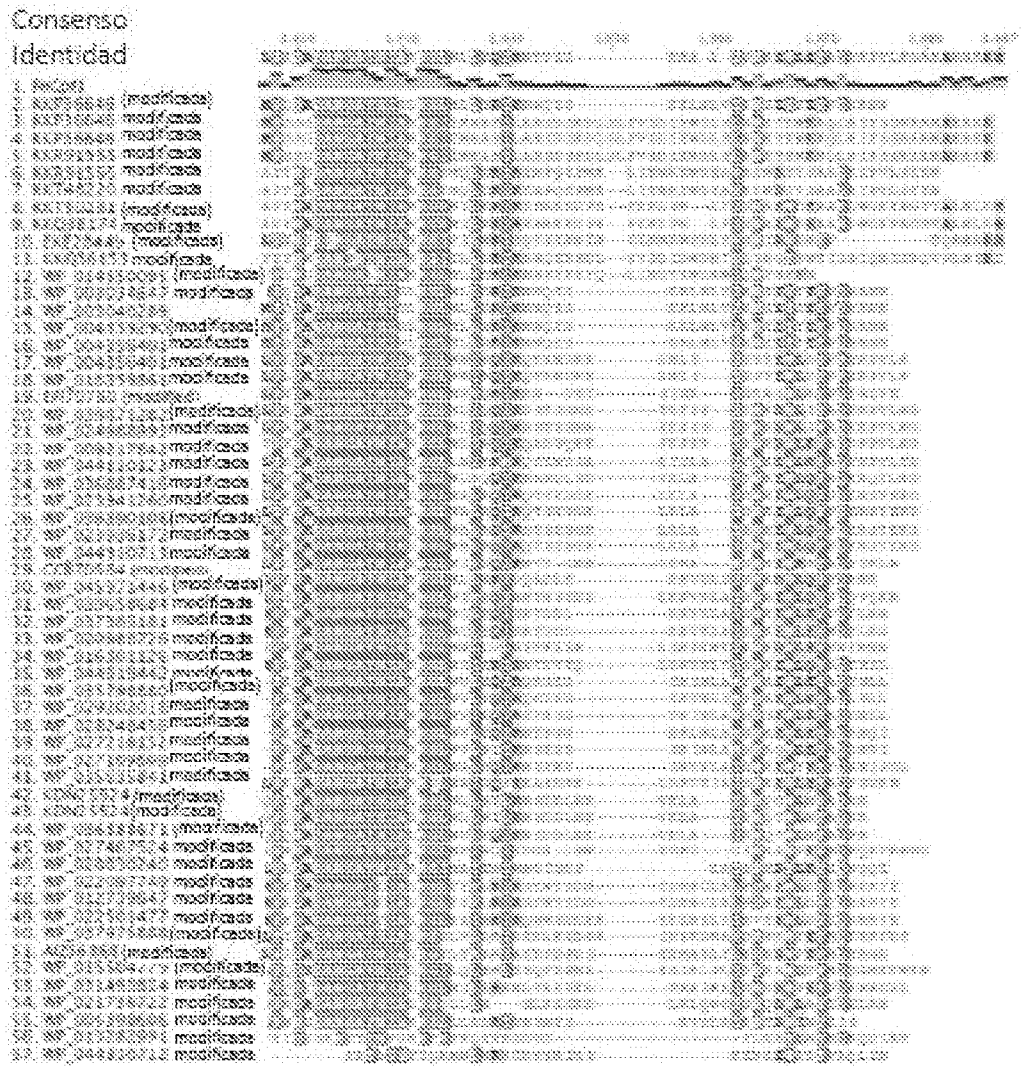


FIG. 38BB

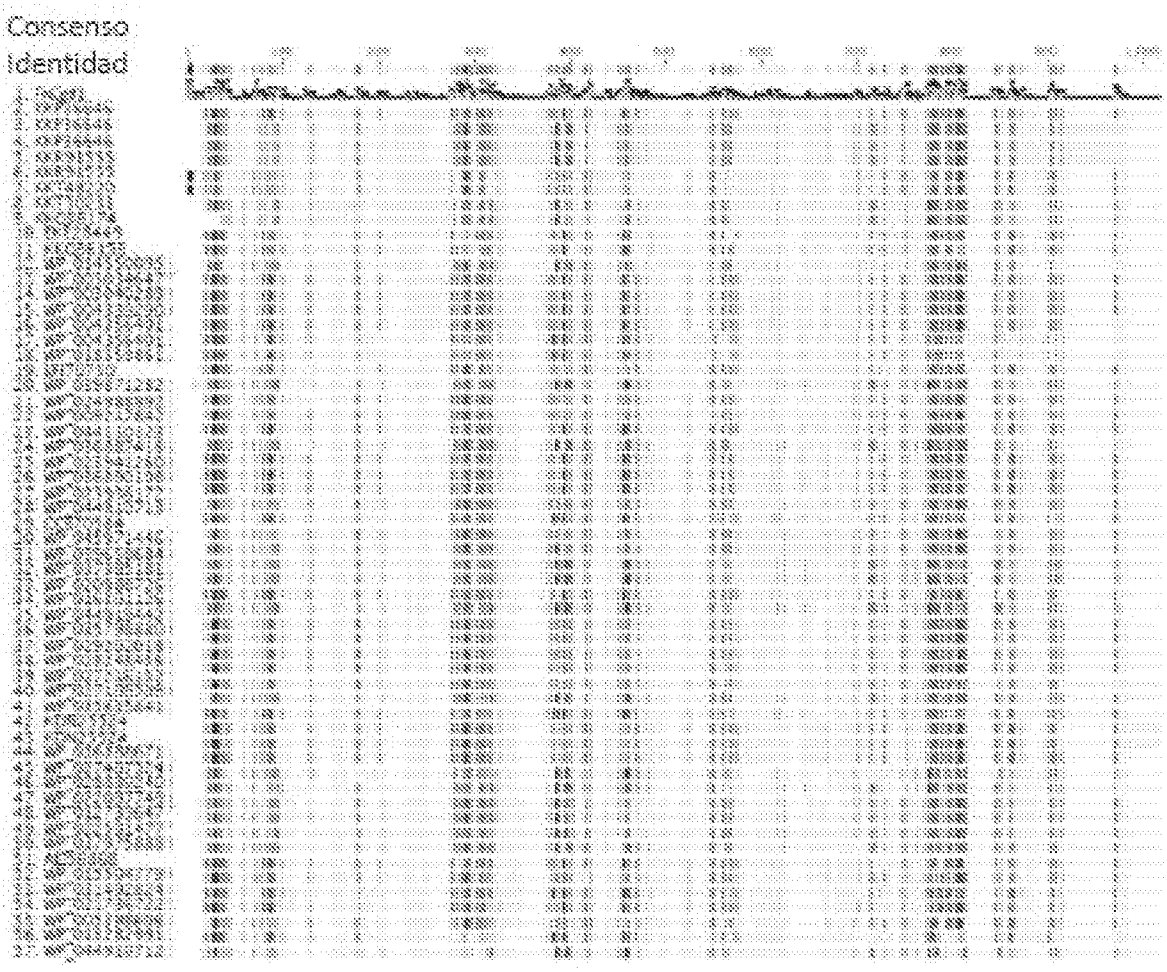


FIG. 39A

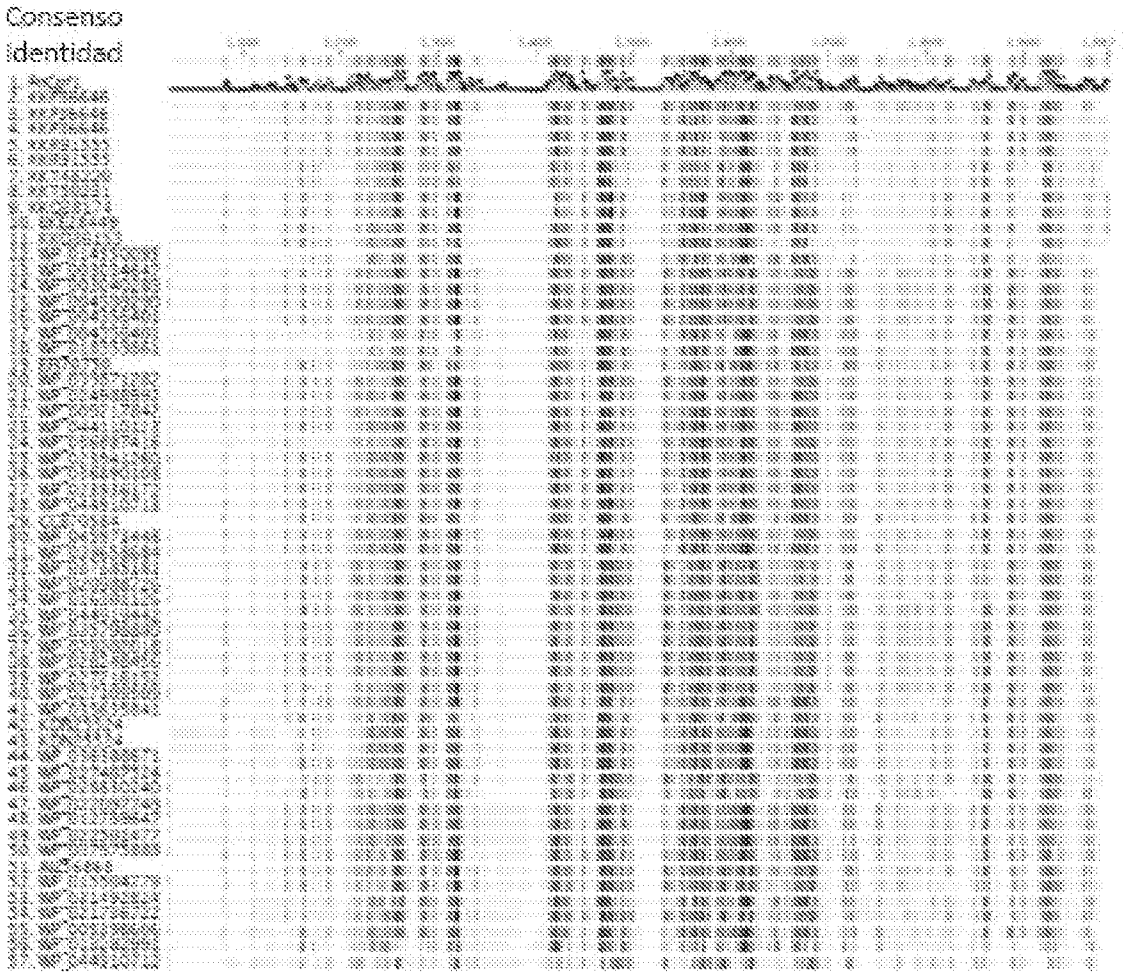


FIG. 39B

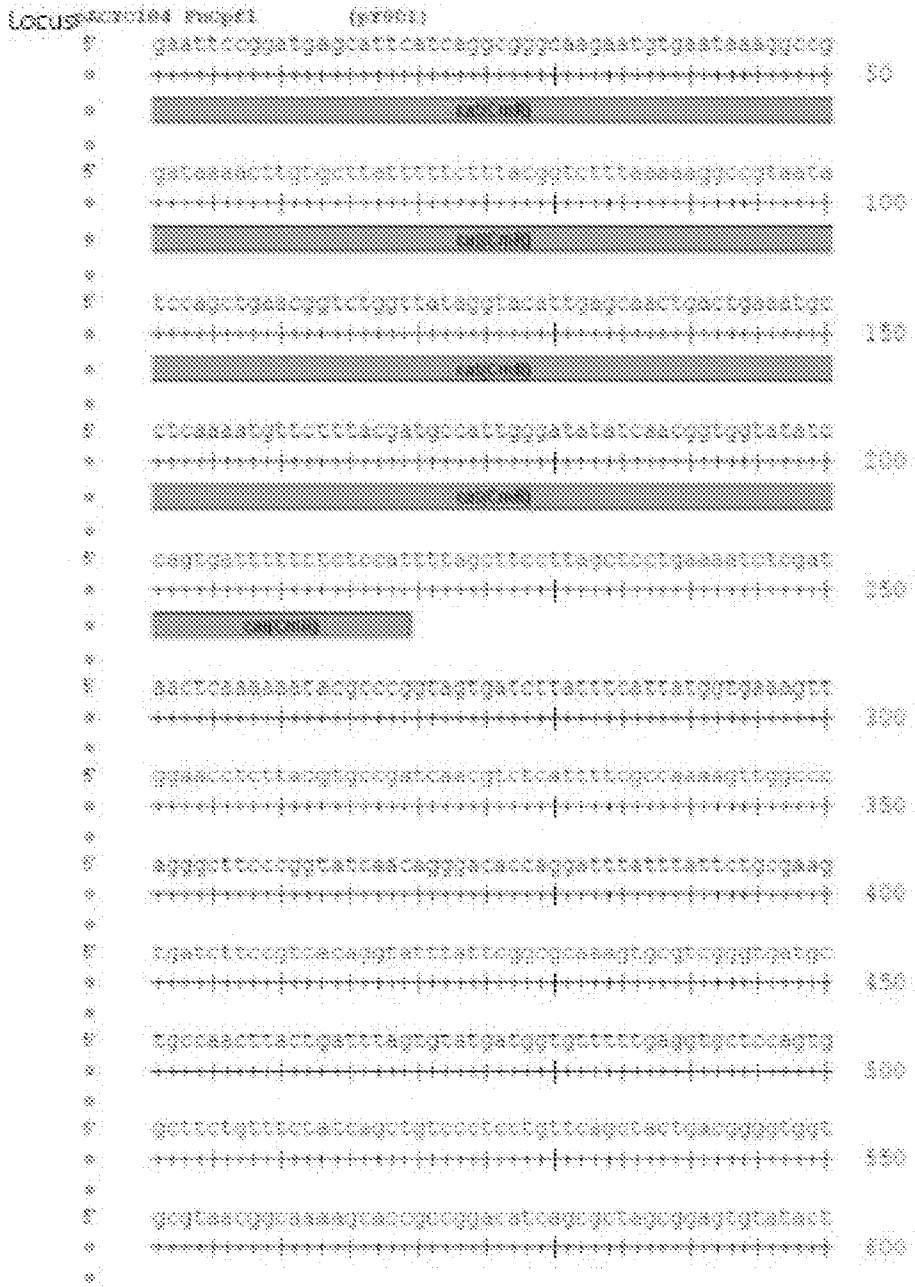


FIG. 40A

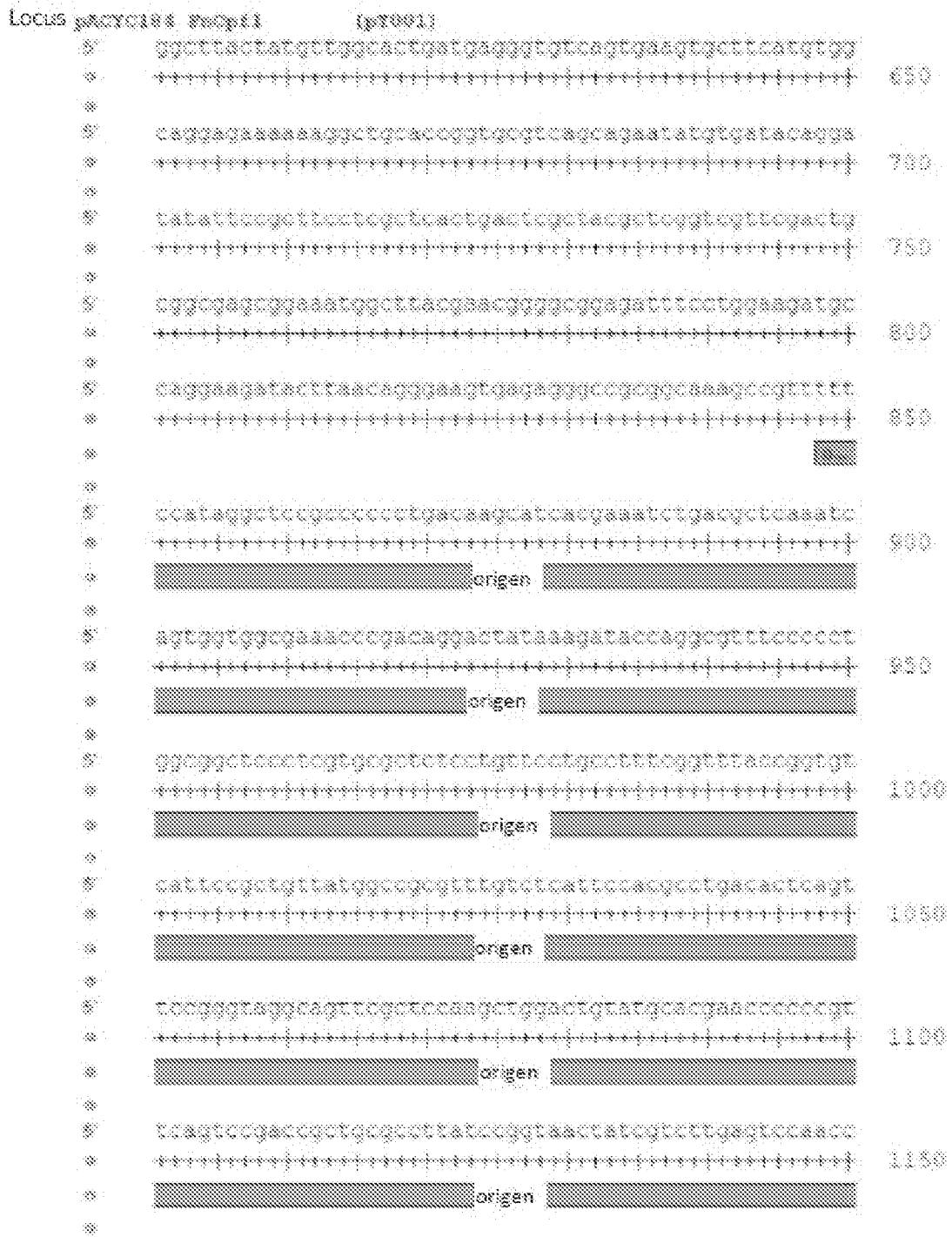

Locus pACYC184 BmOpt1 (pT001)
650
700
750
800
850
900
origen
950
origen
1000
origen
1050
origen
1100
origen
1150
origen

FIG. 40B

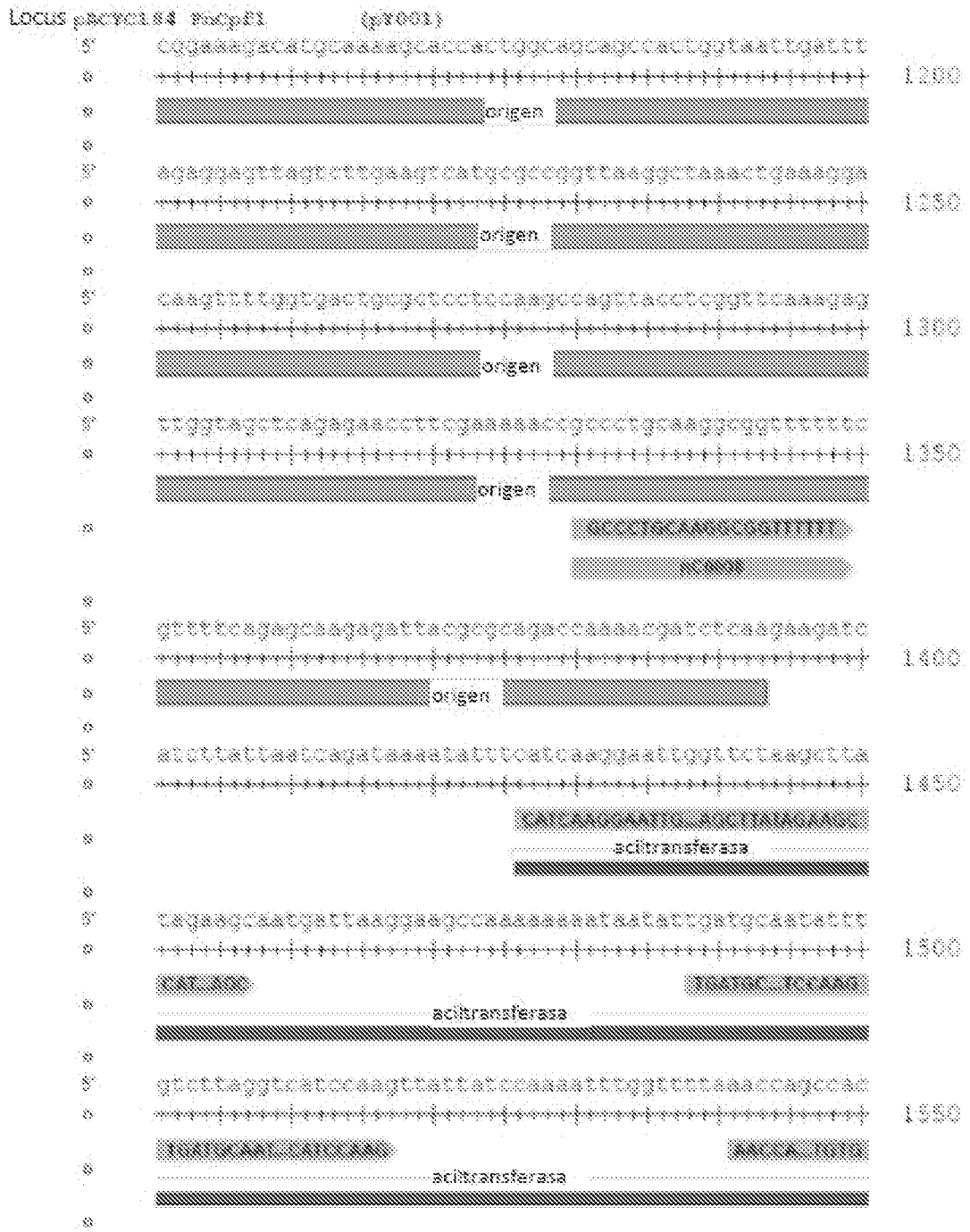

origen
origen
origen
origen
aciltransferasa
aciltransferasa
aciltransferasa

FIG. 40C

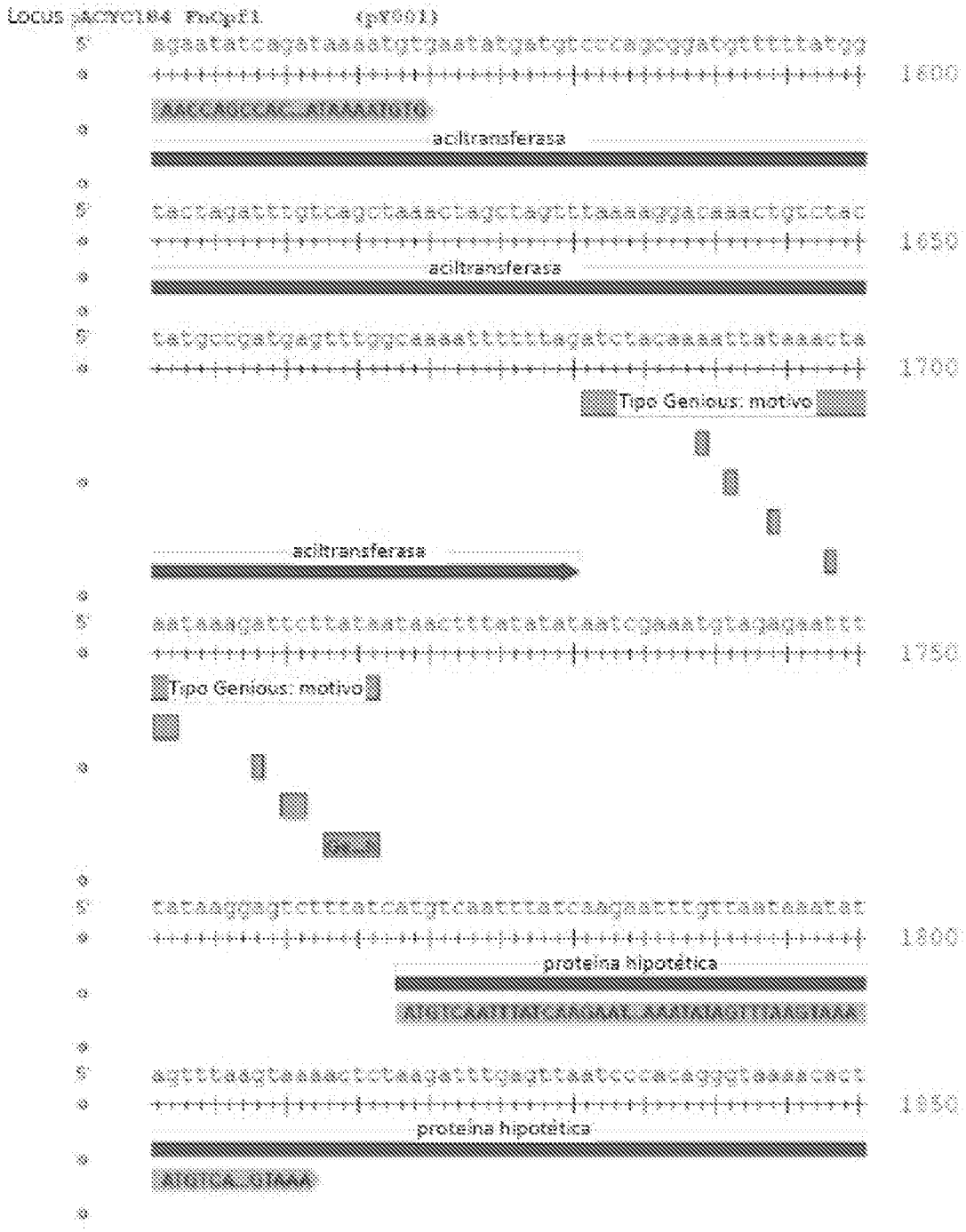

aciltransferasa
aciltransferasa
Tipo Genious: motivo
Tipo Genious: motivo
proteína hipotética
proteína hipotética

FIG. 40D

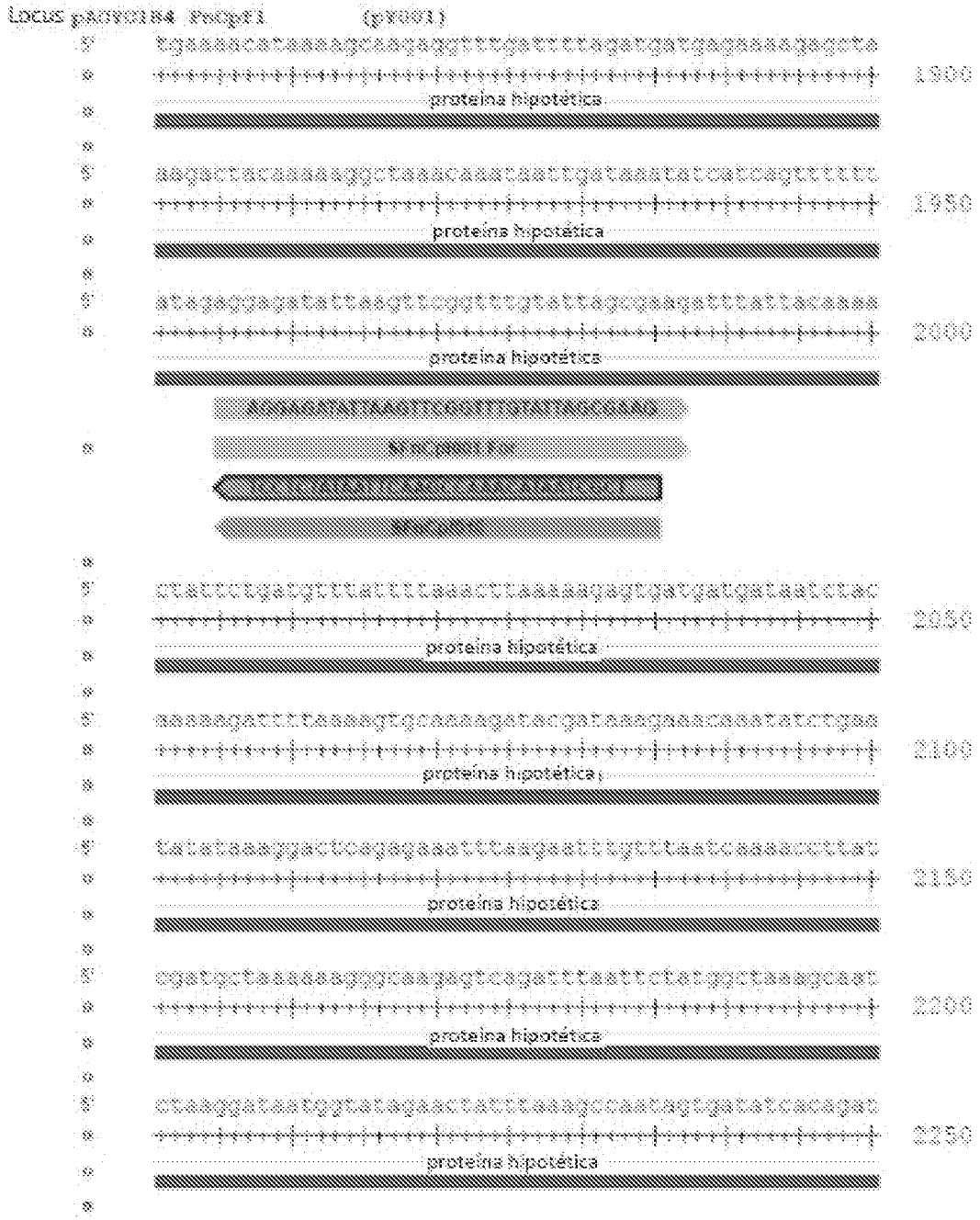

1300
proteína hipotética
1350
proteína hipotética
2000
proteína hipotética
2050
proteína hipotética
2100
proteína hipotética
2150
proteína hipotética
2200
proteína hipotética
2250
proteína hipotética

FIG. 40E

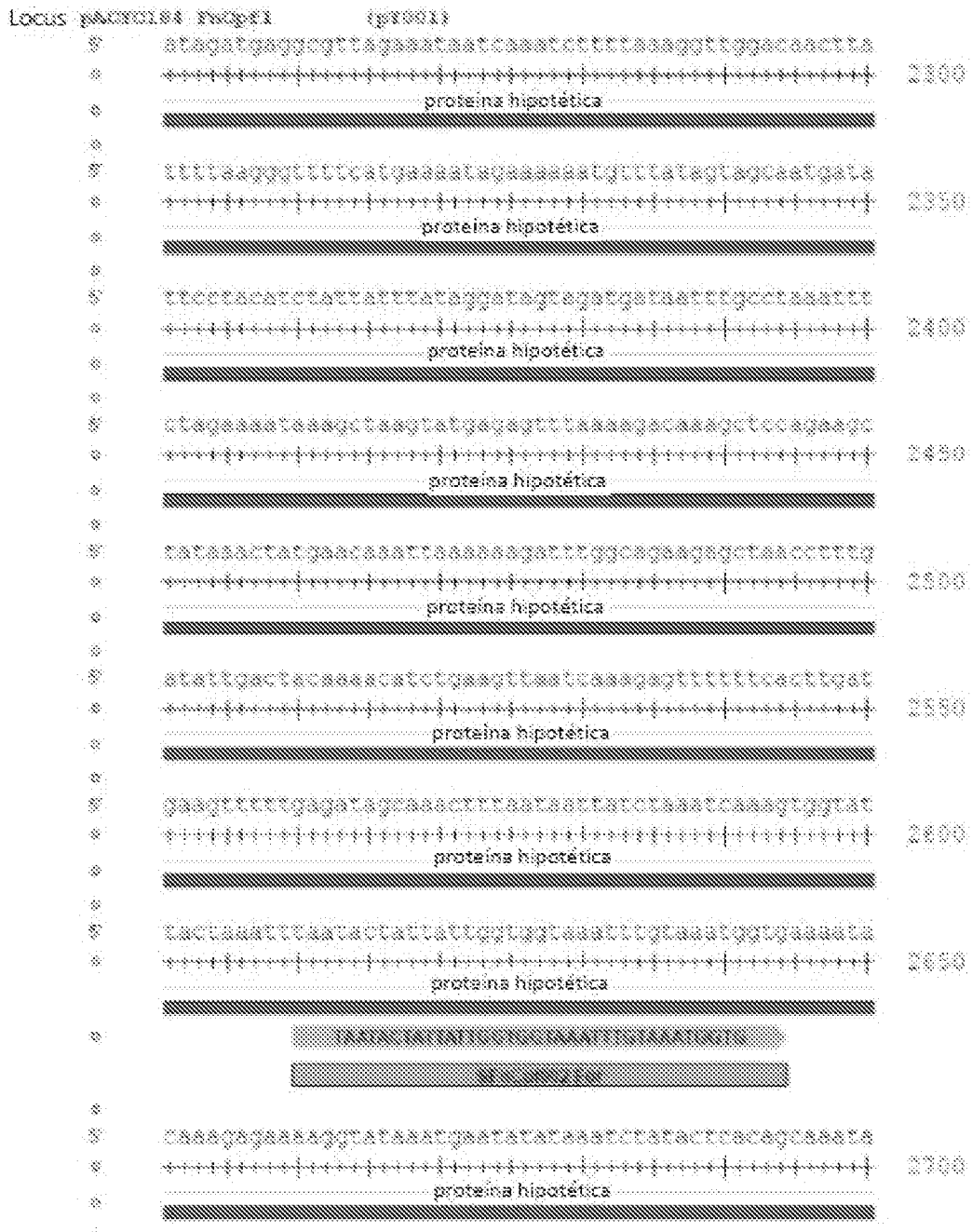

proteína hipotética
2300
2350
2400
2450
2500
2550
2600
2650
2700

FIG. 40F

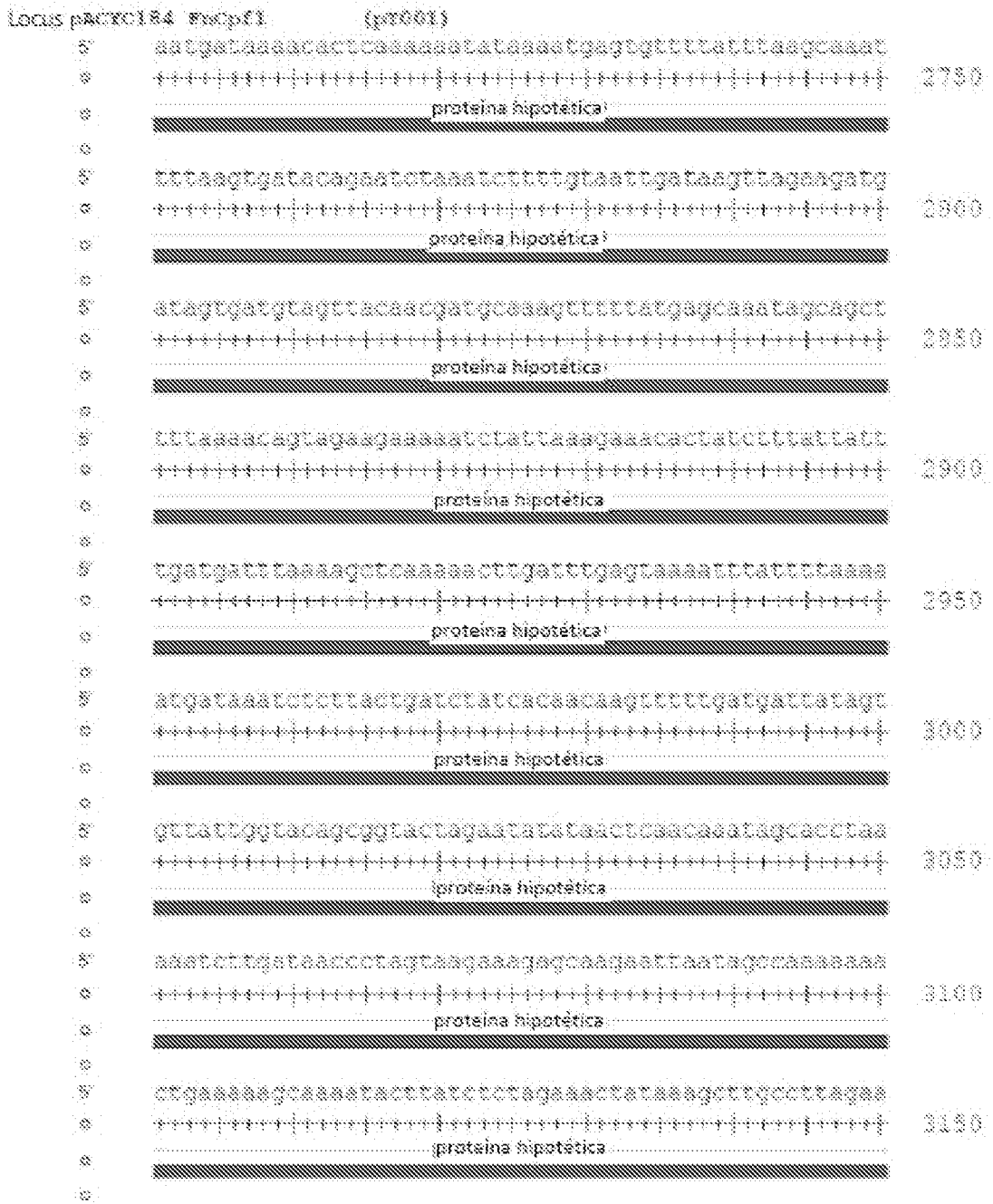

Locus pACYC184 rscpf1 (pt001)
2750
proteína hipotética
2800
proteína hipotética
2850
proteína hipotética
2900
proteína hipotética
2950
proteína hipotética
3000
proteína hipotética
3050
proteína hipotética
3100
proteína hipotética
3150
proteína hipotética

FIG. 40G

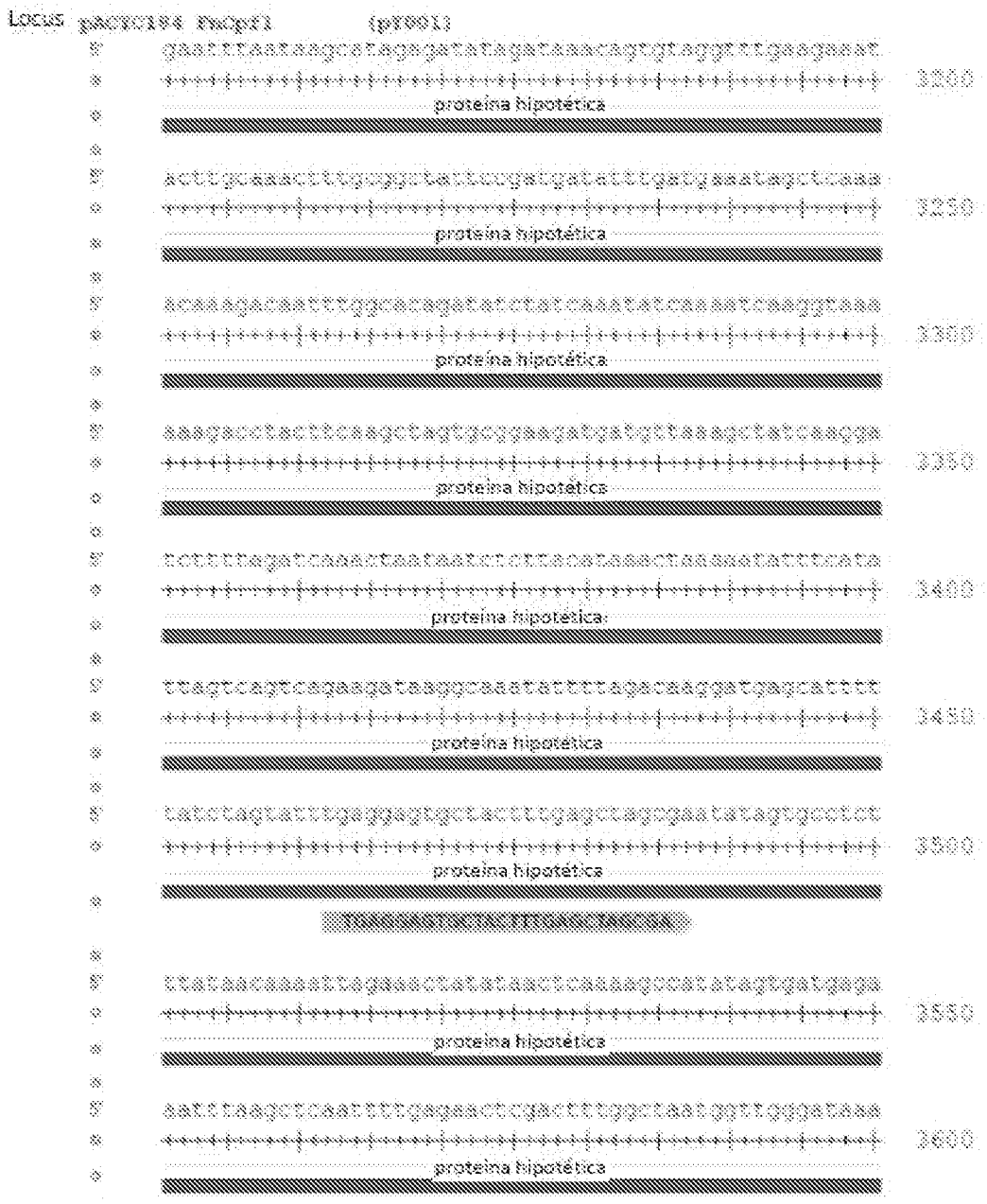

Locus pAC3C194 5462pt1 (pIP501)
3200
proteína hipotética
3250
proteína hipotética
3300
proteína hipotética
3350
proteína hipotética
3400
proteína hipotética
3450
proteína hipotética
3500
proteína hipotética
3550
proteína hipotética
3600
proteína hipotética

FIG. 40H


Locus pACYC184 PnlpF1 (pR001)
3650
proteína hipotética
3700
proteína hipotética
3750
proteína hipotética
3800
proteína hipotética
3850
proteína hipotética
3900
proteína hipotética
3950
proteína hipotética
4000
proteína hipotética

FIG. 401


Locus
proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética

FIG. 40J

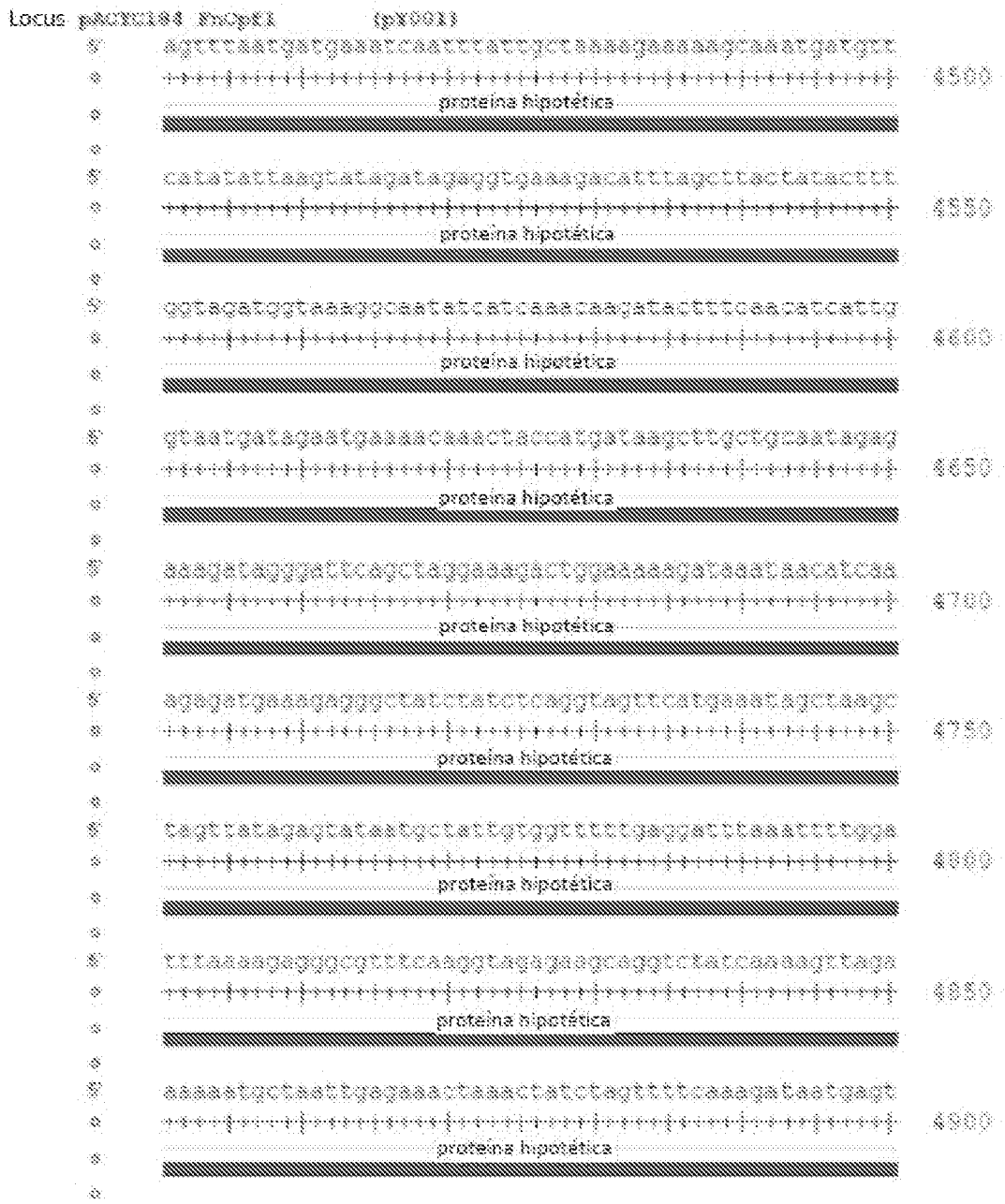

Locus pACYC184 rnlpE1 (pX001)
4500
proteína hipotética
4550
proteína hipotética
4600
proteína hipotética
4650
proteína hipotética
4700
proteína hipotética
4750
proteína hipotética
4800
proteína hipotética
4850
proteína hipotética
4900
proteína hipotética

FIG. 40K

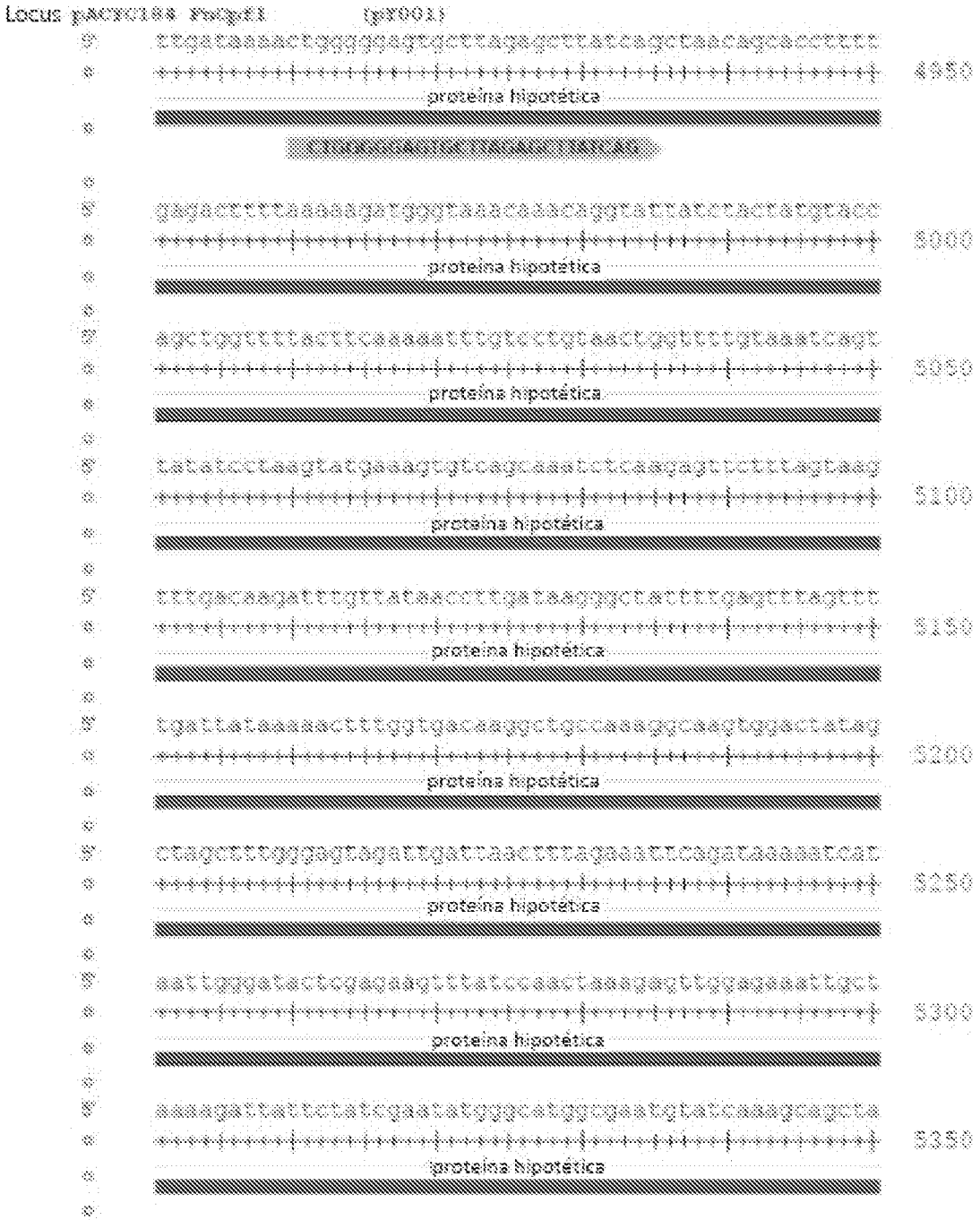

proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética
proteína hipotética

FIG. 40L


Locus pACYC184 PscpE1 (p7001)
5'
3'
5850
proteína hipotética
5900
5950
6000
6050
6100
6150
6200
proteína hipotética

FIG. 40M


Locus pACYC184 PvuII (pr001)
5850
endonuclease
5900
endonuclease
5950
endonuclease
6000
endonuclease
6050
endonuclease
6100
endonuclease
6150
endonuclease
6200
endonuclease
6250
endonuclease

FIG. 40N

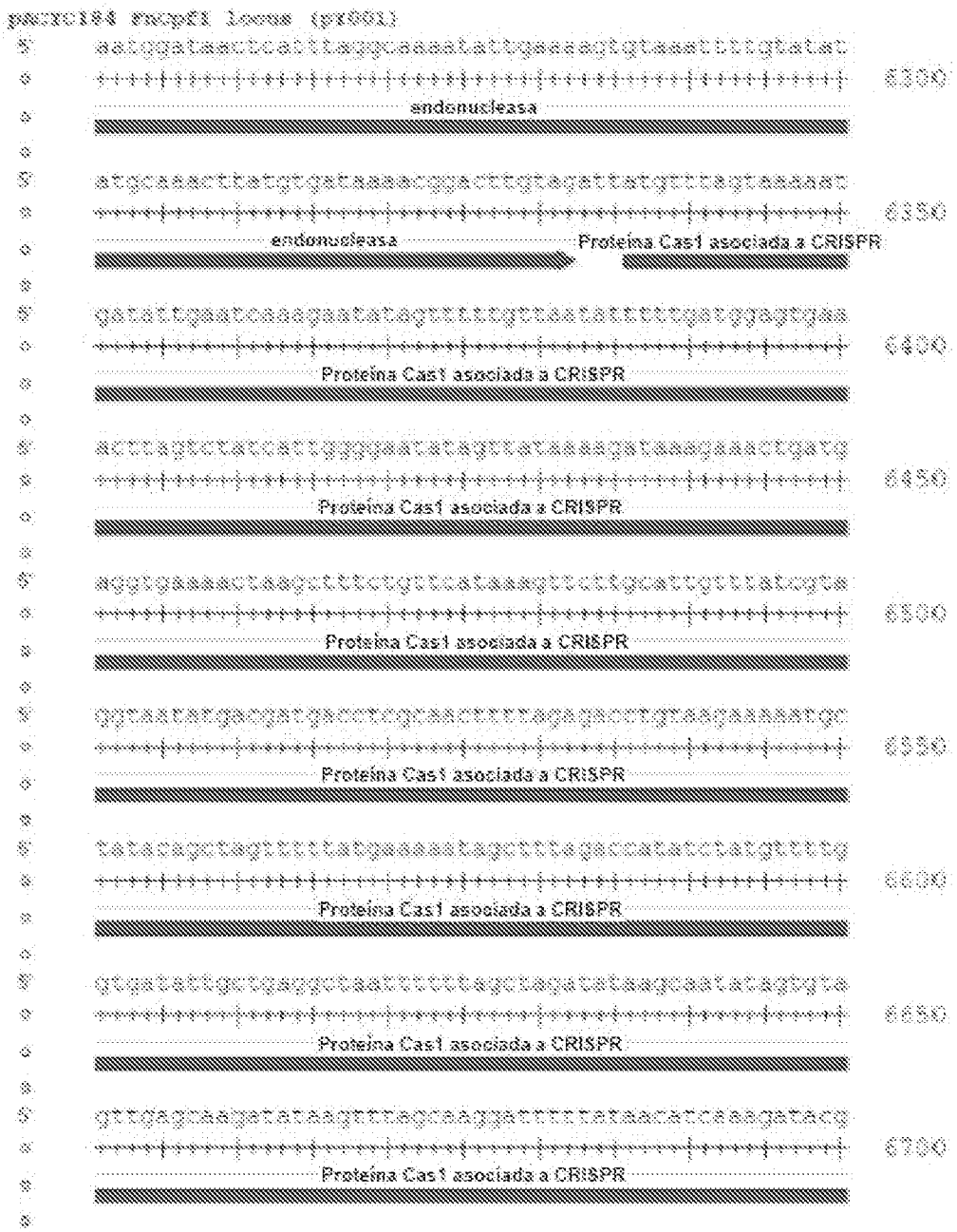

pacrc188 nucpfi locus (pr001)
6300
endonucleasa
6350
endonucleasa
Proteína Cas1 asociada a CRISPR
6400
Proteína Cas1 asociada a CRISPR
6450
Proteína Cas1 asociada a CRISPR
6500
Proteína Cas1 asociada a CRISPR
6550
Proteína Cas1 asociada a CRISPR
6600
Proteína Cas1 asociada a CRISPR
6650
Proteína Cas1 asociada a CRISPR
6700
Proteína Cas1 asociada a CRISPR

FIG. 40O

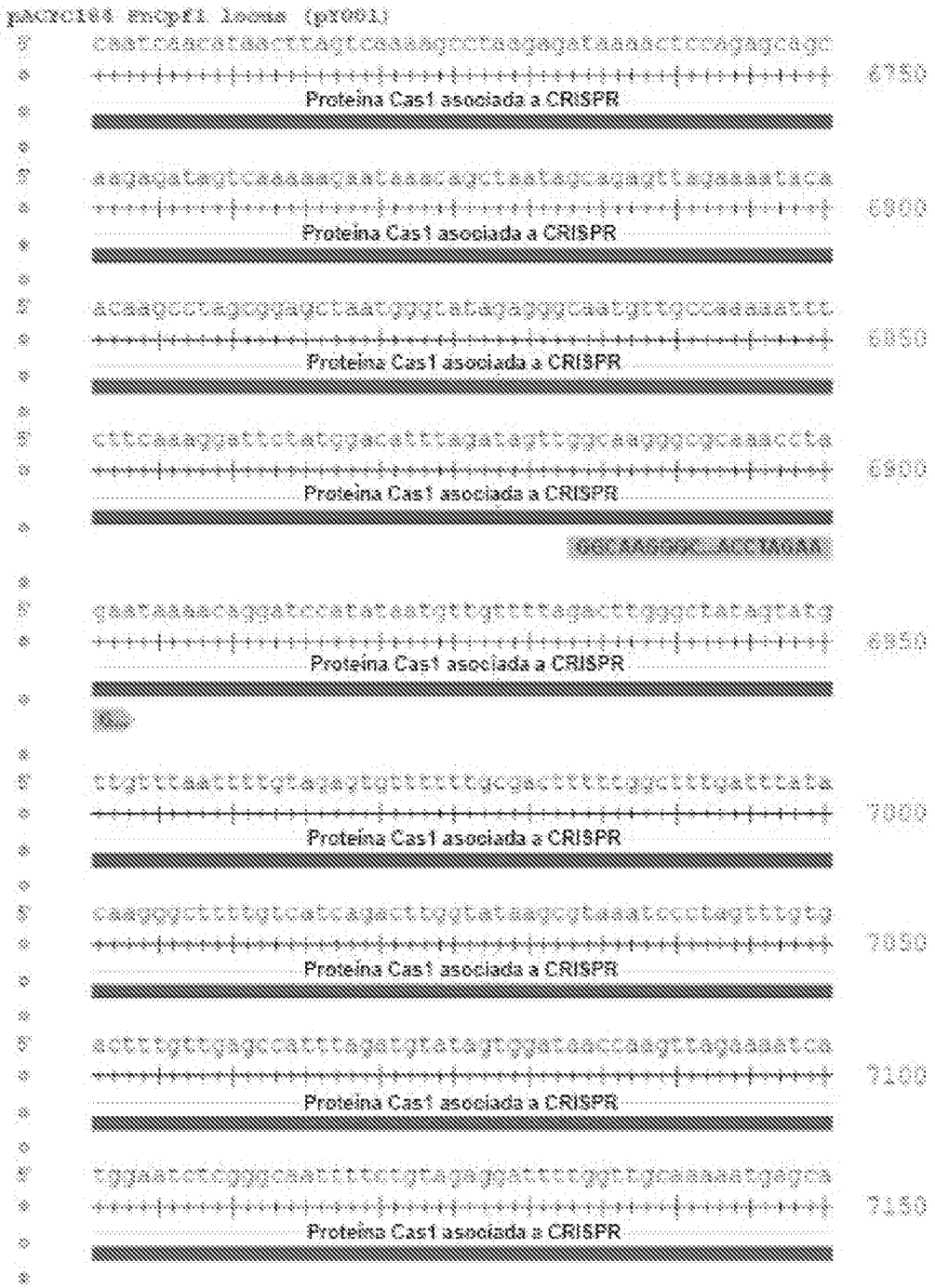

pACYC184 EcoRI locus (pYG01)
6750
6800
6850
6900
6950
7000
7050
7100
7150
Proteína Cas1 asociada a CRISPR

FIG. 40P

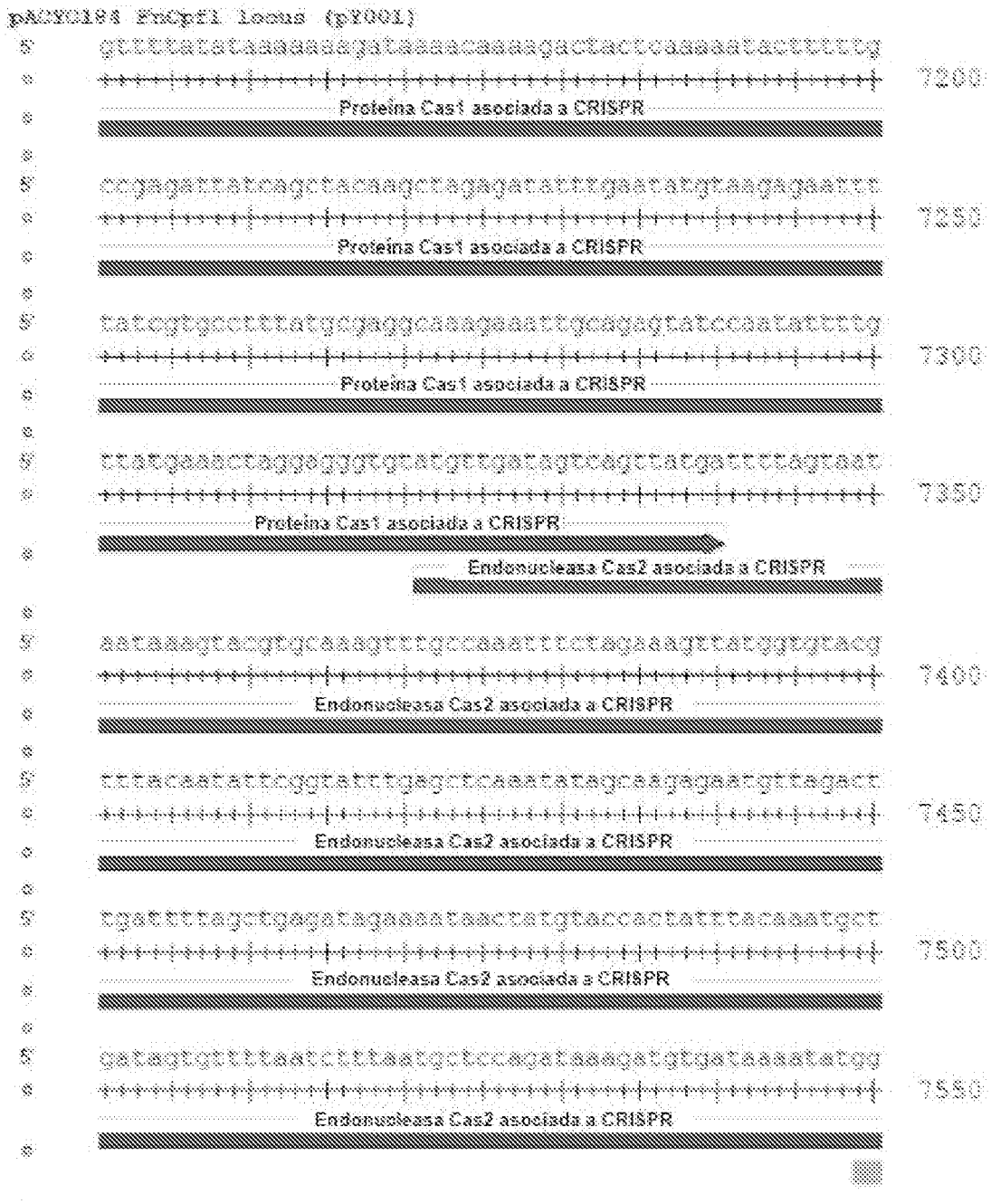

pACYC184 FnCpf1 locus (pY001)
7200
Proteína Cas1 asociada a CRISPR
7250
Proteína Cas1 asociada a CRISPR
7300
Proteína Cas1 asociada a CRISPR
7350
Proteína Cas1 asociada a CRISPR
Endonucleasa Cas2 asociada a CRISPR
7400
Endonucleasa Cas2 asociada a CRISPR
7450
Endonucleasa Cas2 asociada a CRISPR
7500
Endonucleasa Cas2 asociada a CRISPR
7550
Endonucleasa Cas2 asociada a CRISPR

FIG. 40Q

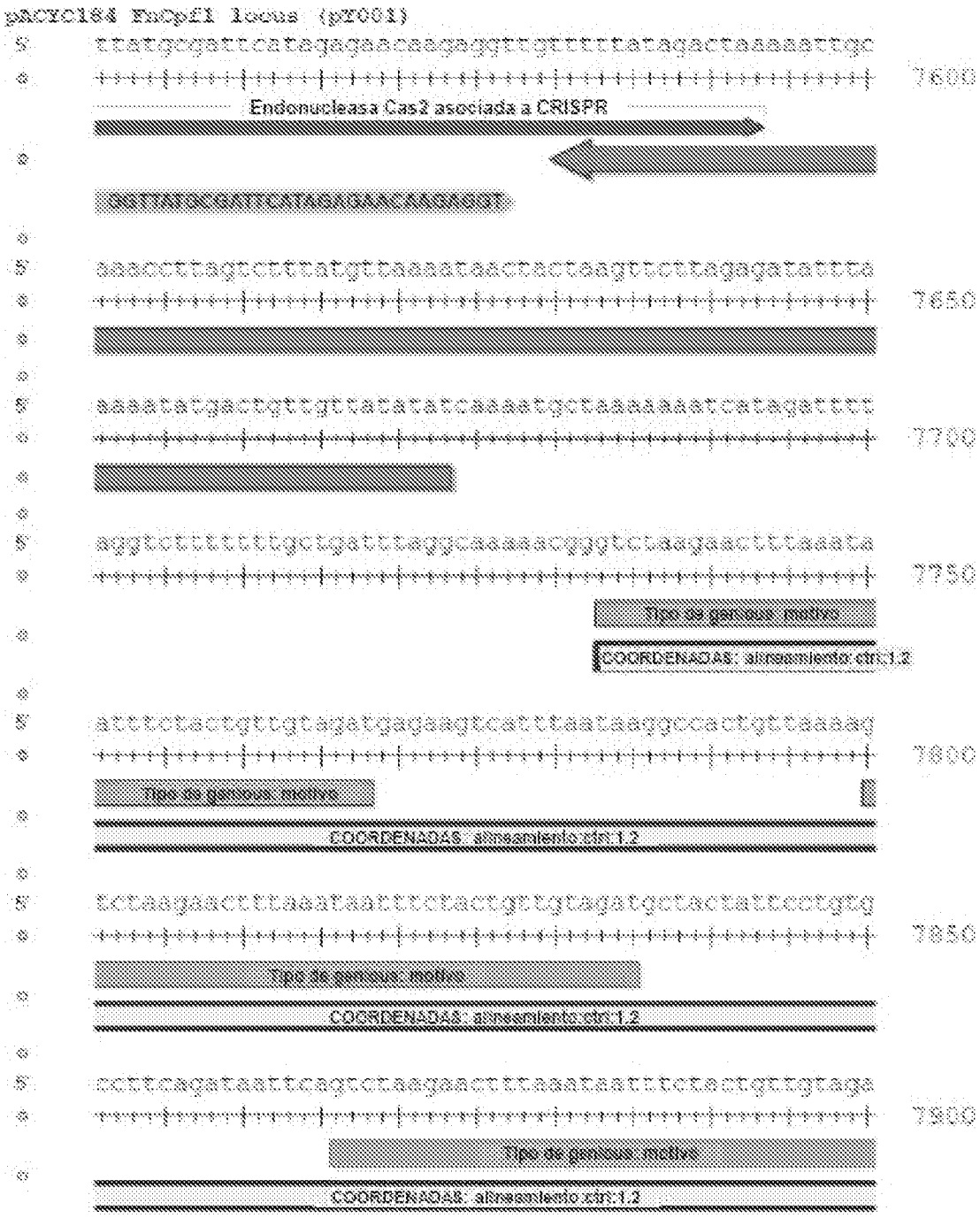

pACYC184 EnCpf1 locus (pY001)
5' ttatgcgattcatagagaacaagaggttgtttttatagactaaaaattgc
7600
Endonucleasa Cas3 asociada a CRISPR
GGTTATACGATTCTATAGAACAAGAGGT
5' aaaccttagtcrttatgttaaaataactactaagttcttagagatattta
7650
5' aaaatatgactgttgttatatatcaaaatgctaaaaaaaatcatagatttt
7700
5' aggtcttttttttgctgatttaggcaaaaacgggtctaagaactttaaata
7750
Tipo de genoux motivo
COORDENADAS: alineamiento ctrl 1.2
5' atttctactgttgtagatgagaagtcatttaataaggccactgttaaaag
7800
Tipo de genoux motivo
COORDENADAS: alineamiento ctrl 1.2
5' tctaagaacttttaaataatttctactgttgtagatgctactattcctgtg
7850
Tipo de genoux motivo
COORDENADAS: alineamiento ctrl 1.2
5' ccttcagataattcagtctaagaacttttaaataatttctactgttgtaga
7900
Tipo de genoux motivo
COORDENADAS: alineamiento ctrl 1.2

FIG. 40R

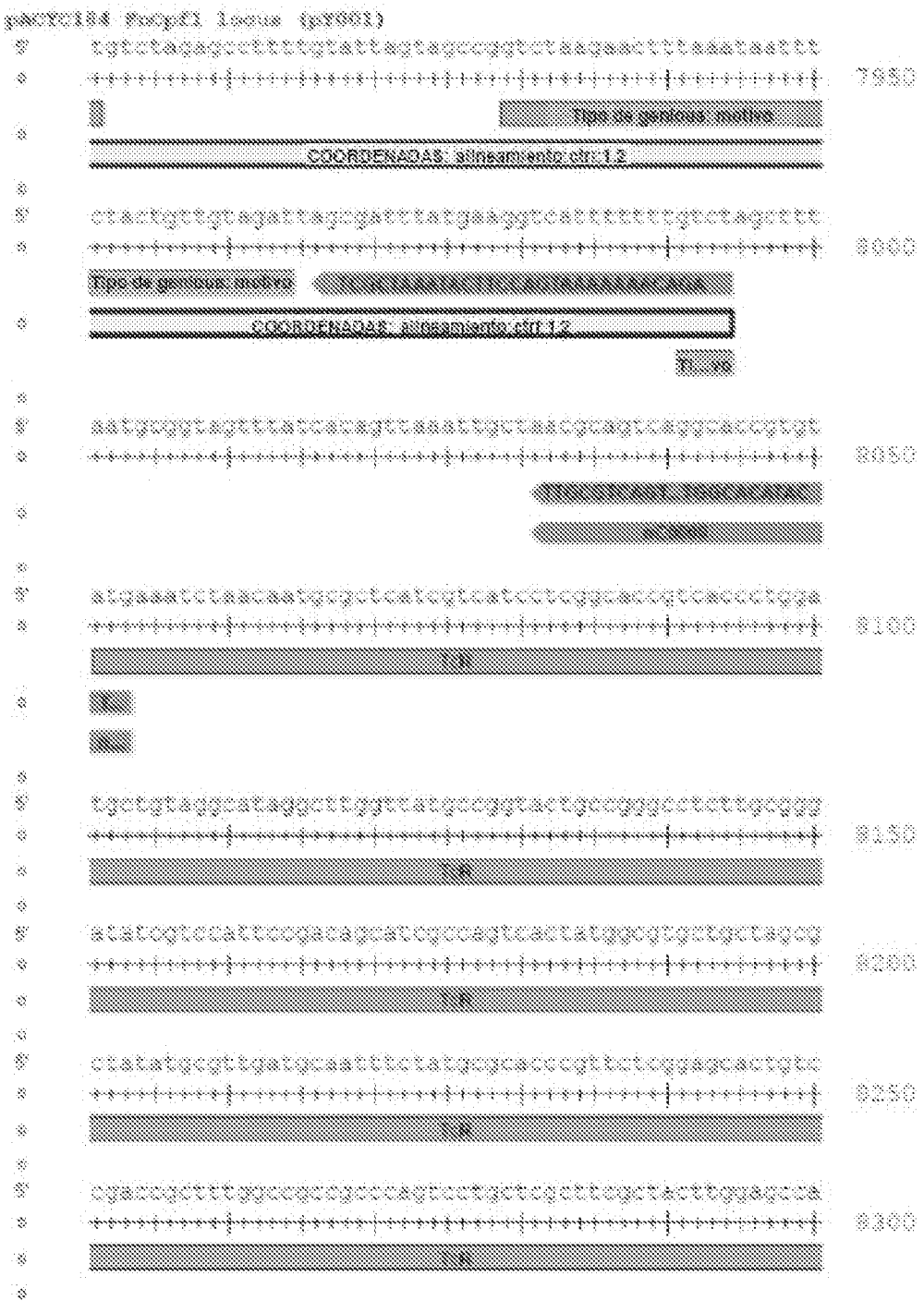

pACYC184 PscpC1 locus (pY001)
7950
8000
8050
8100
8150
8200
8250
8300
COORDENADAS: alineamiento ctrl 1.2
COORDENADAS: alineamiento ctrl 1.2

FIG. 40S

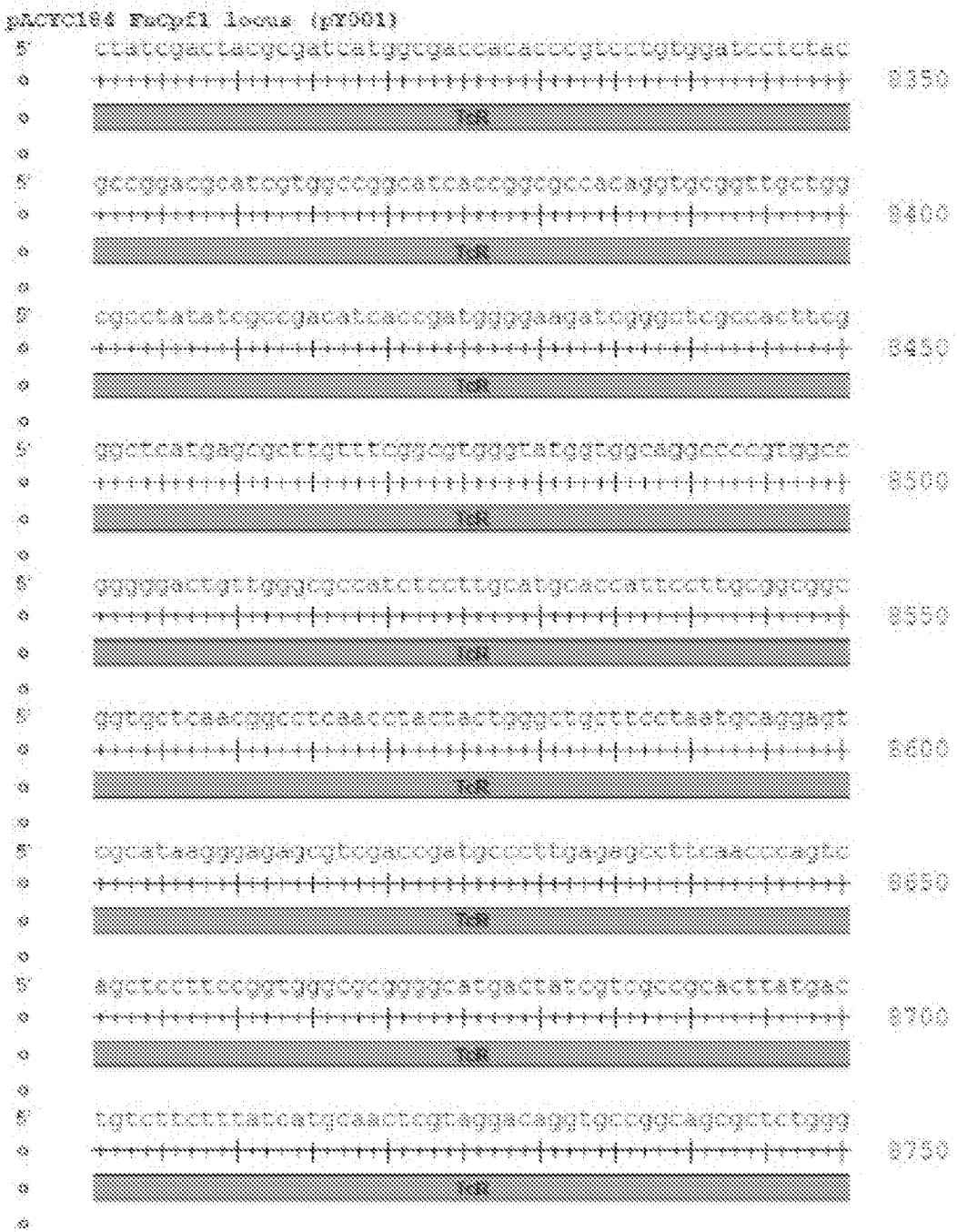

pRCTC184 FnCpf1 locus (pT001)
5'
8350
8400
8450
8500
8550
8600
8650
8700
8750
YGR

FIG. 40T

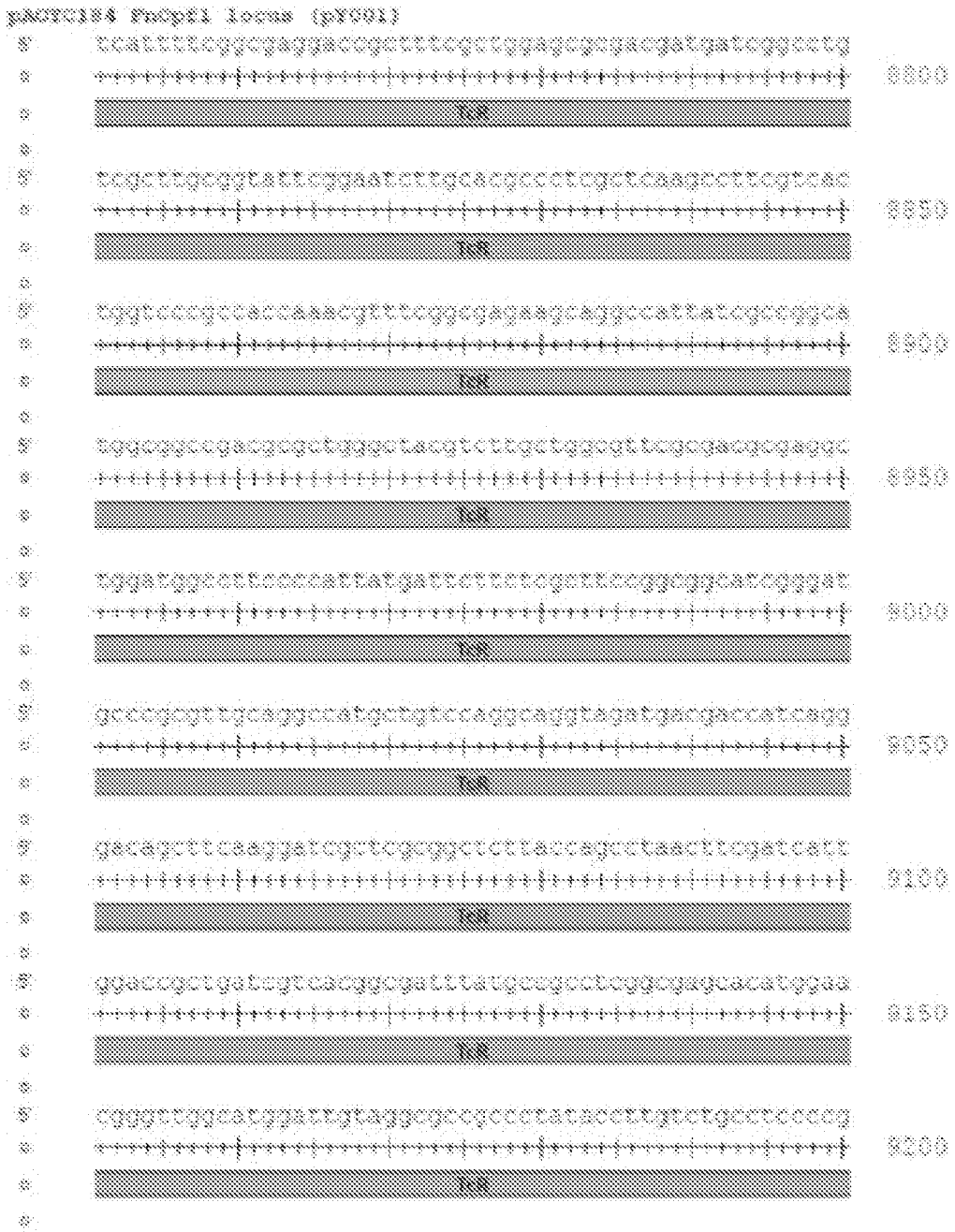

pACYC184 PmCpt1 locus (pF001)
8800
8850
8900
8950
9000
9050
9100
9150
9200
TetR

FIG. 40J

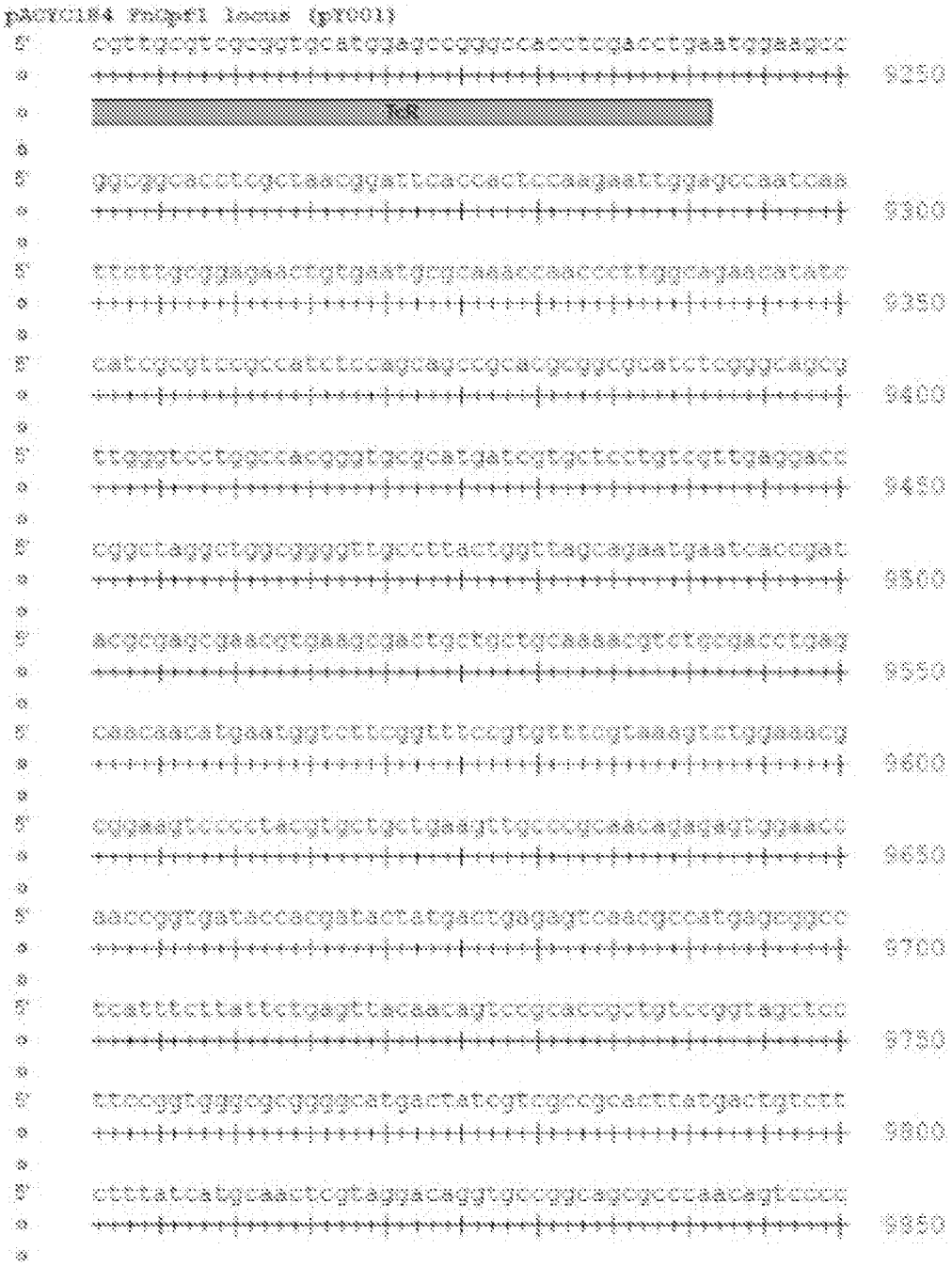

pACYC184 PmgRI locus (pP001)
9250
9300
9350
9400
9450
9500
9550
9600
9650
9700
9750
9800
9850

FIG. 40V

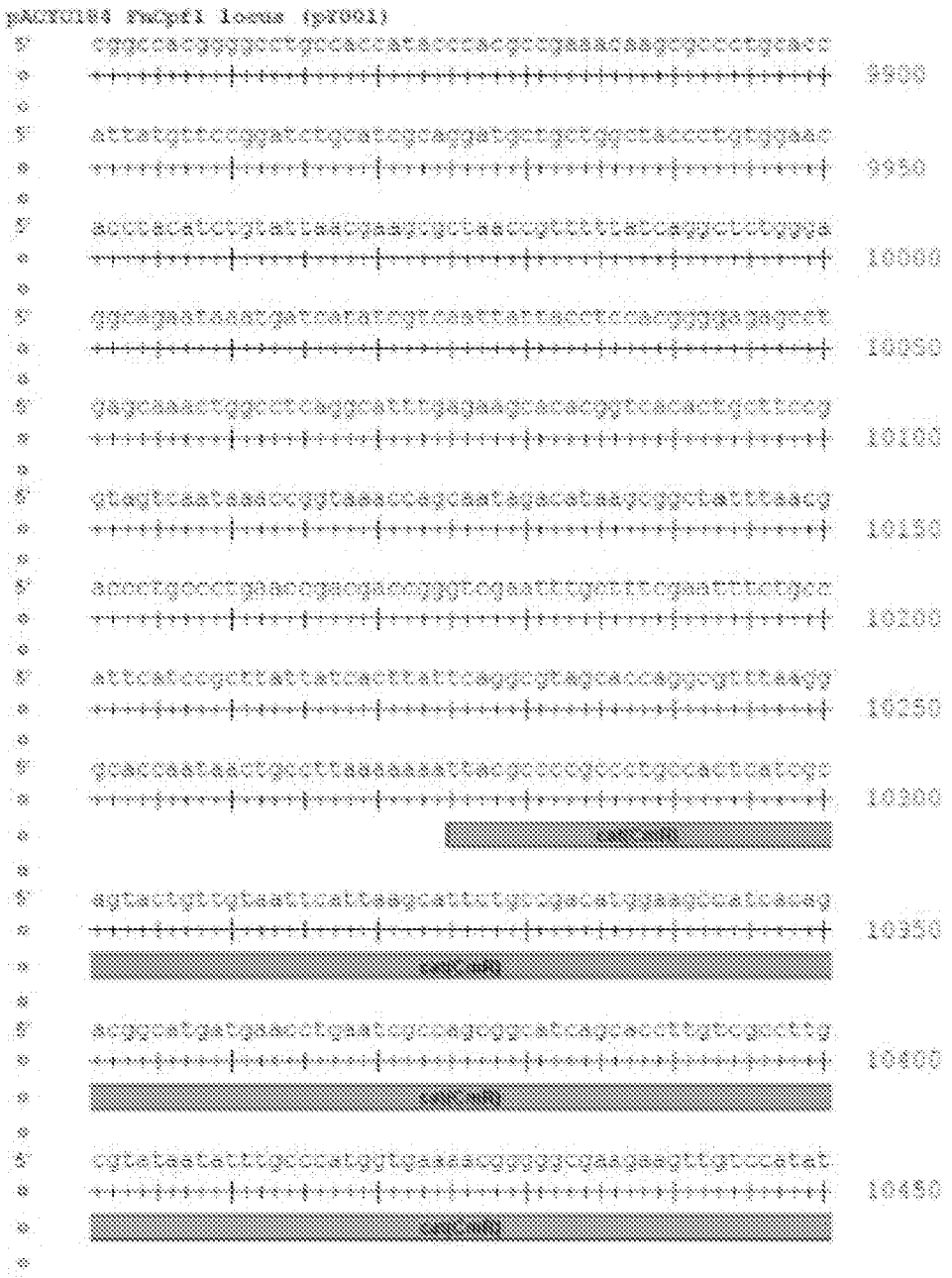


FIG. 40W

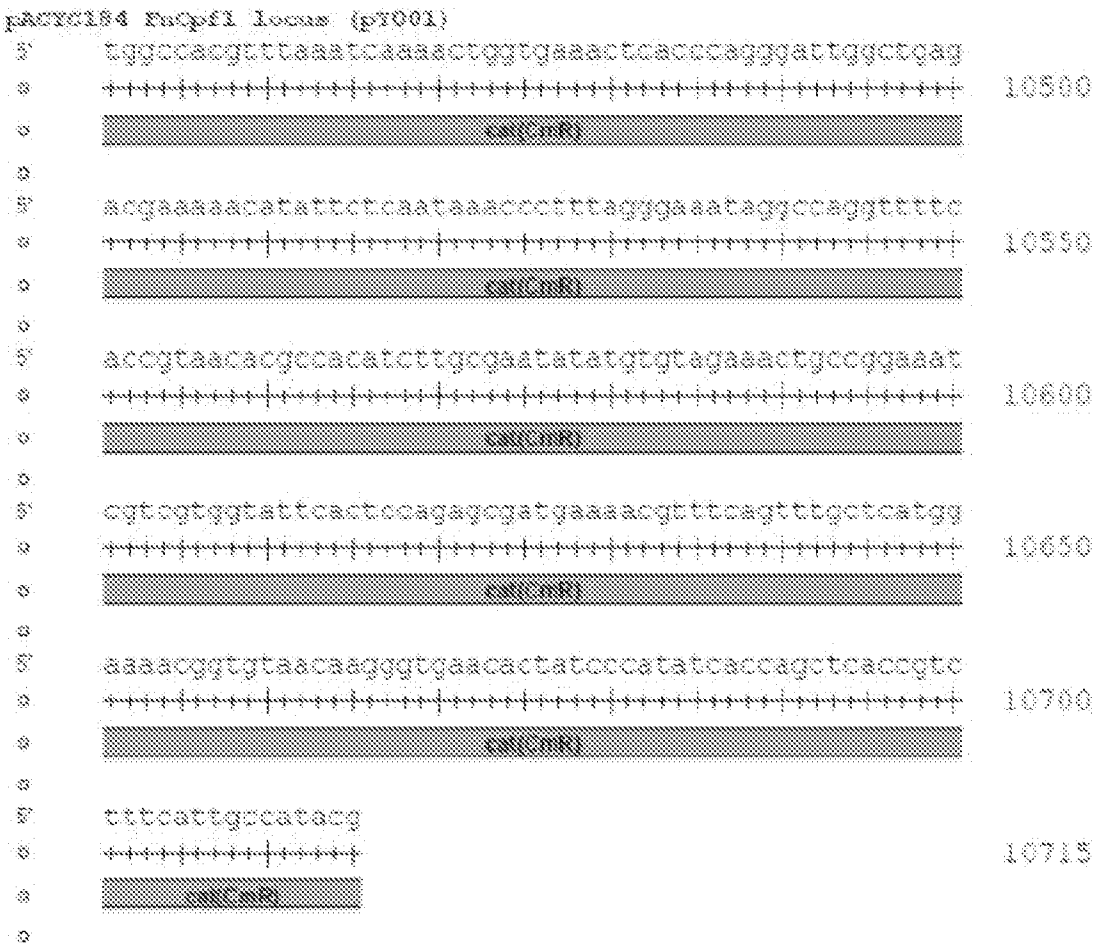

pACYC184 PuCpf1 locus (pT001)
5' tggccacgtttaaatcaaaactggtgaaactcacccagggattggctgag
10500
cat(CmR)
5' acgaaaaacatattctcaataaaccctttagtgaaataggccaggttttc
10550
cat(CmR)
5' accgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaat
10600
cat(CmR)
5' cgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatgg
10650
cat(CmR)
5' aaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtc
10700
cat(CmR)
5' ttcattgccatacg
10715
cat(CmR)

FIG. 40X

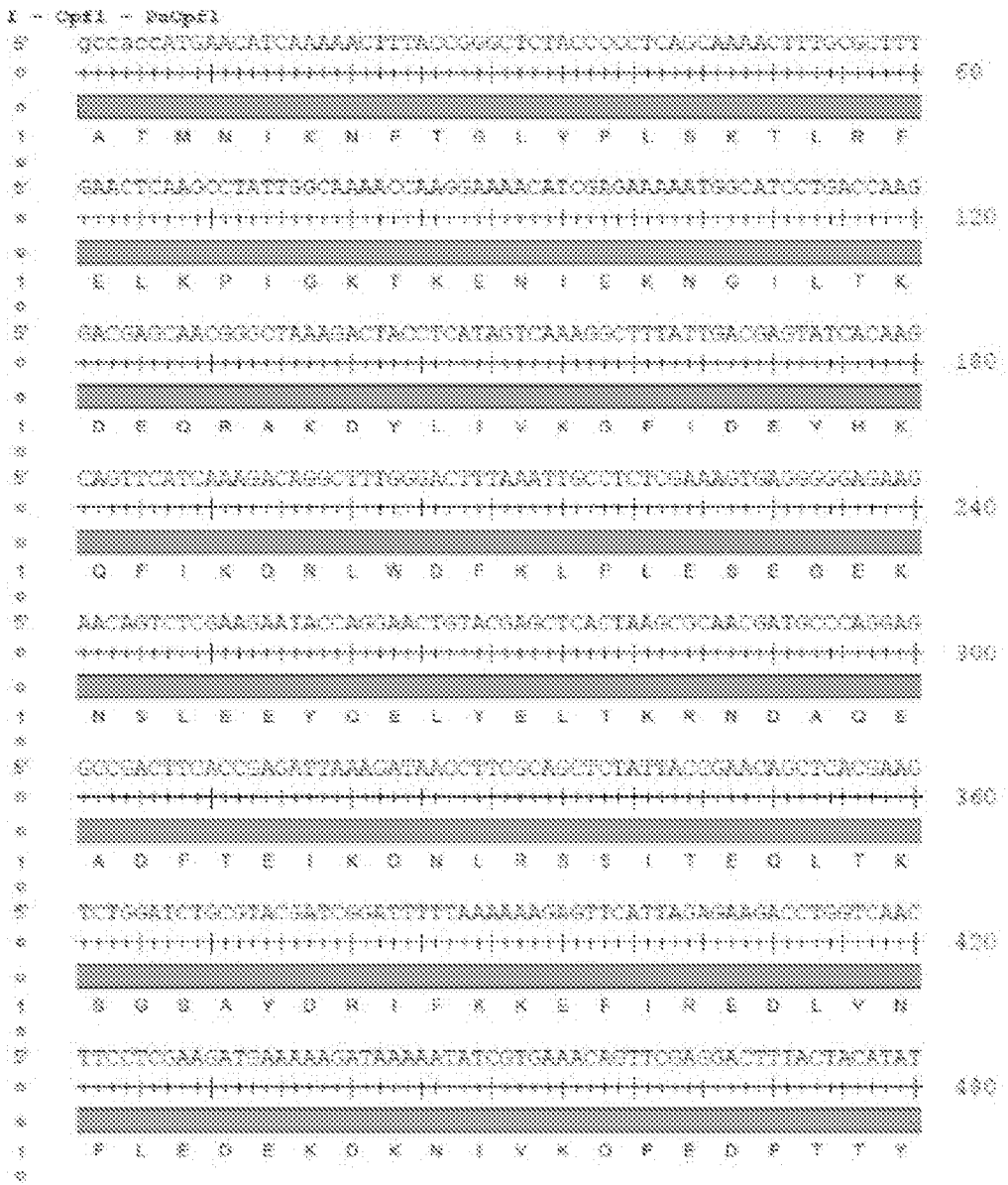


FIG. 41A

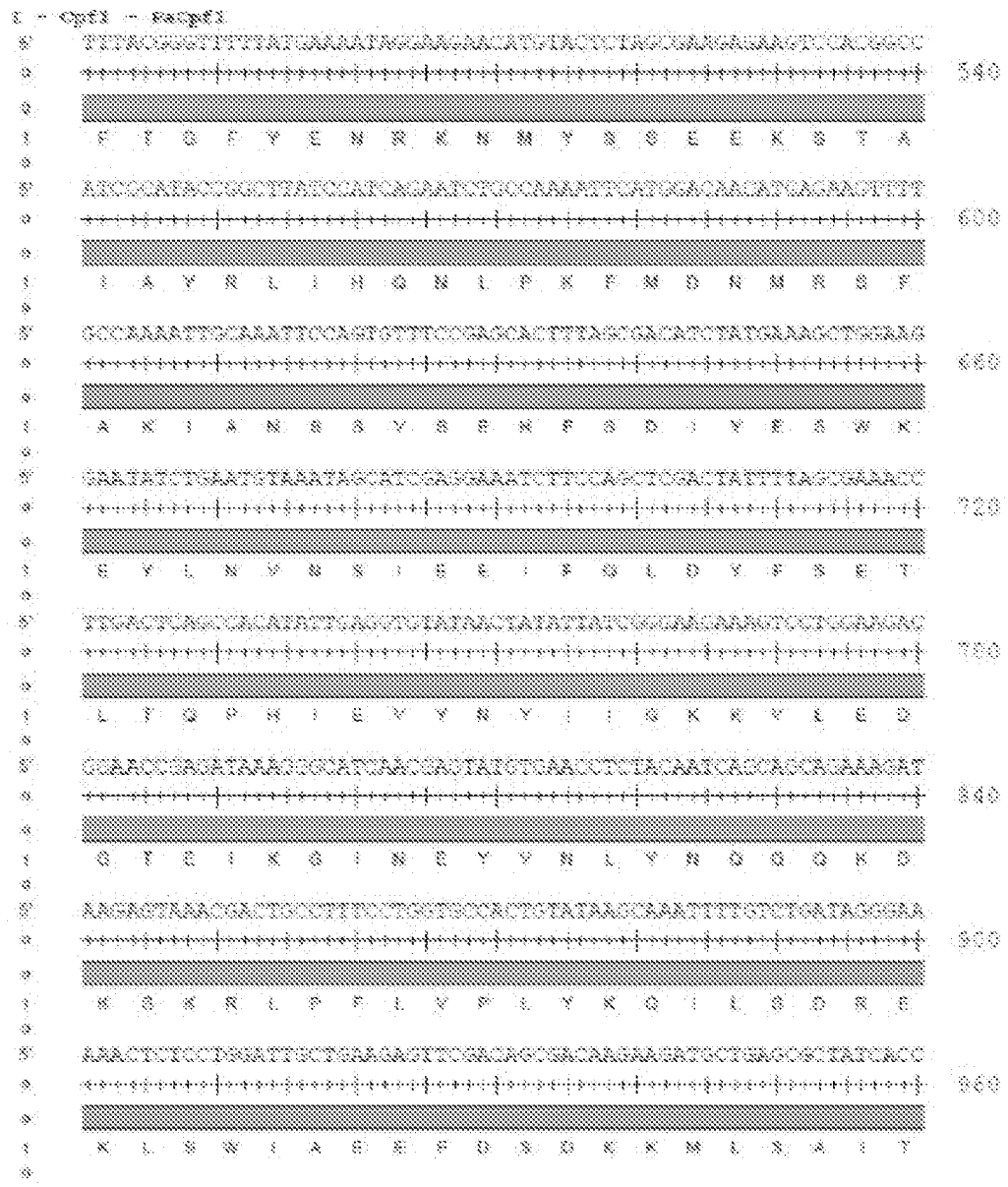


FIG. 41B

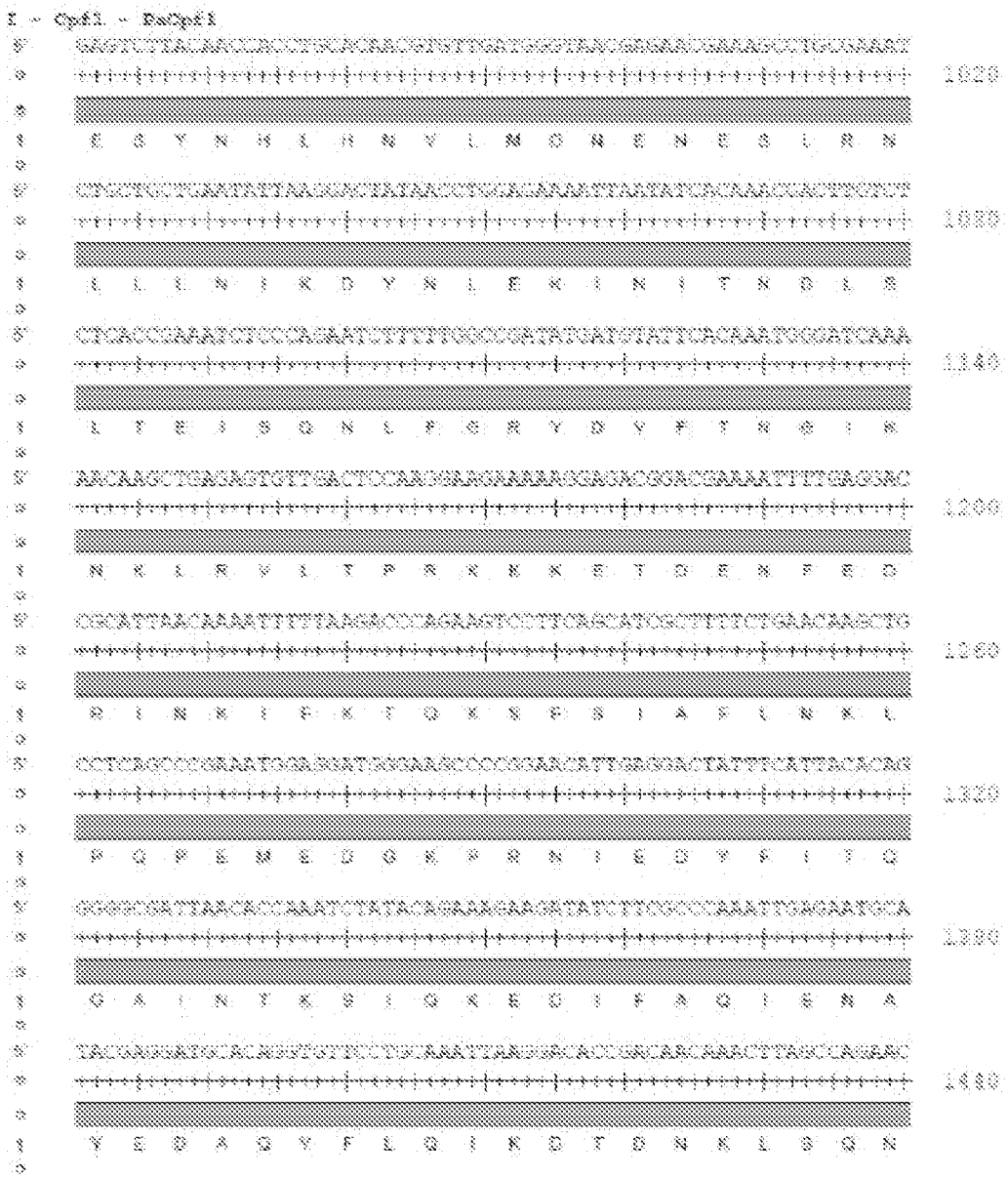


FIG. 41C

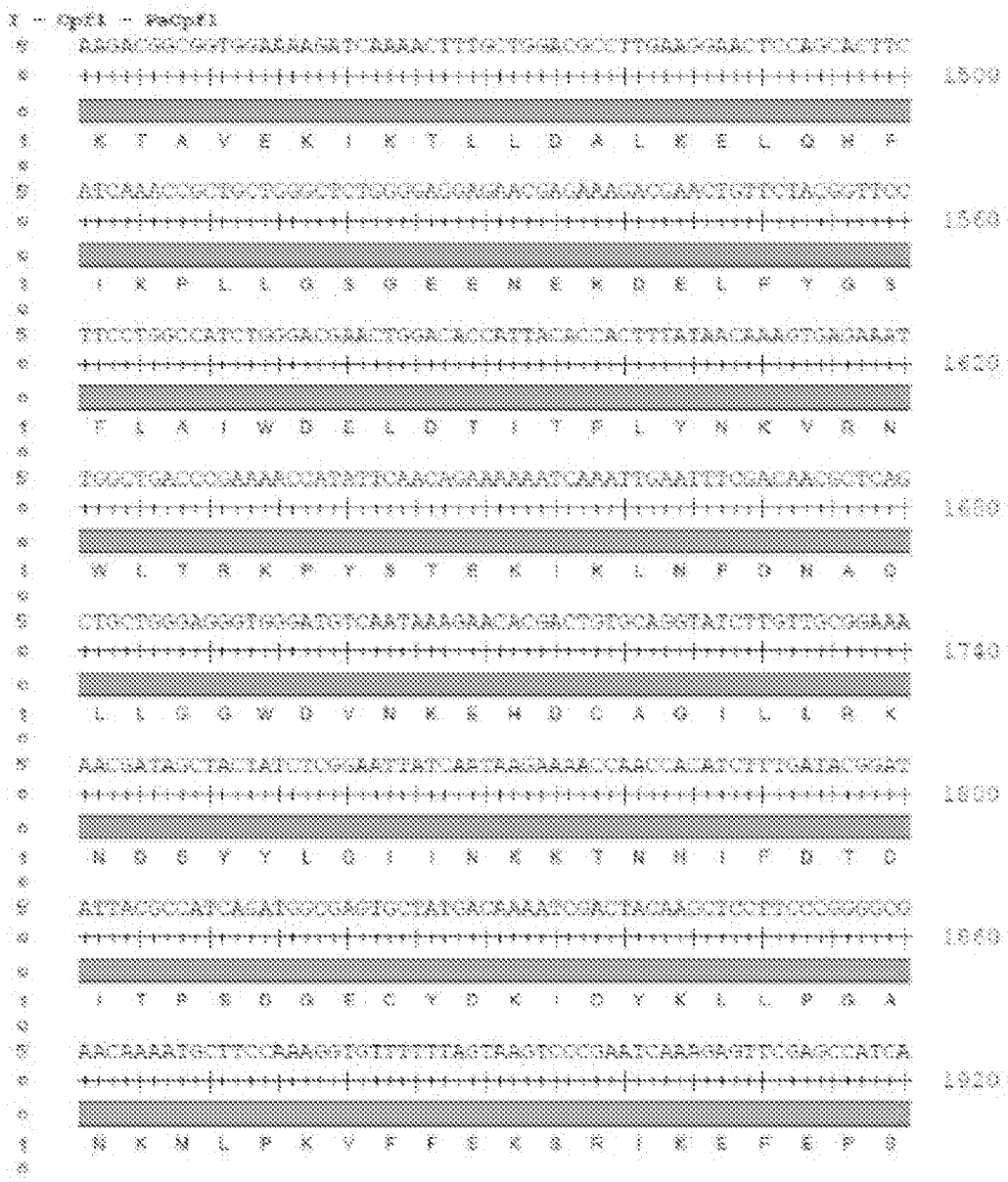


FIG. 41D

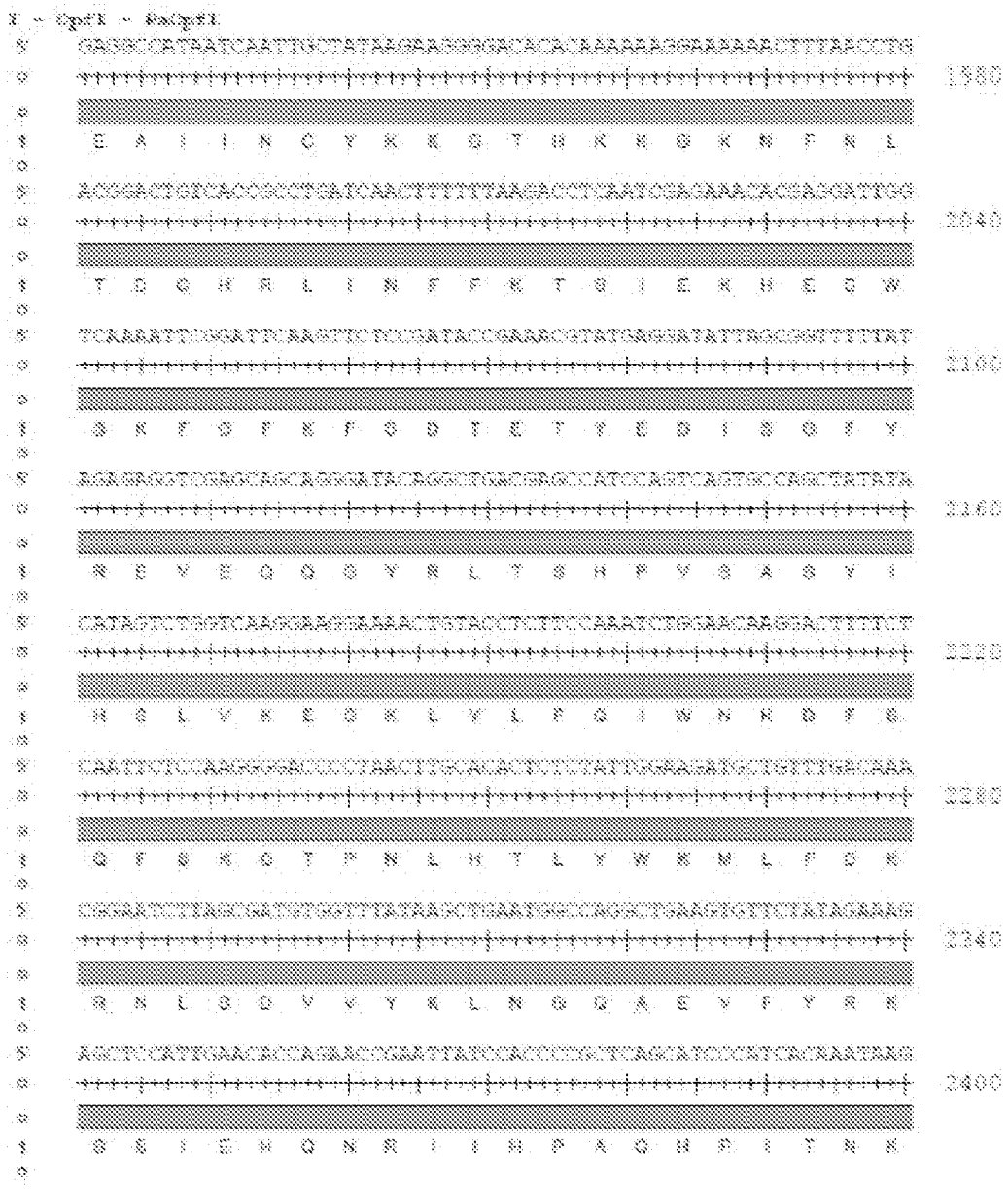


FIG. 41E

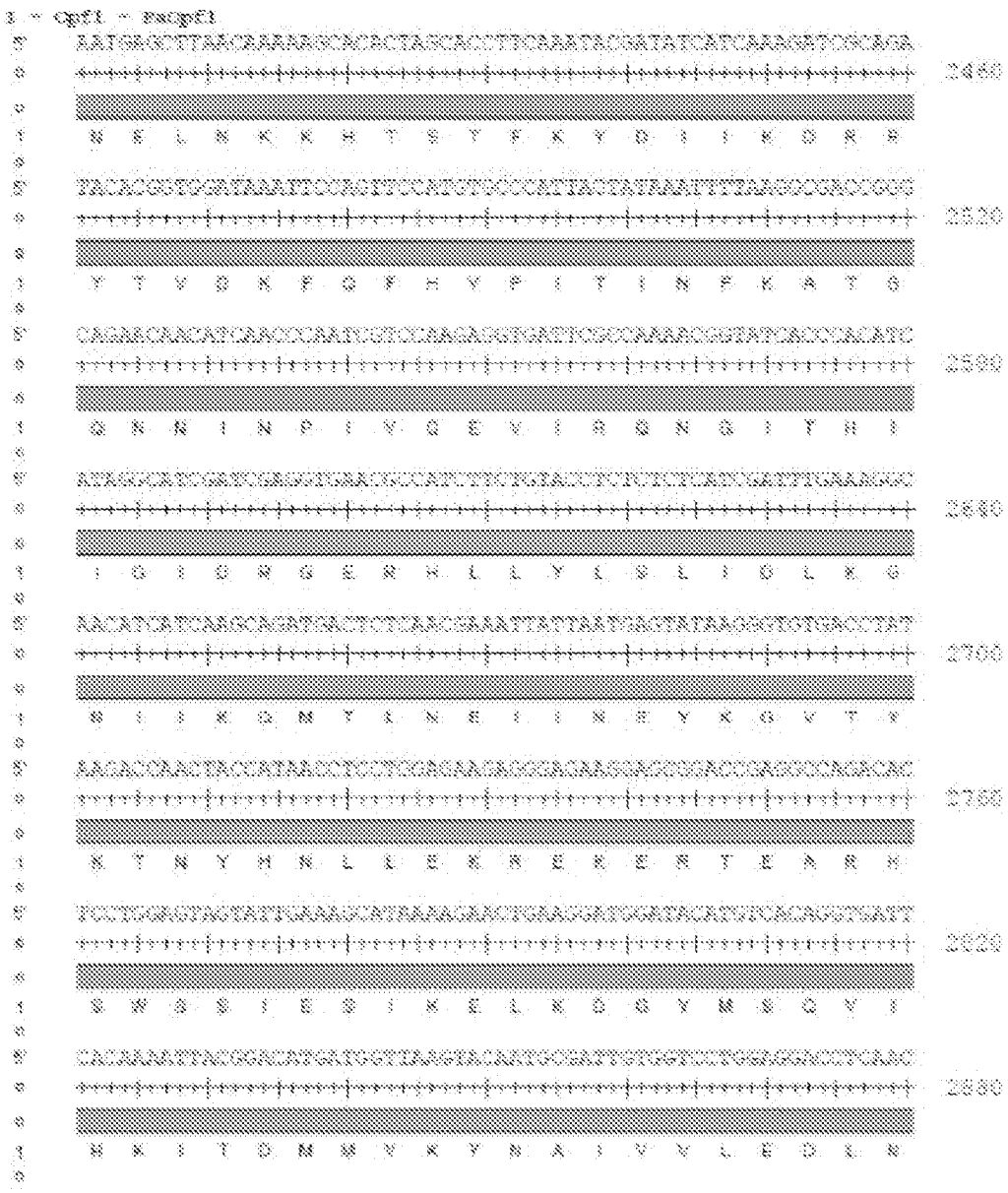


FIG. 41F

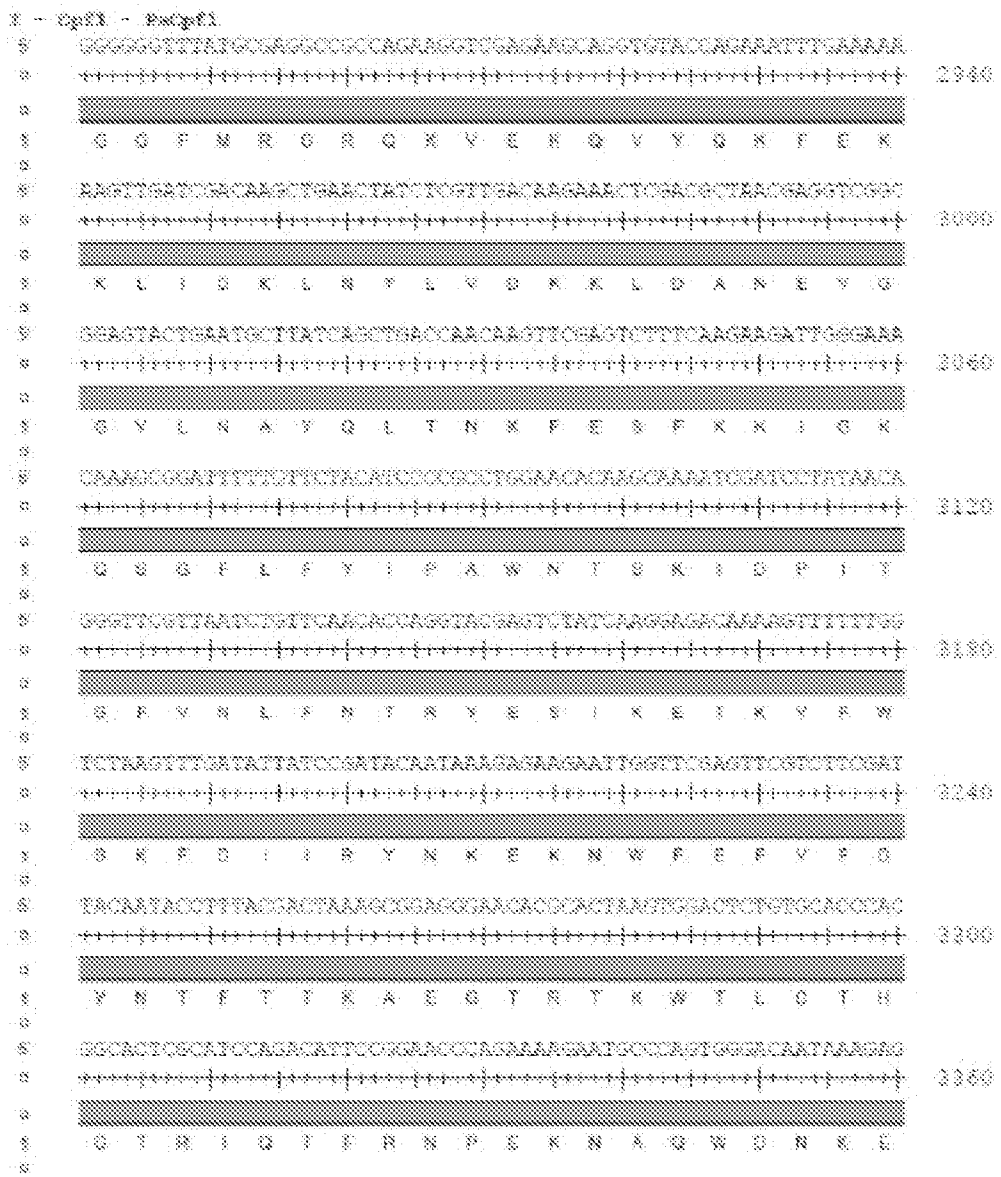


FIG. 41G

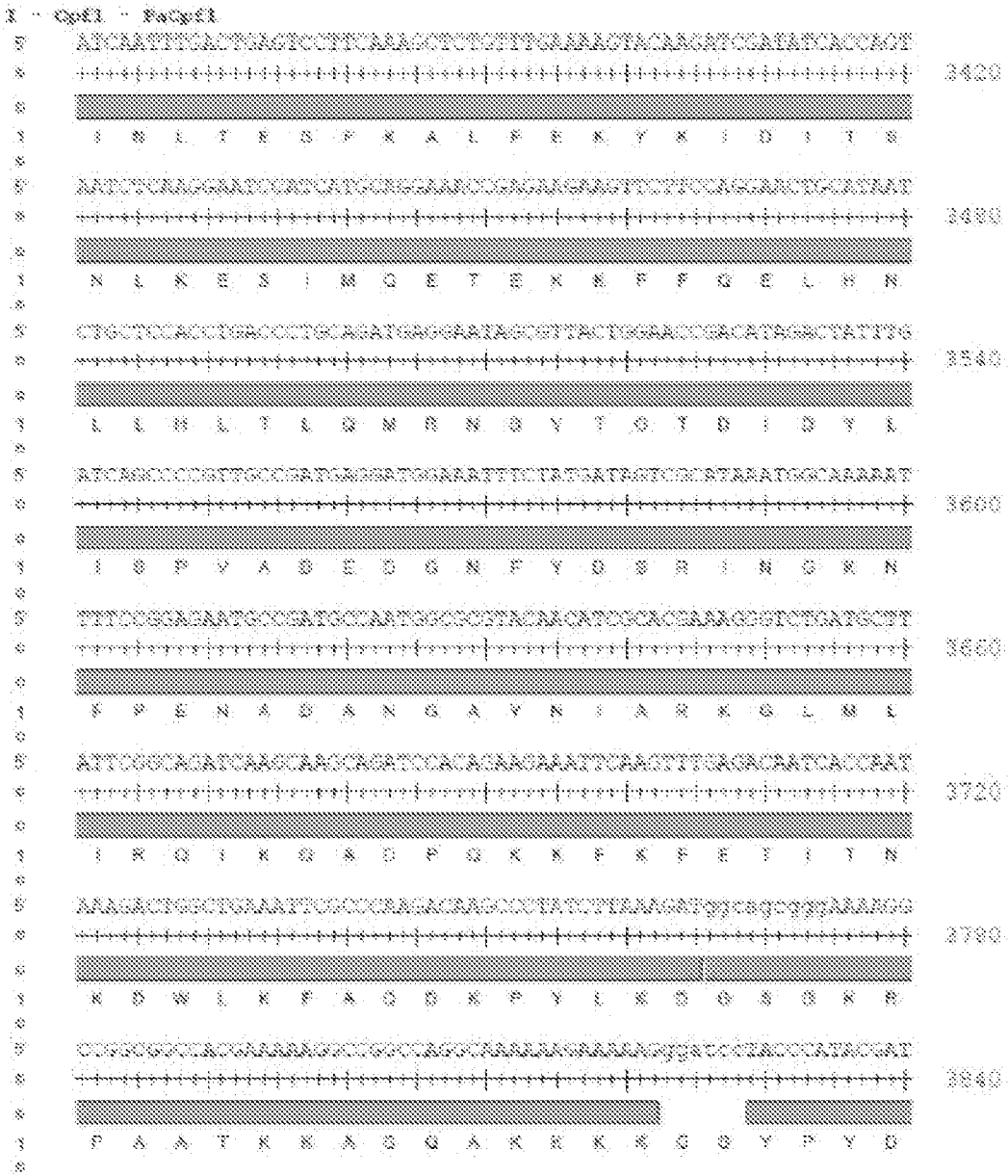


FIG. 41H

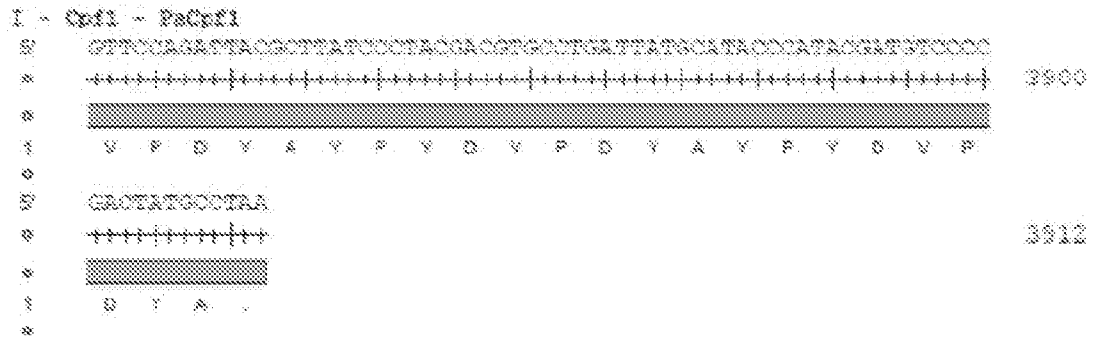


FIG. 41I

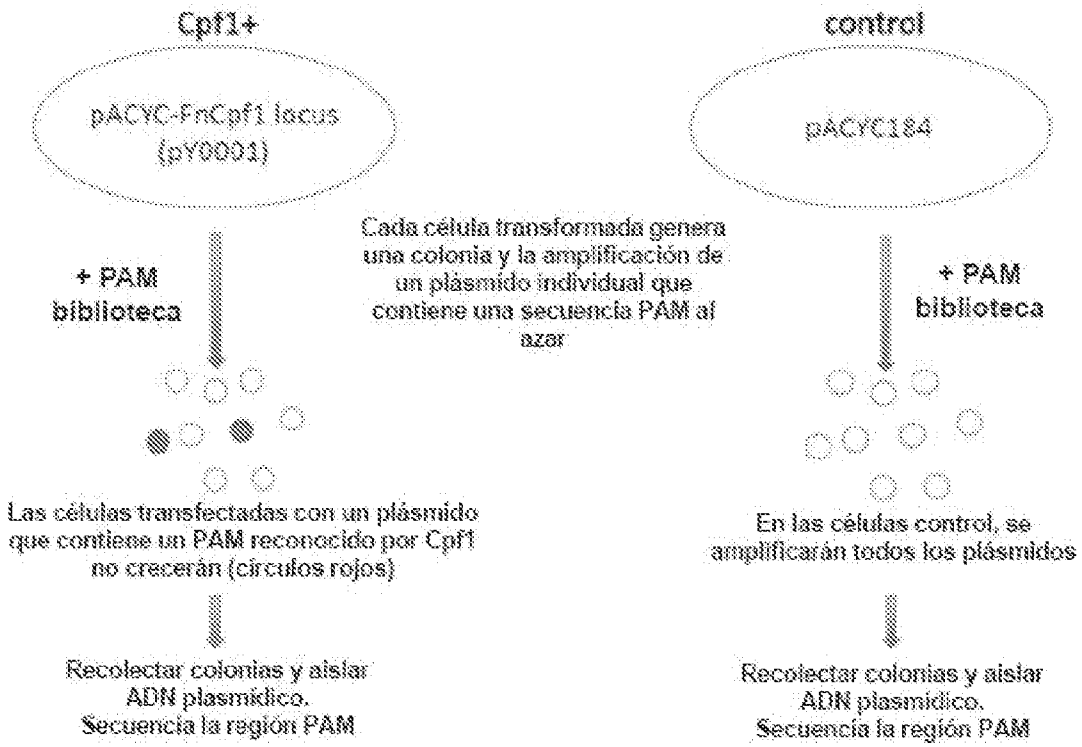

Cpf1+
pACYC-FnCpf1 locus (pY0001)
control
pACYC184
Cada célula transformada genera una colonia y la amplificación de un plásmido individual que contiene una secuencia PAM al azar
+ PAM biblioteca
+ PAM biblioteca
Las células transfectadas con un plásmido que contiene un PAM reconocido por Cpf1 no crecerán (círculos rojos)
En las células control, se amplificarán todos los plásmidos
Recolectar colonias y aislar ADN plasmídico. Secuencia la región PAM
Recolectar colonias y aislar ADN plasmídico. Secuencia la región PAM

FIG. 42

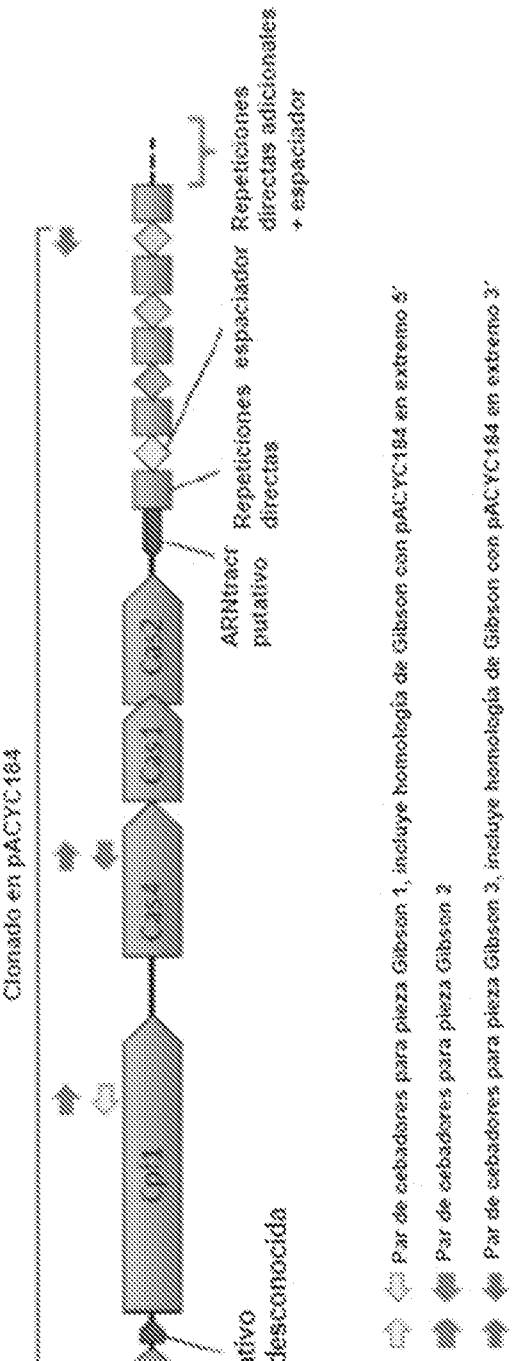

Clonado en pACYC184
Gen antes del locus de Cpf1
Acetiltransferasa
ARN putativo
Función desconocida
Cpf1
ARNtracr putativo
Repeticiones directas
espaciador
Repeticiones directas adicionales + espaciador
Par de cebadores para pieza Gibson 1, incluye homología de Gibson con pACYC184 en extremo 5'
Par de cebadores para pieza Gibson 2
Par de cebadores para pieza Gibson 3, incluye homología de Gibson con pACYC184 en extremo 3'

FIG. 43

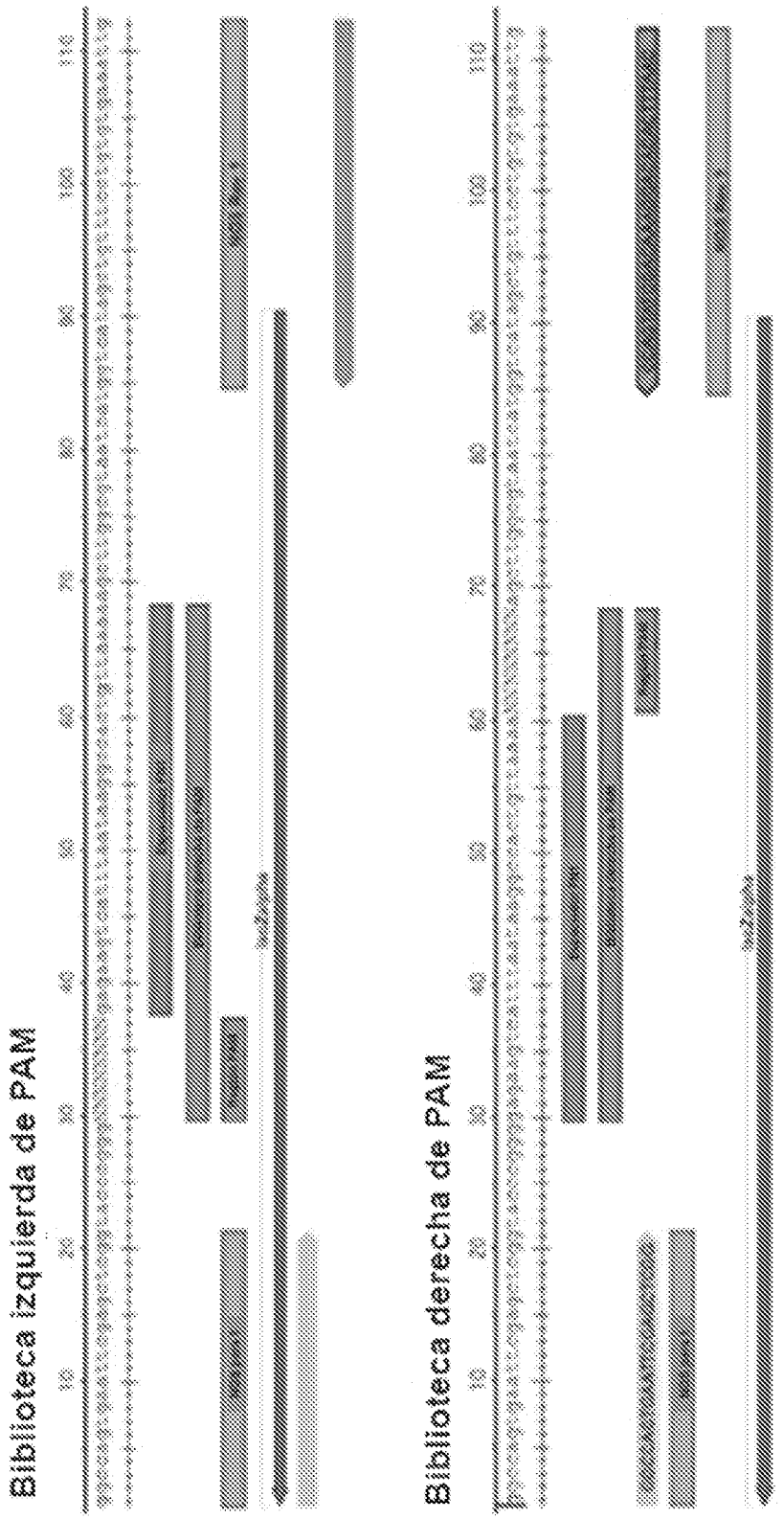

Biblioteca izquierda de PAM
10
20
30
40
50
60
70
80
90
100
110
Biblioteca derecha de PAM
10
20
30
40
50
60
70
80
90
100
110

FIG. 44

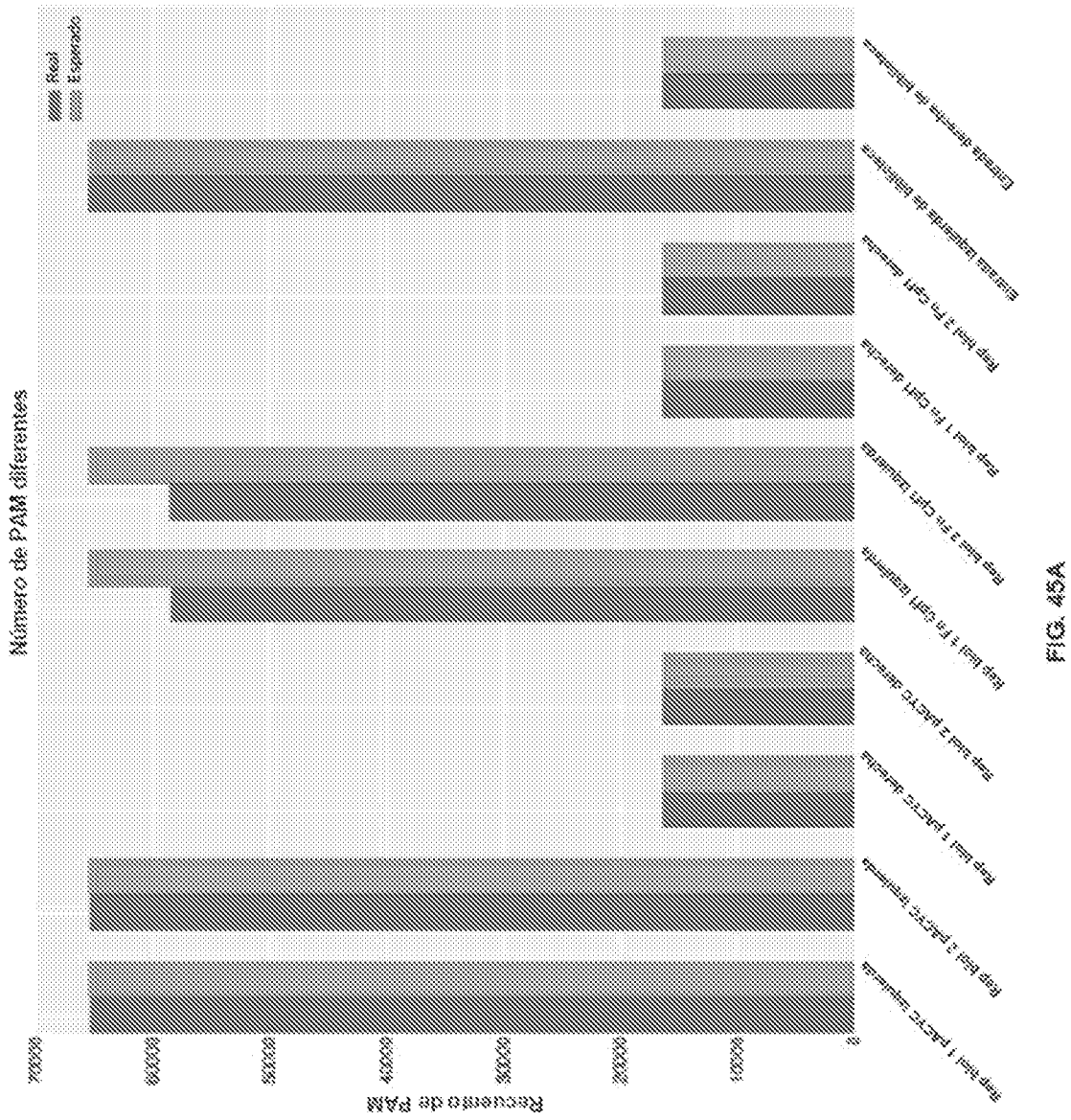

Real
Esperado
Número de PAM diferentes
Recuento de PAM
Emparejamiento de 3 bases PAM 1
Emparejamiento de 3 bases PAM 2
Emparejamiento de 4 bases PAM 1
Emparejamiento de 4 bases PAM 2
Bulge Size 1 3-bp GC Deletion
Bulge Size 1 3-bp GC Insertion
Bulge Size 2 3-bp GC Deletion
Bulge Size 2 3-bp GC Insertion
Bulge Size 1 4-bp GC Deletion
Bulge Size 1 4-bp GC Insertion
Bulge Size 2 4-bp GC Deletion
Bulge Size 2 4-bp GC Insertion
Cuenta desde la izquierda
Cuenta desde la derecha

FIG. 45A

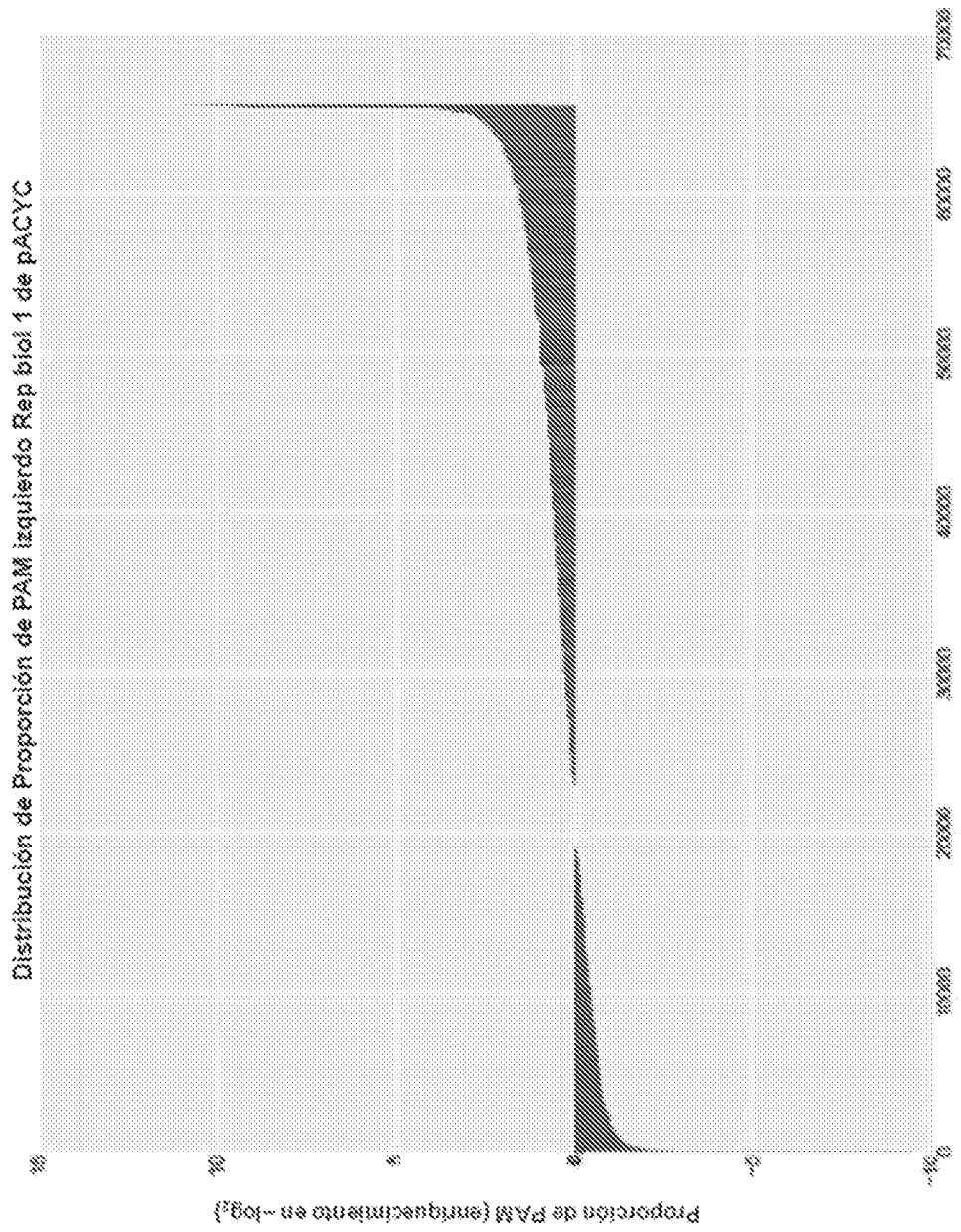

Distribución de Proporción de PAM izquierdo Rep biol 1 de pACYC
Proporción de PAM (enriquecimiento en -log2)
Rango

FIG. 45B


Distribución de Proporción de PAM izquierdo Rep biol 1 de Fn Cpf1
Proporción de PAM (enriquecimiento en -log10)
15
10
5
0
-5
-10
-15
0
10000
20000
30000
40000
50000
60000
70000
Rango

FIG. 45C


Distribución de Proporción de PAM izquierdo Rep biol 2 de Fn Cpf1
Proporción de PAM (enriquecimiento en -log10)
Rango
0
10000
20000
30000
40000
50000
60000
70000

FIG. 45D



FIG. 45E

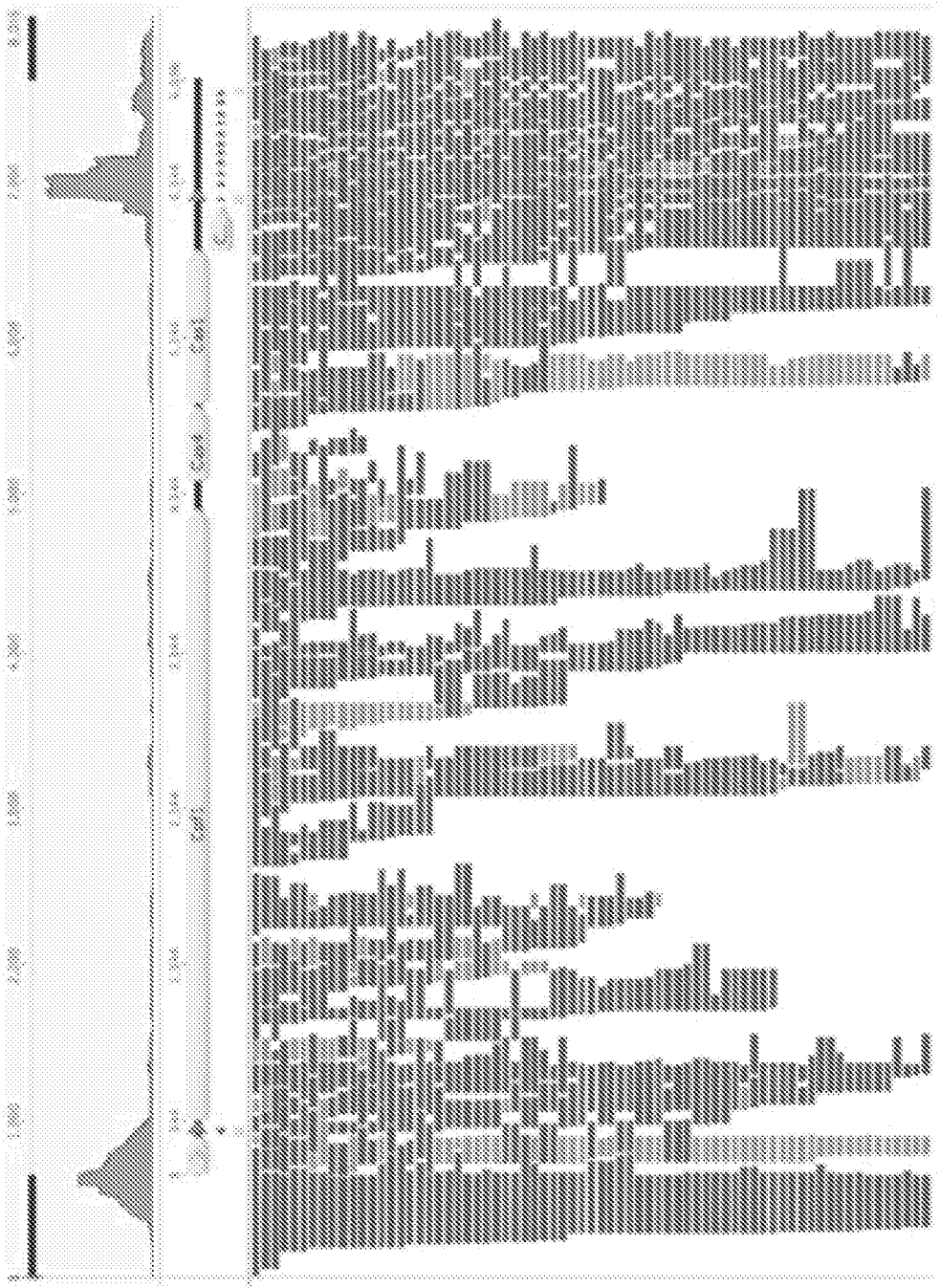


FIG. 46

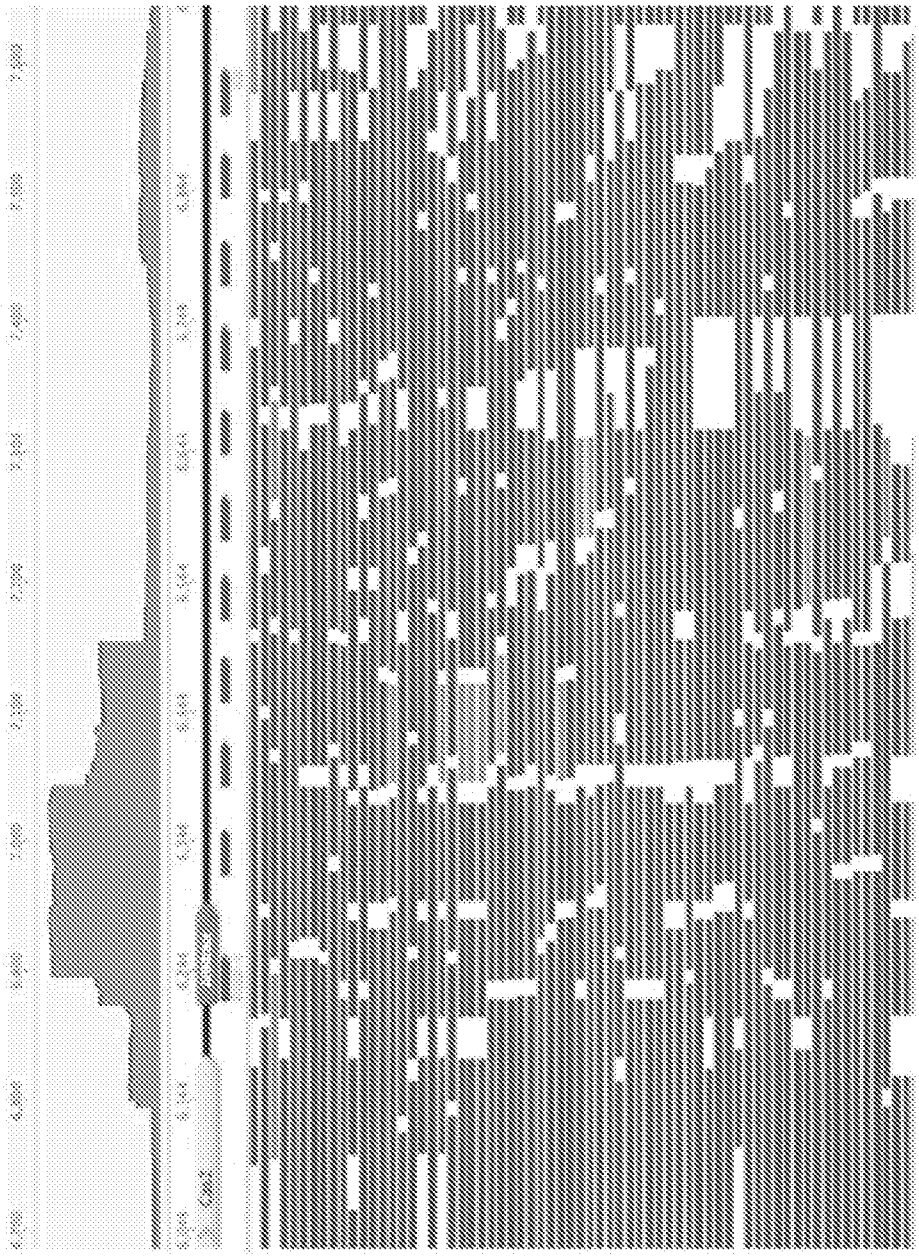


FIG. 47

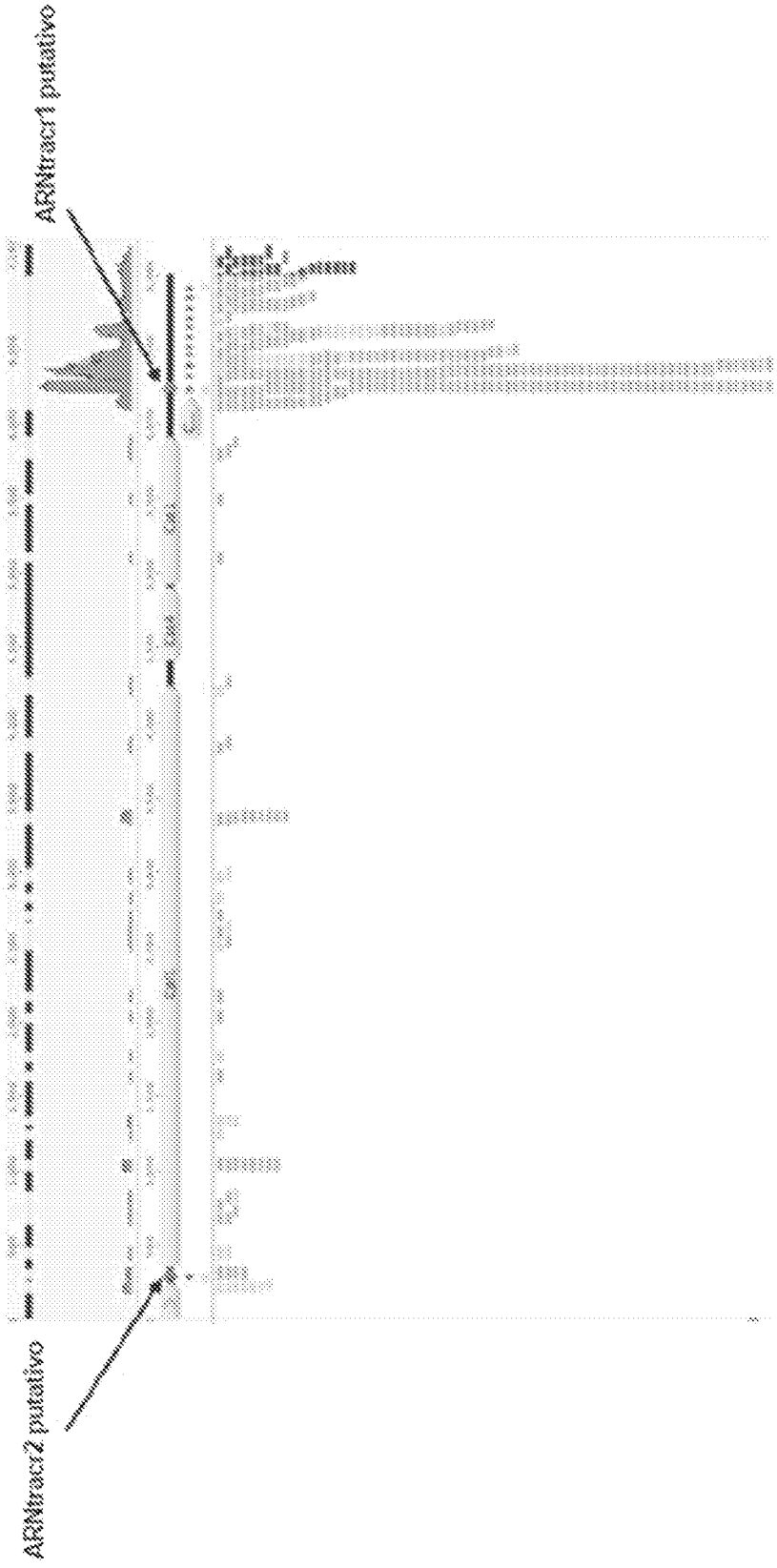


FIG. 48

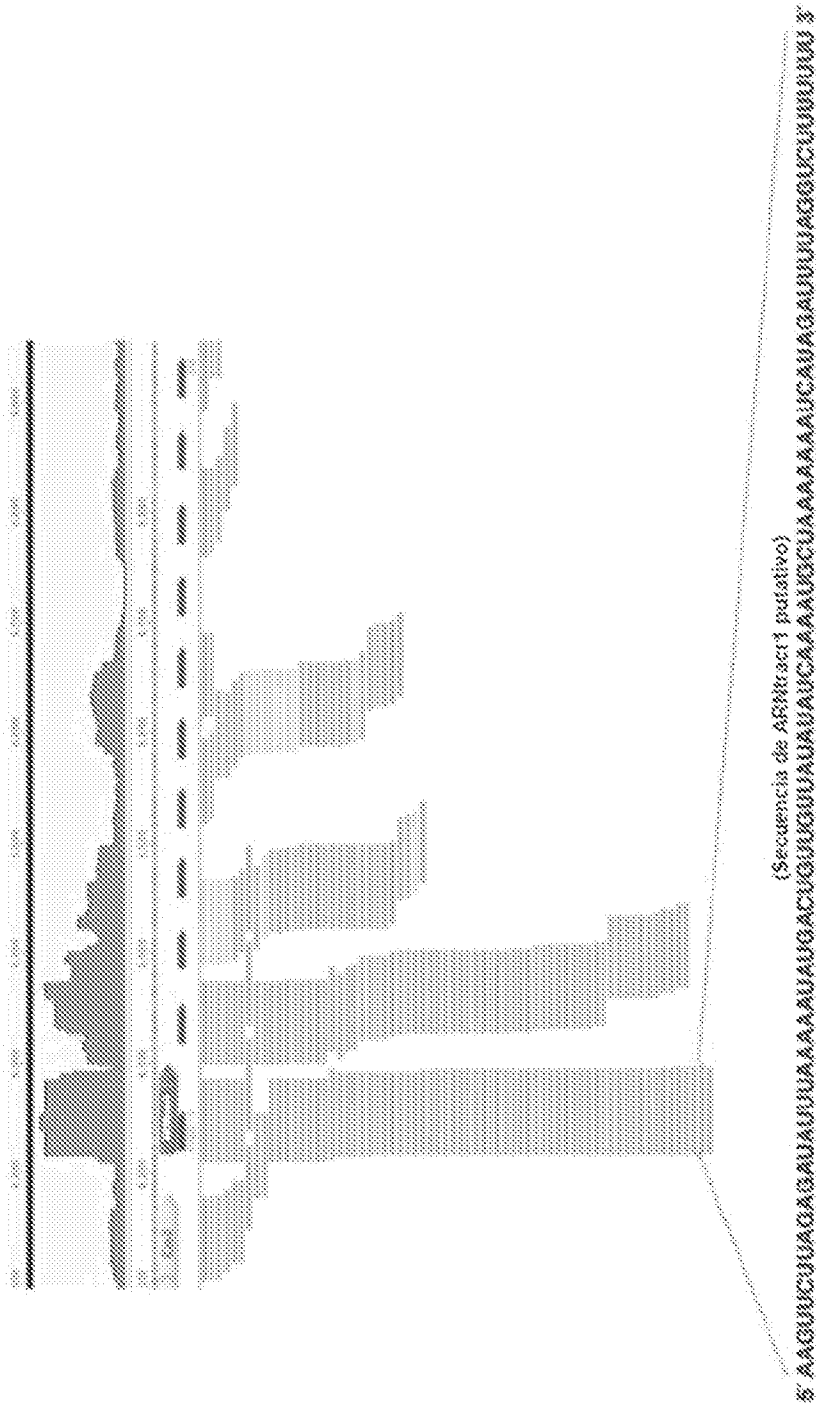

(Secuencia de ARNtract1 putativo)
5' AAGUUCUUAGAUAUUUAAAAUAUGACUGUUGUUAUAUCAAAAUCCUAAAAAAAUCAUAGAUUUAGGUCUUUUUU 3'

FIG. 49

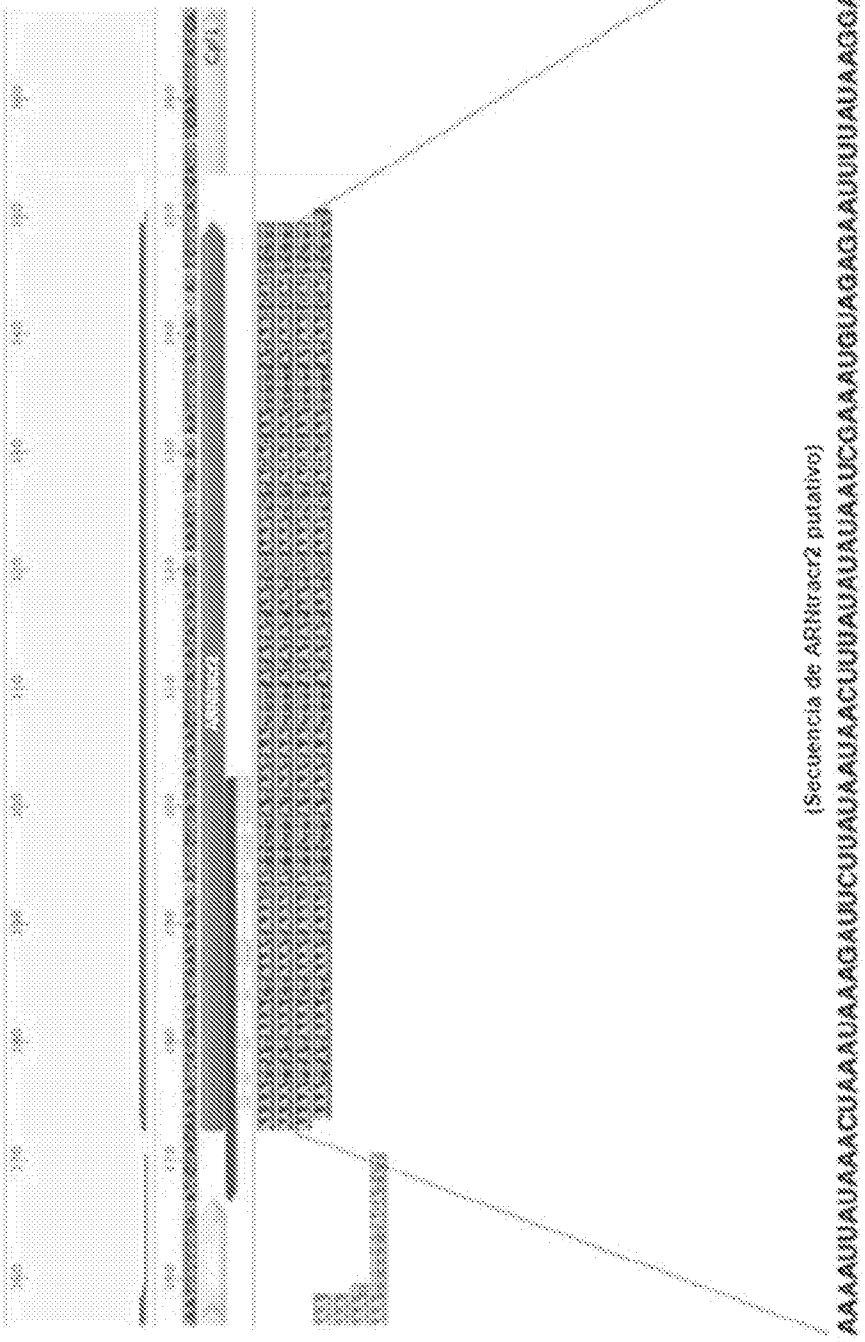

AthtracR2
(Secuencia de AthtracR2 putativo)
5' AAAAUUAUAAACUAAAUAAAGAUUCUUAUAAUAAUAACUUUAUAUAUAAUCGAAAUGUAGAGAAUUUUAUAAAGGAGUCUU 3'

FIG. 50

Versión 1: 5' ACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGGCCACTGTTAAAAGTCTAAGAAACTTTAAATA 3'

Versión 2: 5' CTACTATTCCTGTCCCTTCAGATAATTCAGTCTAAGAACTTTAAATAATTTCTACTGTTGTAGATGTCTAGAAGCCTT 3'

FIG. 51

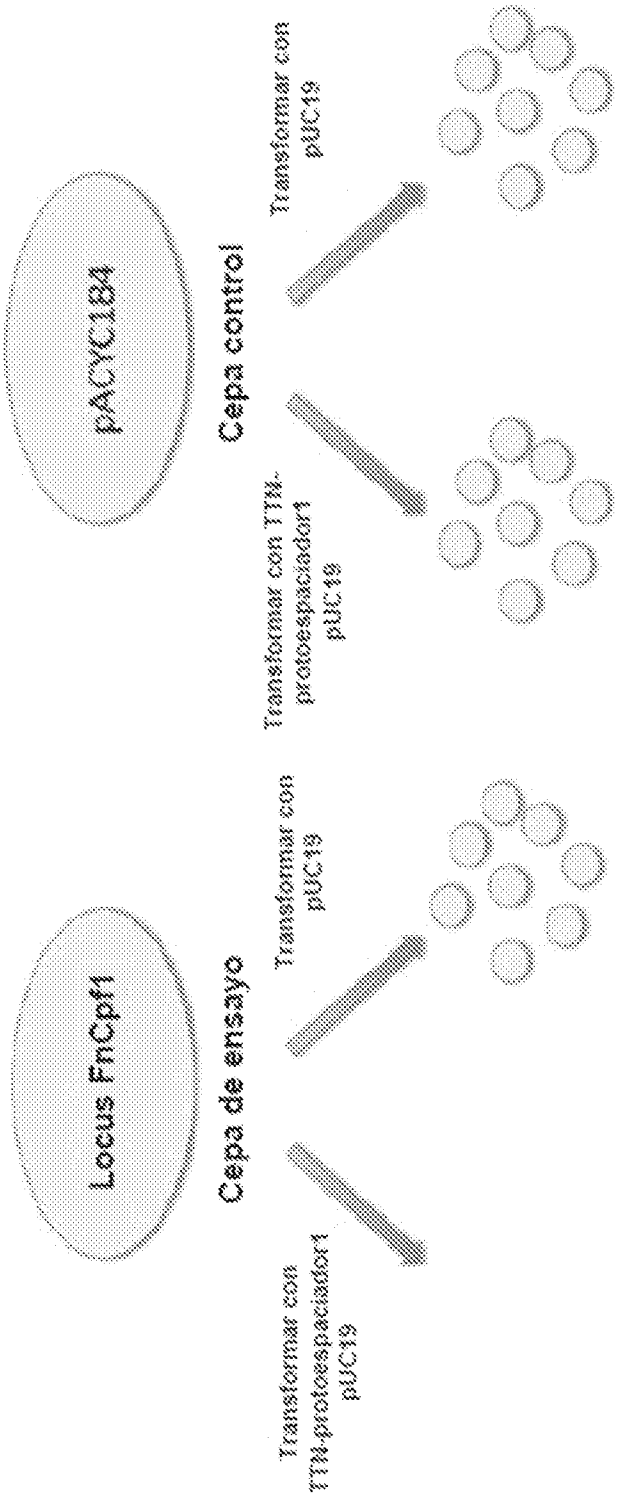

pACYC184
Cepa control
Transformar con pUC19
Transformar con LTM-proteasadeputa pUC19
Locus FnCpf1
Cepa de ensayo
Transformar con pUC19
Transformar con LTM-proteasadeputa pUC19

FIG. 52

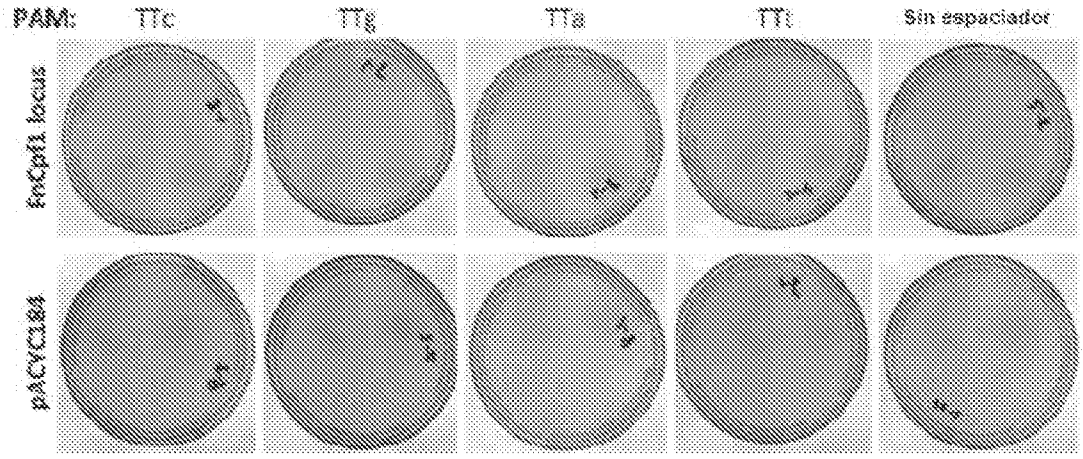


FIG. 53

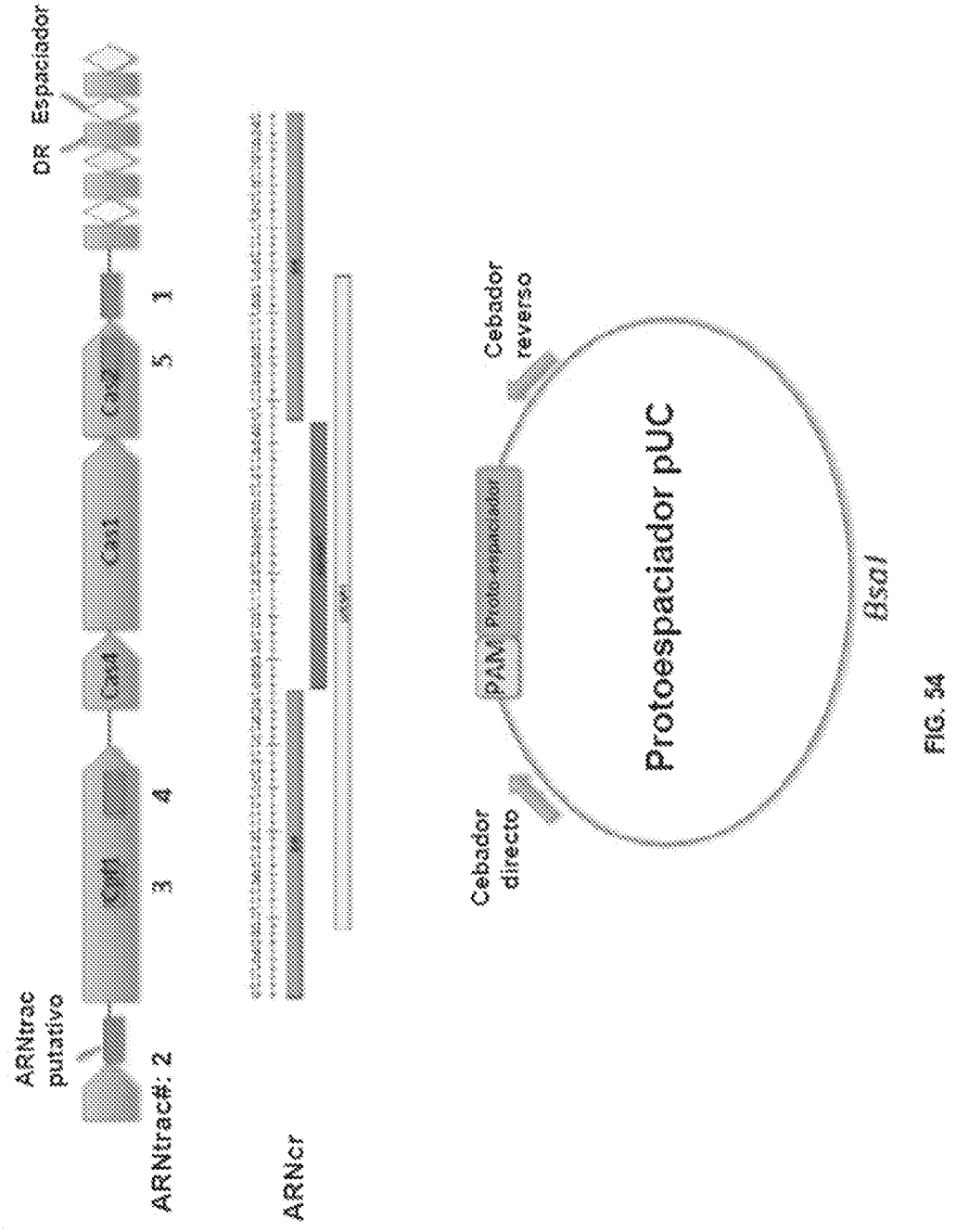

ARNtrac putativo
ARNtrac#: 2
3
4
5
1
DR
Espaciador
ARNcr
Cebador directo
Cebador reverso
Protoespaciador pUC
BsaI

FIG. 34

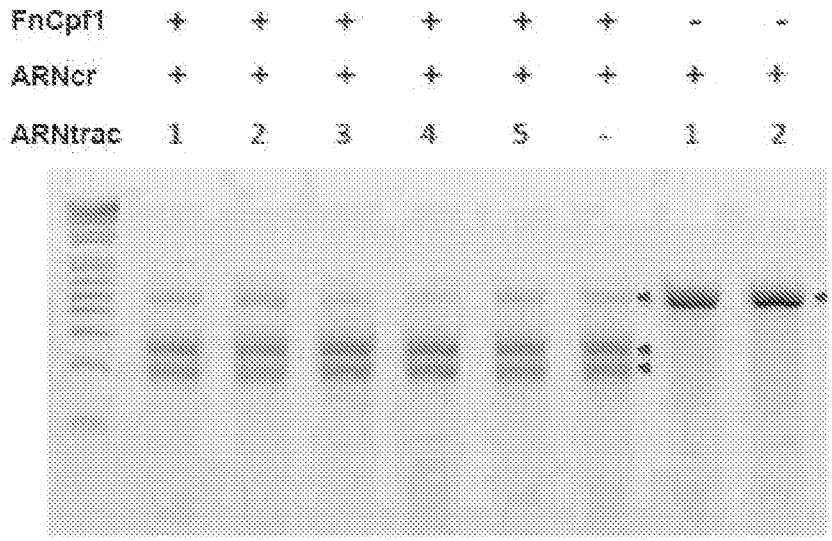

FnCpf1 + + + + + + - -
ARNcr + + + + + + + +
ARNtrac 1 2 3 4 5 - 1 2

FIG. 55

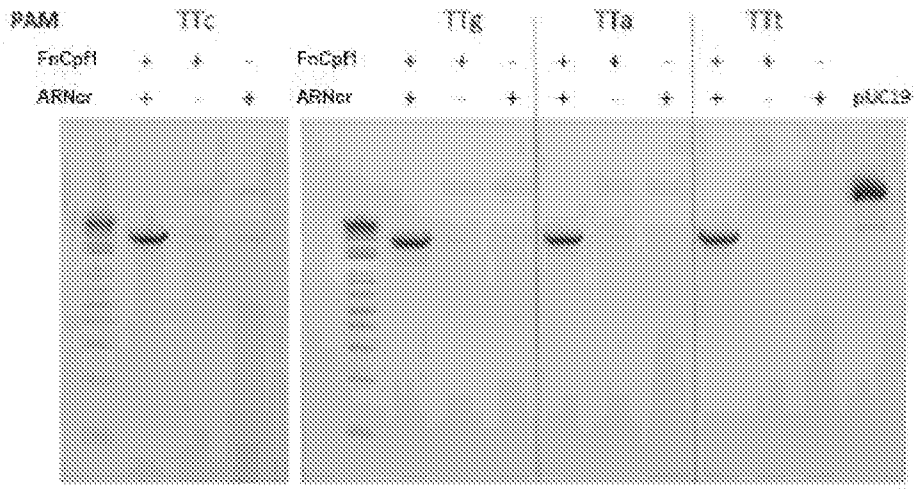

PAM
TTc
TTg
TTa
TTt
FnCpf1
ARNcr
pUC19

FIG. 56

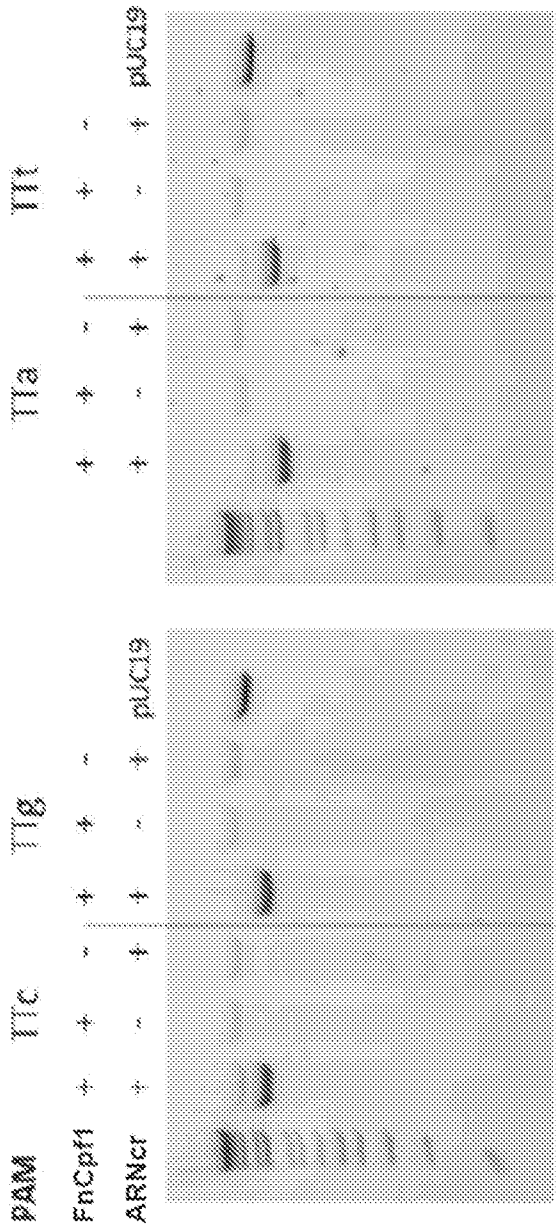

PAM
TTc
TTg
TTa
TTt
FnCpf1
ARNcr
pUC19
Digestión con BscI después de la incubación en lisado celular
Protoespaciador pUC

FIG. 57

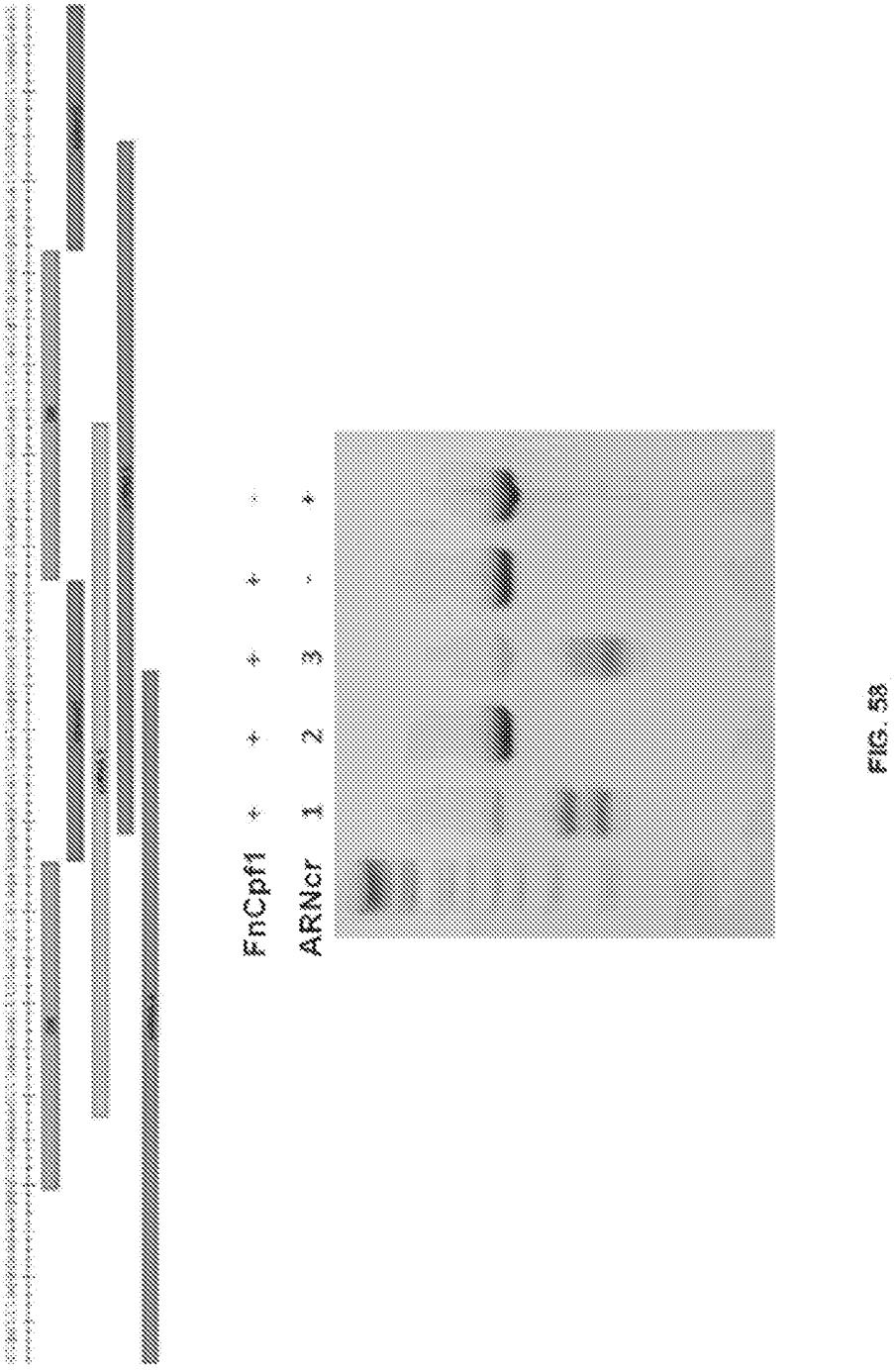
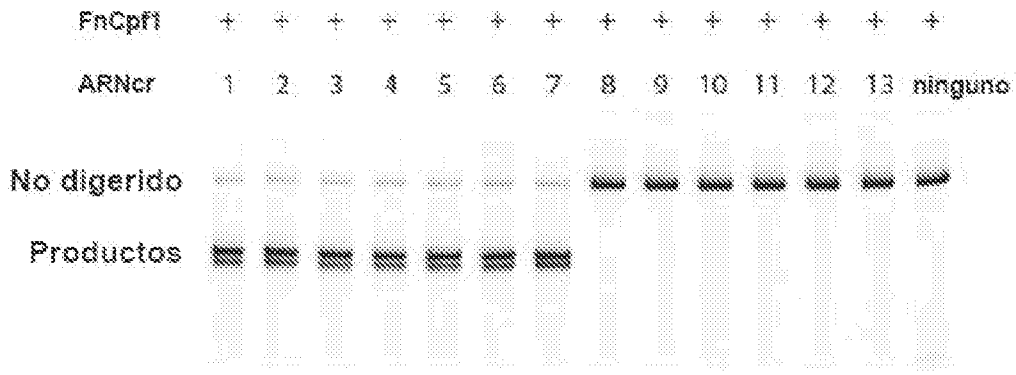


FIG. 53



ARNcr			
1	[illegible]	[illegible]	[illegible]
2	[illegible]	[illegible]	[illegible]
3	[illegible]	[illegible]	[illegible]
4	[illegible]	[illegible]	[illegible]
5	[illegible]	[illegible]	[illegible]
6	[illegible]	[illegible]	[illegible]
7	[illegible]	[illegible]	[illegible]
8	[illegible]	[illegible]	[illegible]
9	[illegible]	[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]
11	[illegible]	[illegible]	[illegible]
12	[illegible]	[illegible]	[illegible]
13	[illegible]	[illegible]	[illegible]

FIG. 59

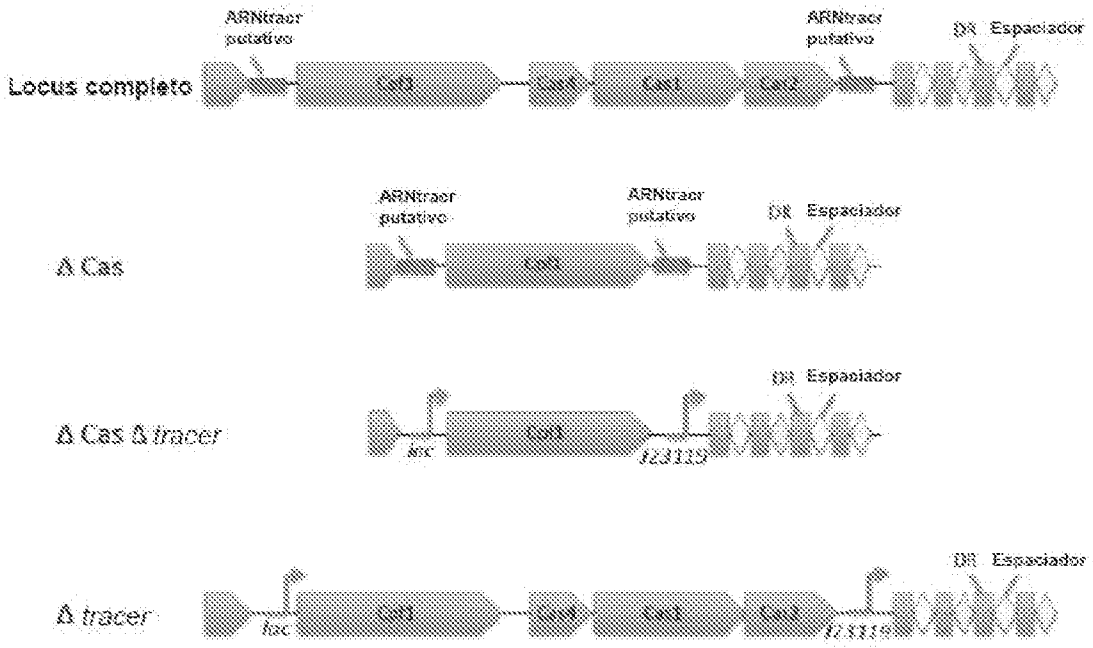

Locus completo
ARNtracr putativo
ARNtracr putativo
DR
Espaciador
Cas3
Cas4
Cas1
Cas2
Δ Cas
ARNtracr putativo
ARNtracr putativo
DR
Espaciador
Cas3
Δ Cas Δ tracer
DR
Espaciador
Cas3
Δ tracer
DR
Espaciador
Cas3
Cas4
Cas1
Cas2

FIG. 60

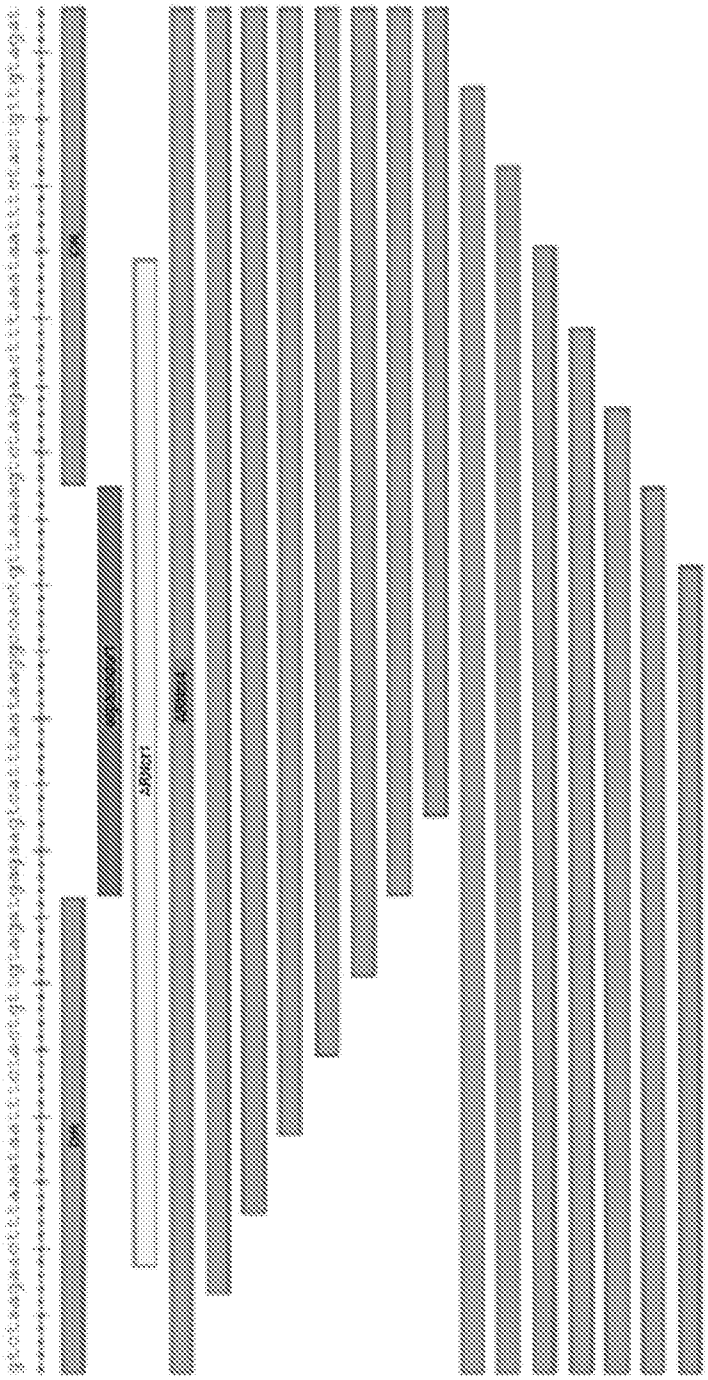


FIG. 61

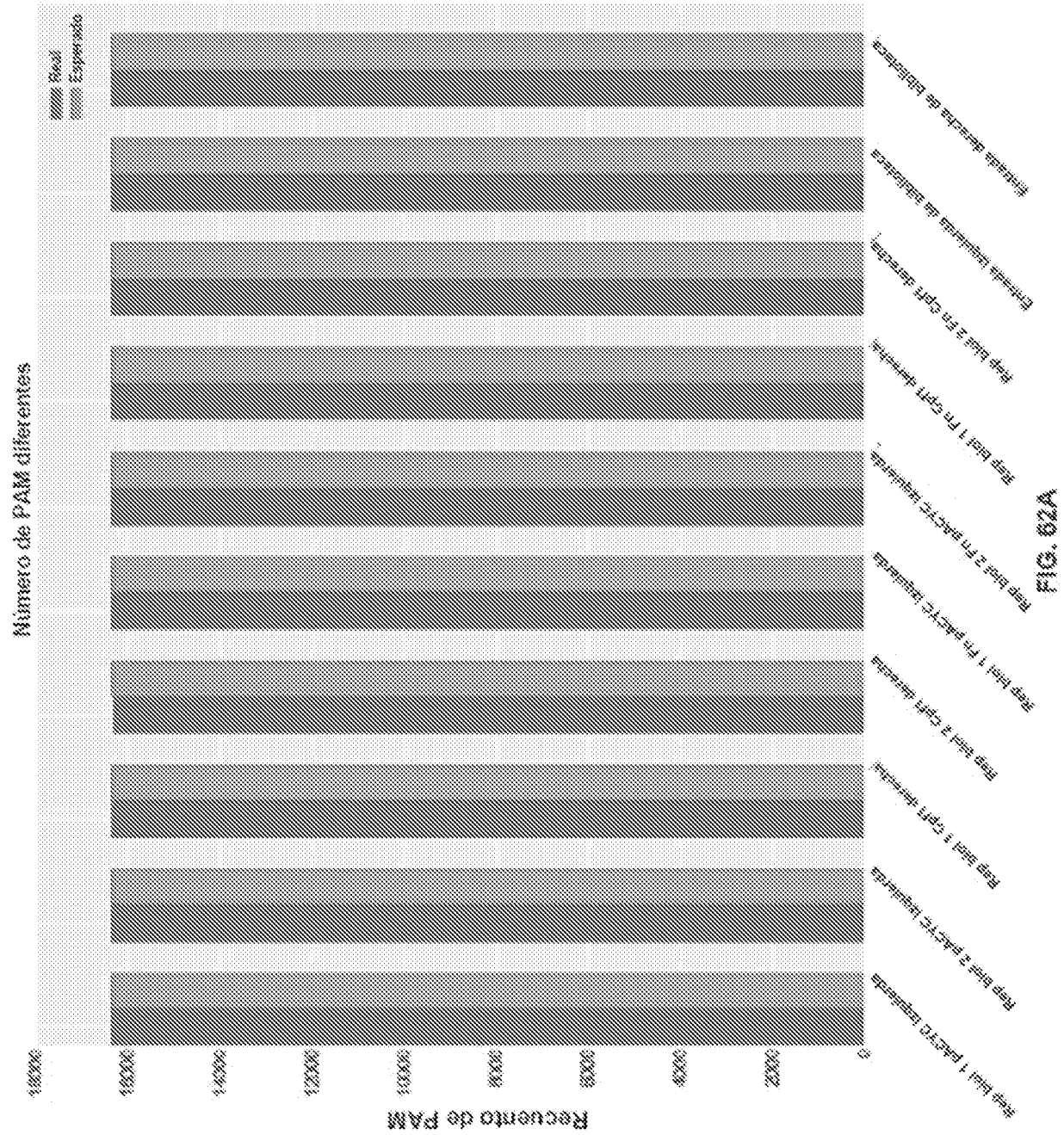

Real
Esperado
Número de PAM diferentes
Recuento de PAM
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Rep biol 1 pacyc tagmentada
Rep biol 2 pacyc tagmentada
Rep biol 1 cpfi ilumina
Rep biol 2 cpfi ilumina
Rep biol 1 fn pacyc tagmentada
Rep biol 2 fn pacyc tagmentada
Rep biol 1 fn cpfi ilumina
Rep biol 2 fn cpfi ilumina
Enriquecida duplicados de biblioteca
Enriquecida duplicados de biblioteca
Enriquecida duplicados de biblioteca

FIG. 62A

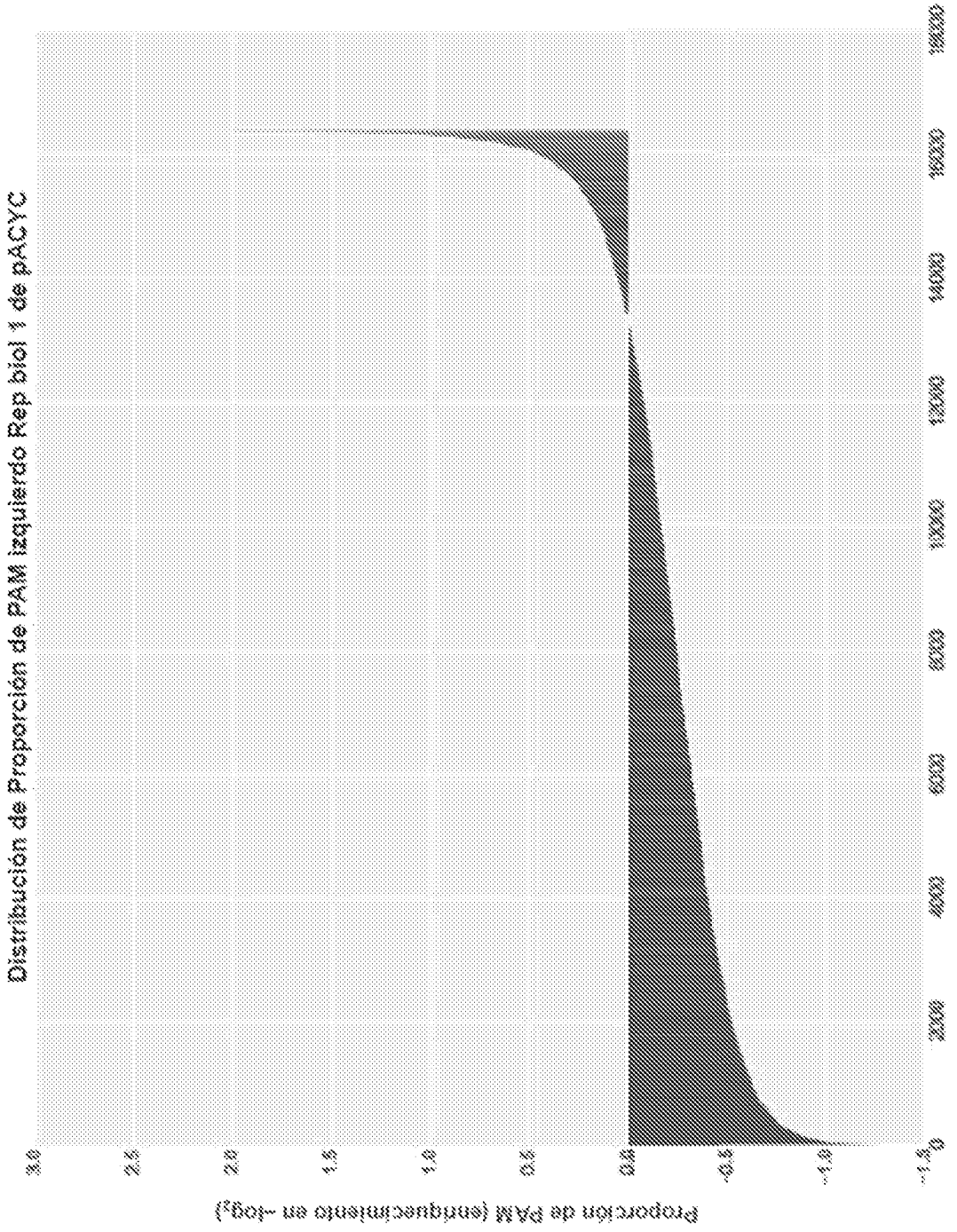

Distribución de Proporción de PAM izquierdo Rep biol 1 de pACYC
Proporción de PAM (enriquecimiento en -log2)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
Rango

FIG. 62B

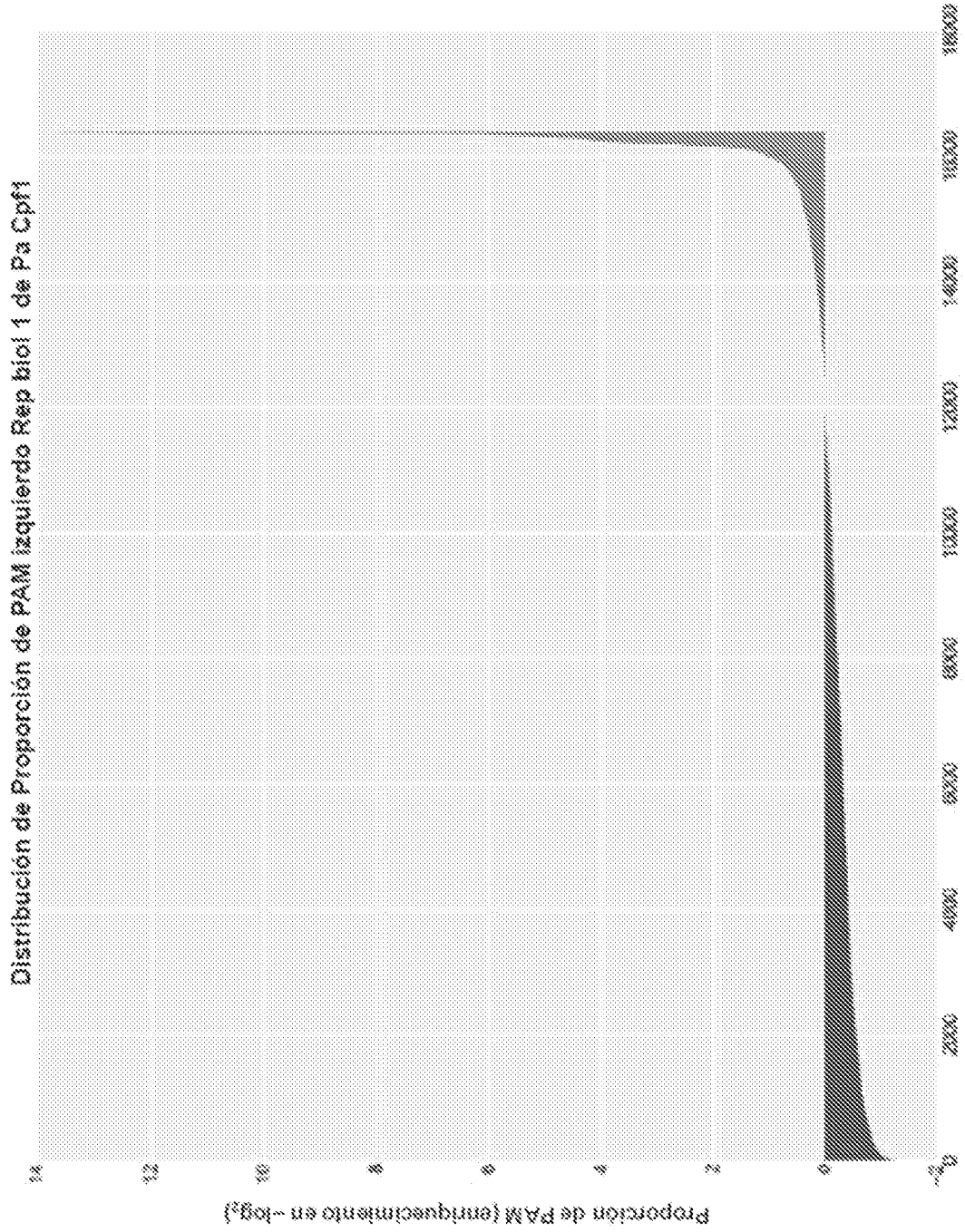

Distribución de Proporción de PAM izquierdo Rep biol 1 de Pa Cpf1
Proporción de PAM (enriquecimiento en -log10)
Rango
0
2000
4000
6000
8000
10000
12000
14000
16000

FIG. 62C

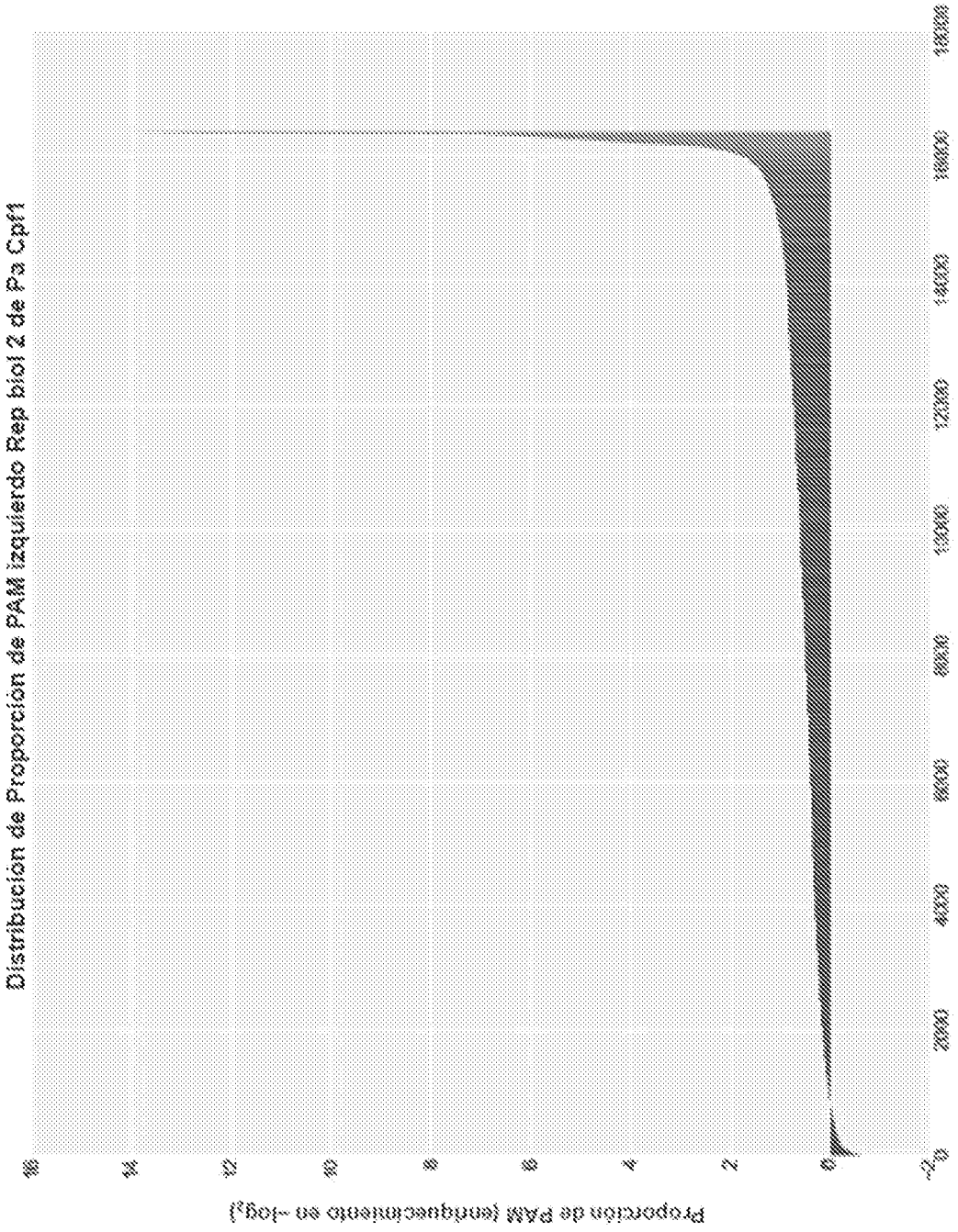

Distribución de Proporción de PAM izquierdo Rep biol 2 de Pa Cpf1
Proporción de PAM (endoguecimiento en -log2)
Rango

FIG. 62D

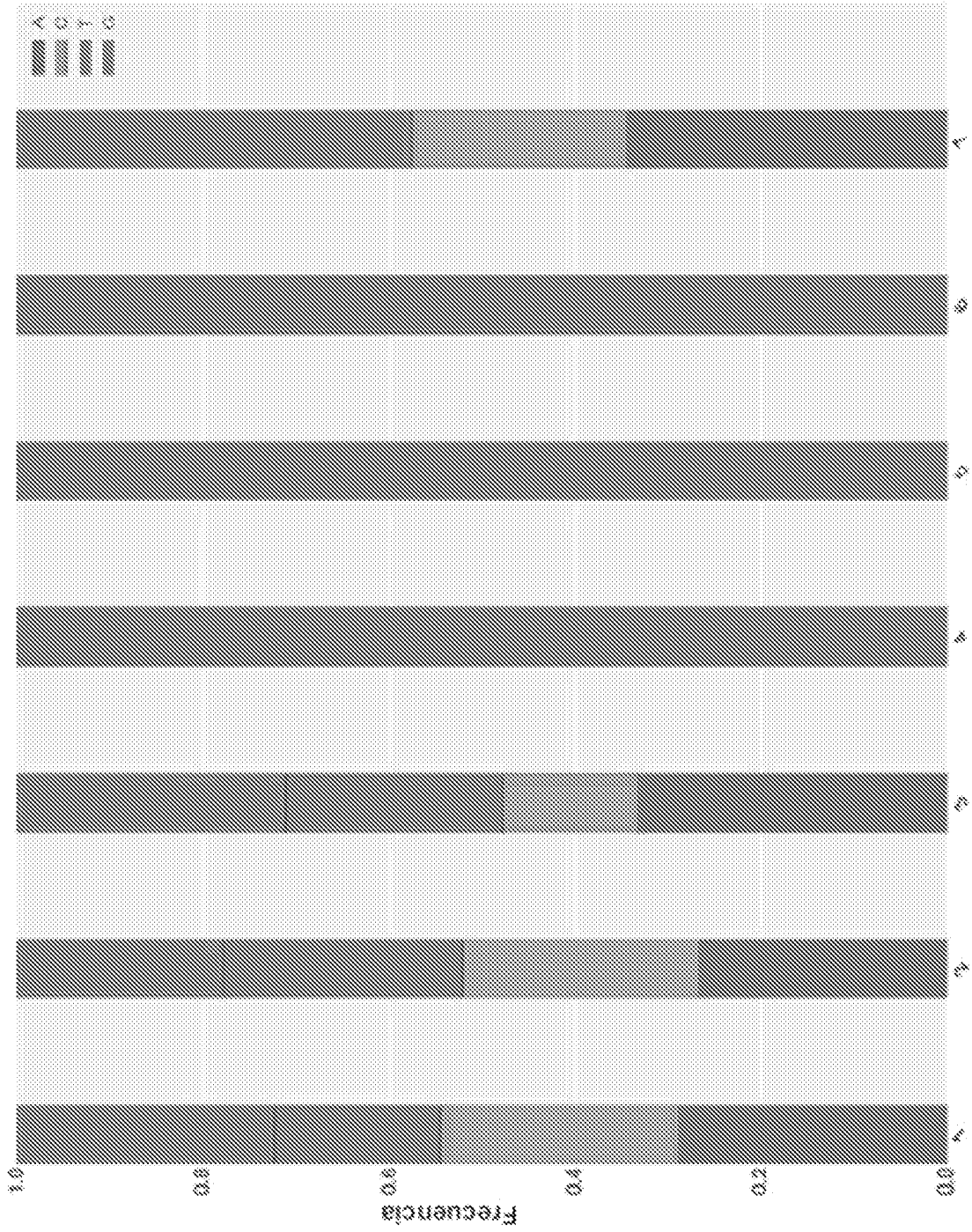


FIG. 62E

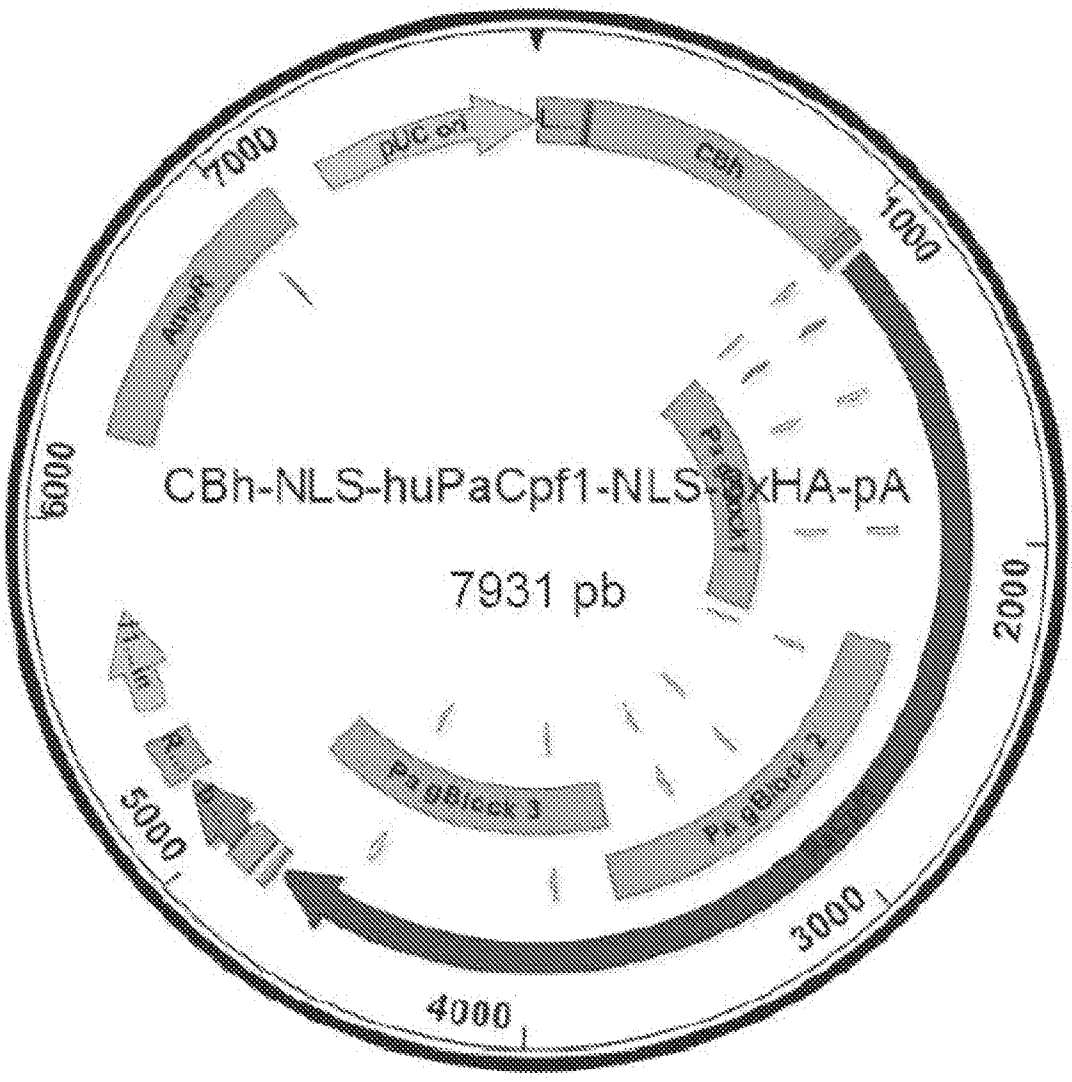

CBh-NLS-huPaCpf1-NLS-3xHA-pA
7931 pb
1000
2000
3000
4000
5000
6000
7000

FIG. 63

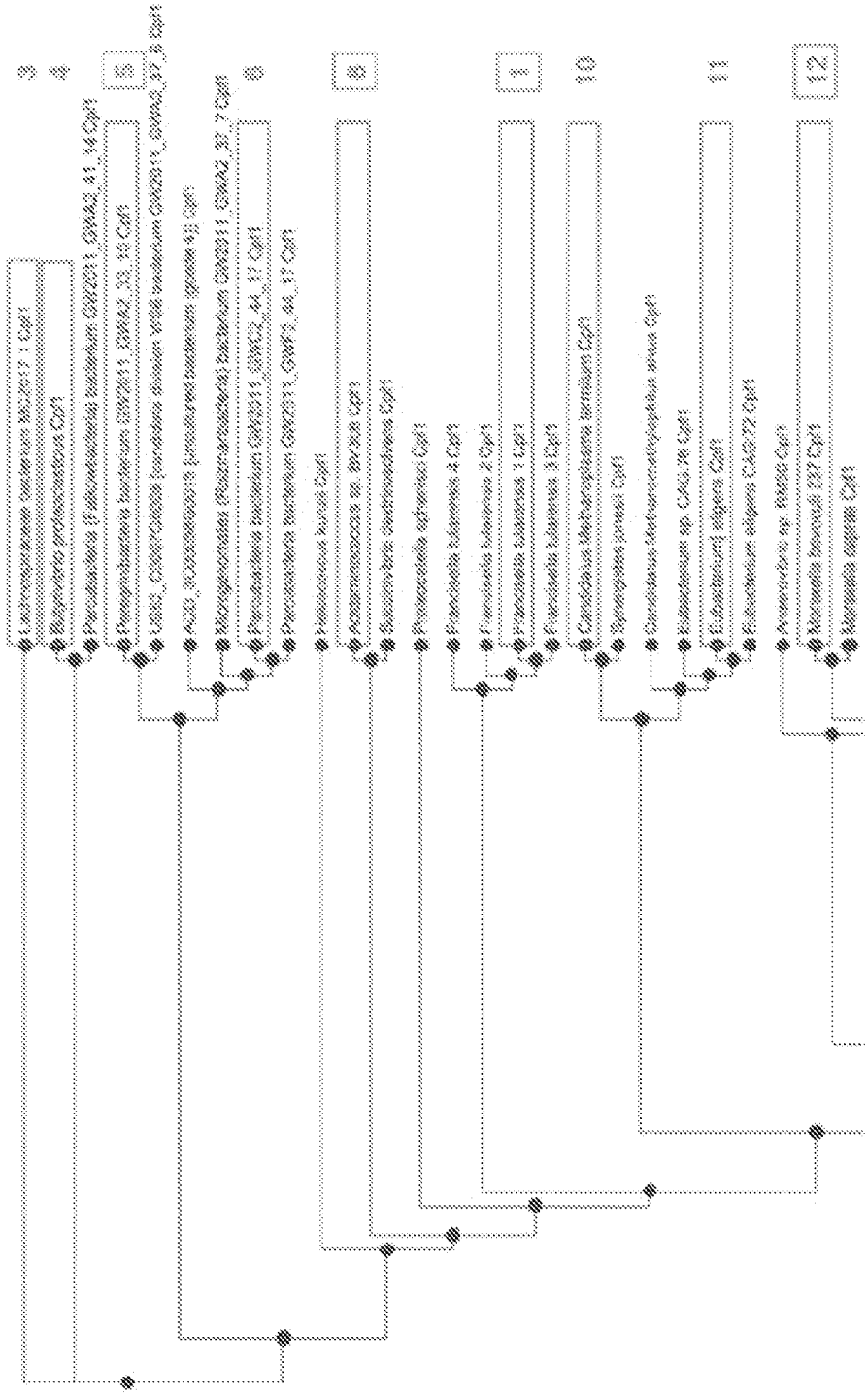


FIG. 64A

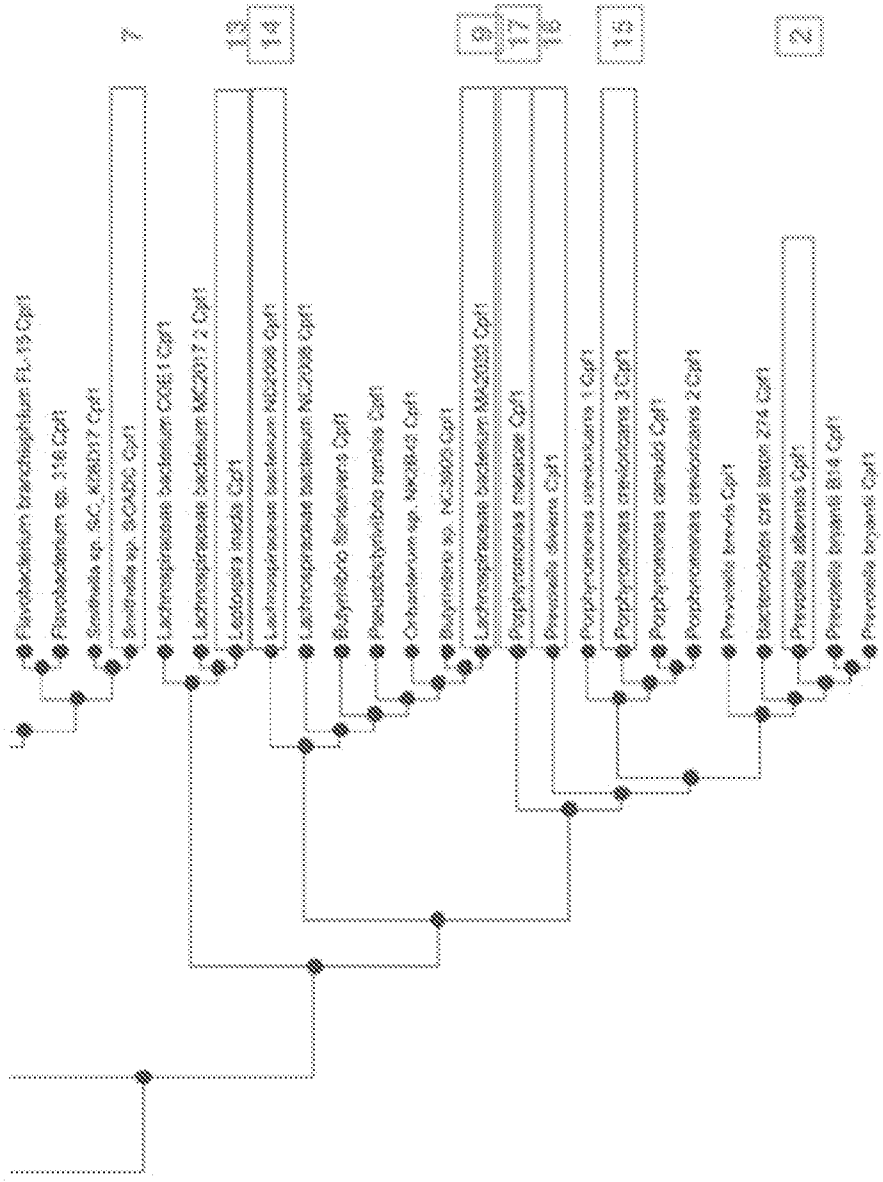

7
12
14
6
17
18
16
2
Ruminococcus bromii L2-63 Cpt1
Ruminococcus sp. K-1 Cpt1
Roseburia sp. 11SE37 Cpt1
Roseburia hominis A2-183 Cpt1
Roseburia intestinalis L1-82 Cpt1
Roseburia inulinivorans DSM 16841 Cpt1
Eubacterium rectale A1-86 Cpt1
Eubacterium eligens ATCC 27750 Cpt1
Eubacterium ventriosum GA1 Cpt1
Eubacterium siraeum 70/3 Cpt1
Faecalibacterium prausnitzii L2-6 Cpt1
Faecalibacterium prausnitzii M21/2 Cpt1
Faecalibacterium prausnitzii A2-165 Cpt1
Faecalibacterium prausnitzii SL3/3 Cpt1
Clostridium sp. L2-50 Cpt1
Clostridium sp. M62/1 Cpt1
Clostridium sp. L2-7 Cpt1
Clostridium sp. SS2/1 Cpt1
Clostridium sp. L2-29 Cpt1
Clostridium sp. M62/29 Cpt1
Clostridium sp. L2-30 Cpt1
Clostridium sp. L2-50 Cpt1
Clostridium sp. M62/1 Cpt1
Clostridium sp. L2-7 Cpt1
Clostridium sp. SS2/1 Cpt1
Clostridium sp. L2-29 Cpt1
Clostridium sp. M62/29 Cpt1
Clostridium sp. L2-30 Cpt1

FIG. 04B

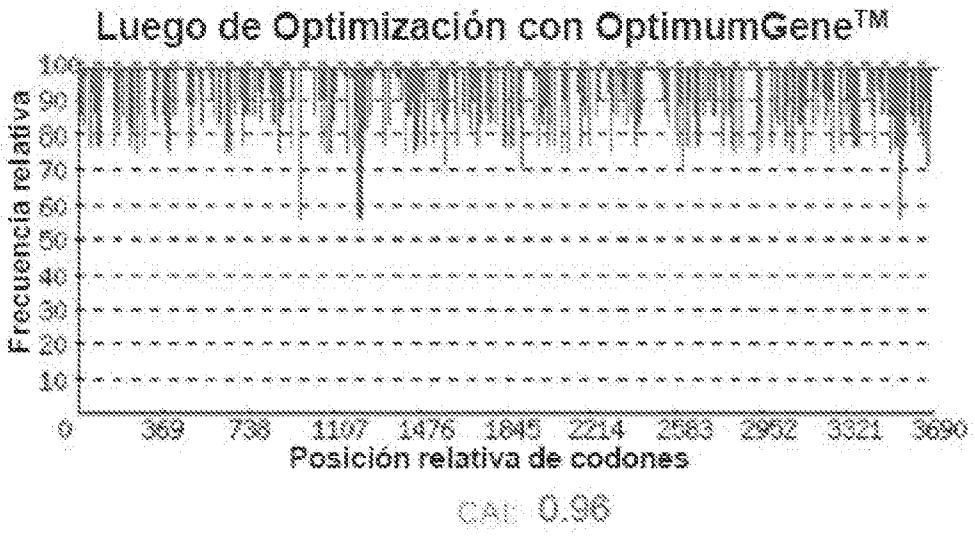


FIG. 65A

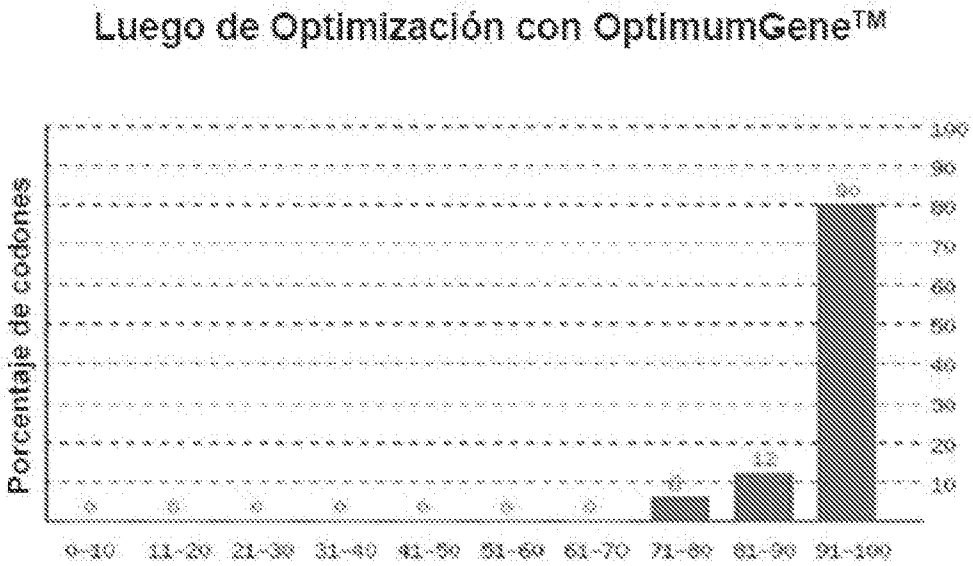


FIG. 65B

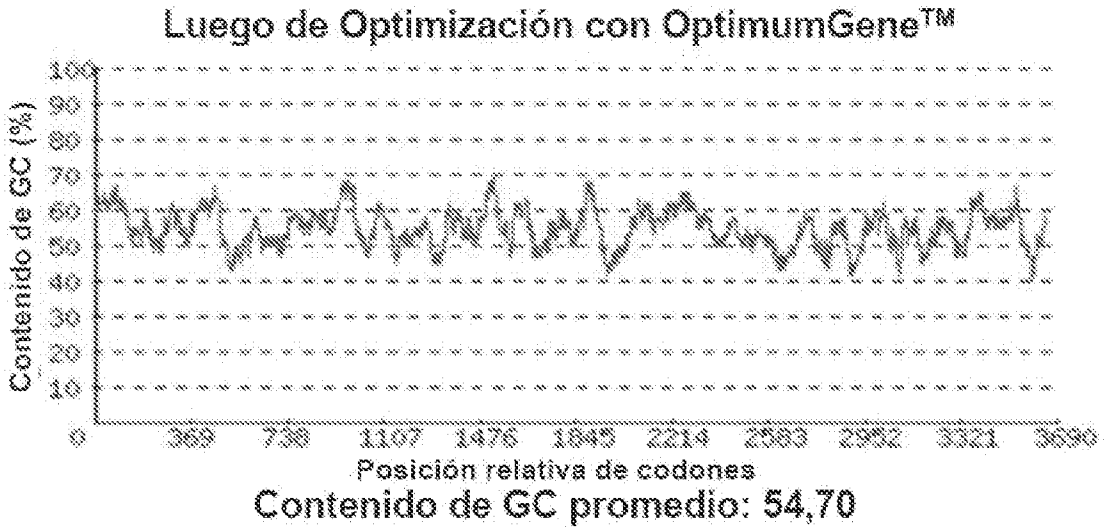

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
369
738
1107
1476
1845
2214
2583
2952
3321
3690
Posición relativa de codones
Contenido de GC promedio: 54,70

FIG. 65C

Enzimas de restricción	Optimizado
* Verde: sitios filtrados, azul: sitios confirmados (no filtrados); Rojo: sitios mantenidos	
BamHI(GGATCC)	1(3754)
EcoRI(GAATTC)	1(3844)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3818)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PoliA(AATAAA)	0
PoliA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PoliT(TTTTTT)	0
PoliA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 65D

Luego de la optimización

Repetición directa Máx: Tamaño:15 Distancia:2688 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de diadas Máx: Tamaño:12 Tm: 49,0 Posiciones de inicio: 1666, 1908

FIG. 65E

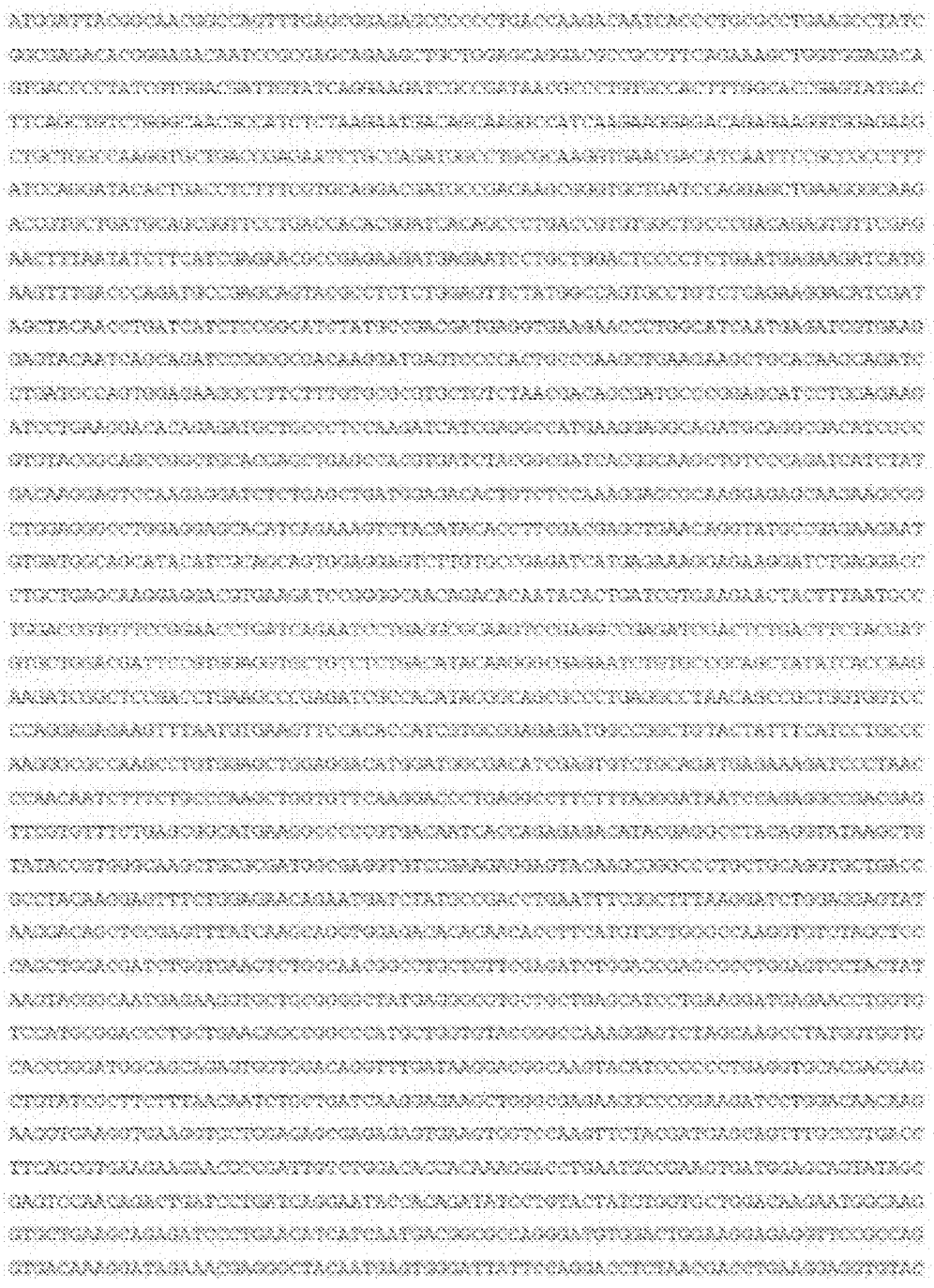


FIG. 65F



FIG. 65G

MDYGNQQPERSAPLIKTITLRLKPIGETRETIRSQKLLEQDAAPRKLVETVTPIVDDCIR
KIADNALCHFGTEYDPSCLQNAISKNDSKAIKKETEKVEKLLAKVLTENLPDGLRKVNDI
NSAAFIQDTLTSFVQDDADKRVLIQELKGKTVLMQRFLTTRITALTVWLPDRVFENFNIF
IENAERNRILLDSPLNPKIMKFDPDAEQYASLEFYGQCLSQKDIDSYNLIISGIYADDEV
KNPGINEIVKEYNQQIRGDFDESPLPKLKKLHKQILMPVEKAPFVRVLSNDSDARSILEK
ILKDTENLPSKIIEAMKEADAGDIAVYGSRLHELSNVIYGDNGKLSQIIYDKESKRISEL
METLSPKERKESKKRLEGLEEHIRESTYTFDELNRYAEKNVMAAYIAAVEESCAEIMRKE
KDLRTLLSKEDVKIPGNRHNTLIVKNYPNAKTVFRNLIKILERKGEARIDSDFYDVLIDS
VEVLSLTYKGENLCRSYITKKIGSDLKPEIATYGSALRPNSRWWSPGEKFNVKPRTTIVRR
DGRLIYFILPKGAKPVELEDMDGDIECLQMRKIPNPTIFLPKLVPKDPEAFFRDNPEADE
FVFLSGMKAPVTITRETYEAYRYKLYTVGKLRDGEVSEEEYIRALLQVLTAYKEFLENRM
IYADLNPGPKDLEEYKDSGEPIKQVETHNTFMCNAKVSSSQLDDLVKSGNGLLFEIWGER
LESTYKYGNEKVLRGYEGVLLSILKDENLVSMRTLLNSRPMLVYRPKEGSSEPMVVHRDGS
RVVDRFDNDGKYIPPEVHDELYRFPRNLLIKEKLGEKAPKILDNEKVKVKVLESERVKNS
KFYDEQFAVTFSVKENADCLDTTKDLNAEVMEQYSESNRLILIRNTTDILYYLVLDKNGK
VLKQRSLNIINDGARDVDWKERFRQVTKDRNEGYNEWDYSRTSNDLKEVYLNYALKEIAE
AVIEYNAILIIEKMSNAFKDKYSPLDDVTFEGFETKLLAKLSDLNFRGIKDGEPCSFTNP
LQLCQNDSNKILQDGVIFNVPNSMTRSLDFDTGFIFAINDHNIRTKKAKLNFLSKFDQLK
VSSEGCLIMKYSGDSLPTHNTDNRVWNCCCMKPITNYDRETKEVEFIEEPVEELSRVLEE
NGIETDTELNKLNEREMVPGKVVDAIYSLVLNYLRGTVSGVAGQRAVTYSPVTGKEYDIS
FIQAMRLNRKCDYYRIGSKERGEWTDPVAQLIN

FIG. 65H

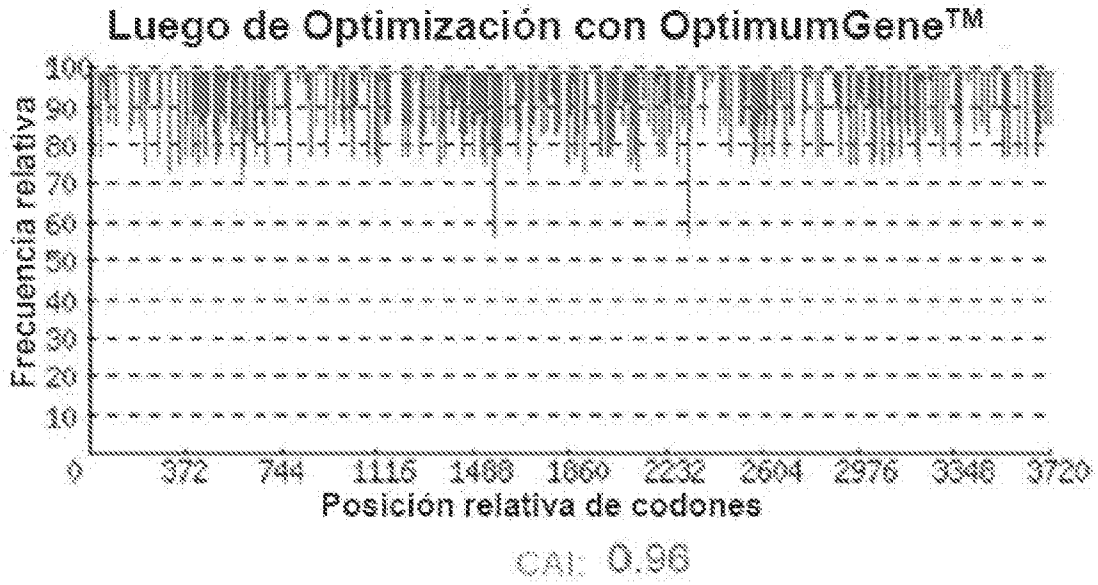


FIG. 66A

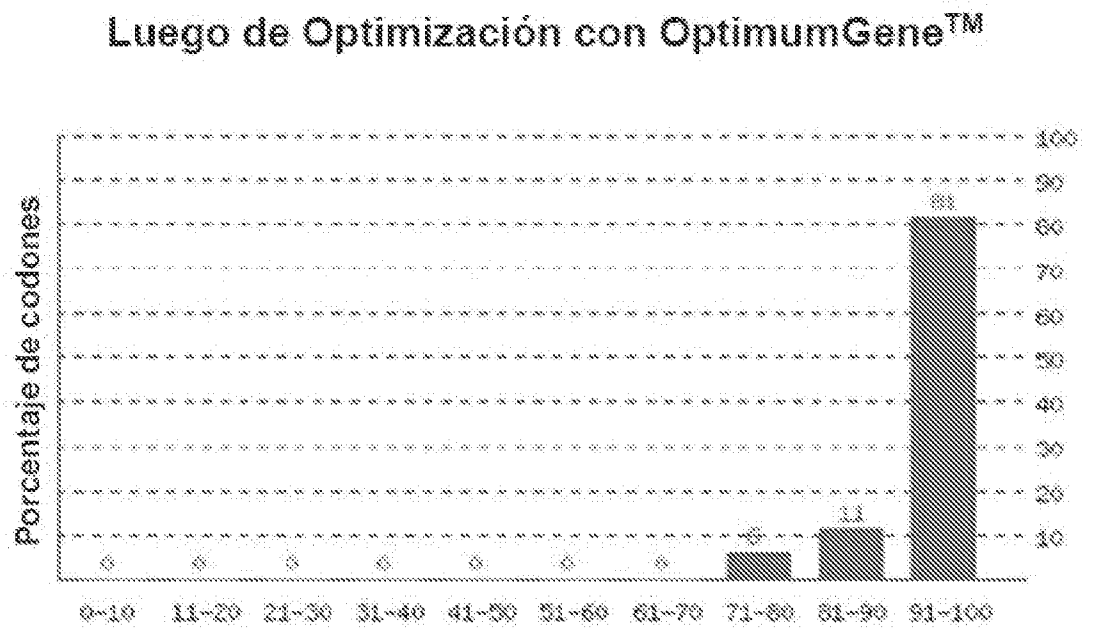


FIG. 66B

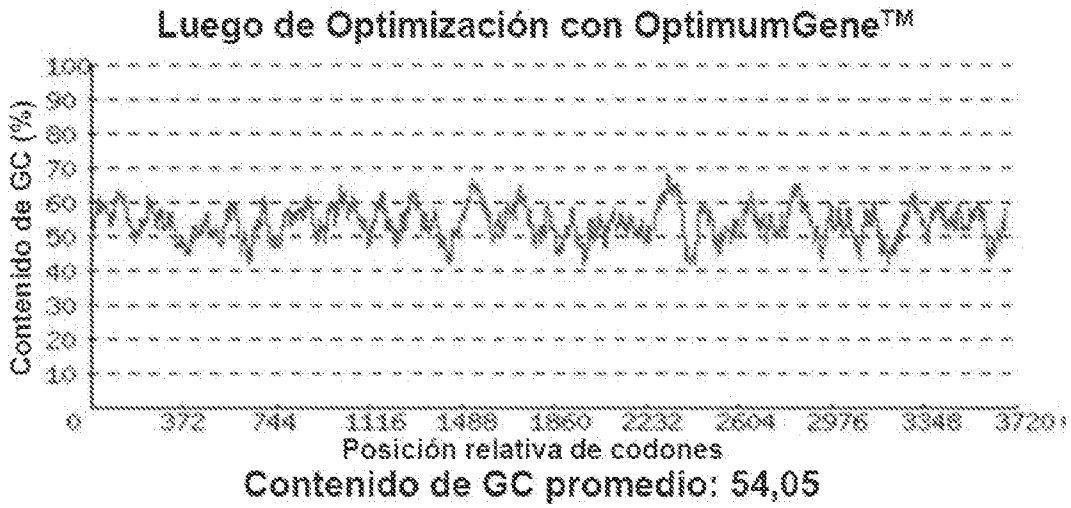

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
372
744
1116
1488
1860
2232
2604
2976
3348
3720
Posición relativa de codones
Contenido de GC promedio: 54,05

FIG. 66C

Enzimas de restricción	Optimizado
* verde: sitios diana, azul: sitios confirmados (no diana), Rojo: sitios eliminados	
BamHI(GGATCC)	1(3778)
EcoRI(GAATTC)	1(3883)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3842)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PolIA(AATAAA)	0
PolIA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PolIT(TTTTTT)	0
PolIA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 66D

Luego de la optimización

Repetición directa Máx: Tamaño:12 Distancia:3285 Frecuencia:2

Repetición invertida Máx: Tamaño: 12 Tm: 43,5 Posiciones de inicio: 1631, 1547

Repetición de diadas Máx: Ninguna

FIG. 66E

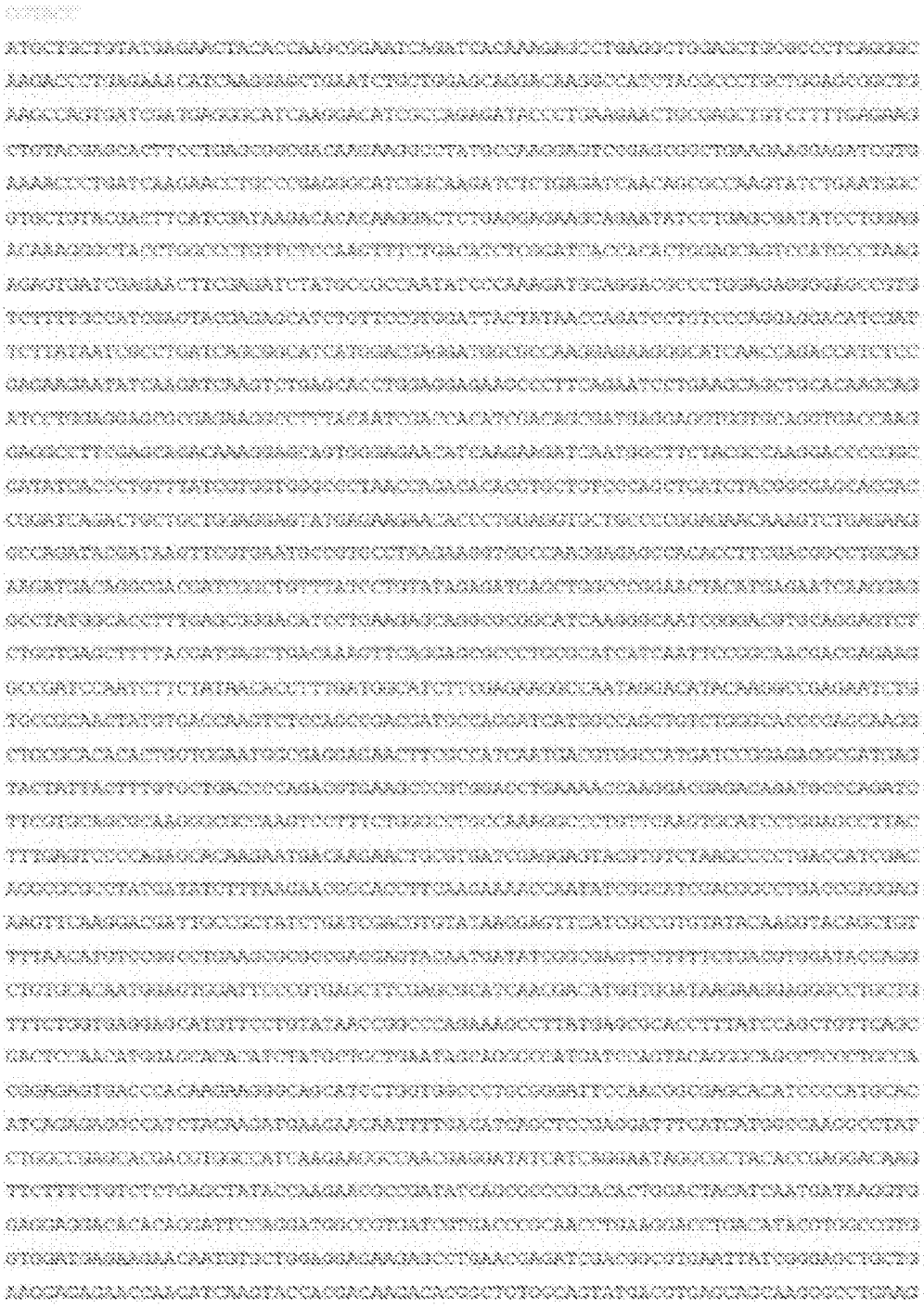


FIG. 66F



FIG. 66G

MLLYENYTKRNQITKSLRLELRPQGKTLRNIEELNLLEQDKAIYALLERLKPVIDRGIKD
IARDTLKNCELSFEELVEHFLSGDKEAVAFESERLKKEIVKTLIKNLPEGIGELSEINSA
KYLNGVLYDFIDKTHKDSEEKQNILSDILETGYLALFSKFLTSRITTLEQSMPKRVIEN
FEIYAANIPKNQDALERGAVSFAIEYESICSVDYYNQILSQEDIDSYNRLISGIMDEDGA
KEKGINQTISENNIKIKSEHLEEKPFRILKQLHKQILEEREKAFTIDNIDSDEEVVQVTK
EAFEQTEEQWENIKKINGFYAKDPGDITLFIVVGPNQTHVLSQLIYGEHDRIKLLLEEYE
KNTLEVLPRRTKSEKARYDKFVNAVPKKVAKESHTFDGLQKMTGDDRLFILYRDELARNY
MRIKEAYGTFERDILKSRPGIKGNRDVQESLVSFYDELTKFRSALRIINSGNDEKADPIF
YNTFDGIFEKANRTYKAENLCRNYVTKSPADDARIMASCLGTPARLRTHWWNGEENFAIN
DVAMIRRGDEYYYFVLTPDVKFVDLKTKDETDAQIFVQRKGAKSFLGLFKALFKCILEPY
FESFEHENDKNCVIEEYVSKPLTIDRRAYDIFKNGTFEKTNIGIDGLTEEKFKDDCRYLI
DVYKEFIAVYTRYSCFNMSGLKRADEYNDIGEFFSDVDTRLCTMEWIPVSFERINDMVDK
KEGLLFLVRSNFLYNEPRKPYERTFIQLPSDSNMENTSMLLNSRAMIQYRAASLFRRVTH
KRGSILVALRDSNGEHIPMSIREAIYKMKNNFDISSEDPIMAKAYLAEHDVAIKRANEDI
IRNRRYTEDKFFLSLSYTKNADISARTLDYINDCVEEDTQDSRMAVIVTRNLKDLTYVAV
VDEKNNVLEEKSLNEIDGVNYRELLKERTKIKYHDKTELWQYDVSSKGLKEAYVELAVTQ
ISKLATKYNAVVVVESMSSTFKDKFSFLDEQIFKAFEARLCARMSDLSFNTIKEGEAGSI
SNPIQVSNNKGNSYQDGVIYFLNNAYTRTLCPDTGFVDVFDETRLITMQSKRQFFAKMKD
IRIDDGEMLFTFNLEEYPTKRLLDREWTVKIAGDGSYFDKDKGEYVYVNDIVKRQIIPA
LLEDKAVFDGNMAEKPLDKTAISGKSVELIYKNFANALYGIITKKDGEFIYRSPITGTRI
DVSKNTTYNFGNKFMFKQEYRGDGDFLDAFLNYNQAQDIAV

FIG. 66H

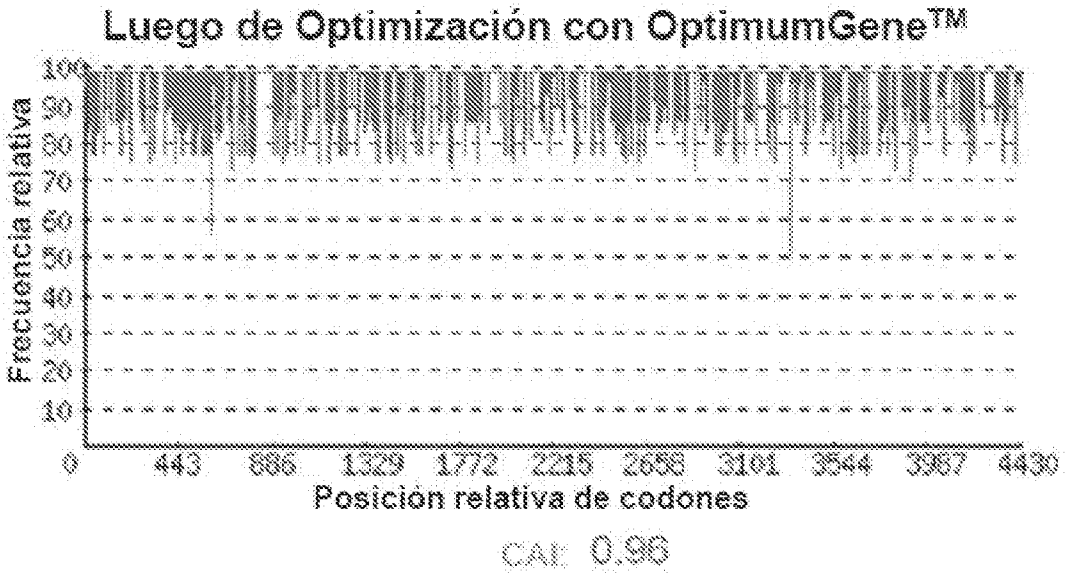


FIG. 67A

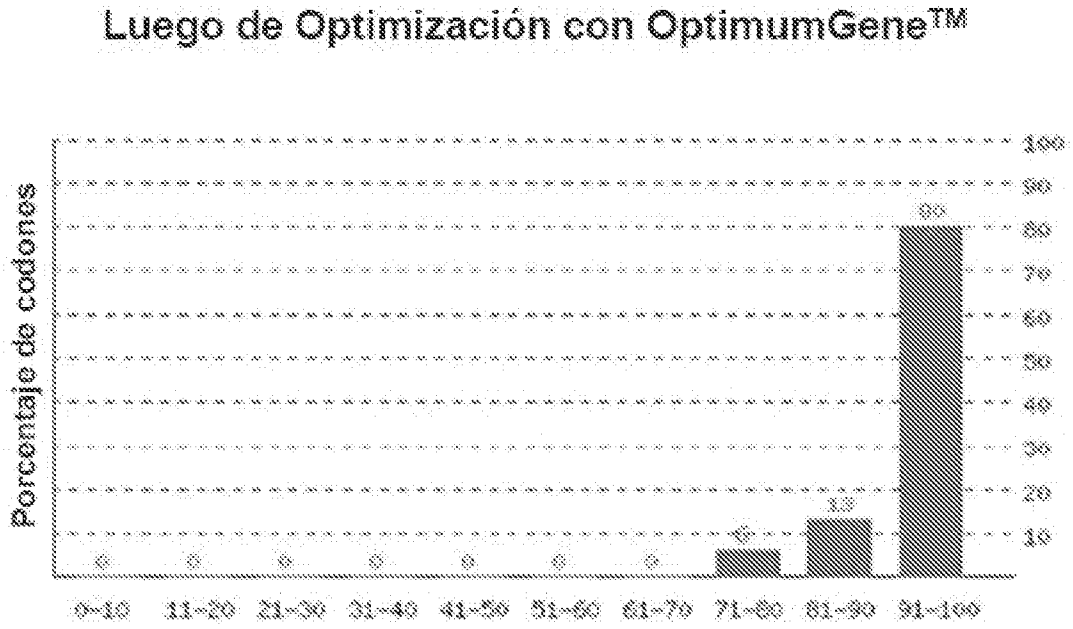


FIG. 67B

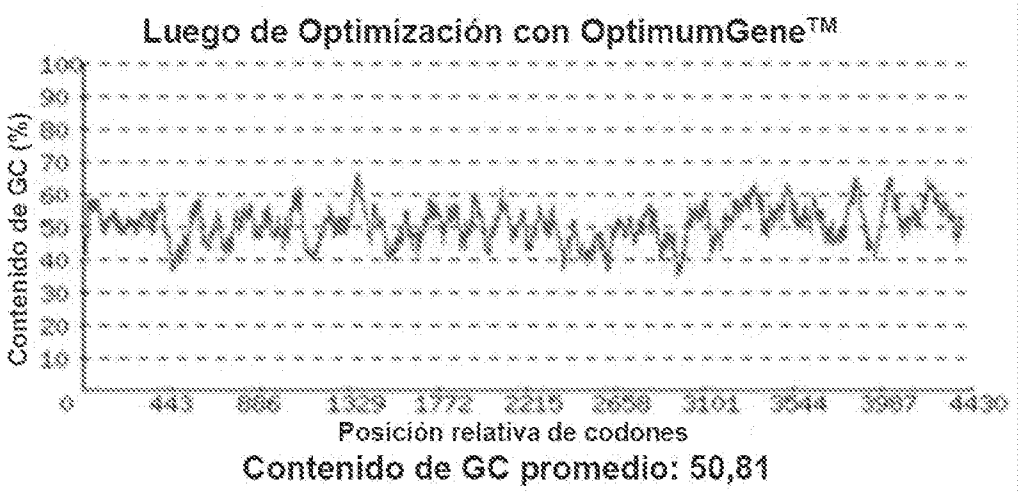

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
443
886
1329
1772
2215
2658
3101
3544
3987
4430
Posición relativa de codones
Contenido de GC promedio: 50,81

FIG. 67C

Enzimas de restricción	Optimizado
* verde: sitios filtrados, azul: sitios confirmados (no filtrados); Rojo: sitios mantenidos	
BamHI(GGATCC)	1(4486)
EcoRI(GAATTC)	1(4576)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4550)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PolIA(AATAAA)	0
PolIA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PolIT(TTTTTT)	0
PolIA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 67D

Luego de la optimización

Repetición directa Máx: Tamaño:15 Distancia:3543 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de díadas Máx: Ninguna

FIG. 67E

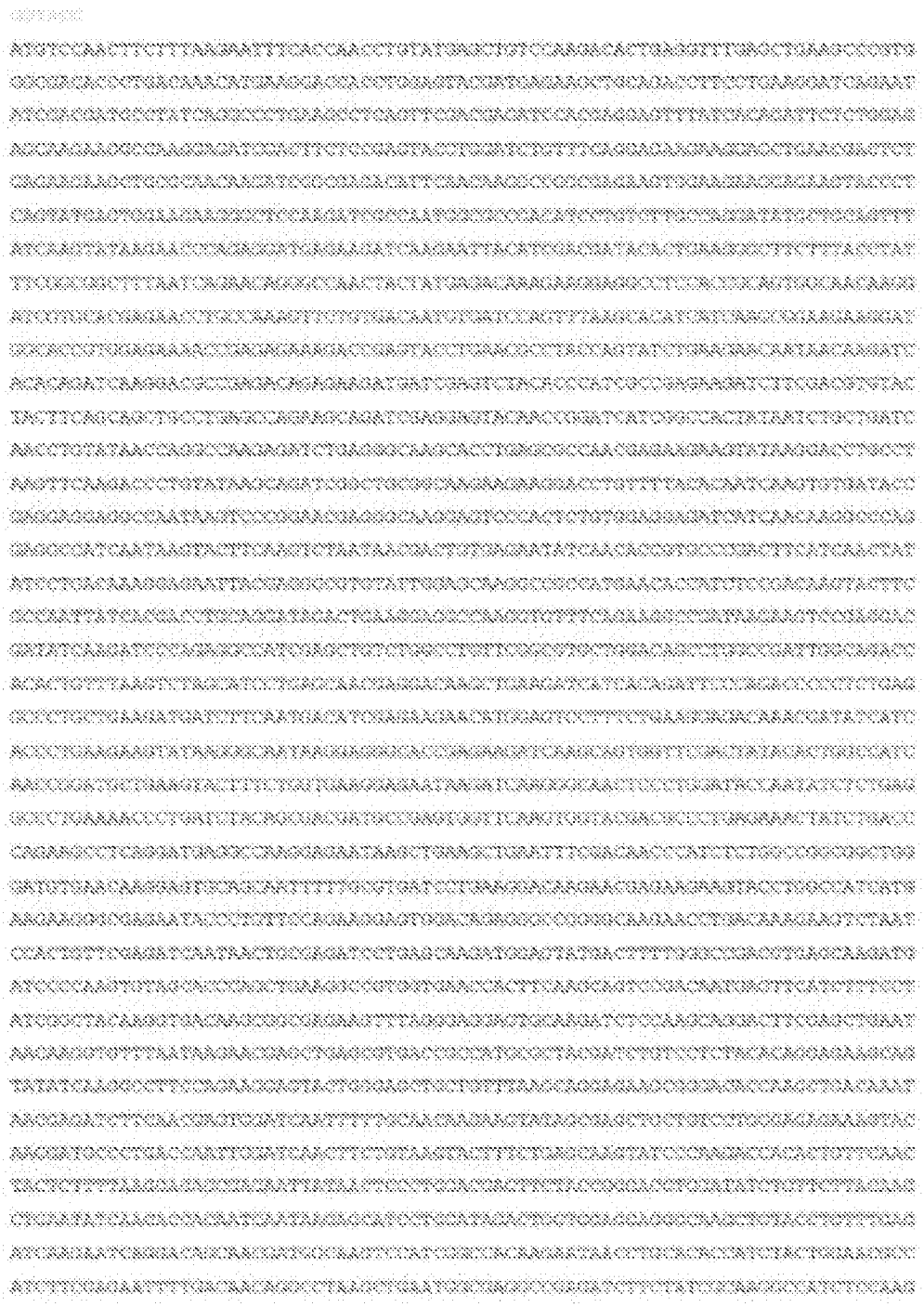


FIG. 67F



FIG.67G

MSNFFENFTNLYELSKTLRFELRPVGDTLTNMEDHLEYDEKLQTFLKDQNIDDAYQALKP
QFDEIHEEFITDSLESKKAKEIDFSEYLDLFQEFKELMDSEKKLRNKIGETFNKAGEKWK
KEKYPQYEWKKGSKIANGADILSCQIMLQFIKYENFEDEKIKNYIDDTLKGFFTYFGGFN
QNRANYYETKKEASTAVATRIVHENLPKFCDNVIQFKHIIKRKKDGTVEKTERKTEYLNA
YQYLKNNNKITQIKDAETKMIESTPIAEKIFDVYYFSSCLSQKQIEEYNRIIGHYNLLI
NLYNQAKRSEGKHLSANEKYKDLPKFKTLYKQIGCGKKDLFYTIKCDTEEEANKSENE
GKESHSVEEIINKAQEAINKYFKSNNDCENINTVPDFINYILTKENYEGVYWSKAAMNTI
SDKYPANYHDLQDRLKEAKVFQKADKKSEDDIKIPEAIELSGLFGVLDSLADWQTTLFKS
SILSNEDKLKIITDSQTPSEALLKMIFNDIEKNMESFLKETNDIITLKKYKGNKEGTEKI
KQWFDYTLAINEMLKYFLVKENKIKGNSLDTNISEALKTLIYSDDAEWFKWYDALRNYLT
QKPQDEAKENKLKLNFDNPSLAGGWDVNKECSNFCVILKDKNEEKYLAIMKKGENTLFQK
ENTEGRGENLTKKSNPLFEINNCEILSKMEYDFWADVSEMIPECSTQLKAVVNHFKQSDN
EFIFPIGYKVTSGEKFREECKISKQDFELINKVFNKNELSVTAMRYDLSSTQEKQYIKAF
QKETWELLFKQEKRDTKLTNNEIFNEWINPCNKKYSELLSWERKYKDALTNWINFCKYFL
SKYPKTTLFNYSFKESENYNSLDEFYRDVDICSYKLNINTTINKSILDRLVEEGKLYLFE
IKNQDSNDGKSIGHKNNLATIYNNAIFENFDNRPKLNGEAEIFYRKAISKDKLGIVKGKK
TKNGTEIIKNYEFSNEKFILHVPITLNFCSNNEYVNDIVNTKFYNFSNLHFLGIDRGEKH
LAYYSLVNKNGEIVDQGTLNLPFTDKDGNQRSIKKEKYFYNKQEDKWEAKEVDCWNYNDL
LDAMASNRDMARKNWQRIGTIKEAKNGYVSLVIRKIADLAVKNERPAFIVLEDLNTGFKR
SRQKIDKQVYQKFELALAKKLNFLVDKNAKRDEIGSPTKALQLTPPVNNYGDIENKKQAG
IMLYTRANYTSQTDPATGWRETIYLKAGPEETTYKKDGEIKNESVKDQIIETFTDIGFDG
KDYYPEYDKGEFVDEKTGEIKPKKWRLYSGENGKSLDRFRGEREKDKYEWKIDKIDIVKI
LDDLPVNPDKNISLLKQLKEGVELTRNNEHGTGESLRPAIMLIQQIRNTGNNERDNDFIL
SPVRDENGKHFDSREYWDEETKGEKISMPSSGDANGAFNIARKGIIMNAHILANSDSKDL
SLPVSDEEWDLHLANKTEWKKQLNIFSSRKAMAKRKK

FIG. 67H

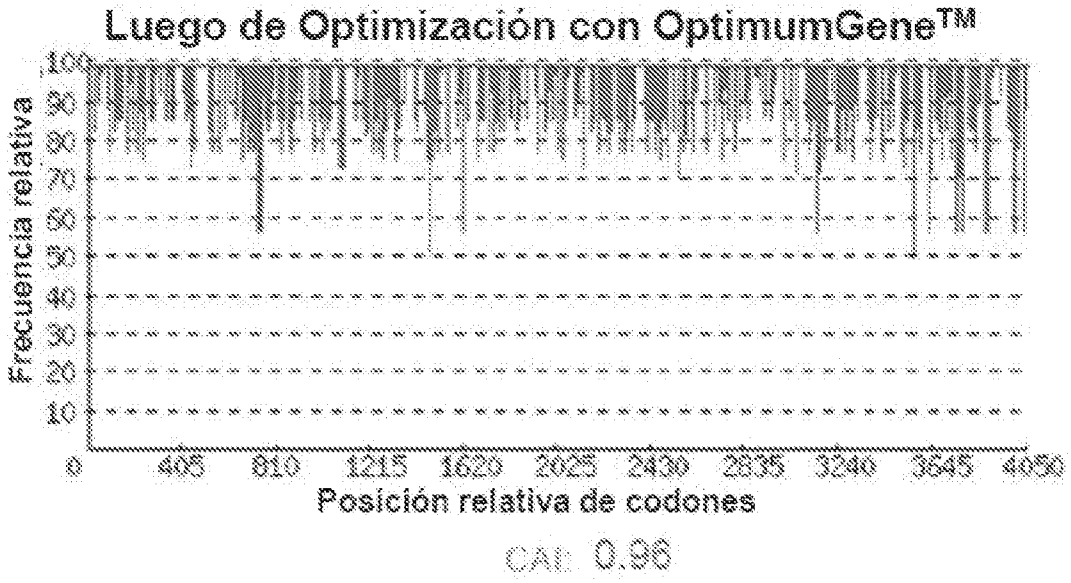


FIG. 68A

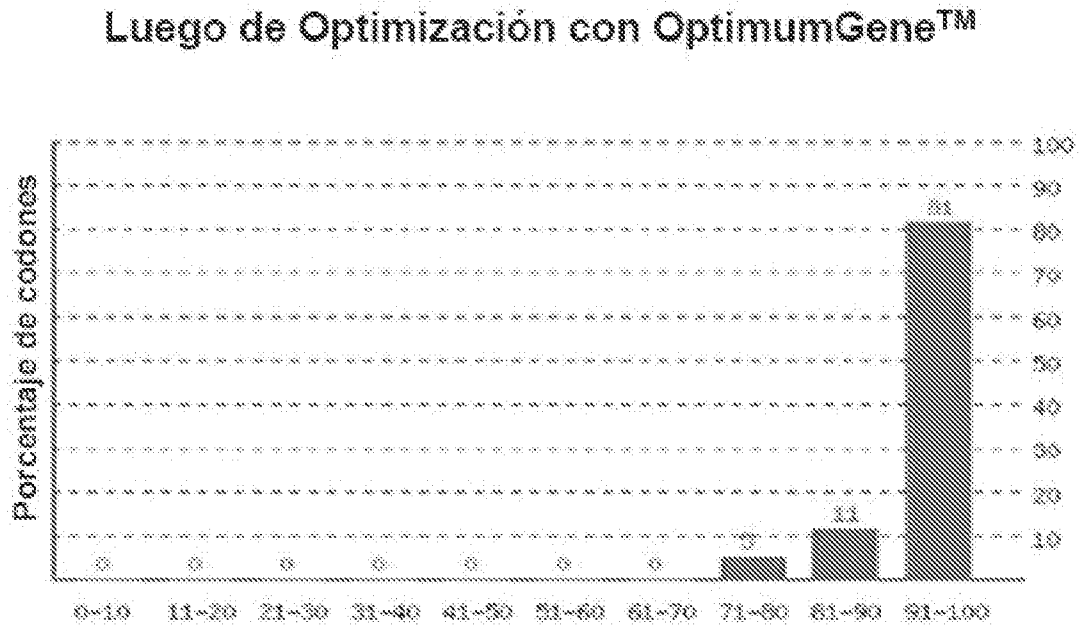


FIG. 68B

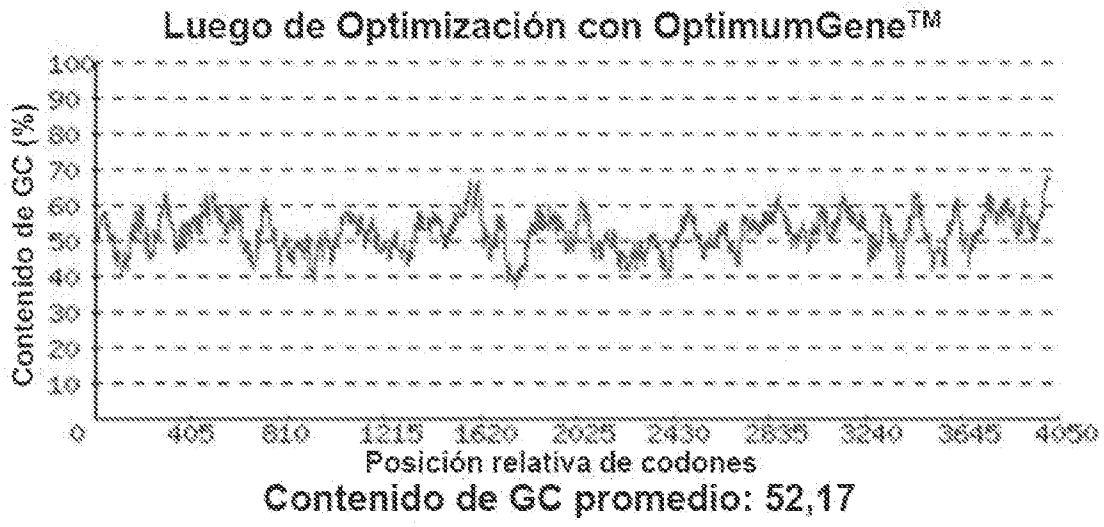

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
405
810
1215
1620
2025
2430
2835
3240
3645
4050
Posición relativa de codones
Contenido de GC promedio: 52,17

FIG. 68C

Enzimas de restricción	Optimizado
* verde: sitios filtrados, azul: sitios confirmados (no filtrados), Rojo: sitios ocultados	
BamHI(GGATCC)	1(4111)
EcoRI(GAATTC)	1(4201)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4176)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
SgrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PolIA(AATAAA)	0
PolIA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PolIT(TTTTTT)	0
PolIA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 68D

Luego de la optimización

Repetición directa Máx: Tamaño:14 Distancia:2001 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de diadas Máx: Ninguna

FIG. 68E

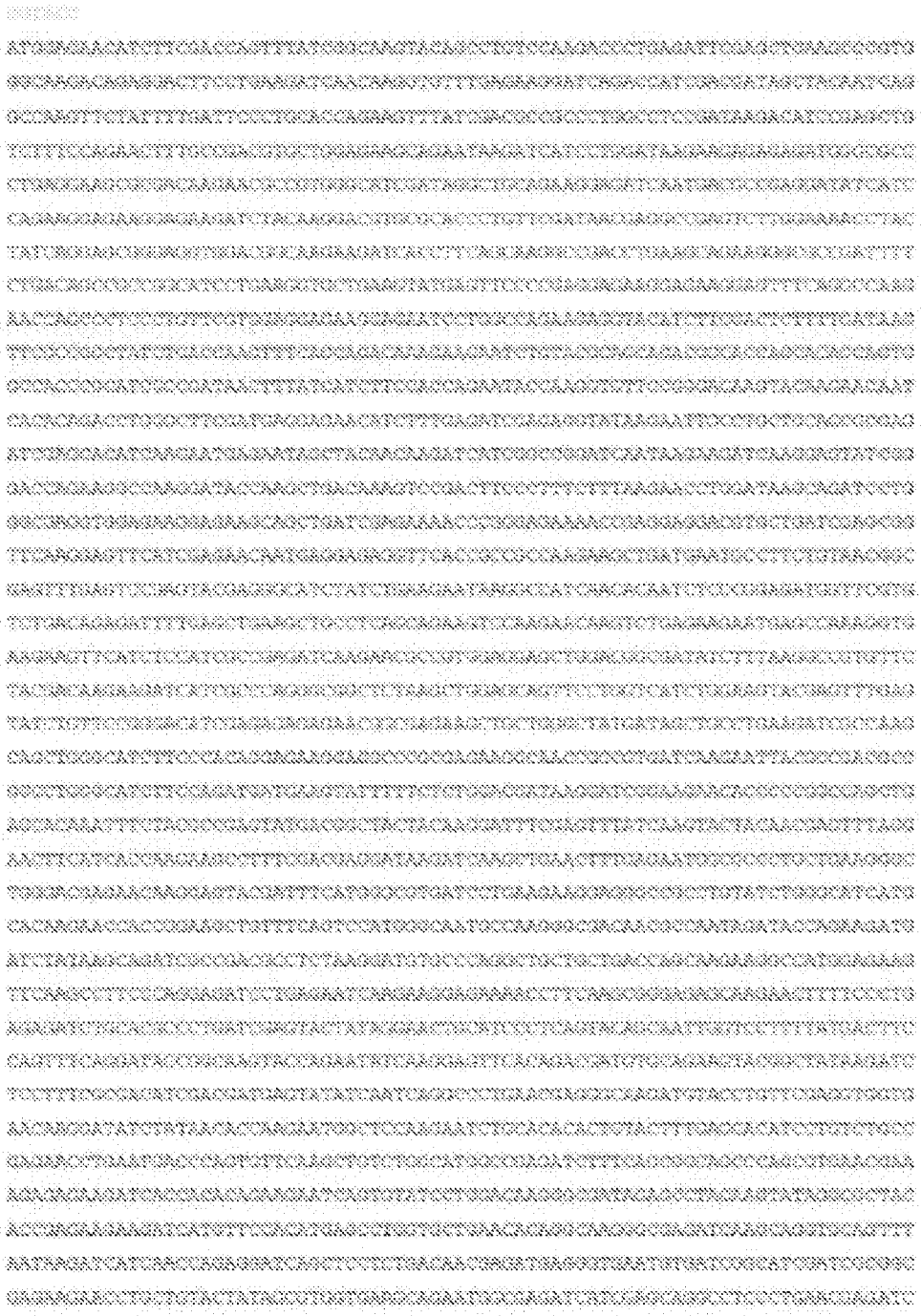


FIG. 68F

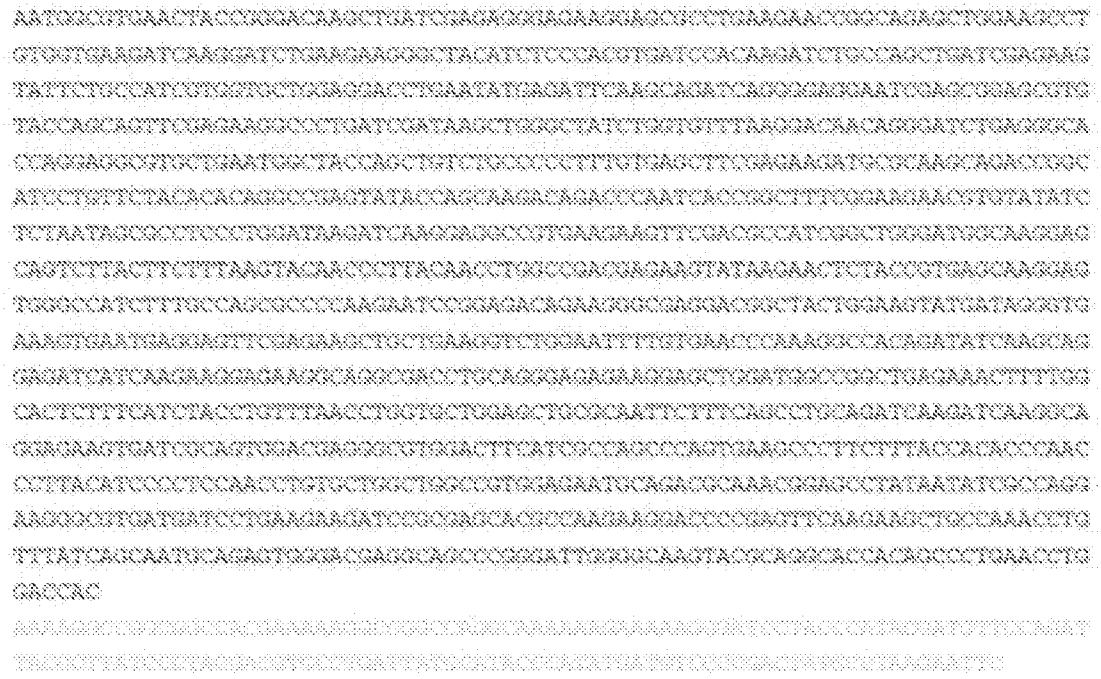


FIG. 68G

MENIPDQPIGKYSLSKTLRFELKPVGKTEDFLKINKVFEKDQTIDDGYNQAKFYFDSLHQ
KFIDAALASDKTSELSFQNPADVLEEQNKIILDKKREMGALEKRDENAVGIDRLQKEIND
AEDIIQKEKEKIYEDVRTLFDNEAESWKTYYQERRVDGKKITPSKADLKQKGADFLTAAG
ILEVLKYEFPEEKEKEFQAKNQPSLFVEEKENPQQKEYIFDSFDKFAGYLTKFQQTKKNL
YAADGTSTAVATRIADNFIIFNQNTKVFRDKYKNNHTDLGFDEENIPETERYKNCLLQRE
IEHIKNENSYNKIIGRINKKIKEYRDQKAKDTKLTKSDFPFFKNLDKQILGEVEKEKQLI
EKTREKTEEDVLIERFEEFIENNEERFTAAKKLMNAFCNGEPESEYEGIYLKNKAINTIS
RRWFVSDRDFELKLPQQKSKNKGEKNEPKVKKFISIAEIKNAVEELDGDIFKAVFYDKKI
IAQGGSKLEQFLVINKYEFEYLFRDIERENGEKLLGYDSCLKIAKQLGIFPQEKEARERKA
TAVIKNYADAGLGIPQMNKYPSLGDKDRENTPGQLSTNFYAEYDGYYKDFEPIKYVNEFR
NFITKKPFDEDKIKLNFENGALLKGWDENKEYDFMGVILKKEGRLYLGIMHKNHRKLFQS
MGNAKGDNANRYQKMIYEQIADASKDVPRLLLTSKKAMEKFKPSQEILRIKKEKTFERES
KNFSLRDLHALIEYYRNCIPQYSNWSFYDFQFQDTGKYQNIKEFTDDVQKYGYKISFRDI
DDEYINQALNEGKMYLFEVVNKDIYNTENGSKNLHTLYFEHILSAENLNDPVFKLSGMAE
IFQRQPSVNEREKITTQENQCILDKGDRAYKYRRYTEKKIMFHMSLVLNTGKGEIKQVQF
NKIINQRISSSDNEMRVNVIGIDRGEKNLLYYSVVKQNGEIIEQASLNEINGVNYRDKLI
EREKERLKNRQSWKPVVKIKDLKKGYISHVIHKICQLIEKYSAIVVLEDLNMRFKQIRGG
IERSVYQQFEKALIDKLGYLVFKDNEDLRAPGGVLNGYQLSAPFVSPEDMRKQTGILFYT
QAEYTSKTDPITGFRKNVYISNSASLDKIKEAVKEFDAIGNDGKEQSYFFKYNPYNLADE
KYKNSTVSKEWAIFASAPRIERQRGEDGYWKYDRVKVNEEFEKLLKVWNFVNPKATDIKQ
EIIKKEKAGDLQGEKELDGRLRNFWNSFIYLFNLVLELRNSFSLQIKIKAGEVIAVDEGV
DFIASPVKPFFFTTPNPYIPSNLCWLAVENADANGAYNIARKGVMILKKIREHAKEDPEFK
KLPNLFISNAEWDEAAEDWGKYAGTTALNLDN

FIG. 68H

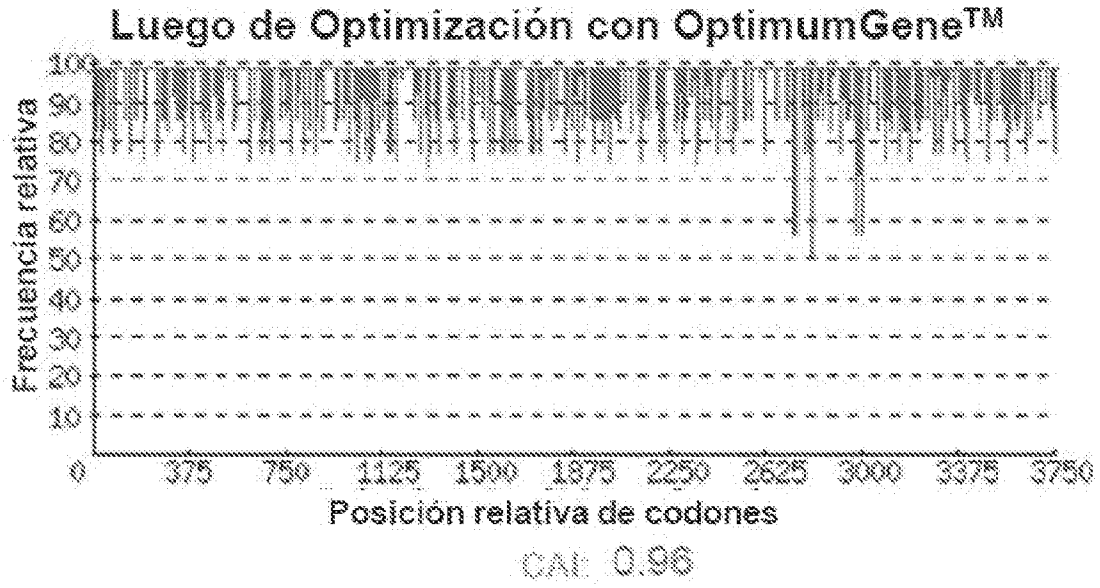


FIG. 69A

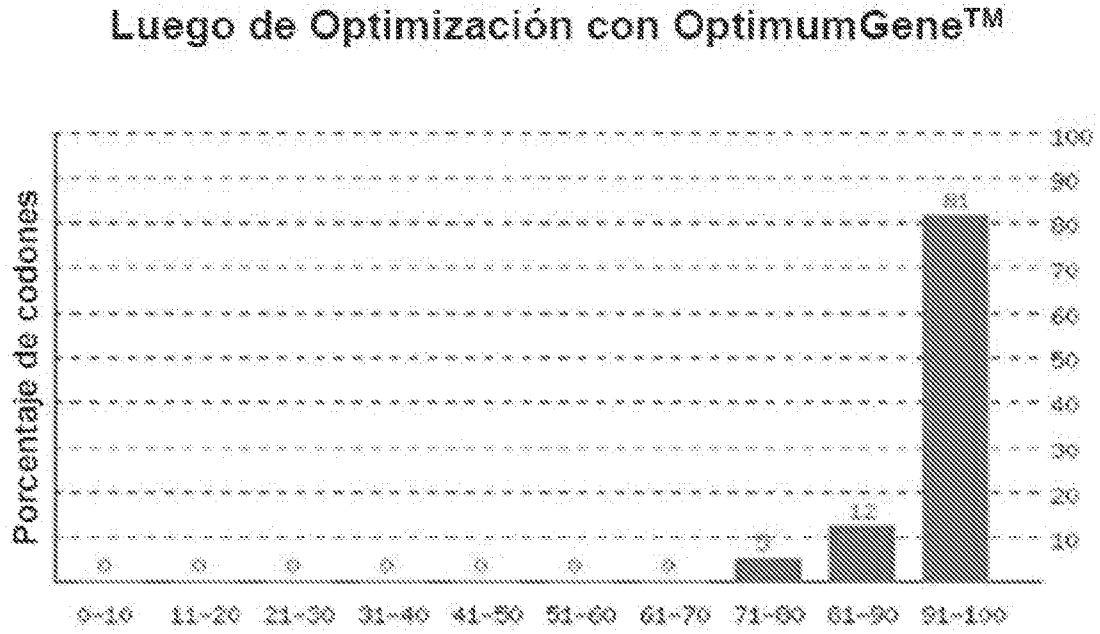


FIG. 69B

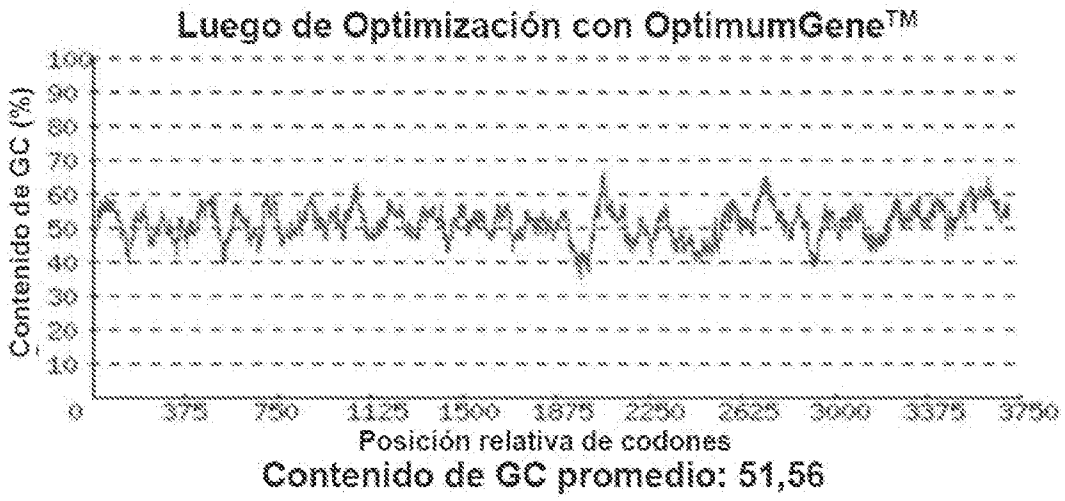

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
375
750
1125
1500
1875
2250
2625
3000
3375
3750
Posición relativa de codones
Contenido de GC promedio: 51,56

FIG. 69C

Enzimas de restricción	Optimizado
* Verde: sitios filtrados, azul: sitios confirmados (no filtrados); Rojo: sitios mantenidos	
BamHI(GGATCC)	1(3805)
EcoRI(GAATTC)	1(3895)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3869)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PoliA(AATAAA)	0
PoliA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PoliT(TTTTTT)	0
PoliA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 69D

Luego de la optimización

Repetición directa Máx: Tamaño:15 Distancia:93 Frecuencia:2
Repetición invertida Máx: Tamaño:12 Tm:39,3 Posiciones de inicio: 842,2867
Repetición de diadas Máx: Ninguna

FIG. 69E

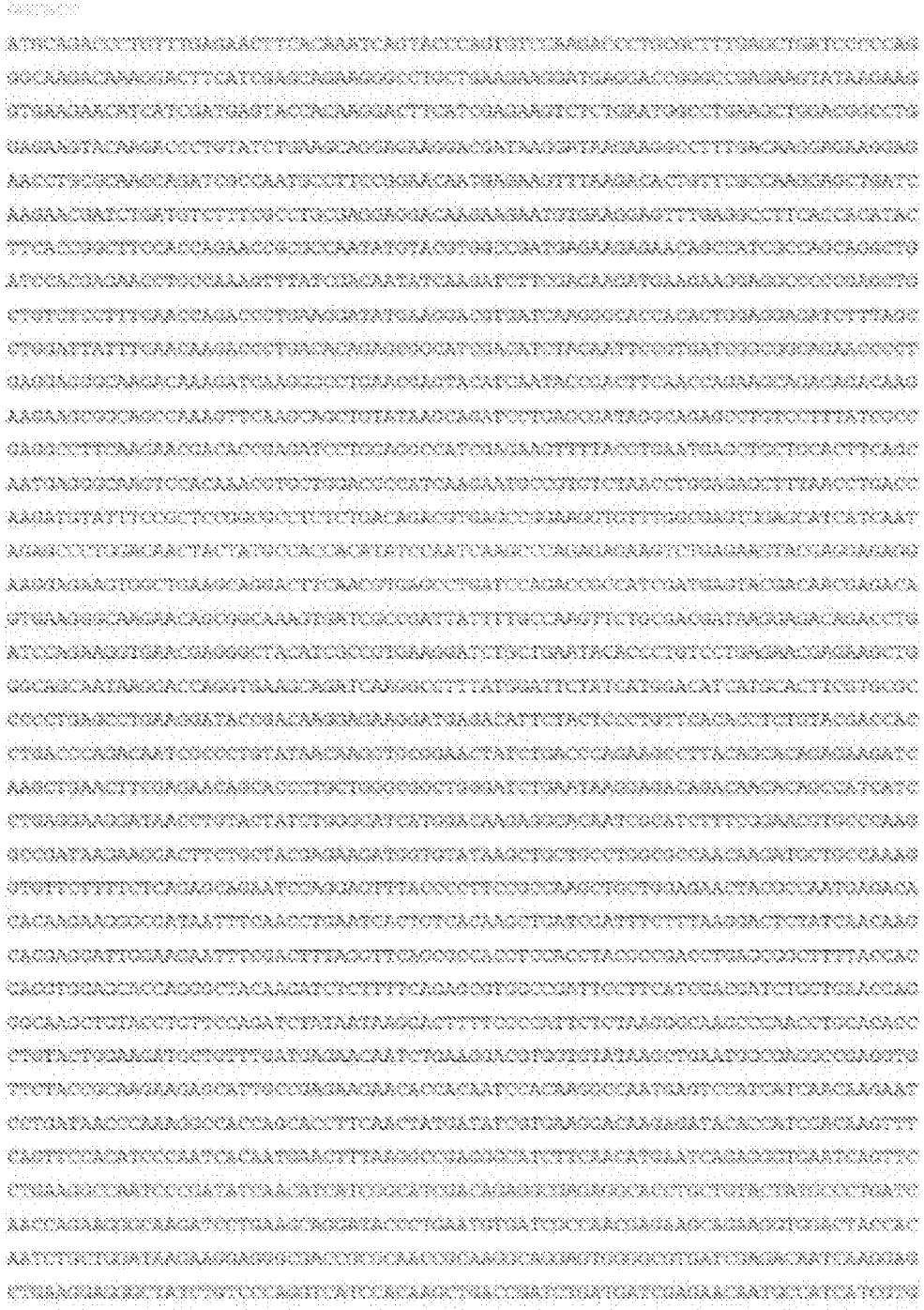


FIG. 69F



FIG. 69G

MQTLFENPTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKDEDRAEKYKKVKNIIDEYHK
DFIEKSLNGLKLDGLEKYKTLYLKQEKDDKDKEAFDKEKENLRKQIANAFRNNEKFKTLF
AKELIKNDLMSFACEEDKKNVERFRAFTTYFTGFHQNRANMYVADSKFTAIAGRLIKENL
PKFIDNIKIFEKMKKEAFELLSFFNQTLKDNKDVIKGTTLEEIFSLDYFNKTLTQSGIDI
YNSVIGGRTPEDGKTKIKGLNEYINTDFNQEQTDKKKRQFKFKQLYKQILSDRQSLSFIA
EAFKNDTEILEAIEKFYVNELLHFSNEGKSTNVLDAIKNAVSNLESFNLTKMYFRSGASL
TDVSRKVFGEWSIINRALDNYYATTYPIEFREKSEKYEERKEKWLKQDFNVSLIQTAIDE
YDNETVEGKNSGKVIADYFAKFCDDKETDLIQKVNEGYIAVKDLLNTPCFENEKLGSNKD
QVKQIKAFMDSIMDINHFVRPLSLKDTDKERDETFYSLFTPLYDRLTQTIALYNKVRNYL
TQKPYSTEKIKLNFENSTLLGGWDLNKETDNTAIILEKDNLYYLGINDKRHNRIFRNVPK
ADKKDFCYEKMVYKLLPGANKMLPKVFFSQSRIQEFTPSAKLLENYANETHKKGDNFNLN
HCHKLIDFFKDSINKHEDWKNFDFRFSATSTYADLSGFYHEVEHQGYKISFQSVADSFID
DLVNEGKLYLFQIYNKDFSPFSKGKPNLHTLYWKMLFDENNLKDVVYKLNGEAEVFYRKK
SIAENITTIHKANESIINKNPDNFKATSTFNYDIVKDKRYTIDKFQFHIPITMNFKAEGI
FNMNQRVNQFLKANPDINIIGIDRGERHLLYYALINQKGKILKQDTLNVIANEKQKVDYH
NLLDKKEGDRATARQEWGVIETIKELKEGYLSQVIHKLTDLMIENKAIIVMEDLNFGFKR
GRQKVEKQVYQKFEKMLIDKLNYLVDKNKKANELGGLLNAFQLANKFESFQKMGKQNGFI
FYVPAWNTSKTDPATGFIDFLKPRYENLNQAKDFFEKFDSIRLNSKADYFEFAFDFKNFT
EKADGGRTKWTVCTTNEDRYAWNRALNNNRGSQEKYDITAELKSLFDGKVDYKSGKDLKQ
QIASQESADFFKALMKNLSITLSLRHNNGEKCDNEQDYILSPVADSKGRFFDSEKADDDM
PKNADANGAYHIALKGLWCLEQISETDDLKKVKLAISNKEWLEFVQTLKG

FIG. 69H

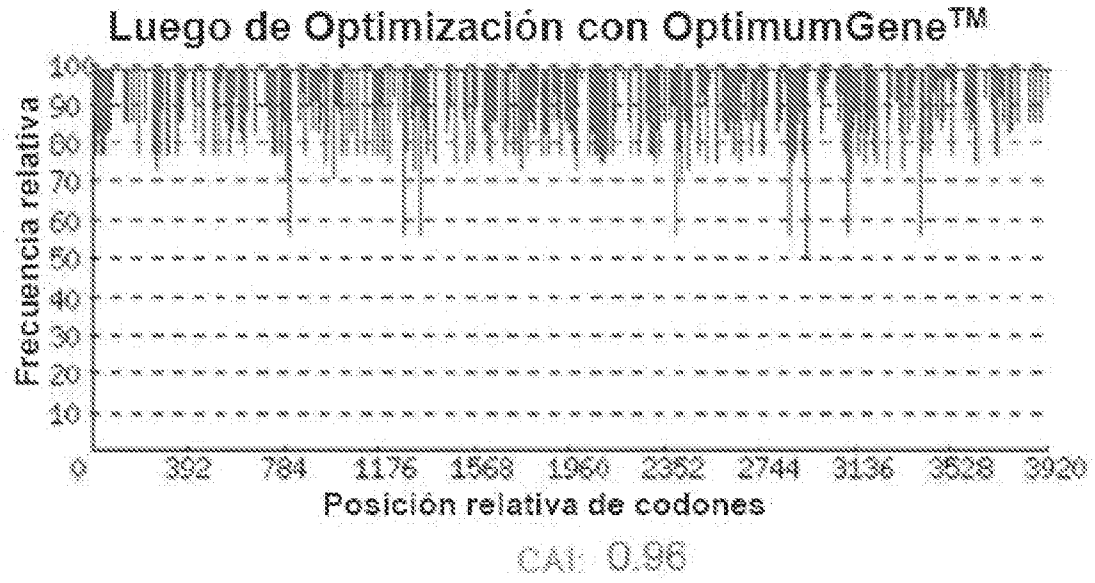


FIG. 70A

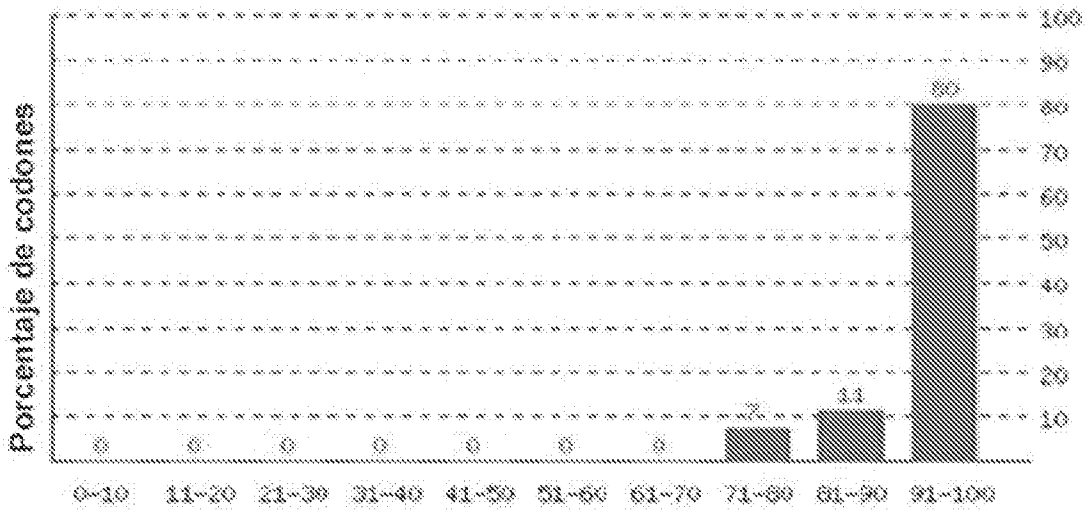


FIG. 70B

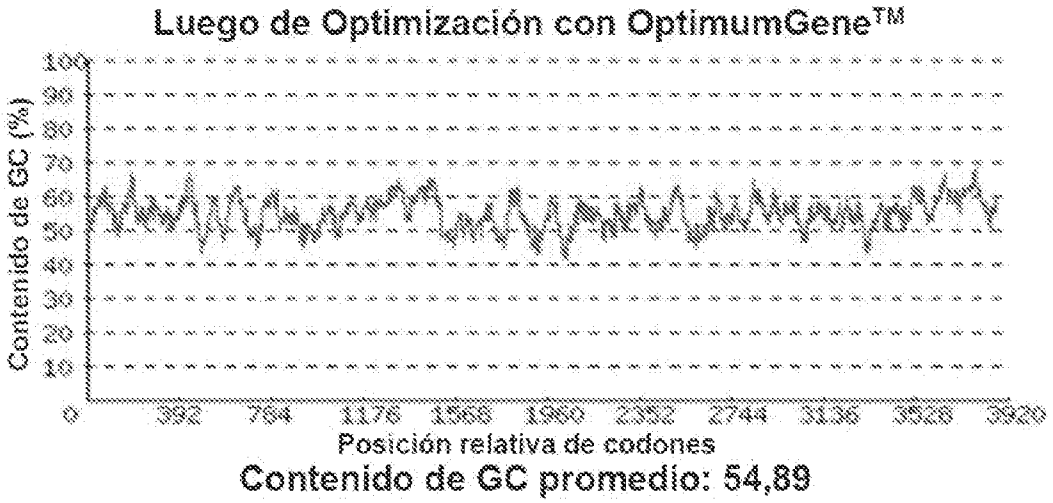

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
392
784
1176
1568
1960
2352
2744
3136
3528
3920
Posición relativa de codones
Contenido de GC promedio: 54,89

FIG. 70C

Enzimas de restricción	Optimizado
* verde: sitios filtrados, azul: sitios confirmados (no filtrados), Rojo: sitios mantenidos	
BamHI(GGATCC)	1(3975)
EcoRI(GAATTC)	1(4068)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4040)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PoliA(AATAAA)	0
PoliA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PoliT(TTTTTT)	0
PoliA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 70D

Luego de la optimización

Repetición directa Máx: Tamaño:15 Distancia:36 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de diadas Máx: Ninguna

FIG. 70E

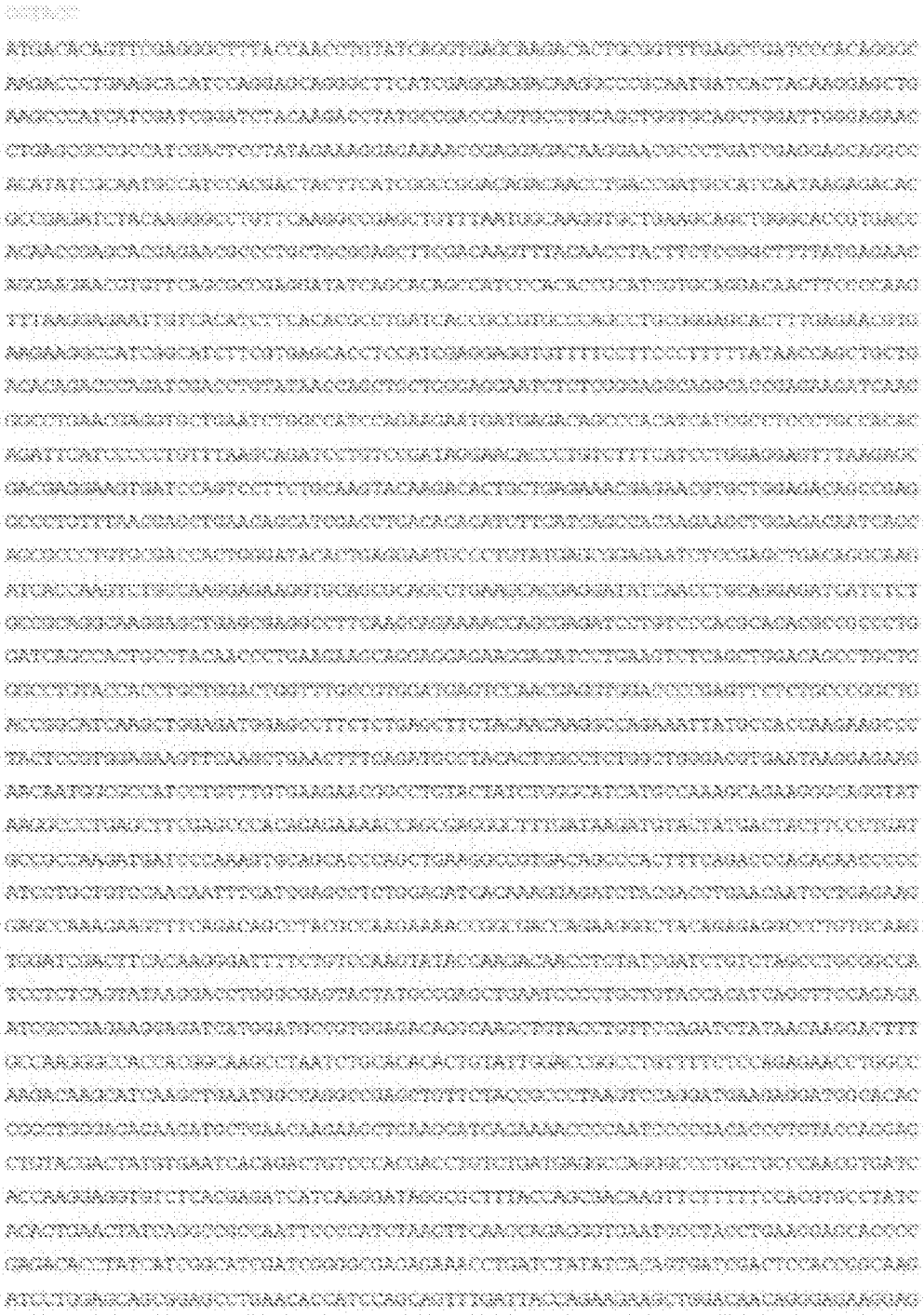


FIG. 70F

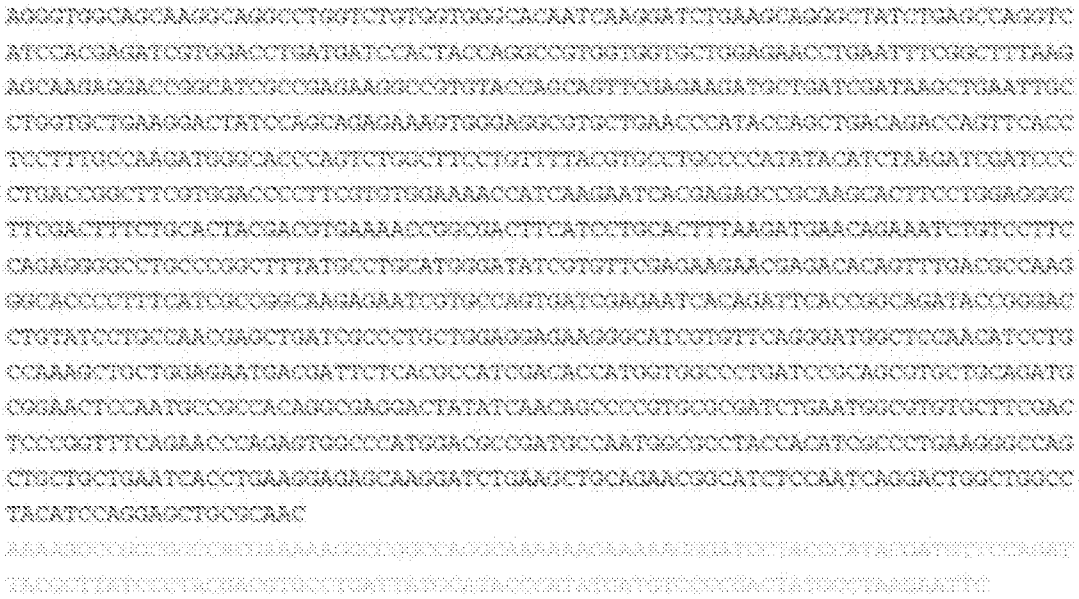


FIG. 70G

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKT
YADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDA
INKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVF
SAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEV
FSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPH
RFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSID
LTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINL
QEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEEILKSQLDSLLGLYHL
LDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTL
ASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPD
AAEMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYA
KKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYH
ISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIK
LNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSD
EARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHP
ETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSV
VGTIKDLKQGYLSQVINEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLI
DKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFV
DPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVF
EKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNIL
PKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM
DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN

FIG. 70H

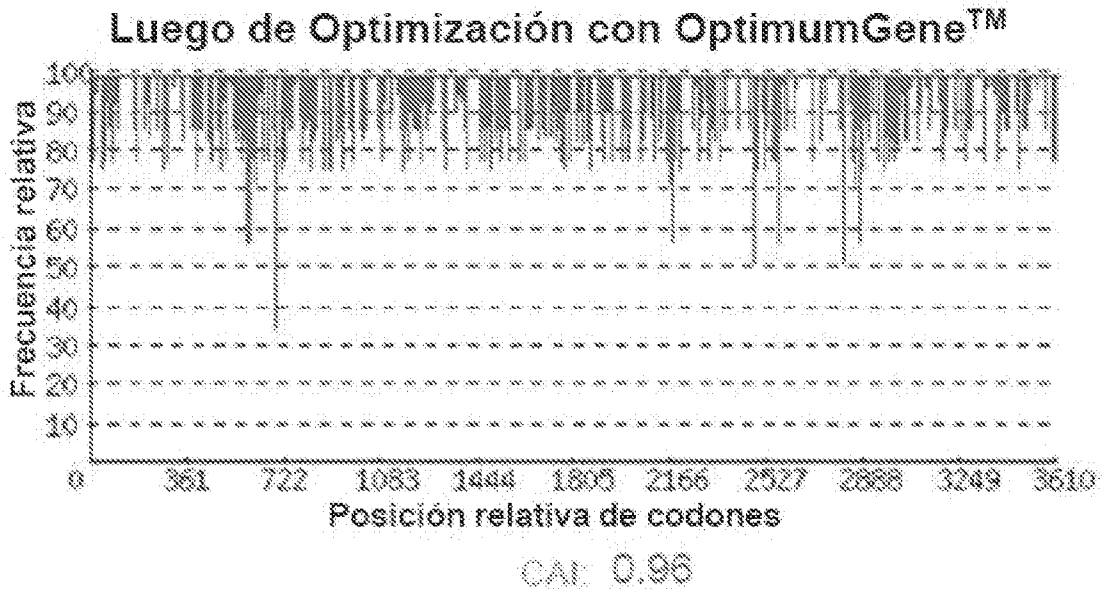


FIG. 71A

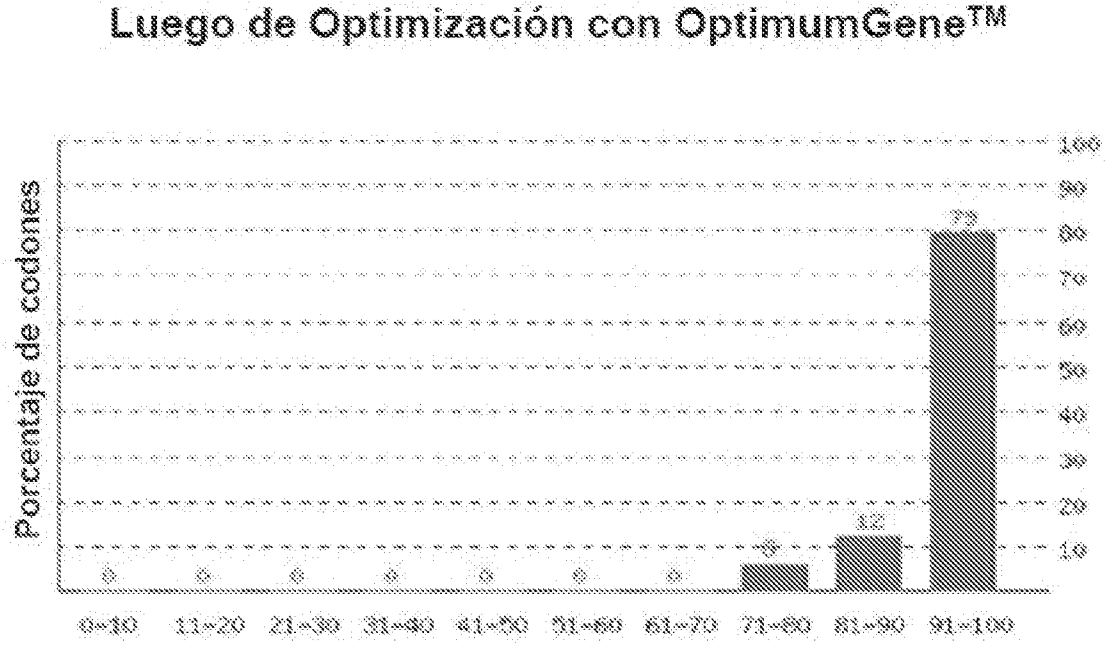


FIG. 71B

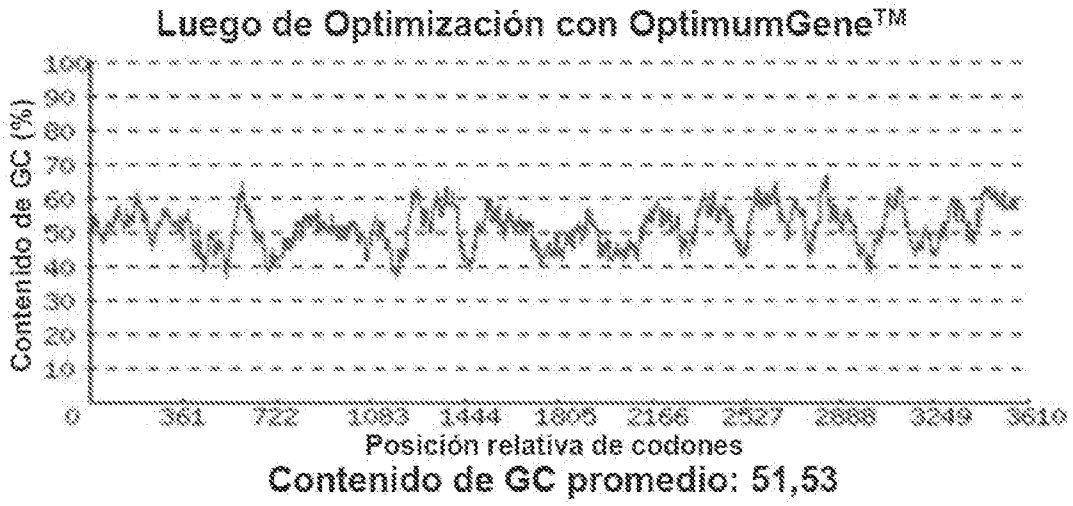

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
361
722
1083
1444
1805
2166
2527
2888
3249
3610
Posición relativa de codones
Contenido de GC promedio: 51,53

FIG. 71C

Enzimas de restricción	Optimizado
* verde: sitios filtrados, azul: sitios confirmados (no filtrados), Rojo: sitios inestables	
BamHI(GGATCC)	1(3673)
EcoRI(GAATTC)	1(3783)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(CGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3737)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PoliA(AATAAA)	0
PoliA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PoliT(TTTTTT)	0
PoliA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 71D

Luego de la optimización

Repetición directa Máx: Tamaño:17 Distancia:1800 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de díadas Máx: Ninguna

FIG. 71E

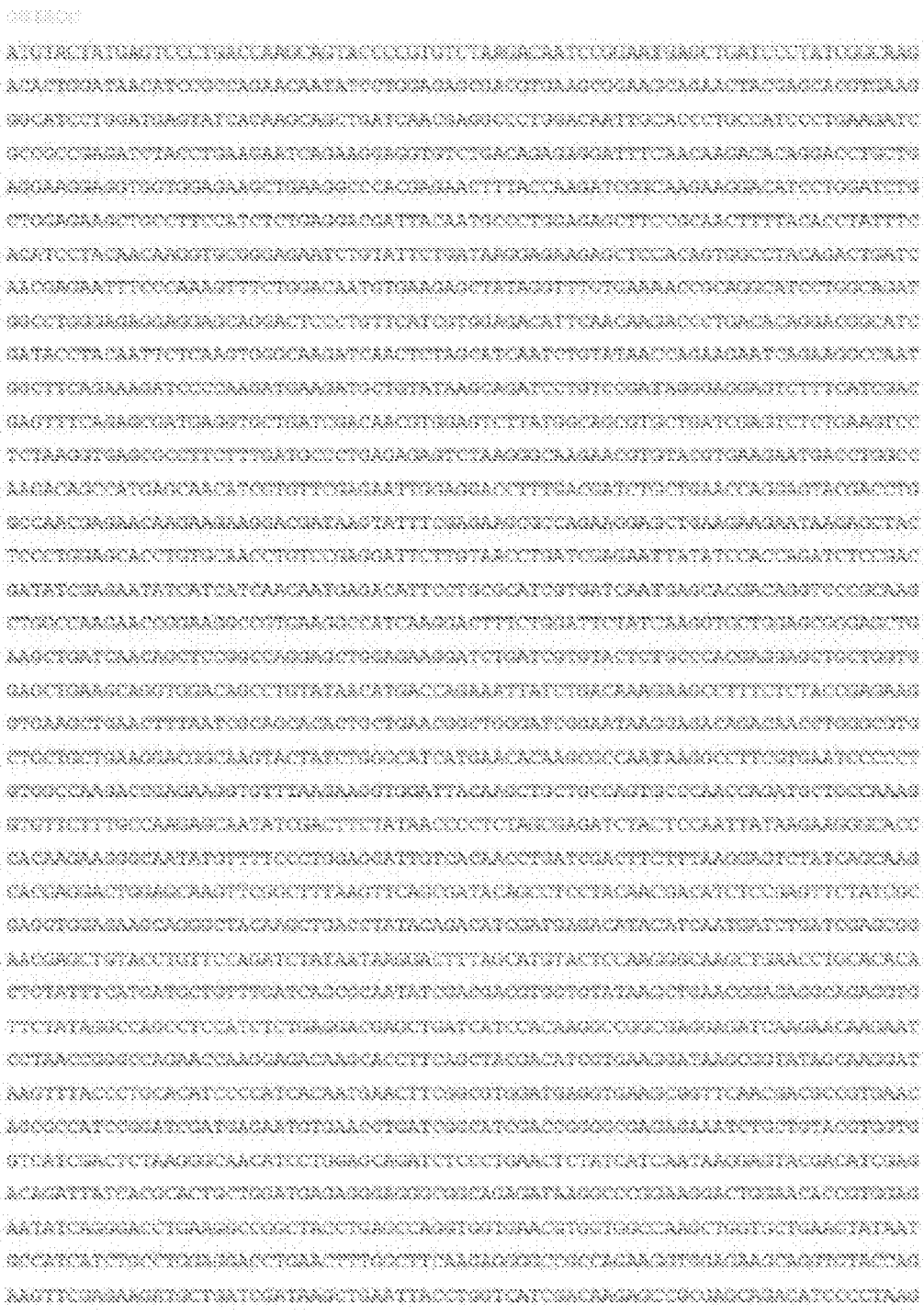


FIG. 71F



FIG. 71G

MYYESLTKQYPVSKTIRNELIPIGKTLDNIRQNNILESDVKRKQNYEHVKGILDEYHKQL
INEALDNCTLPSLKIAAEIYLKNQKEVSDREDFNKTQDLLRKEVVEKLKAHENFTKIGKK
DILDLLEKLPSISEDDYNALESFRNFYTYFTSYNKVRENLYSDKEKSSTVAYRLINENFP
KFLDNVKSYRFVKTACILADGLGEEEQDSLFIVETFNKTLTQDGIDTYNSQVGKINSSIN
LYNQKNQKANGFRKIPKMKMLYKQILSDREESFIDKFQSDEVLIDNVESYGSVLIESLKS
SKVSAFFDALRESKGENVYVKNDLAKTAMSNIVFENWRTFDDLLNQEYDLANENKKDDK
YFEKRQKELKKNKSYSLEHLCNLSEDSCNLIENYIHQISDDIENIIINNETFLRIVINEN
DRSEKLAFNRKAVKAIKDFLDSIKVLERELKLINSSGQELEKDLIVYSAKEELLVELKQV
DSLYNMTENYLTKKPFSTEKVKLNFNRSTLLNGWDRNKETDNLGVLLLKDGKYYLGIMNT
SANKAFVNPPVAKTEKVFKKVDYKLLPVPNQMLPKVFFAKSNIDFYNPSSEIYSNYKKGT
HKGNMFSLEDCHNLIDFFKESISKHEDWSKFGFKFSDTASYNDISEFYREVEKQGYKLT
YTDIDETYINDLIERNELYLFQIYNKDFSMYSKGKLNLHTLYFMMLFDQRNIDDVVYKLN
GEAEVFYRPASISEDELIIHKAGEEIKNENPNEARTLETSTFSYDIVKDKRYSKDKFTLH
IPITMNFGVDEVKRFNDAVNSAIRIDENVNVIGIDRGERNLLYVVVIDSKGNILEQISLN
SIINKEYDIETDYHALLDERESGRDEARKDWNTVENIRDLKAGYLSQVVNVVAKLVLKYN
AIICLEDLNFGFKRGRQKVEKQVYQKFEKMLIDKLNYLVIDKSREQTSPKELGGALNALQ
LTSKFKSFKELGKQSGVIYYVPAYLTSKIDPTTGFANLFYMKCENVEKSKEFFDGFDFIR
FNALENVFEFGFDYRSFTQRACGINSKWTVCTNGERIIKYRNPDKNNMFDEKVVVVTDEM
KNLFEQYKIPYEDGRNVKDMIISNEEAEFYRRLYRLLQQTLQMRNSTSDGTRDYIISPVK
NKREAYFNSELSDGSVPKDADANGAYNIARKGLMVLEQIRQKSEGEKINLAMTNAEWLEY
AQTHLL

FIG. 71H

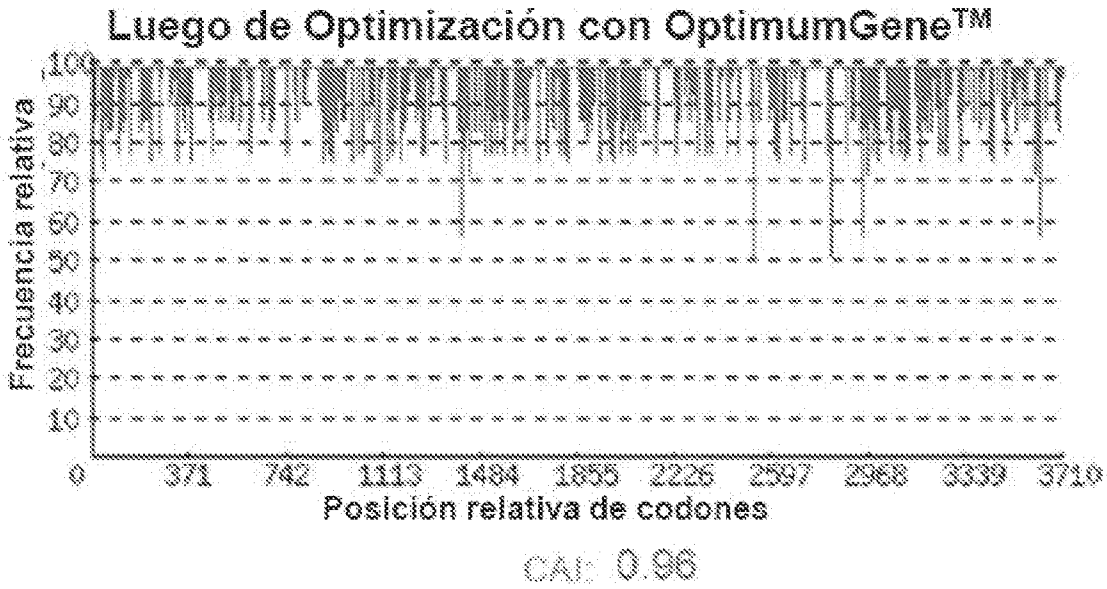


FIG. 72A

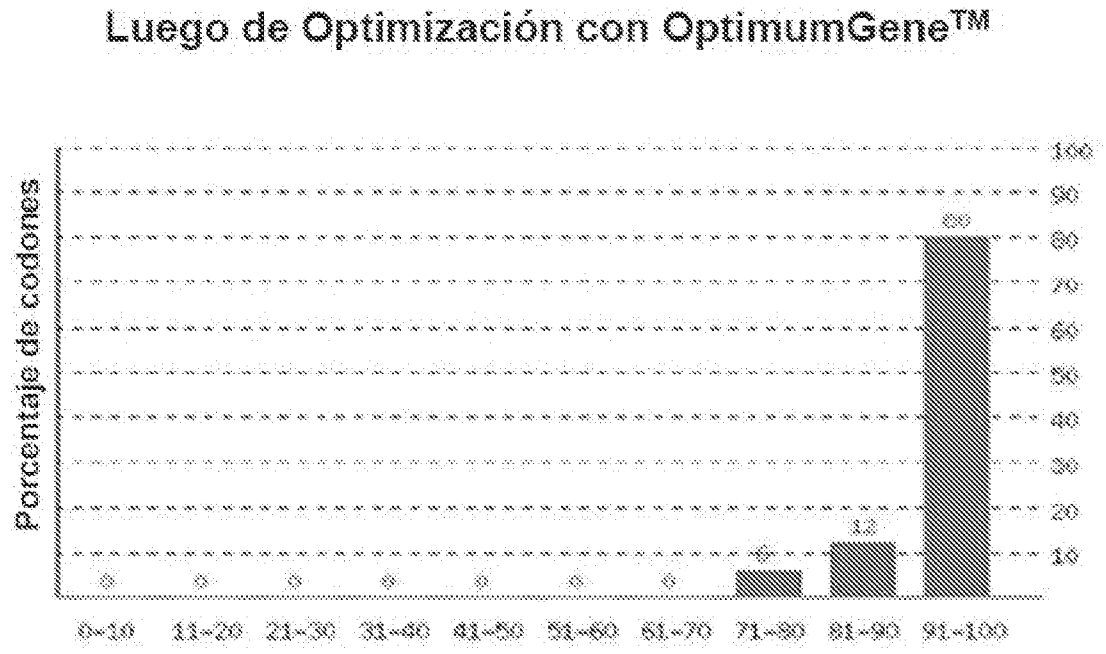


FIG. 72B

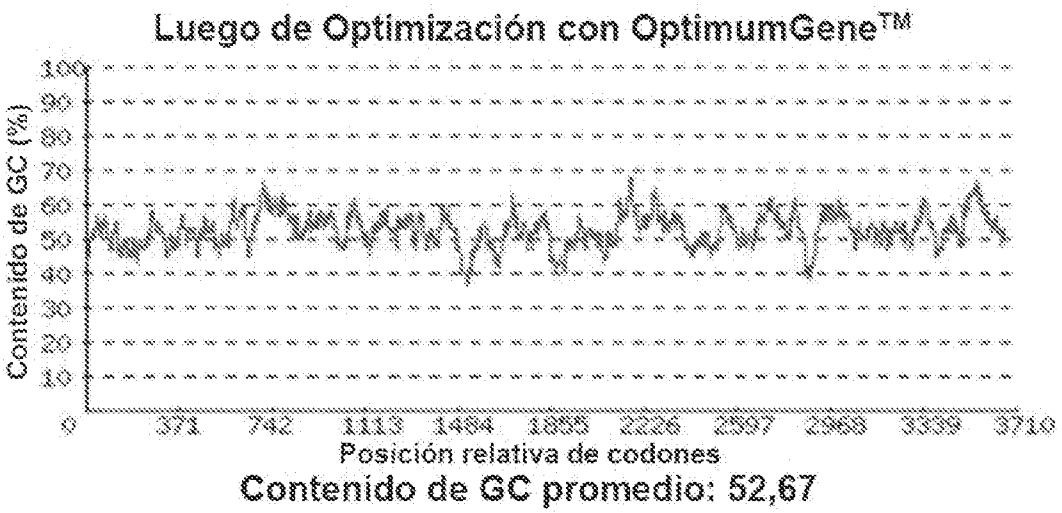

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
371
742
1113
1484
1855
2226
2597
2968
3339
3710
Posición relativa de codones
Contenido de GC promedio: 52,67

FIG. 72C

Enzimas de restricción	Optimizado
* verde: sitios filtrados, azul: sitios confirmados (no filtrados), Rojo: sitios inaceptables.	
BamHI(GGATCC)	1(3789)
EcoRI(GAATTC)	1(3859)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3833)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PoliA(AATAAA)	0
PoliA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PoliT(TTTTTT)	0
PoliA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 72D

Luego de la optimización

Repetición directa Máx: Tamaño:13 Distancia:1506 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de diadas Máx: Ninguna

FIG. 72E

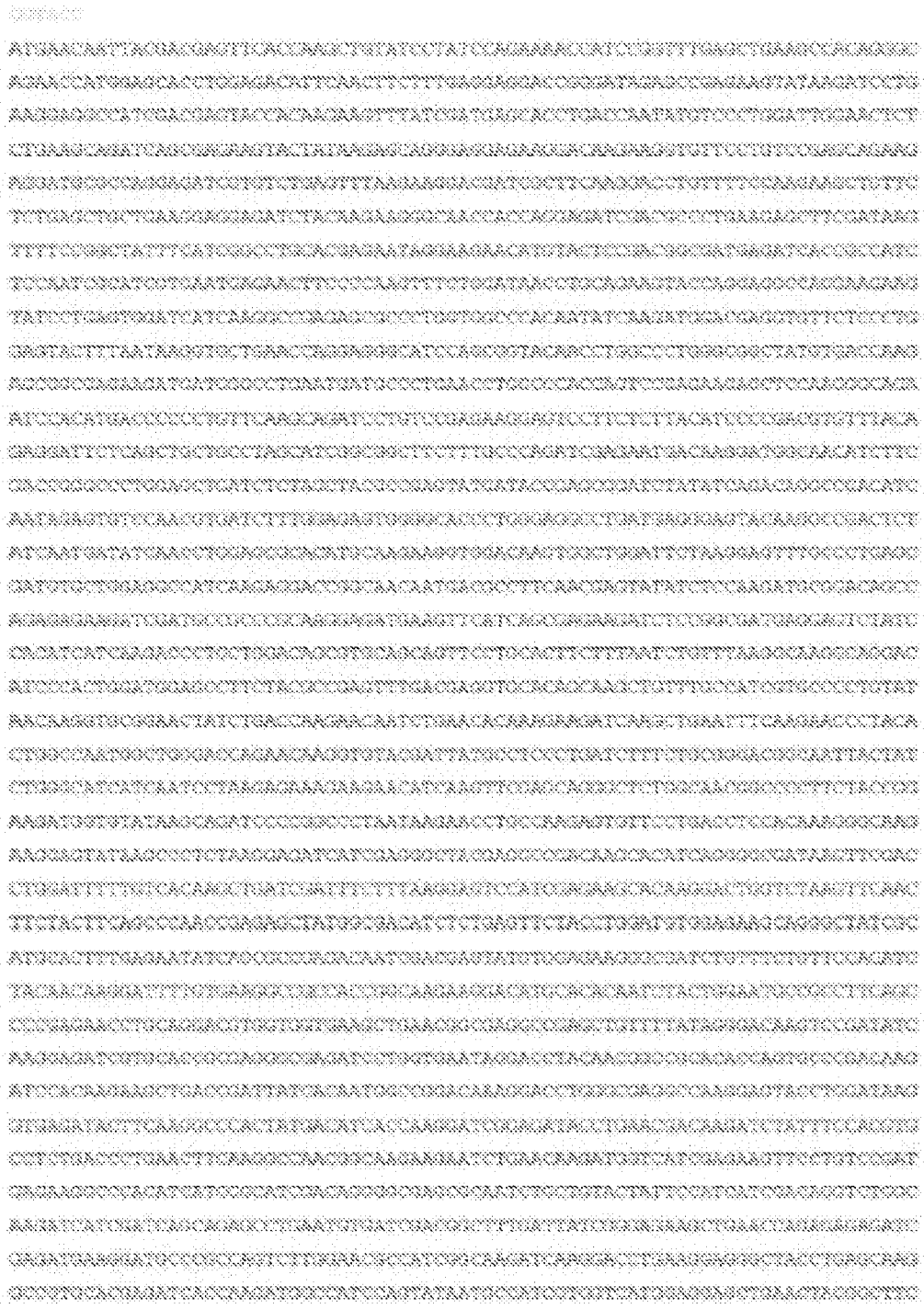


FIG 72F



FIG. 72G

MDNYDEPTKLYPIQKTIRFELKPQGRTMEHLETFNFFEEDRDRAEKYKILKEAIDEYHKK
FIDENLTNMSLDWNSLKQISEKYYKSREEKDKKVFLSEQKRMRQEIVSEFKKDDRFKDLF
SKKLPSELLKEEIYKKGNHQEIDALKSFDKFSGYFIGLHENRKNMYSDGDEITAISNRIV
NENFPKFLDNLQKYQEARKKYPEWIIKAESALVAHNIKMDEVFSLEYFNKVLNQEGIQRY
NLALGGYVTKSGERMMGLNDALNLAHQSEKSSKGRIHMTPLFKQILSEKESPSYIPDVFT
EDSQLLPSIGGFFAQIENDKDGNIFDRALELISSYAEYDTERIYIRQADINRVSNVIFGE
WGTLGGLMREYKADSINDINLERTCKKVDKWLDSKEFALSDVLEAIERTGNNDAFNEYIS
KMRTAREKIDAARKEMKFISEKISGDEESINIIKTLLDSVQQFLHFFNLFKARQDIPLDG
AFYAEFDEVHSKLFAIVPLYNKVRNYLTKNNLNTKKIKLNFKNPTLANGWDQNKEVYDYAS
LIFLRDGNYYLGIINPKRKKNIKFEQGSGNGFFYERMVYKQIPGPNENLFEVFLTSTKGK
KEYKPSKEIIEGYEADKHIRGDKFDLDFCHKLIDFFKESIEKHKDWSKFNFYFSPTESYG
DISEFYLDVENQGYEMHFENISAETIDEYVEKGDLFLFQIYNKDFVKAATGKKDMHTIYW
NAAFSPENLQDVVVKLNGEAELFYRDKSDIKEIVHEEGEILVNRTYNGRTPVPDKIHKKL
TDYHNGRTKDLGEAKEYLDKVRYFKAHYDITKDERYLNDKIYFHVPLTLNFKANGKKNLN
KMVIEKFLSDEKAHIIGIDRGERNLLYYSIIDRSGKIIDQQSLNVIDGFDYRERLNQREI
EMKDARQSWNAIGKIKDLKEGYLSKAVHEITKMAIQYNAIVVMEDLNYGFKRGRFKVEKQ
IYQKFENMLIDKMNYLVFKDAPDESPGGVLNAYQLTNPLESFAKLGKQTGILFYVPAAYT
SKIDPTTGFVNLFNTSSKTNAQERKEFLQKFESISYSAKDGGIFAFAFDYRKFGTSKTDH
KNVWTAYTNGERMRYIKEKERNELFDPSKEIKEALTSSGIKYDGGQNILFDILRSNNNGL
IYTMYSSFIAAIQMRVYDGKEDYIISPIKNSKGEFFRTDPKERELPIDADANGAYNIALR
GELTMRAIAEKFDPDSEKMAKLELKHKDWFEFMQTRGD

FIG. 72H

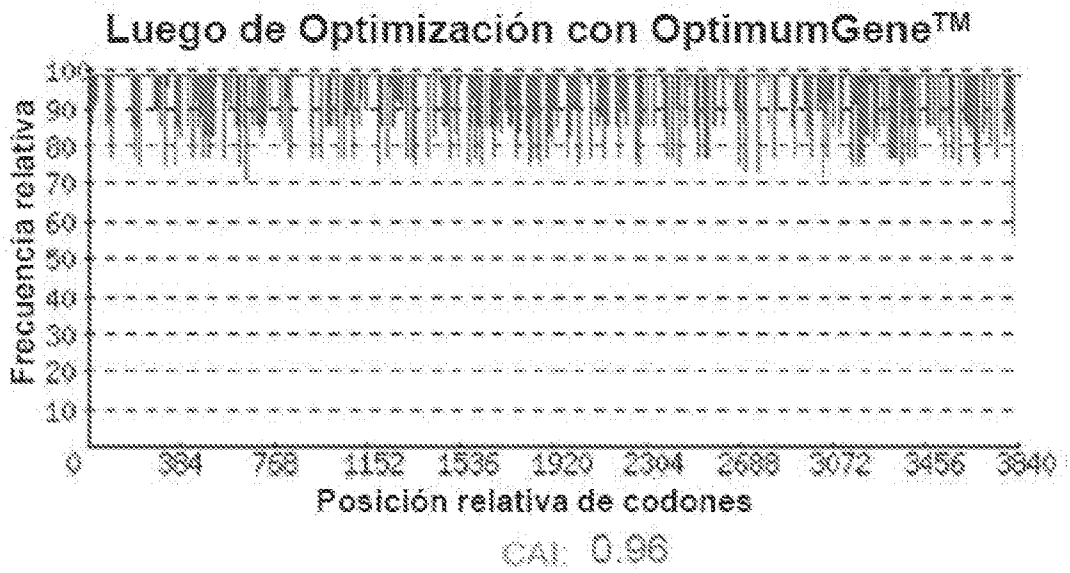


FIG. 73A

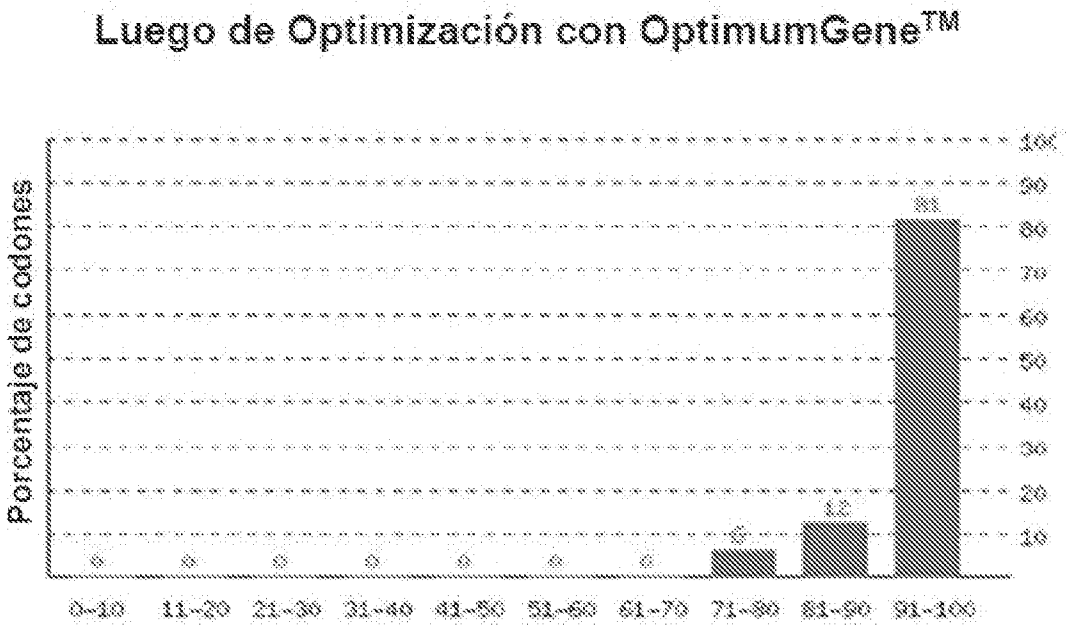


FIG. 73B

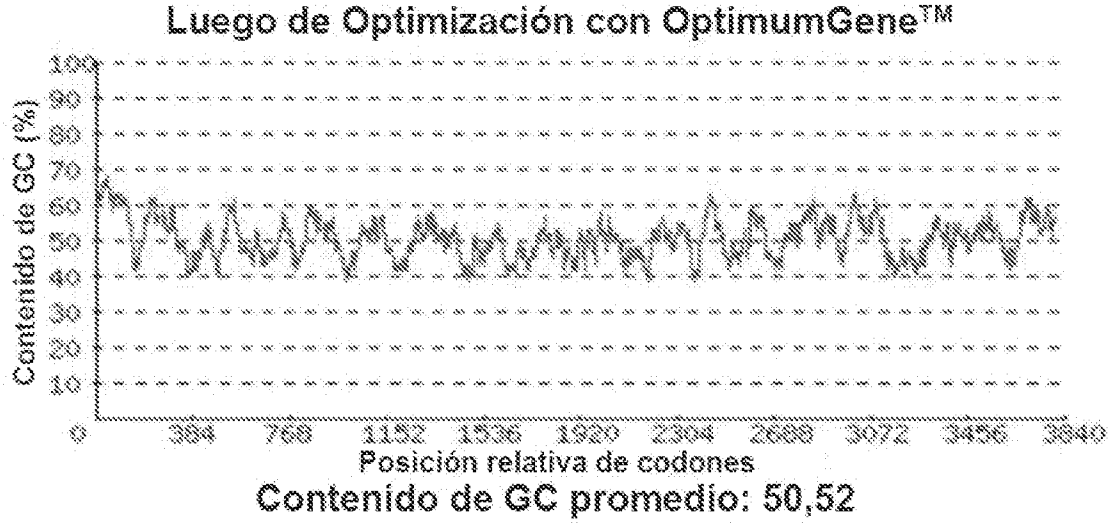

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
384
768
1152
1536
1920
2304
2688
3072
3456
3840
Posición relativa de codones
Contenido de GC promedio: 50,52

FIG. 73C

Enzimas de restricción	Optimizado
BamHI(GGATCC)	1(3901)
EcoRI(GAATTC)	1(3991)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3905)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PolIA(AATAAA)	0
PolIA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PolIT(TTTTTT)	0
PolIA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 73D

Luego de la optimización

Repetición directa Máx: Tamaño:14 Distancia:894 Frecuencia:2

Repetición invertida Máx: Ninguna

Repetición de diadas Máx: Tamaño:13 Tm: 40,1 Posiciones de inicio: 2322, 3831

FIG. 73E

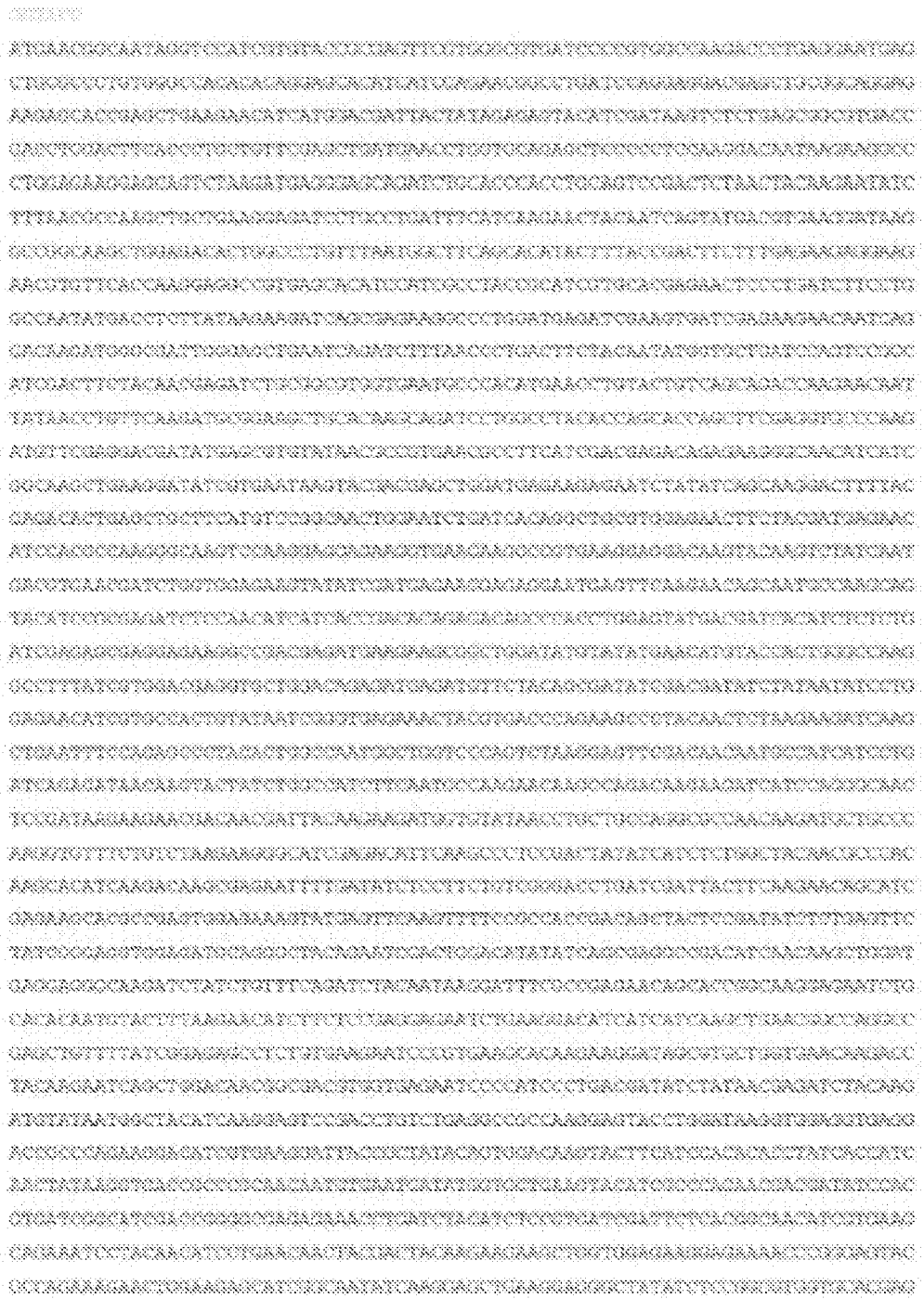


FIG. 73F



FIG. 73G

MRGNRSIVYREFVGVIPVAKTLRNELRPVGHTQEHIIQNGLIQEDELRQEKSTELKNIMD
DYYREYIDKSLSGVTDLDPTLLPELMNLVQSSPSKDNKKALEKEQSKMREQICTHLQSDS
NYKNIFNAKLLKEILPDFIKNYNQYDVKDKAGKLETLALFRGFSTYFTDFFEKRKNVFTK
EAVSTSIAYRIVHENSLIFLANMTSYKFISEKALDEIEVIEKNNQDKMGDWELNQIFNPD
FYRNVLIQSGIDFYNEICGVVNARMSLYCQQTKNNYNLFKMRKLHEQILAYTSTSFEVPK
MFEDDMSVYNAVNAFIDSTEKGNIIGKLEDIVNKYDELDEKRIYISKDFYETLSCFMSGN
WNLITGCVENFYDENINAKGKSKEEKVKKAVKEDKYKSINDVNDLVEKYIDEKERNEFKN
SNAKQYIREISNIITDTETAHLEYDDHISLIESEEKADEMKKFLDMYMSMYHWAKAFIVD
EVLDRDEMFYSDIDDIYNILENIVPLYNRVRNYVTQKPYNSKKIKLNFQSPTLANGWSQS
KEFDNNAIILIRDNKYYLAIFNAKNKPDKKIIQGNSDKKNDNDYKKMVYNLLPGANKMLP
KVFLSKKGIETFKPSDYIISGYNAHKHIKTSENFDISFCRDLIDYFKNSIEKHAEWKKYE
FKFSATDSYSDISEFYREVEMQGYRIDWTYISEADINKLDEEGKIYLFQIYNKDFAENST
GKENLHTMYFKNIFSEENLKDIIIKLNGQAELFYRKASVKNPVKHKKDSVLVNKTYENQL
DNGDVVRIPIPDDIYNEIVKMYNGYIKESDLSEAAKEYLDKVEVRTAQKDIVKDYRYTVD
KYFIHTPITINYKVTARNNVNDMVVKYIAQNDDIHVIGIDRGERNLIYISVIDSHGNIVK
QESYNILNNYDYKKKLVEKEKTREYARKNWKSIGNIKELKEGYISGVVHEIAMLIVEYNA
IIAMEDLNYGFKRGRFKVERQVYQKFESMLINKLNYFASKEKSVDEPGGLLKGYQLTYVP
DNIKNLGRQCGVIFYVPAAFTSKIDPSTGFISAFNFKSISTNASRKQFFMQFDEIRYCAE
KDMFSFGFDYNNFDTYNITMGKTQWTVYTNGERLQSEFNNARRTGKTKSINLTETIKLLL
EDNEINYADGHDIRIDMERMDEDKKSEFFAQLLSLYKLTVQMRNSYTEAEEQENGISYDK
IISPVINDEGEFFDSDNYKESDDKECPMPKDADANGAYCIALKGLYEVLKIKSEWTEDGF
DRNCLKLPHAEWLDFIQNKRYE

FIG. 73H

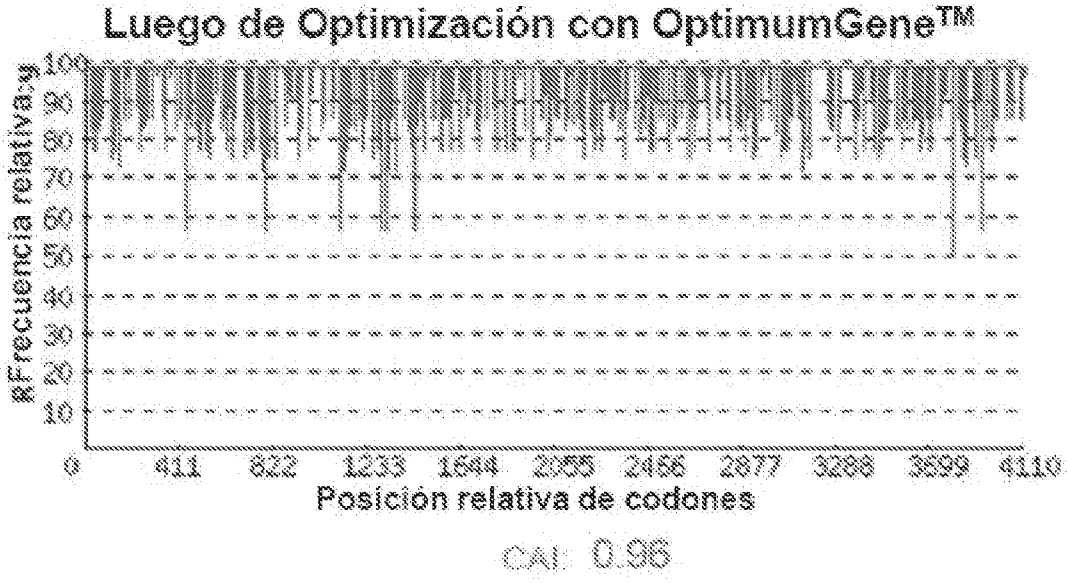


FIG. 74A

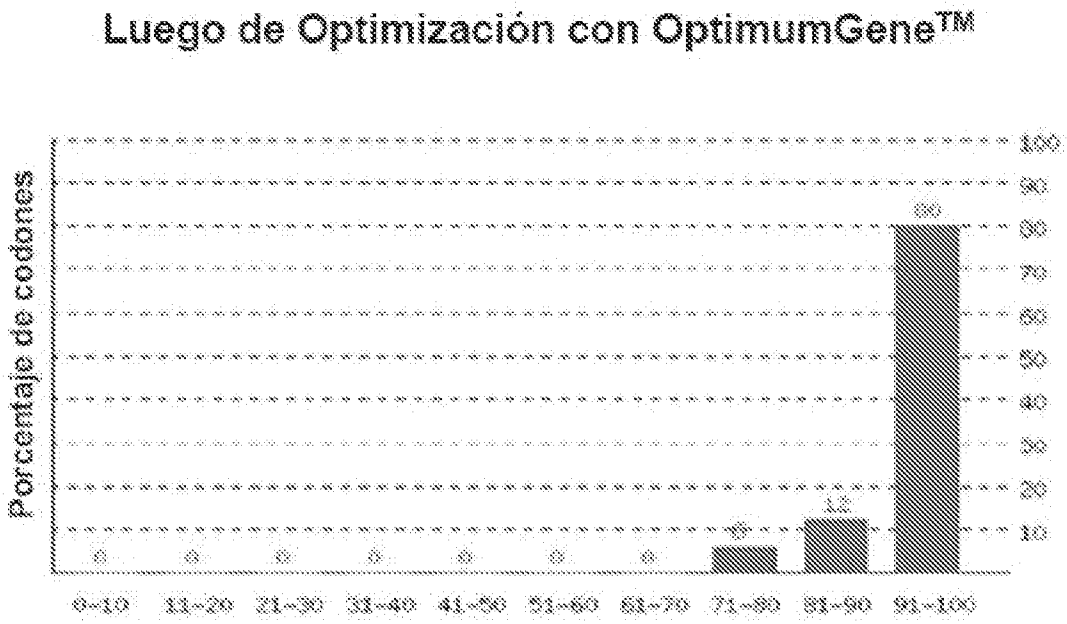


FIG. 74B

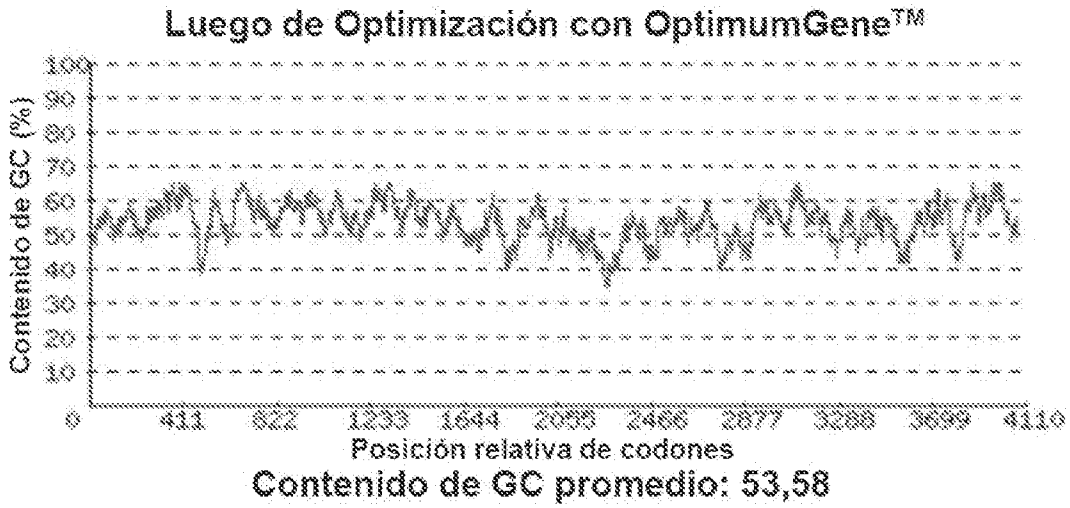

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
411
822
1233
1644
2055
2466
2877
3288
3699
4110
Posición relativa de codones
Contenido de GC promedio: 53,58

FIG. 74C

Enzimas de restricción	Optimizado
* verde: sitios filtrados, azul: sitios confirmados (no filtrados); Rojo: sitios mantenidos	
BamHI(GGATCC)	1(4174)
EcoRI(GAATTC)	1(4284)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4238)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PoliA(AATAAA)	0
PoliA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PoliT(TTTTTT)	0
PoliA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 74D

Luego de la optimización

Repetición directa Máx: Tamaño:14 Distancia:525 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de díadas Máx: Ninguna

FIG. 74E

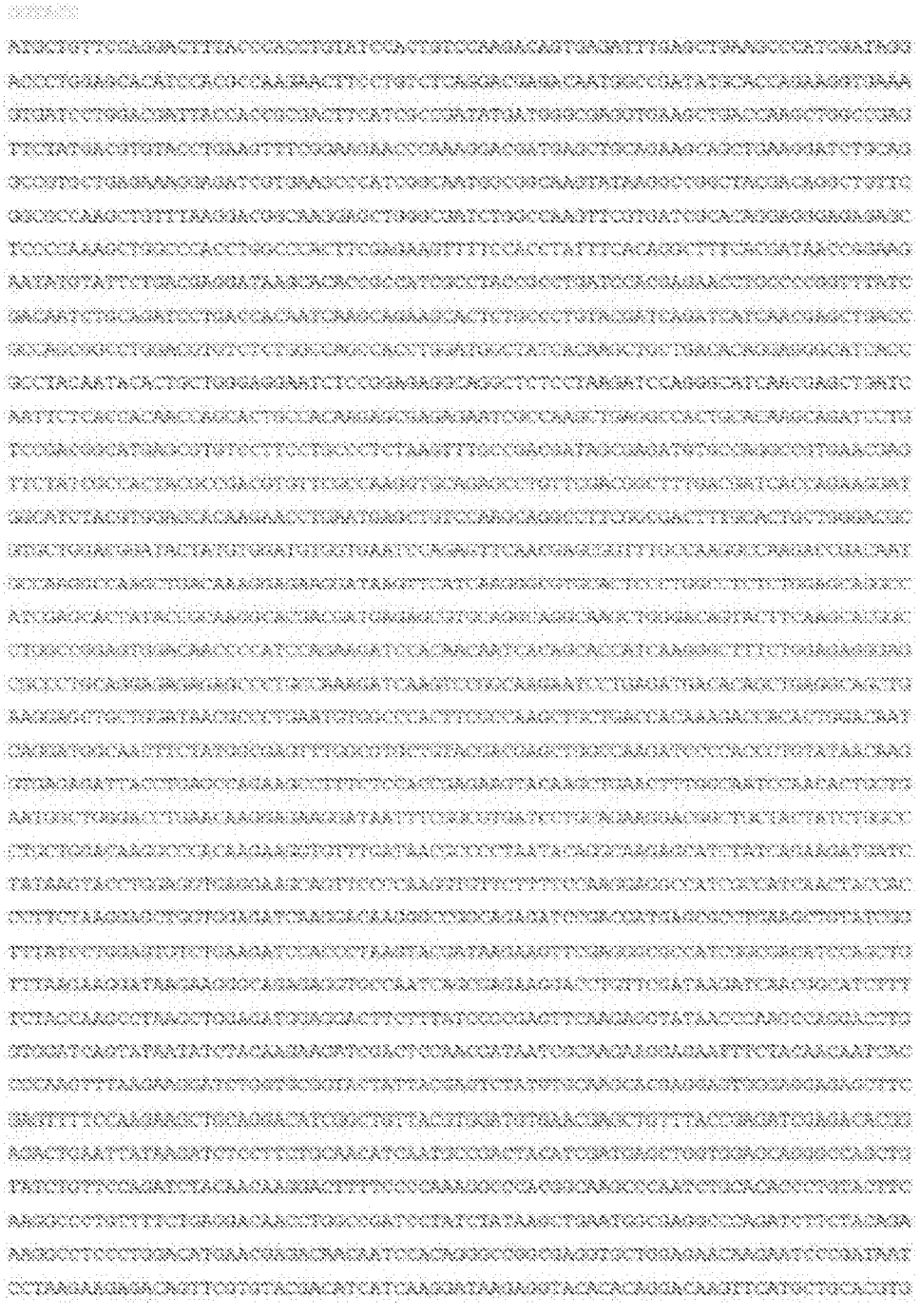


FIG. 74F



FIG. 74G

MLFQDPTHLYPLSETVRPELKPTDRTLEHIHAKNPLSQDETMADMHQKVKVILDDYHRDF
IADMNGEVKLTKLASFYDVYLKFRKNPKDDELQKQLKDLQAVLRKEIVKPIGNGGKYKAG
YDRLFGAKLFRDGKELGDLAKFVIAQEGESSPKLAHLAHFEKFSTYPTGFHDNRKNMYSD
EDKNTAIAYRLIHENLPRFIDNLQILTTIKQKNSALYDQIINELTASGLDVSLASHLDGY
HKLLTQEGITAYNTLLGGISGEAGSFKIQGINELINSHHNQHCHKSERIAELRPLHKQIL
SDGMSVSFLPSEFADDSEMCQAVNEFYRNYADVFAKVQSLFDGFDDNQKDGIYVENENLN
ELSKQAFGDFALLGEVLDGYYVDVVNPEFNERFAKAKTDNAKAKLITKENKEFIKGVHSLA
SLEQAIEHYTAHHDDESVQAGKLGQYFKNGLAGVDNFIQKIHNNHSTIKGFLERERPAGE
RALPKIKSGKNFSNTQLRQLKELLDNALNVAHFAKLLTTKTTLDNQDGNFYGEFGVLYDE
LAKIPTLYNKVRDYLSQKPFSTEKYKLNFGNPTLLNGWDLNKEKDNFGVILQKDGCYYLA
LLDKAHKKVFDNAPNTGKSIYQKMIYKYLEVSKQFPKVFFSKEAIAINYHPSKELVEIKD
KGRQRSDDERLKLYRFILECLKIHPKYDKKFEGAIGDIQLFKKDKKGREVPISEKDLFDK
INGIFSSKPKLEMEDFFIGEFKRYNPSQDLVDQYNIYKKIDSNDNREKENFYNNMPKFKK
DLVRYYYESMCKHEEWEESFRFSKKLQDIGCYVDVNELFTEIETRELNYKISFCNINADY
IDELVEQGQLYLFQIYNKDFSPKAHGKPNLHTLYFKALFSEDNLADPIYKLNGEAQIFYR
KASLDMNETTIHRAGEVLENKNPDNPKRQPVYDIIKDKRYTQDKFMLHVPITMNFGVQG
MTIKEFNKKVNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILEQCSLNDITTASANG
TQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVVHQISQLMLKYNAIVVLE
DLNPGFKRGRFKVERQIYQNFENALIKKLNHLVLKDKADDEIGSYKNALQLTNNFTDLKS
IGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENIAQSQAFFGKFDKICYNADKDYFEF
HIDYAKFTDKAKNSRQIWTICSHGDKRYVYDKTANQNKGAAKGINVNDELKSLFARHHIN
EKQFNLVNDICQNNDKEFHKSLMYLLKTLLALRYSNASSDEDFILSPVANDEGVFFNSAL
ADDTQPQNADANGAYHIALKGLWLLNELKNSDDLNKVKLAIDNQTWLNFAQNR

FIG. 74H

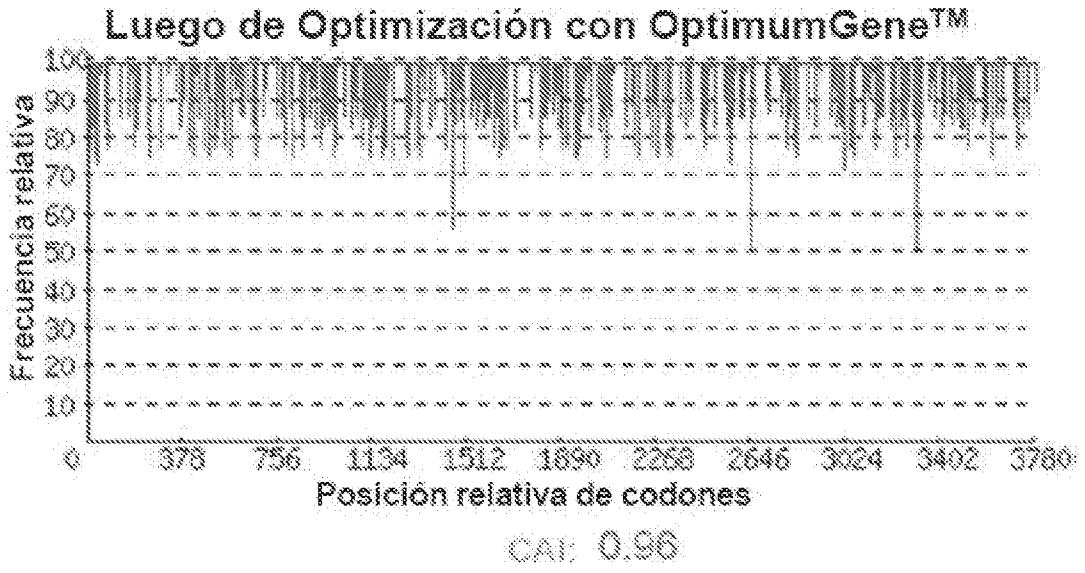


FIG. 75A

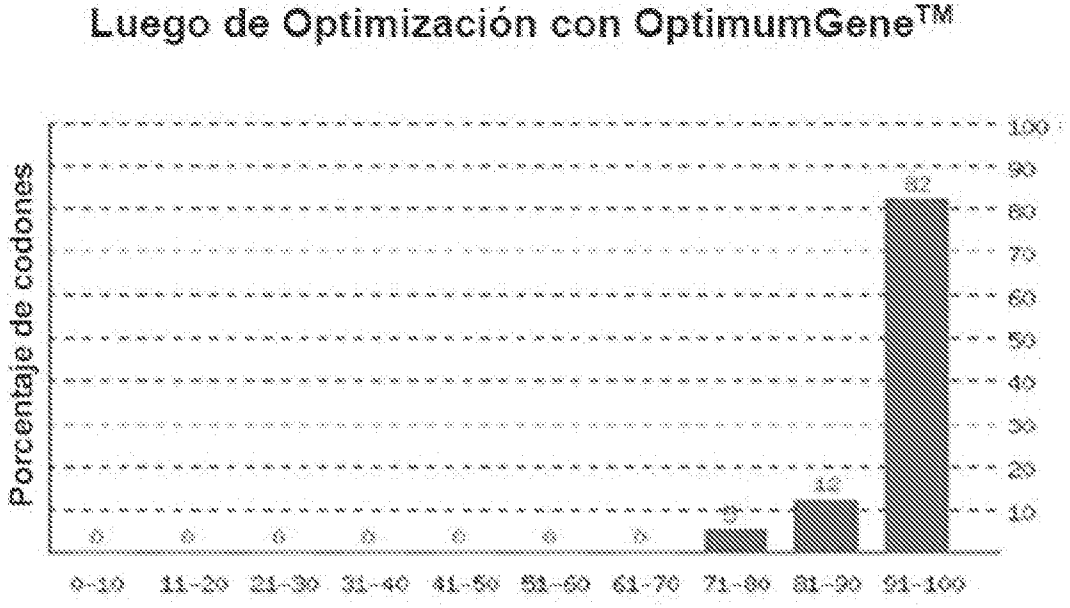


FIG. 75B

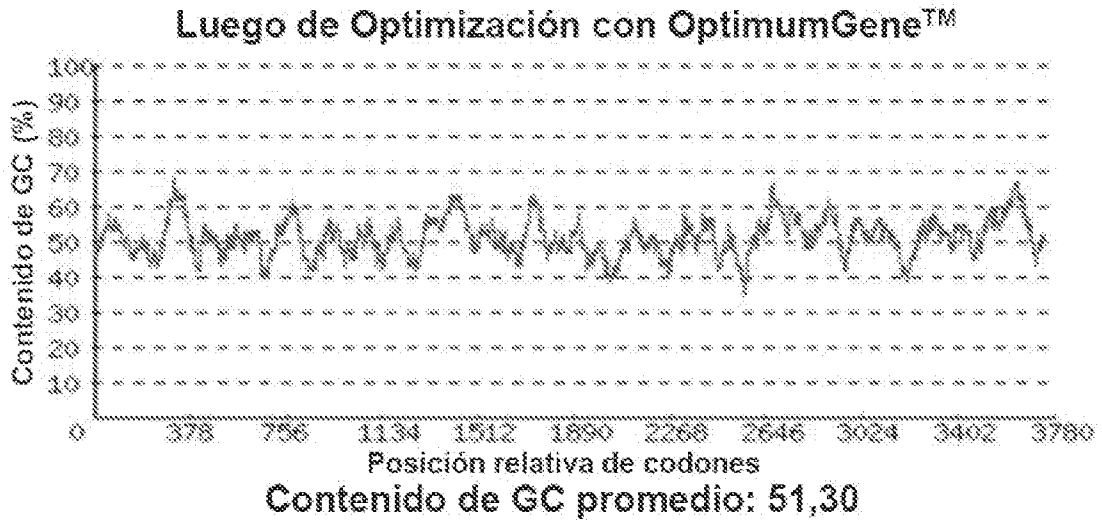

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
378
756
1134
1512
1890
2268
2646
3024
3402
3780
Posición relativa de codones
Contenido de GC promedio: 51,30

FIG. 75C

Enzimas de restricción	Optimizado
Verde: sitios únicos, azul: sitios confirmados (no únicos); Rojo: sitios mantenidos	
BamHI(GGATCC)	1(3844)
EcoRI(GAATTC)	1(3934)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3908)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PolIA(AATAAA)	0
PolIA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PolIT(TTTTTT)	0
PolIA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 75D

Luego de la optimización

Repetición directa Máx: Tamaño:14 Distancia:111 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de diadas Máx: Tamaño:13 Tm: 35,4 Posiciones de inicio: 1027, 221

FIG. 75E

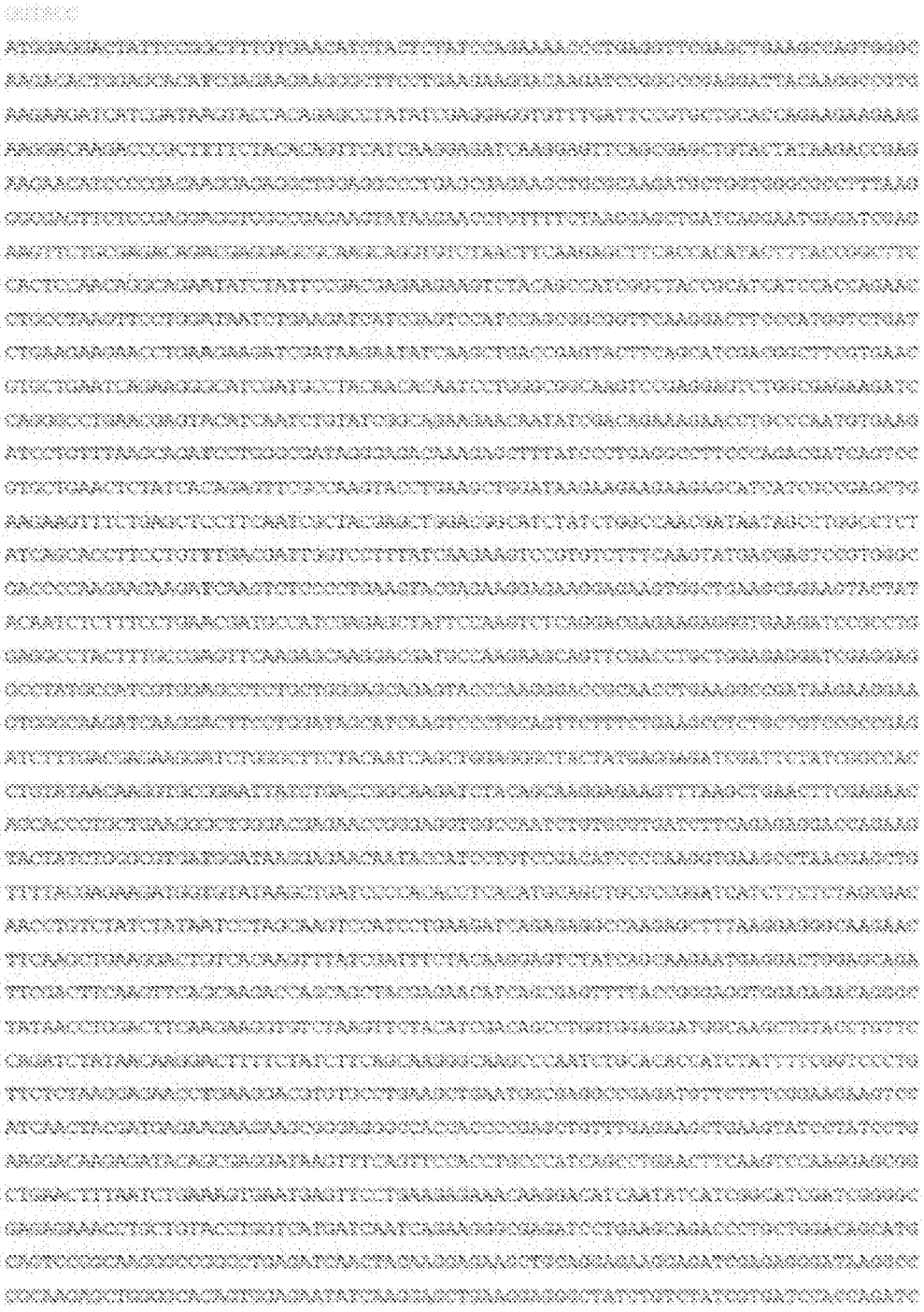


FIG. 75F



FIG. 75G

MEDYSGFVNIYSIQKTLRFELKPVGKTLEHIEKKGFLKKDKIRAEDYKAVKKIIDKYHEA
YIEEVFDSVLHQKKKEDKTRFSTQFIKEIKEFSELYYKTEKNIPDKEHLEALSEKLREML
VGAFKGEFSEEVAEEYKNLFSKELIRNEIEKFCETDEERKQVSNFKSFTTYFTGFHSNRQ
NIYSDEKKSTAIGYRIIHQNLPKFLDNLKIIESIQREFKDFFWSDLKKNLKKIDKNIKLT
EYPSIDGFVNVLNQKGIDAYNTILGGKSEESGEKIQGLNEYINLYRQENNIDRENLPNVK
ILFKQILGDRETKSFIPEAFFDDQSVLNSITEFAKYLKLDKKKKSIIAELKKFLSSFNRY
ELDGIYLANDNSLASISTFLFDDWSFIKKSVSFKYDESVGDFEKKIKSPLEYEKEKEKWL
KQKYYTISFLNDAIESYSKSQDEKEVKIRLEAYFAEFKSKDDAKKQFDLLERIEEAYAIV
EPLLGAEYPRDENLKADKKEVGKIKDFLDSIKSLQFFLKPLLSAEIFDEKDLGFYNQLEG
YYEEIDSIGHLYNKVRNYLTGKIYSKEKFKLNFENSTLLKGWDENREVANLCVIFREDQK
YYLGVMDKENNTILSDIPEVKPNELFYEKMVYKLIPTPNMQLPRIIFSSDNLSIYNPSES
ILKIRKAKSFKEGKNFKLKDCHKFIDFYKESISKHEDWSKFDFEFSKTSSYENISEFYRE
VERQGYNLDFEKVSKFYIDSLVEDGKLYLFQIYNKDFSIFSKGKPNLHTIYFRSLFSKEN
LKDVCLKLNGRAEMFFRKKSINYDKEKREGHHPELFEKLEYPILKDKRYSEDKFQFHLP
ISLNFKSKERLNFNLKVNEFLKRNKDINIIGIDRGERNLLYLVMINQKGEILKQTLLDSM
QSGKGRPEINYKEKLQEKEIERDKARKSWGTVENIKELKEGYLSIVIHQISKLMVENNAI
VVLEDLNIGFKRGRQKVERQVYQKFEEMLIDKLNFLVFKENKPTEPGGVLKAYQLTDEFQ
SFEKLSKQTGFLFYVPSWNTSKIDPETGFIDFLHPAYENIEKAKQWINKFDSIRFNSEMD
WFEFTADTRKFSERNLMLGKNRVWVICTTNVERYFTSKTANSSIQYNSIQITEKLKELFVD
IPFSNGQDLKPEILKNDAVFFKSLLFYIKTTLSLRQNNGFKGEEEKDFILSPVVDSKGR
FFNSLEASDDEPKDADANGAYHIALKGLMNLLVLNETKEENLSRFDWRICNKEDWLEFVQE
RNR

FIG. 75H

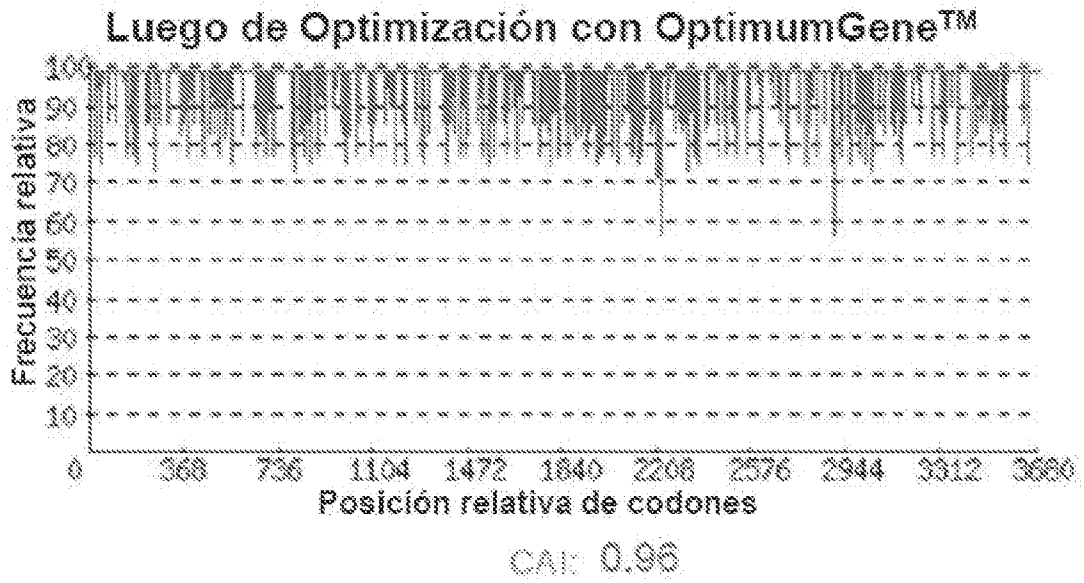


FIG. 76A

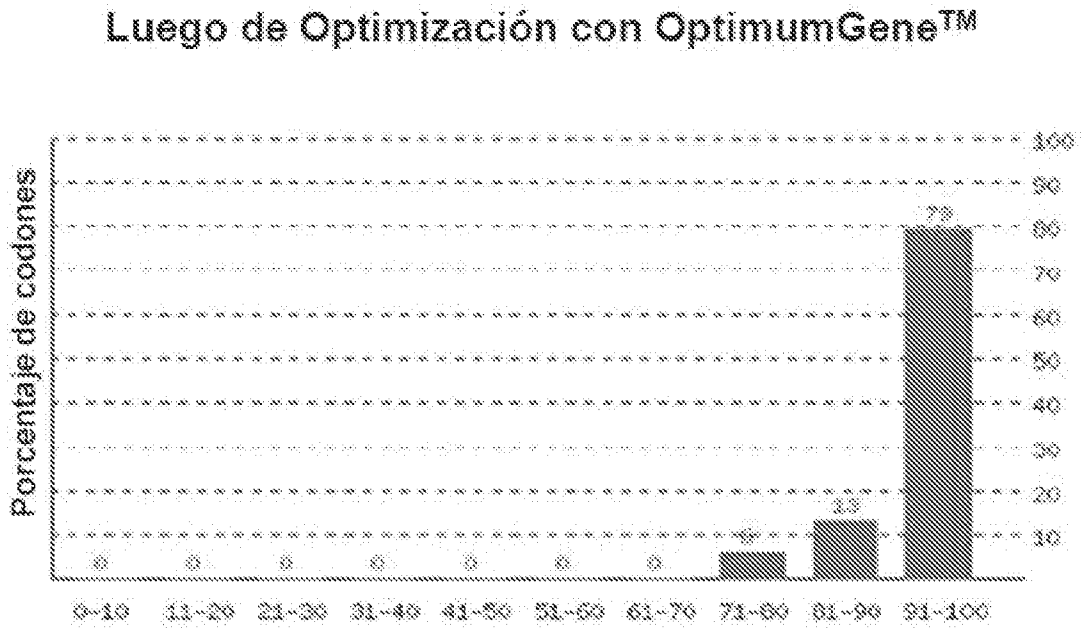


FIG. 76B

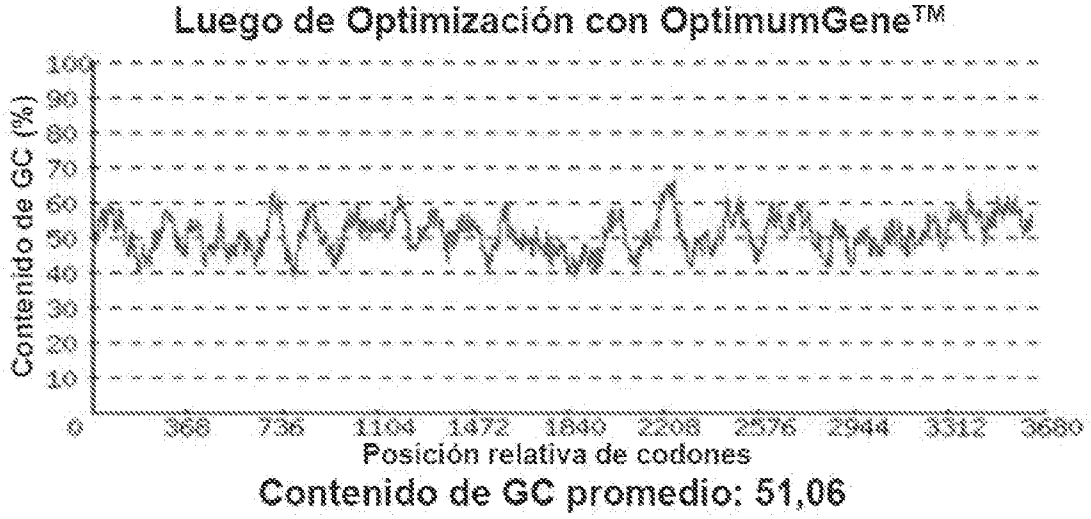

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
368
736
1104
1472
1840
2208
2576
2944
3312
3680
Posición relativa de codones
Contenido de GC promedio: 51,06

FIG. 76C

Enzimas de restricción	Optimizado
* verde: sitios filtrados, azul: sitios conservados (no filtrados); Rojo: sitios eliminados	
BamHI(GGATCC)	1(3736)
EcoRI(GAATTC)	1(3829)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3803)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PolIA(AATAAA)	0
PolIA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PolIT(TTTTTT)	0
PolIA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 76D

Luego de la optimización

Repetición directa Máx: Tamaño:14 Distancia:150 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de díadas Máx: Ninguna

FIG. 76E

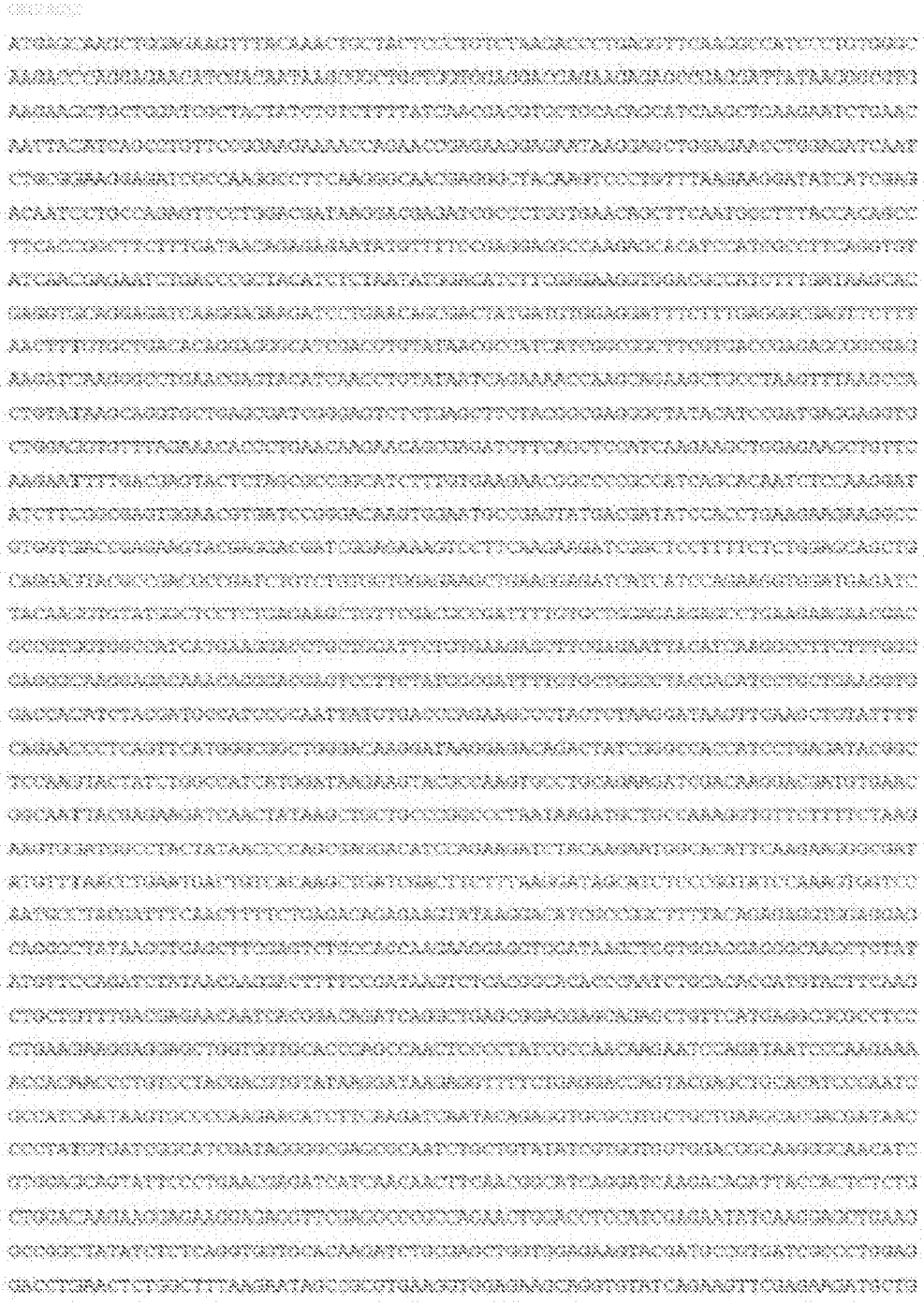


FIG. 76F



FIG. 76G

MSKLEKFTNCYSLSKTLRFKAIPVGETQENIDNKRLLVEDEKRAEDYKGVKKLLDRYYLS
FINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNEGYKSLFK
KDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEEAKSTSIAFRCINENL
TRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEFFNFVLTQEGIDVYNAI
IGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYTSDEEV
LEVFRNTLNKNSEIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRD
KWNAEYDDIHLKKKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQ
KVDEIYKVYGSSEKLFDADFVLEKSLKKNDAVVAIMKDLLDSVKSFENYIKAFFGEGKET
NRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQFMGGWDKDKET
DYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSK
KWMAYYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSET
EKYKDIAGFYREVEEQGYKVSFESASKKEVDKLVEEGKLYMFQIYNKDFSDKSHGTPNLH
TMYFKLLFDENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNPDNPKETTTLS
YDVYKDKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLLY
IVVVDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELK
AGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNYMVDK
KSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTS
IADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTDADYIKKWKLYSYGNRIRIFRNPKK
NNVFDWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMSLMLQMRNS
ITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKK
AEDEKLDKVKIAISNKEWLEYAQTSVKH

FIG. 76H

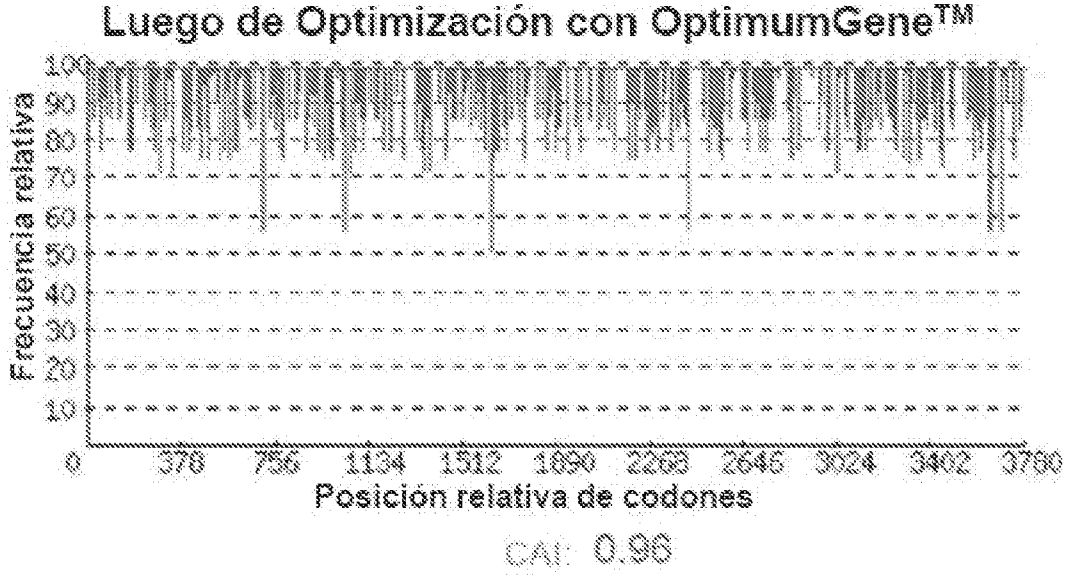


FIG. 77A

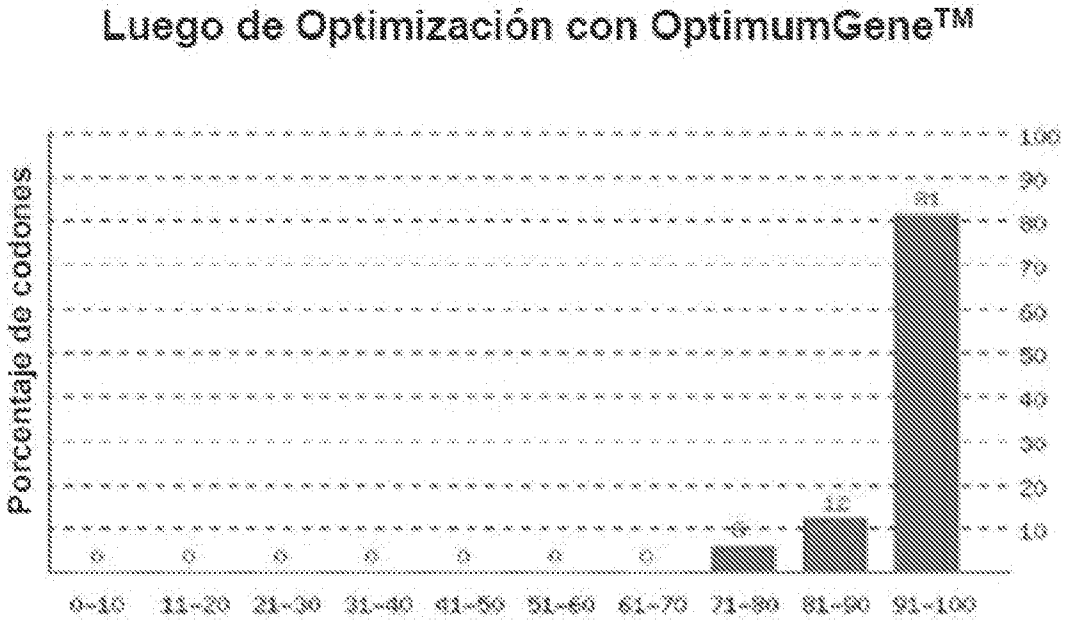


FIG. 77B

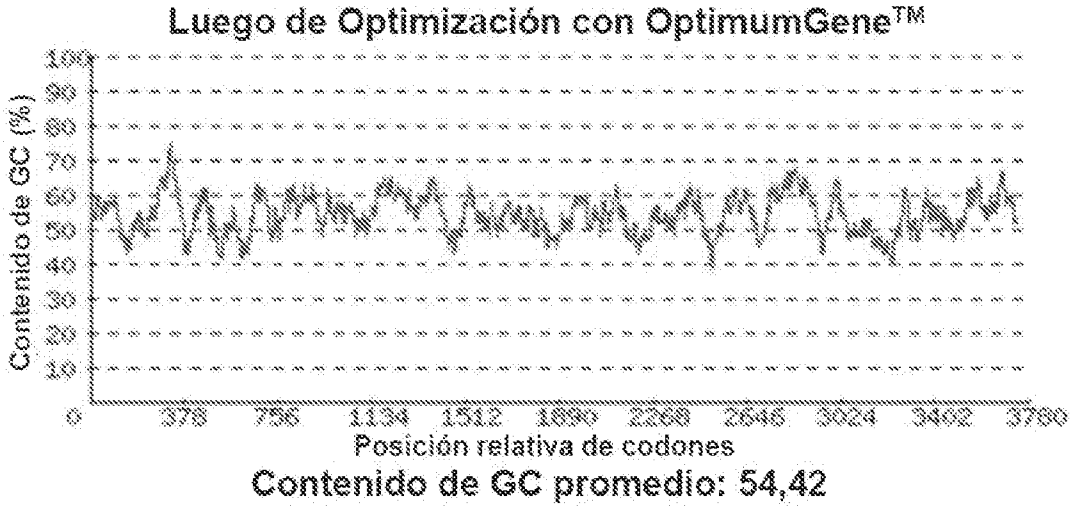

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
378
756
1134
1512
1890
2268
2646
3024
3402
3780
Posición relativa de codones
Contenido de GC promedio: 54,42

FIG. 77C

Enzimas de restricción	Optimizado
verde: sitios filtrados, azul: sitios confirmados (no filtrados), Rojo: sitios mantenidos	
BamHI(GGATCC)	1(3836)
EcoRI(GAATTC)	1(3925)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3890)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PoliA(AATAAA)	0
PoliA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PoliT(TTTTTT)	0
PoliA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 77D

Luego de la optimización

Repetición directa Máx: Tamaño:13 Distancia:2043 Frecuencia:2
Repetición invertida Máx: Tamaño:13 Tm: 41,3 Posiciones de inicio: 1469, 2693
Repetición de diadas Máx: Ninguna

FIG. 77E

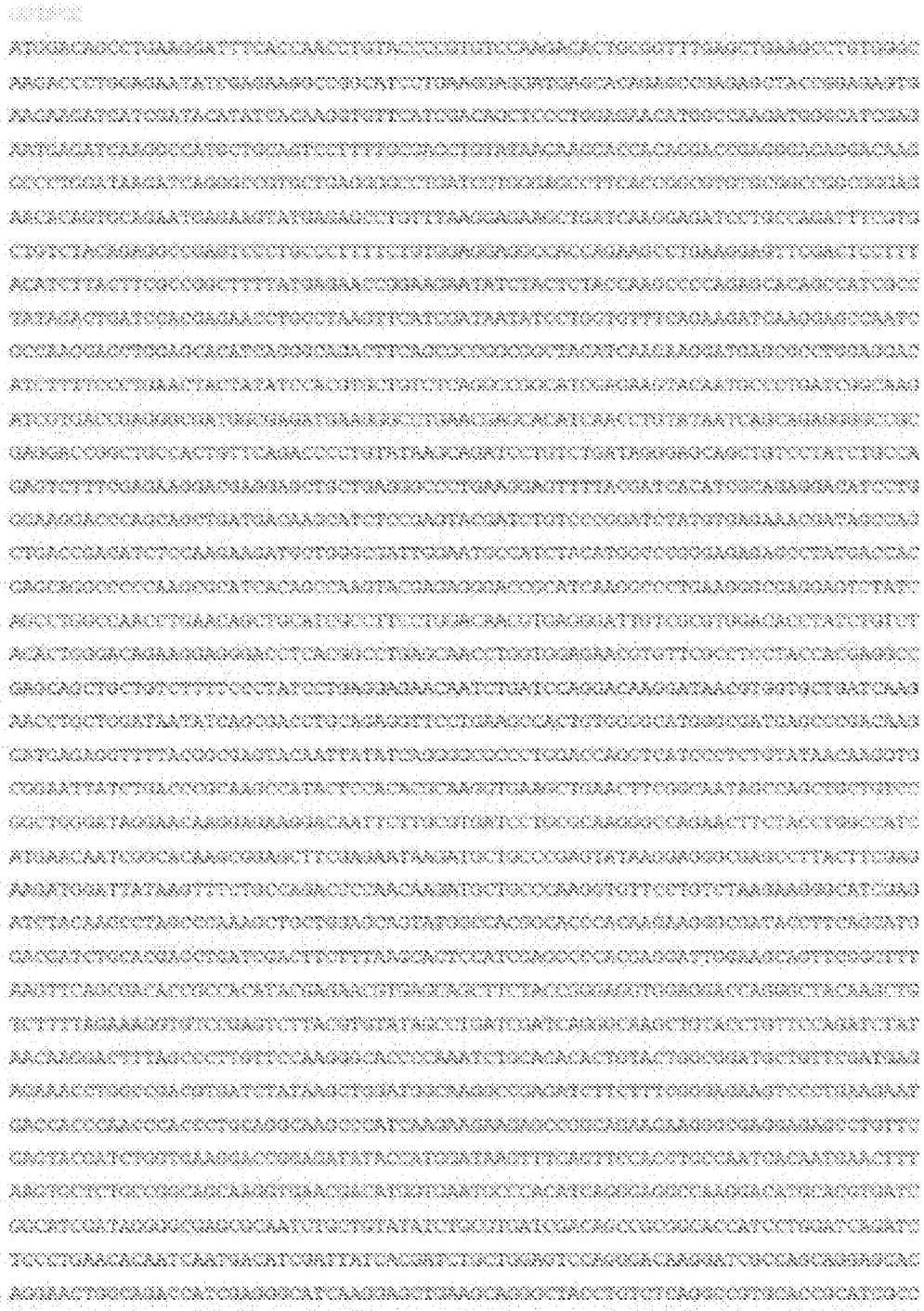


FIG. 77F



FIG. 77G

MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRAESYRRVKKIIDTYHKV
FIDSSLENMADMGIENEIKAMLQSFCELYKKDHRTEGEDKALDKIRAVLRGLIVGAFTGV
CGRRENTVQNEKYESLFKEKLIKEILPDFVLSTEAESLPFSVEEATRSLKEFDSFTSYFA
GFYENRKNIYSTKPQSTAIAYRLINENLPKFIDNILVFQKIKEFIAKELENIKADFSAGG
YIKEDERLEDIFSLNYYIHVLSQAGIEKYNALIGKIVTEGDGEMKGLNEEINLLYNQGRG
EDRLPLFEPLYKQILSDREQLSYLPESFEKDEELLEALKEFYDHIAEDILGRTQQLMTSI
SEYDLSEIYVRNDSQLTDISKEMLGDWNAIYMARERAYDHEQAPKRITAKYERDRIKALK
GEESISLANLNSCIAPLDNVRDCRVDTYLSTLGQKEGPHGLSNLVENVFASYHEAEQLLS
FPYPEENNLIQDKDNVVLIKNLLDNISDLQRFLKPLMGMGDEPDKDERFYGEYNYIRGAL
DQVIPLYNKVRNYLTRKPYSTRKVKLNFGNSQLLGGWDENKEKDNSCVILRKGQNFYLAI
MNNRHKRSFEHFMLPEYKEGEPYFEKMDYKFLPDPNKMLPKVFLSKRGIEIYKPSPKLLE
QYGHGTHKKGDTFSMDDLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENVSSFYREVED
QGYKLSFRKVSESYVYSLIDQGELYLFQIYNKDFSPCSKGTPNLHTLYWRMLFDERNLAD
VIYKLDGKAEIFFREKSLKNDHPTHPAGKPIKYKSRQKKGEESLFEYDLVKDRRYTMDKF
QFHVPITMNFKCSAGSEVNDMVNAHIREAEDMHVIGIDRGERNLLYICVIDSRGTILDQI
SLNTINDIDYHDLLESRDKDRQQEHRNWQTIEGIKELKQGYLSQAVHRIAELMVAYKAVV
ALEDLNMGFKRGRQKVEKSVYQQFEKQLIDKLAYLVDKEKRPEDIGGLLRAYQFTAPFES
FKEMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHVQYENVDKAKSFFQKFDSISYNPKKDW
FEFAFDYKNFTKKAEGSRSMWILCTHGSRIENFRNSQKNQQWDSEEFALTEAFKSLFVRY
EIDYTADLKTAIVDEKQKDFFVDLLKLFKLTVQMRNSWKEKDLDYLISPVAGADGRFFDT
REGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGKLKLAISNKEWLQFVQERSYEED

FIG. 77H

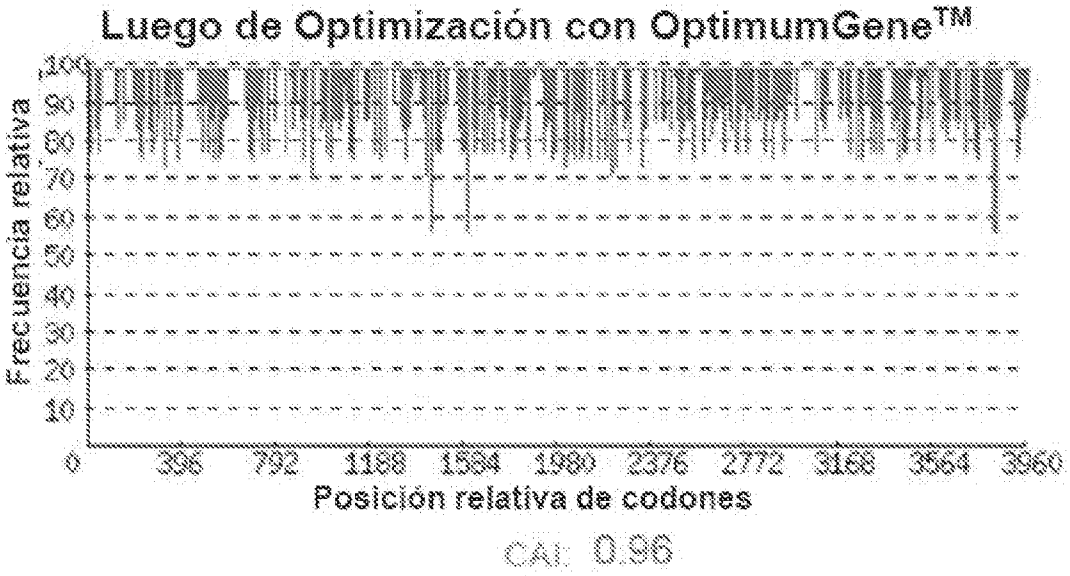


FIG. 78A

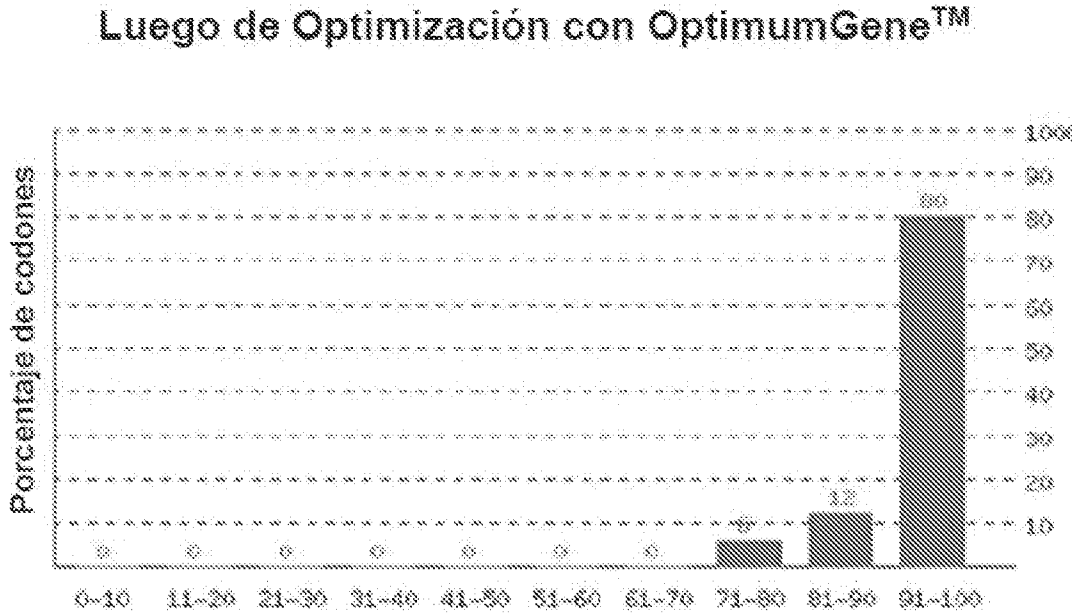


FIG. 78B

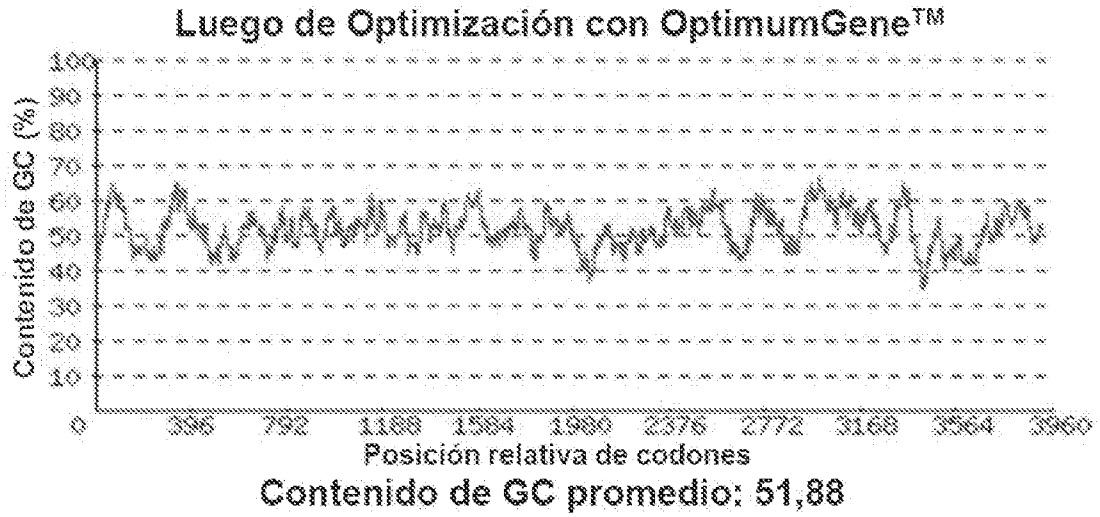

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
396
792
1188
1584
1980
2376
2772
3168
3564
3960
Posición relativa de codones
Contenido de GC promedio: 51,88

FIG. 78C

Enzimas de restricción	Optimizado
* verde: sitios filtrados, azul: sitios confirmados (no filtrados), Rojo: sitios mantenidos	
BamHI(GGATCC)	1(4024)
EcoRI(GAATTC)	1(4114)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(4083)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PolIA(AATAAA)	0
PolIA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PolIT(TTTTTT)	0
PolIA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 78D

Luego de la optimización

Repetición directa Màx: Tamaño:15 Distancia:1842 Frecuencia:2
Repetición invertida Màx: Ninguna
Repetición de díadas Màx: Ninguna

FIG. 78E

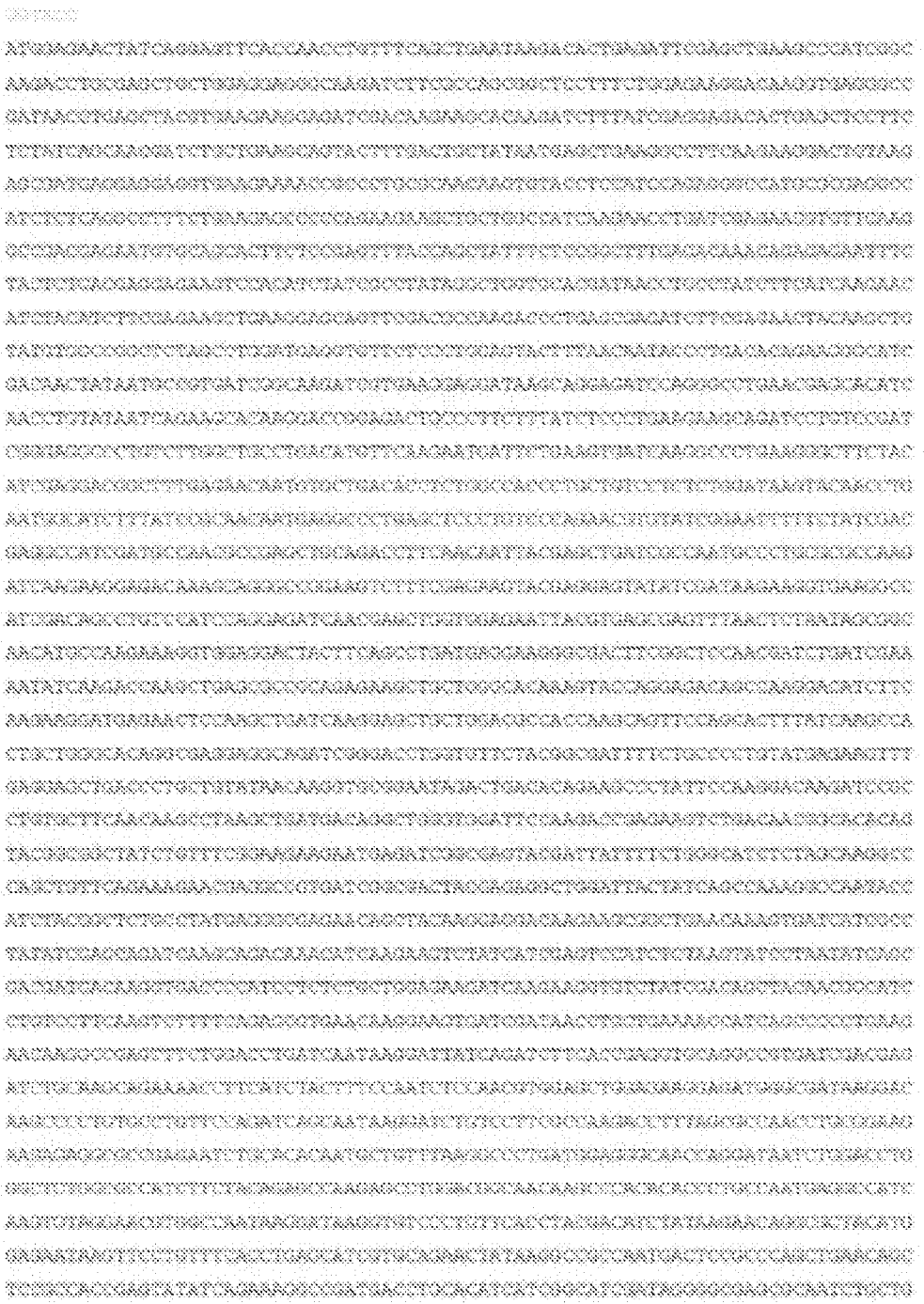


FIG. 78F



FIG. 78G

MENYQEPTNLFQLNKTLPFELKPIQKTCELLEEGKIFASGSPLEKDKVRADNVSYVKKEI
DKFNKIPIEETLSSPSISNDLLKQYFDCYNELKAFKFDCKSDEEEVKKTALRNKCTSIQR
AMREAISQAPLKSFQKELLAIKNLIENVFKADENVQHFSEPTSYFSGFETNEENFYSDEE
KSTSIAYRLVHDNLPIFIKNIYIPEKLKEQFDAKTLSEIFENYKLYVAGSSLDEVFSLEY
FNNTLTQKQIDNYNAVIGKIVKEDKQEIQGLNEHINLYNQKHKDKRLPFFISLKKQILSD
REALSWLPDMFENDSEVIKALKGFYIEDGFENNVLTPLATLLSSLDKYNLNGIFIRNNEA
LSSLSQNVYRNFSIDEAIDANAELQTFNNYELIANALRAKIKKETKQGRKSFEKYEEYID
KKVKAIDSLSIQEINELVRNYVSEPNSNSQNNPREVEDYFSLMRKGDFGSNDLIERIKTK
LSAAEKLLGTKYQETAKDIFKKDENSKLIKELLDATKQFQHFIKPLLGTGEEADRDLVFY
GDFLPLYEKFEELTLLYNKVRNRLTQKPYSEDKIRLCFNKFKLMTGWVDSKTEKSDMQTQ
YGGYLFREKNEIGEYDYFLGISSKAQLFRKNEAVIGDYERLDYYQPKANTIYGSAYEGEN
SYKEDKFRLNKVIIAYIEQIKQTNIKKSIIESISKYFNISDDGKVTPSSLLEKIKKVSID
SYNGILSFESPQSVNKEVIDNLLKTISPLENKAEFLDLIKEDYQIFTEVQAVIDEICKQK
TFIYFPISNVELEKEMGDKDKPLCLFQISNKDLSFAETFSANLREKRGAENLHTMLFKAL
MEGNQDNLDLGSGAIFYRAKSLDGNKPTHPANEAICENVANEDKVSLPTYDIYQNERYM
ENKFLFHLSIVQNYKAANDSAQLNSSATEYIREADDLHIIGIDRGERNLLYYSVIDMKGN
IVEQDSLNIIRNNDLETDYHDLLDKREKERKANRQNWEAVEGIKDLKEGYLSQAVHQIAQ
LMLKYNAIIALEDLQMPVTRQQKIEKAVYQQFEKSLVDKLSYLVDKKRPVNRLGGILKA
YQLASSITKNNSDKQNQFLFYVPANNTSKIDPVTGFTDLLRPKAMTIKEAQDFFQKFDNI
SYNDEGYFEFETNYDKFEIRMKSAQTRWTICTFGNRIKRKKDKNYWNYREVELTEEFKKL
FKDSNIDYENCNLKEEIQNKDNREFFDDLIKLLQLTLQMRNSDKGNDYIISPVANAEGQ
FFDSRNGDEKLPLDADANGAYNIARKGLWNIRQIKQTKDDKKLNLSISSTEWLDFVREKP
YLK

FIG. 78H

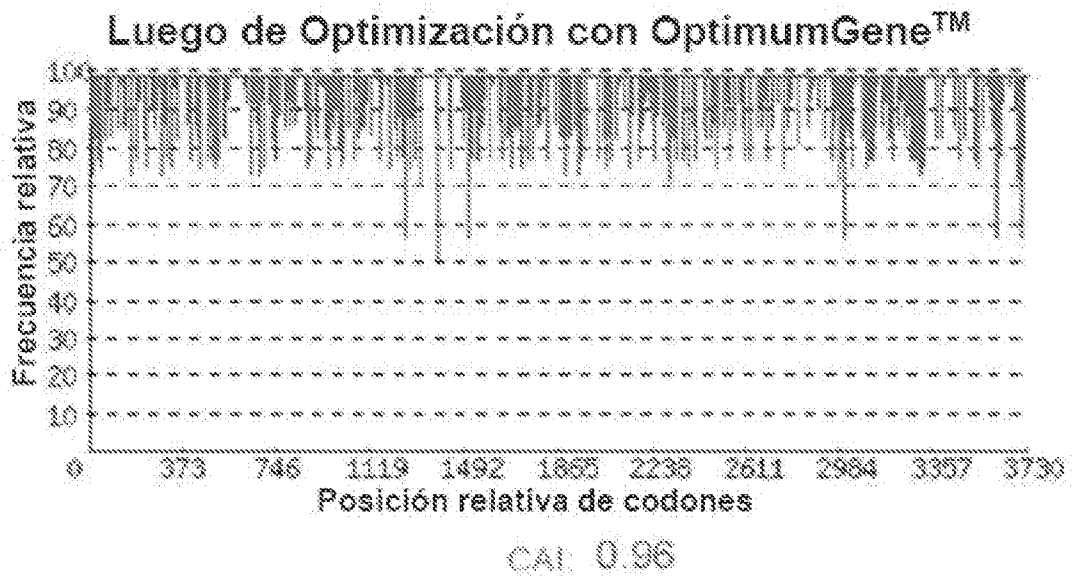


FIG. 79A

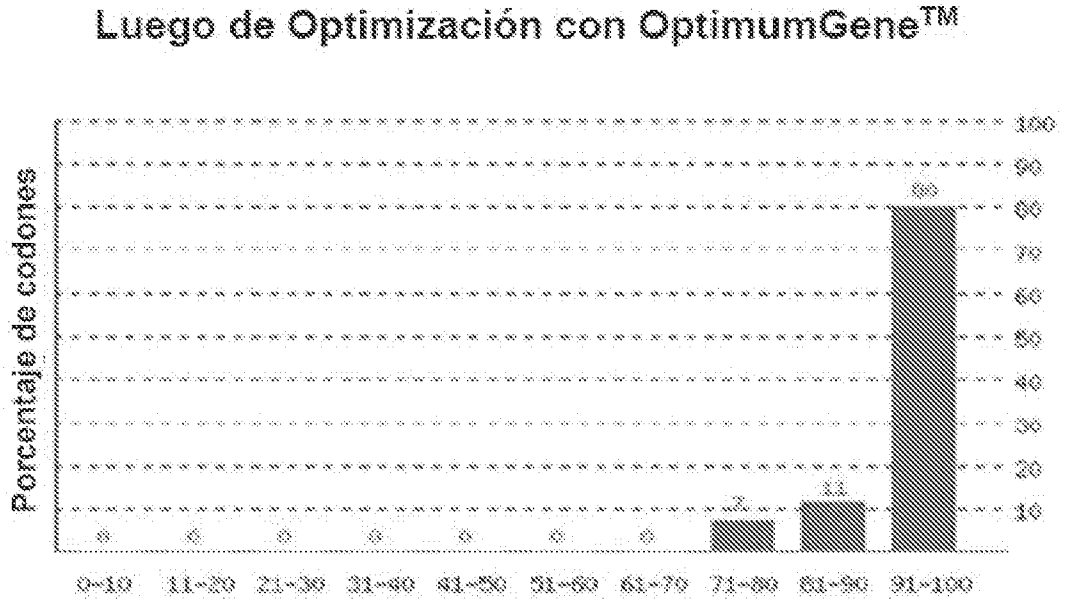


FIG. 79B

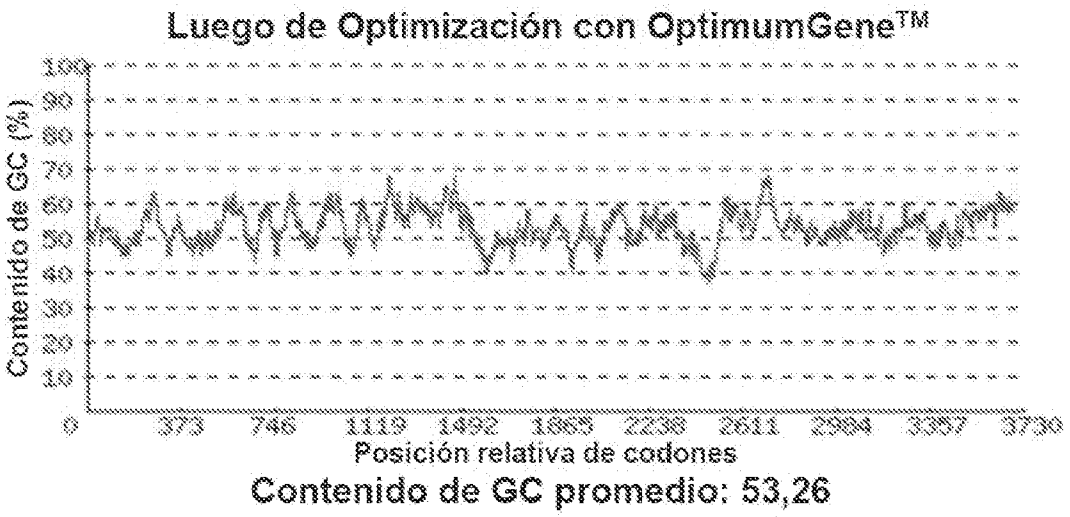

Luego de Optimización con OptimumGene™
Contenido de GC (%)
100
90
80
70
60
50
40
30
20
10
0
373
746
1119
1492
1865
2238
2611
2984
3357
3730
Posición relativa de codones
Contenido de GC promedio: 53,26

FIG. 79C

Enzimas de restricción	Optimizado
BamHI(GGATCC)	1(3793)
EcoRI(GAATTC)	1(3863)
HindIII(AAGCTT)	0
BsmBI(CGTCTC)	0
BsaI(GGTCTC)	0
BbsI(GAAGAC)	0
AgeI(ACCGGT)	0
XhoI(CTCGAG)	0
NdeI(CATATG)	1(3857)
NotI(GCGGCCGC)	0
KpnI(GGTACC)	1(1)
BsrGI(TGTACA)	0
SpeI(ACTAGT)	0
XbaI(TCTAGA)	0
NheI(GCTAGC)	0
ARE	0

Elementos de acción CIS	Optimizado
Corte y empalme (GCTAAG)	0
Corte y empalme (GGTGAT)	0
PoliA(AATAAA)	0
PoliA(ATTAAA)	0
Desestabilizante(ATTTA)	0
PoliT(TTTTTT)	0
PoliA(AAAAAAA)	0

Motivos antivirales	Optimizado
	0

FIG. 79D

Luego de la optimización

Repetición directa Máx: Tamaño:13 Distancia:99 Frecuencia:2
Repetición invertida Máx: Ninguna
Repetición de díadas Máx: Ninguna

FIG. 79E

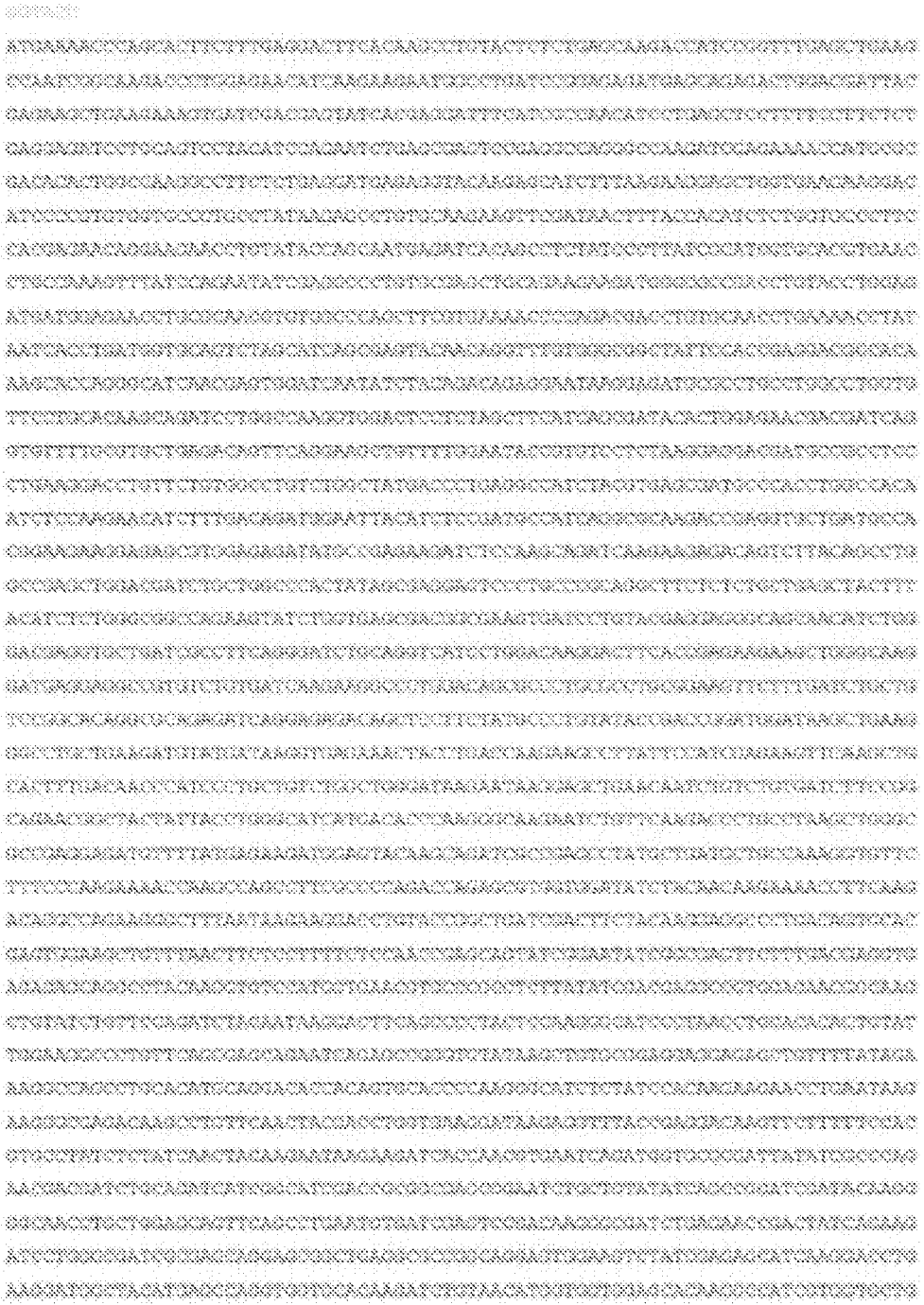


FIG. 79F

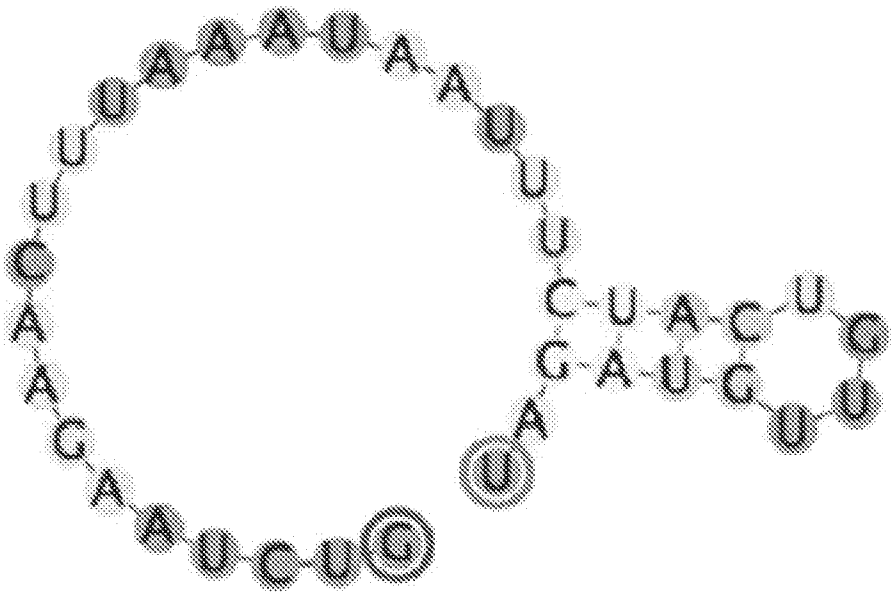


FIG. 79G

MRTQHFFEDPTGLYSLSKTIRFELKPIGKTLENIKKNGLIRRDEQRLDDYEKLKKVIDEY
HEDFIANTLSSPSFSREILQSYIQNLSESEARAKIEKTMRDTLAKAFSEDERYKSIFKE
LVKRDIPVWCPAYKSLCKKFDNPTTSLVPFHENRKNLYTSNEITASIPYRIVHVNLPKFI
QNIEALCELQKKMGADLYLEMMENLENVWPSFVKTPDDLCNLKTVNHLMVQSSISEYNEP
VGGYSTEDGTKHQGINEWINIYRQRNKEMRLPGLVFLHKQILAKVDSSSFISDTLENDDQ
VFCVLRQPRKLFWNTVSGKRDDAASLKDLPCGLSGYDPEAIYVSDAHLATISKNIFDRWN
YISDAIRRKTEVLMPRKFESVERYAEKISKQIKKRQSYSLAELDDLLAHYSEESLPAGPS
LLSYPTSLGGQKYLVSDGEVILYEEGGNIWDEVLIAFRDLQVILDEDFTEKKLGKDEEAV
SVIKKALDSALRLEKFFDLLSGTGAEIRRDSSFYALYTDRMDKLEGLLKMYDKVRNYLTK
KPYSIEKFKLNFDNPSLLSGWDKNKELNNLGVIFRQNGYYYLGIMTPEGKNLFKTLPKLG
AEEMFYEKMEYKQIAEPMLMLPKVFFPEKTKPAFAPDQSVVDIYNKKTFKTGQEGFNKKD
LYRLIDFYKEALTVHEWKLFNFSFSPTEQYENIGEFFDEVREQAYKVSMVNVPASYIDEA
VENGKLYLFQIYNKDFSPYSKGIPNLHTLYWKALFSEQNQSRVYKLCGGAELFYRKASLN
MQDTTVHPKGISIHKKNLNKKGETSLFNYDLVKDKRFTEDKFYFHVPISINYKNKKITNV
NQMVRDYIAQNDDLQIIGIDRGERNLLYISRIDTRGNLLEQPSLNVIESDKGDLRTDYQK
ILGDREQERLRRRQENKSIEGIKDLKDGYMSQVVHEICNMVVEHKAIVVLENLNLGFMKG
RKKVEKSVYEKFERMLVDKLNYLVVDKKNLSNEPGGLYAAYQLTNPLFSFEELHKYPQSG
ILFFVDPWNTSLIDPSTGFVNLLGKIKYTNVGDARKFFDEFNAIRYDGKGNILFDLDLSR
FDVRVETQRKLWTLTTFGSRIARSKKSGKWMVERIENLSLCFLELFEQFNIGYRVEKDLK
KAILSQDKEFYVELIYLFNLMMQIRNSDGEEDYILSPALNENNLQFDSRLIEAKDLPVD
ADANGAYNVARKGLMVVQRIKEGDHESIHRIGRAQWLRYVQEGIVE

FIG. 79H

>1 DR
GUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAU



>2 DR
GCUAUAAUGCCUAUAUAAUUUCUACUAUUGUAGAU

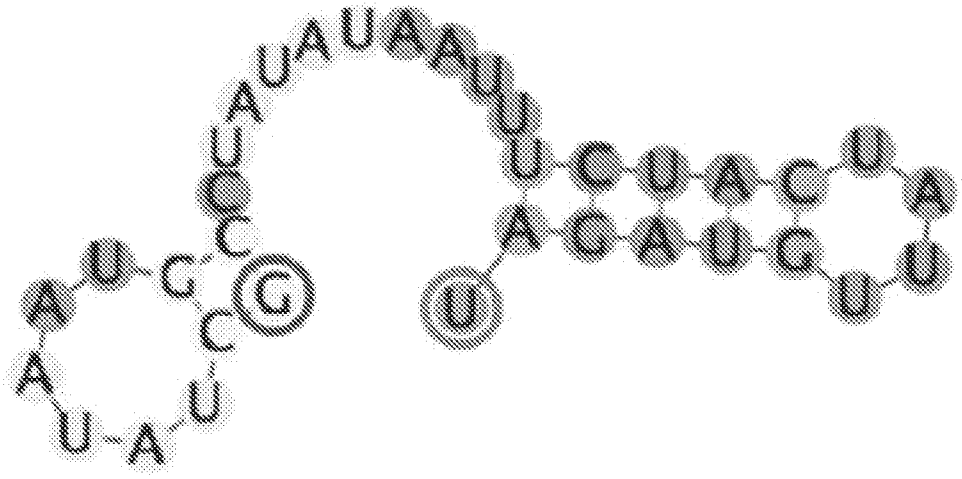
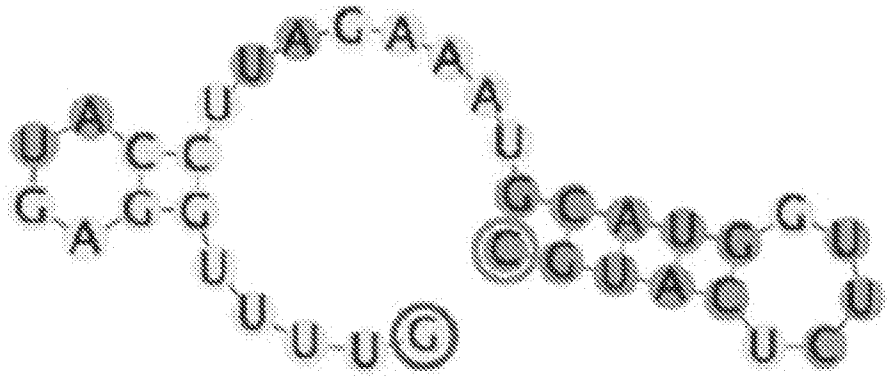


FIG. 80A

>3 DR
GCCUAUAAGGCUUUAGUAAUUUCUACUAUUGUAGAU



>4 DR
GUCAAUAAGACUCAUUUAAUUUCUACUUCGGUAGAU

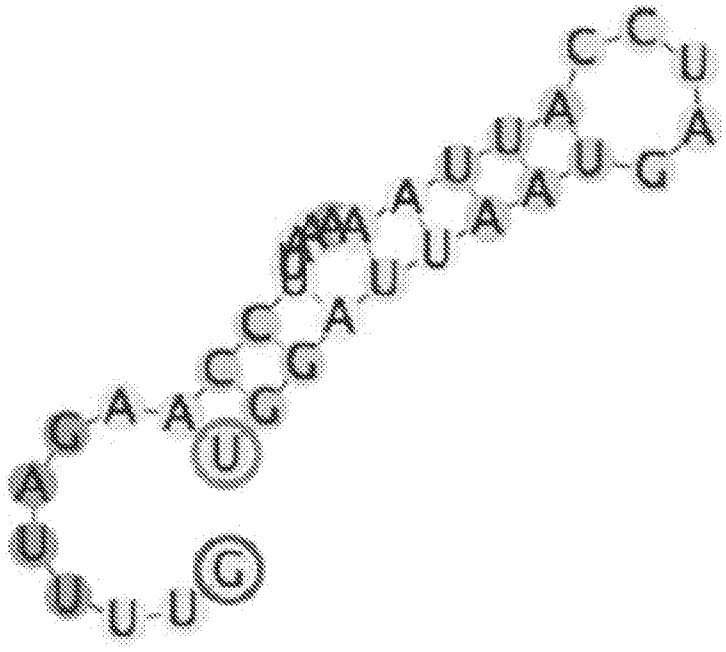
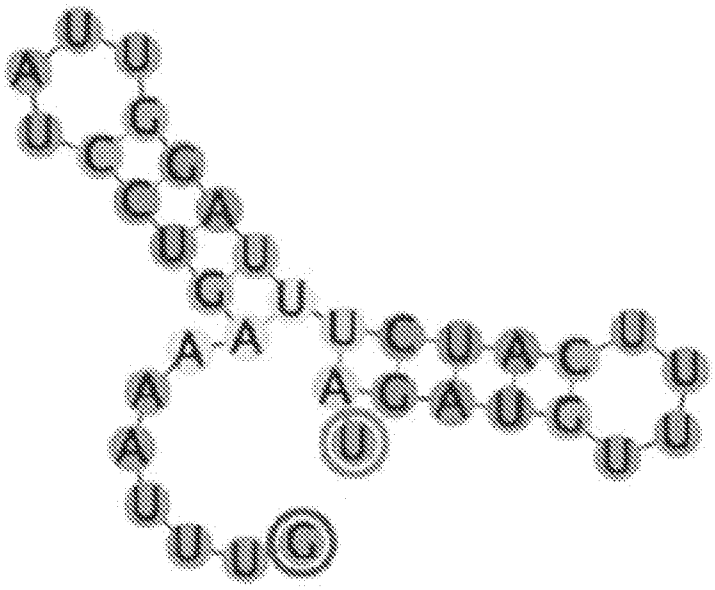


FIG. 80B

>5 DR
GUCUAGGUACUCUCUUUAAUUUCUACUAUUGUAGAU



>6 DR
GUUUCAAAGAUUAAAUAAUUUCUACUAAGUGUAGAU

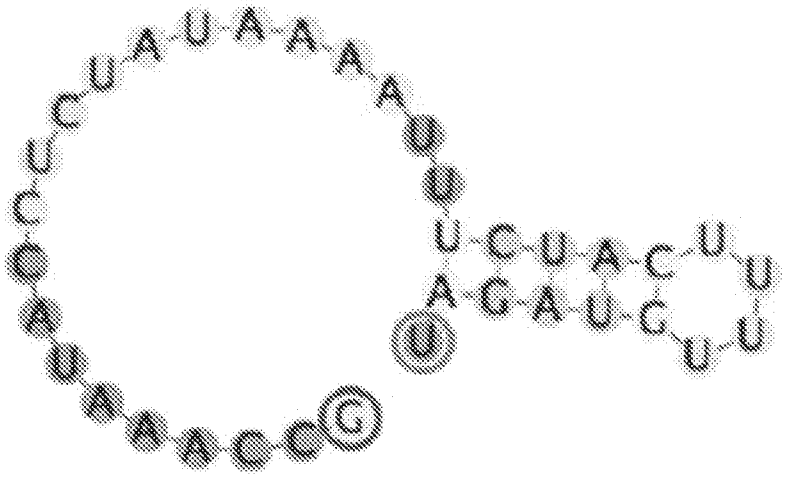
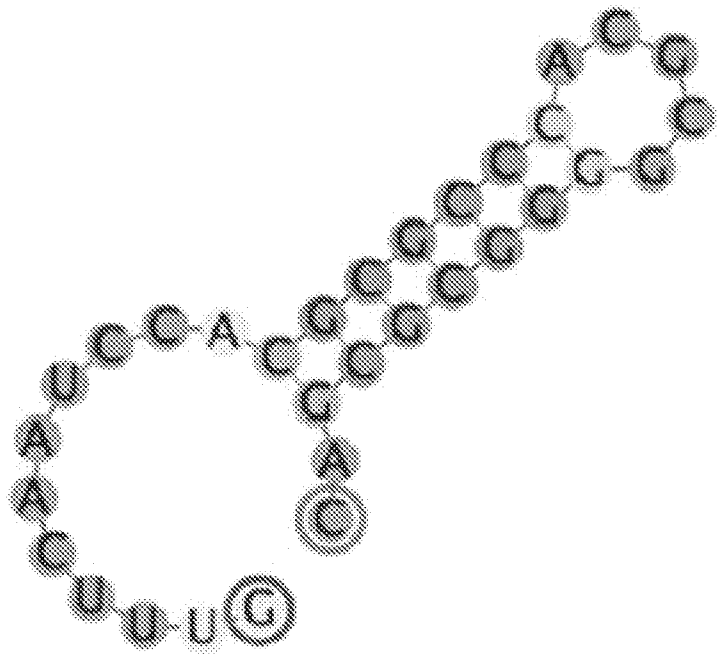


FIG. 80C

>7 DR
CUCUAAAGAGAGGAAAGAAUUUCUACUUUUGUAGAU



>8 DR
GUCUAACGACCUUUUAAAUUUCUACUGUUUGUAGAU

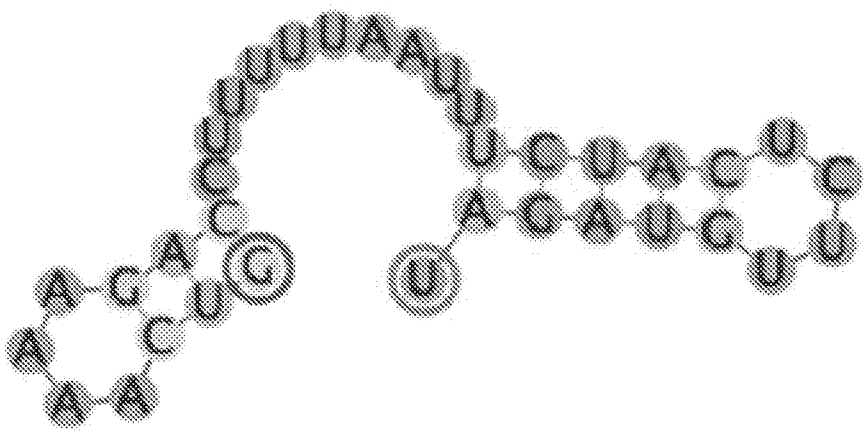
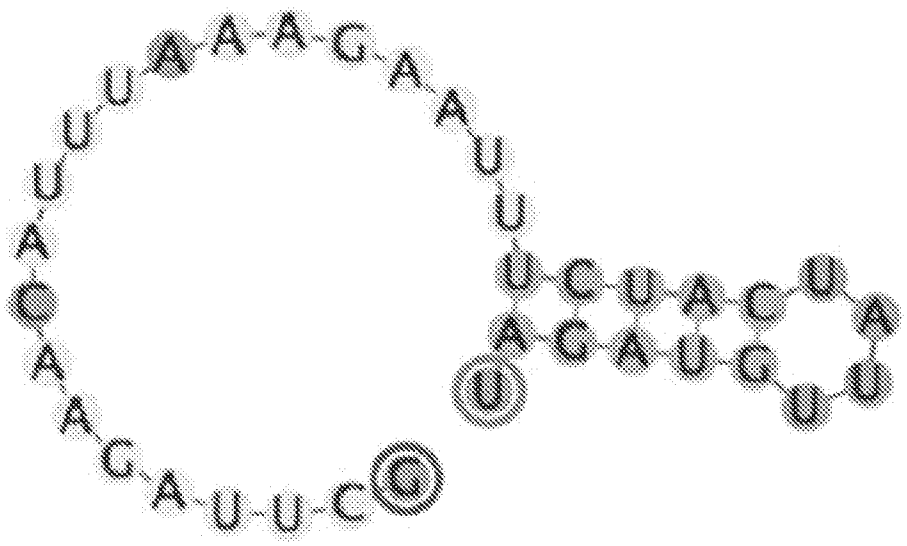


FIG. 80D

>9 DR
GUUUGAAUAACCUUAAAUAAUUUCUACUUUGUAGAU



>10 DR
CUCAAAACUCAUUCGAAUCUCUACUCUUUGUAGAU

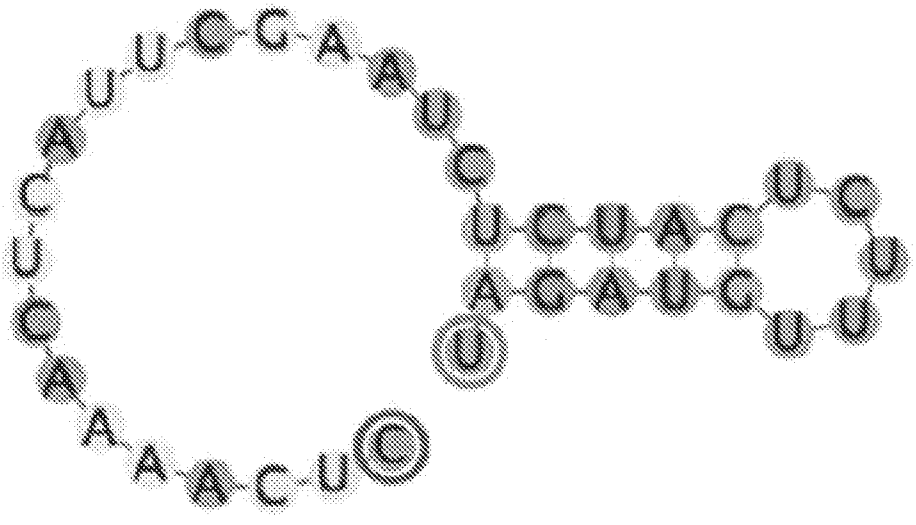
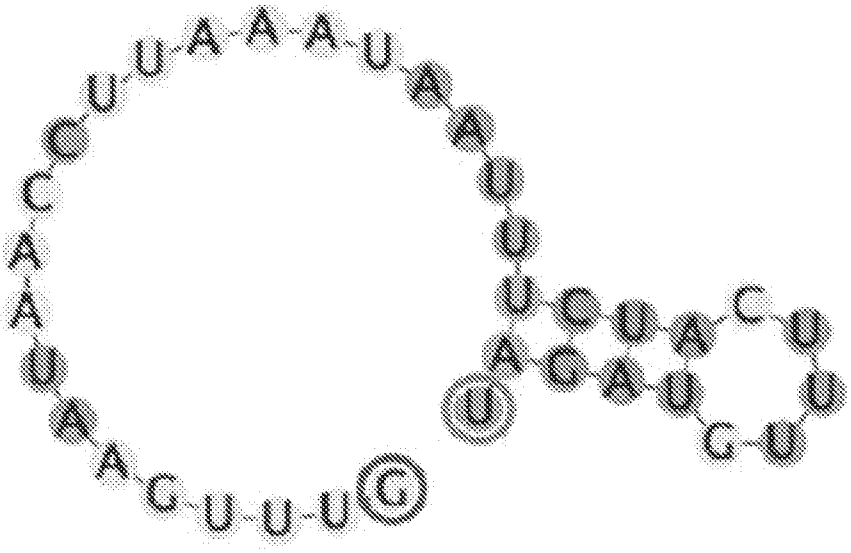


FIG. 80E

>11 DR
GCUUAGAACAUUUAAAGAAUUUCUACUAUUGUAGAU



>12 DR
GUCAAAAGACCUUUUUUAAUUUCUACUCUUGUAGAU

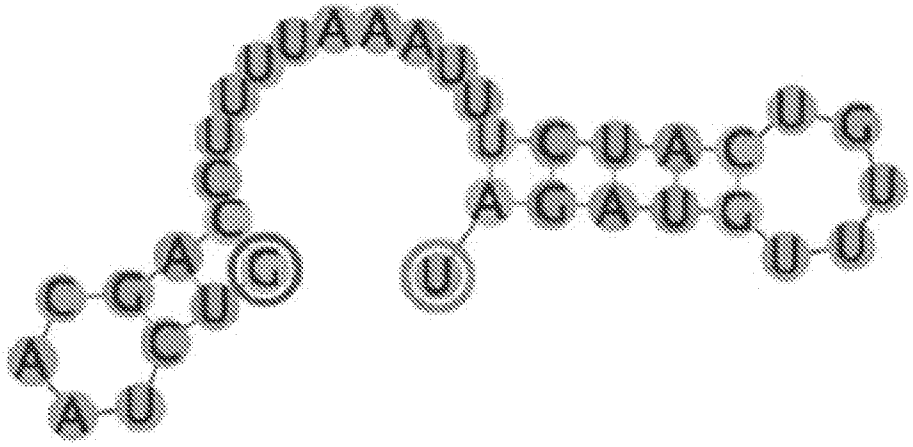
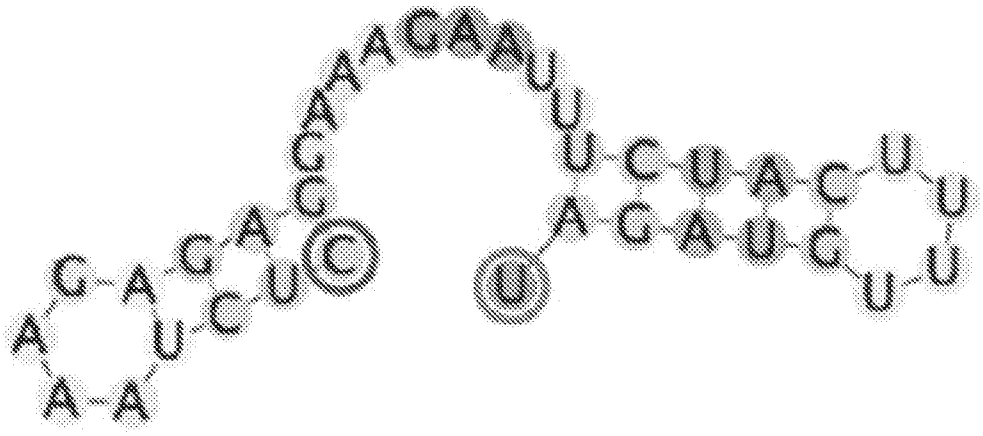


FIG. 80F

>13 DR
GUUUCAAUCCACGCGCCCACGCGGGGCGCGAC



>14 DR
GCCAAAUACCUCUAUAAAAUUUCUACUUUUGUAGAU

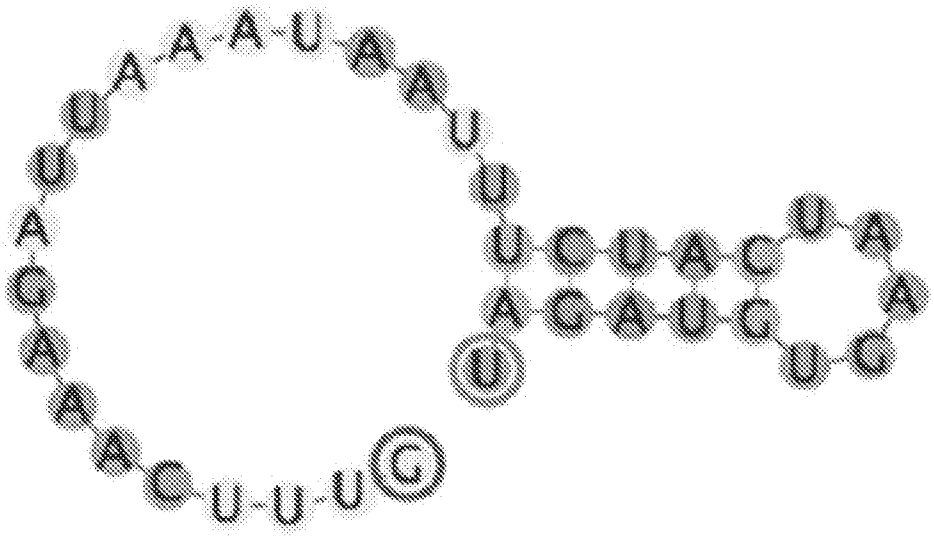
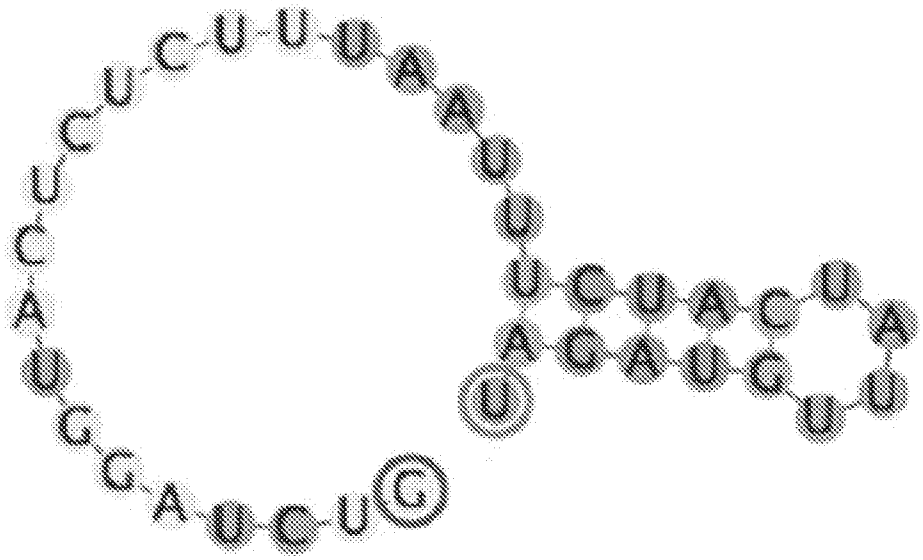


FIG. 80G

>15 DR
GUUUAAAAGUCCUAUUGGAUUUCUACUUUUGUAGAU



>16 DR
GUUUUAGAACCUUAAAAAUUACCUAGUAAUUAGGU

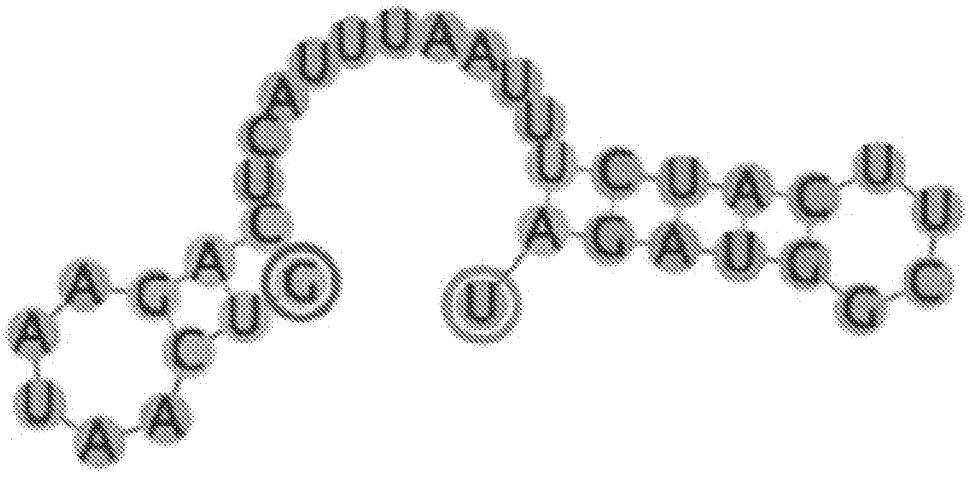


FIG. 80H

>17 DR
GUUUUGGAGUACCUUAGAAAUGCAUGGUUCUCAUGC

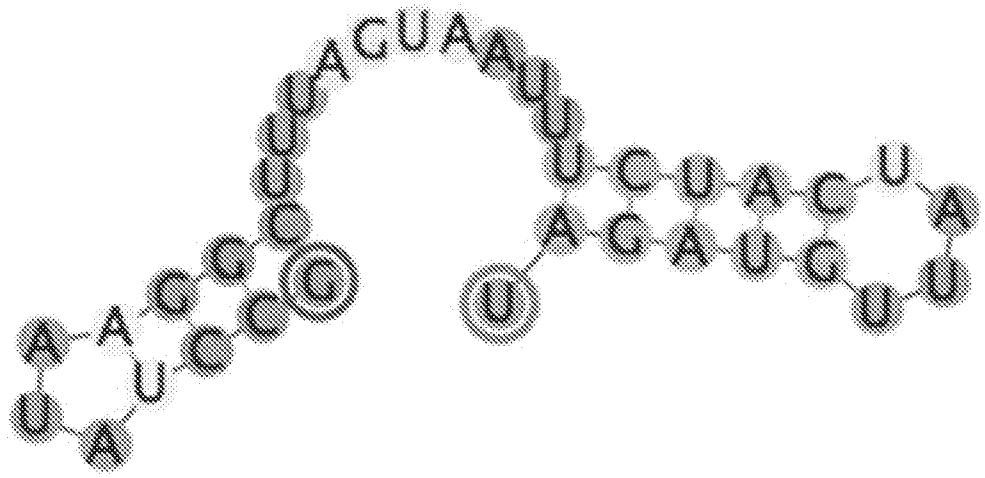
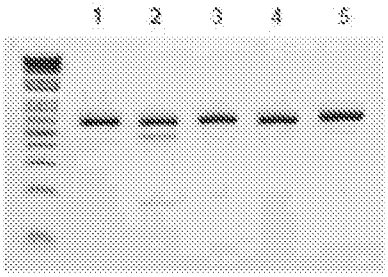


FIG. 80I



Calle	ARNcr	sustrato de ADN
1	Emx1 crRNA1	EMX1 amplicon
2	Emx1 crRNA2	EMX1 amplicon
3	Emx1 crRNA10	EMX1 amplicon
4	Emx1 crRNA13	EMX1 amplicon
5	ninguno	EMX1 amplicon

Secuencias de ARNcr (5´a 3´)

ARNcr1



UAAAU

ARNcr2



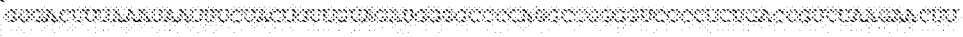
UAAAU

ARNcr10



UAAAU

ARNcr13



UAAAU

Secuencia de amplicón EMX1:

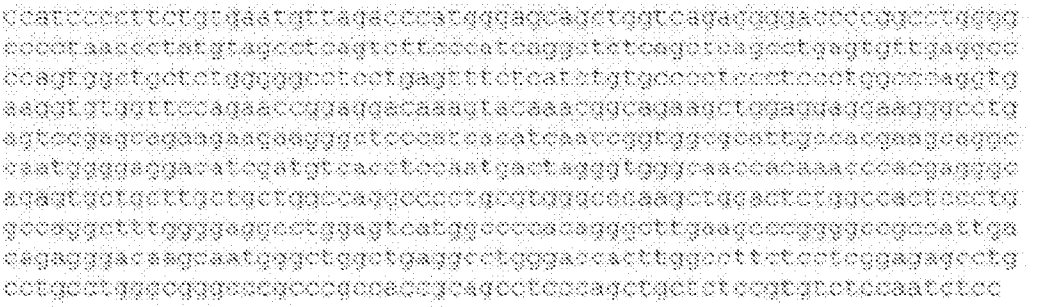


FIG. 81

Calle	ADN	ARNcr
1	pUC10-Espaciador1 (TTC)	deltaDR1
2	pUC10-Espaciador1 (TTC)	deltaDR2
3	pUC10-Espaciador1 (TTC)	deltaDR3
4	pUC10-Espaciador1 (TTC)	deltaDR4
5	pUC10-Espaciador1 (TTC)	deltaDR5

Secuencias de ARNcr (5' a 3'):

deltaDR1
GUUUUAGAGCUGUGUUGUUCUCCAUAUUGUCAGCAUUCACACAGGCCGAUUGUGAAAGUGCCGUUGCAAACAGGCAUAGCAAGUUGAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU

FIG. 82A

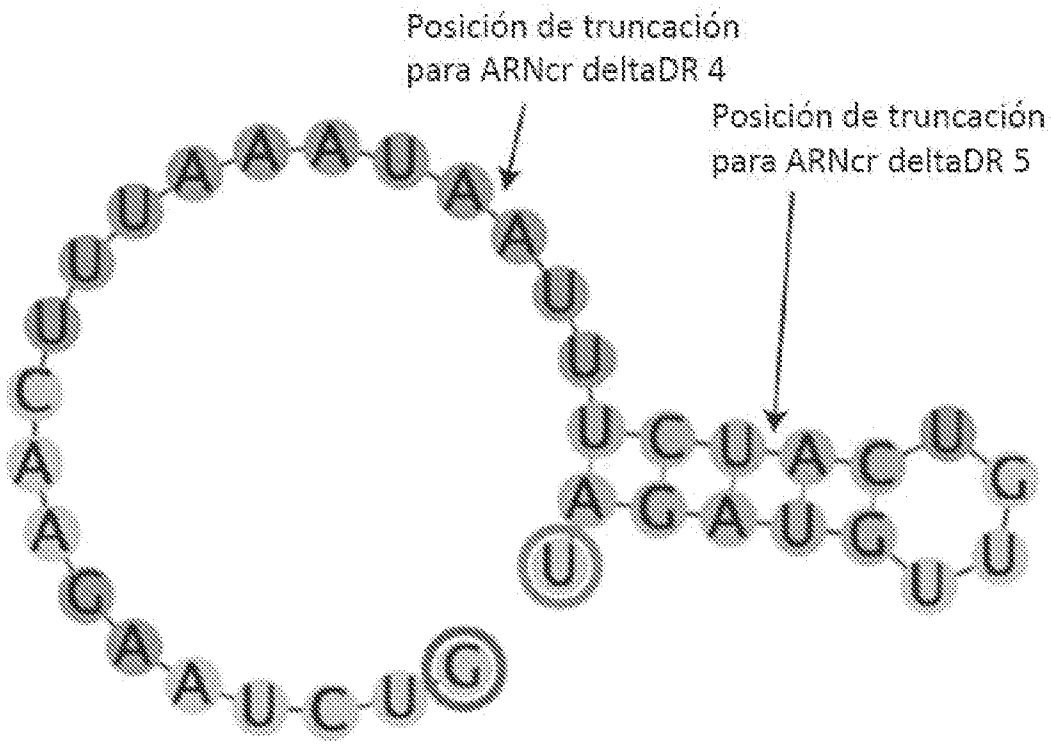
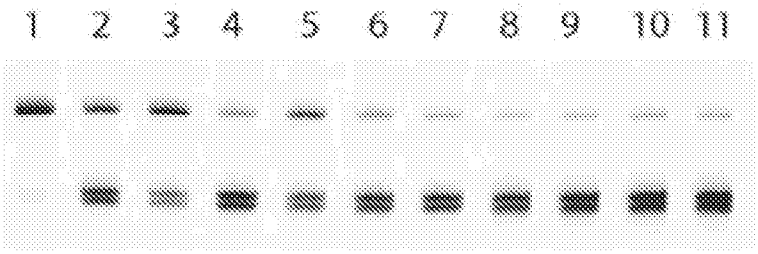

Posición de truncación
para ARNcr deltaDR 4
Posición de truncación
para ARNcr deltaDR 5

FIG. 82B



Cada calle consiste en lisado de células que contienen Cpf1, pUC19 con protoespaciador TTc, y el correspondiente ARNcr, indicado como 1-11

secuencias de ARNcr (5' a 3', base no coincidente subrayada)

1 GGGACTTTAAATAATTTCTACTGTTGTAGATGAAAGTCATTTAATAAGCCACTGTTAAAAGTCTAAGAACTTTAAAT

2 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAGTCATTTAATAAGCCACTGTTAAAAGTCTAAGAACTTTAAAT

3 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGCATTTAATAAGCCACTGTTAAAAGTCTAAGAACTTTAAAT

4 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCAGTTAATAAGCCACTGTTTAAAAGTCTAAGAACTTTAAAT

5 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTGATAAGCCACTGTTAAAAGTCTAAGAACTTTAAAT

6 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATGAGCCACTGTTAAAAGTCTAAGAACTTTAAAT

7 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGCCACTGTTAAAAGTCTAAGAACTTTAAAT

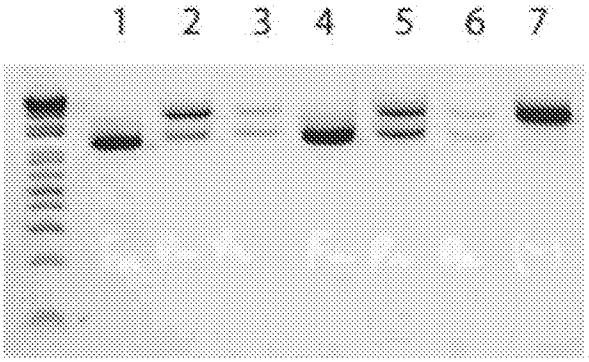
8 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGCTGCTGTTAAAAGTCTAAGAACTTTAAAT

9 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGCCACTGTTAAAAGTCTAAGAACTTTAAAT

10 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGCCACTGTTTAAAAGTCTAAGAACTTTAAAT

11 GGGACTTTAAATAATTTCTACTGTTGTAGATGAGAAGTCATTTAATAAGCCACTGTTTAAAGTCTAAGAACTTTAAAT

FIG. 83



Calle	Sol. de proteína purificada	ARNcr
	ADN sustrato	
1	FnCpf1	espaciador1- ARNcr de direccionamiento
	pUC con protoespaciador TTc	
2	PaCpf1	espaciador1- ARNcr de direccionamiento
	pUC con protoespaciador TTc	
3	PaCpf1	espaciador1- ARNcr de direccionamiento
	pUC con protoespaciador TTc	
4	GST~FnCpf1	espaciador1- ARNcr de direccionamiento
	pUC con protoespaciador TTc	
5	GST~PaCpf1	espaciador1- ARNcr de direccionamiento
	pUC con protoespaciador TTc	
6	MBP~PaCpf1	espaciador1- ARNcr de direccionamiento
	pUC con protoespaciador TTc	
7	ninguno	espaciador1- ARNcr de direccionamiento
	pUC con protoespaciador TTc	

espaciador 1- secuencia de ARNcr de direccionamiento (5' a 3')

acuuuaaaaaauuucuacuguuguagaugagaacucauuaauaaggcacuguuaaaaguauaagaacu uaaaau

FIG. 84

Estructura secundaria de DR de FnCpf1 (tallo y bucle resaltados):

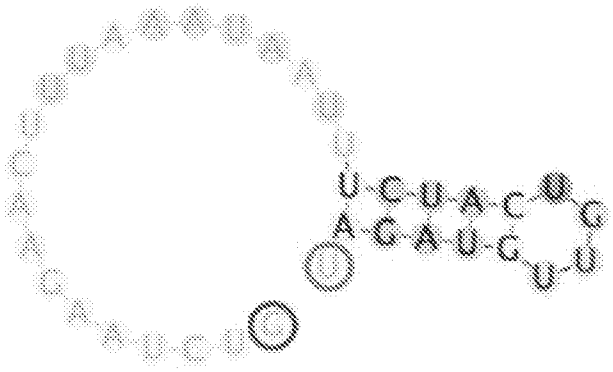


FIG. 85A

Estructura secundaria de DR de PaCpf1 (tallo y bucle resaltados, idénticos excepto por una diferencia de base individual en la región del bucle):

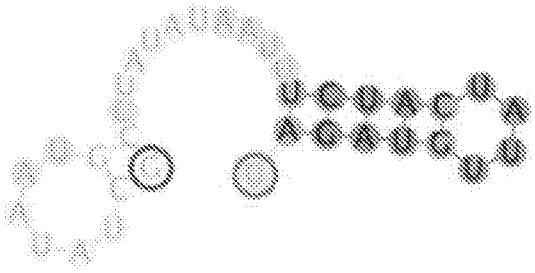


FIG. 85B

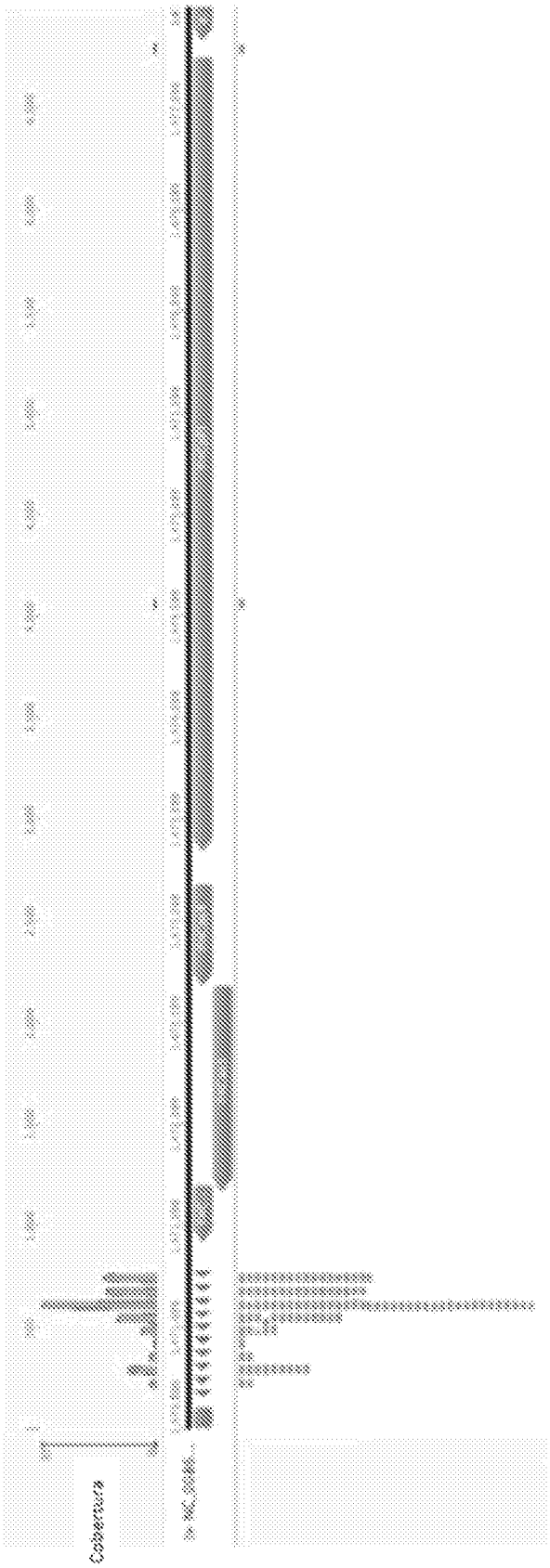


FIG. 66

Las secuencias de ARNcr maduros para FnCpf1 tienen la forma de :

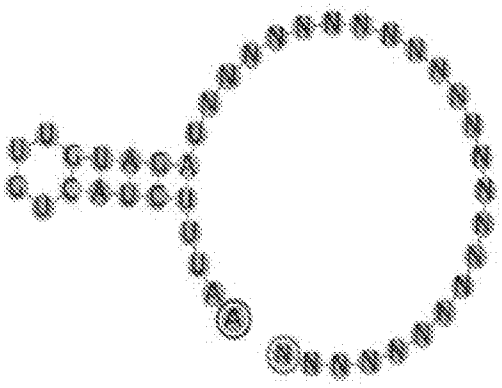


FIG. 87A

Las secuencias de ARNcr maduros para PaCpf1 tienen la forma de :

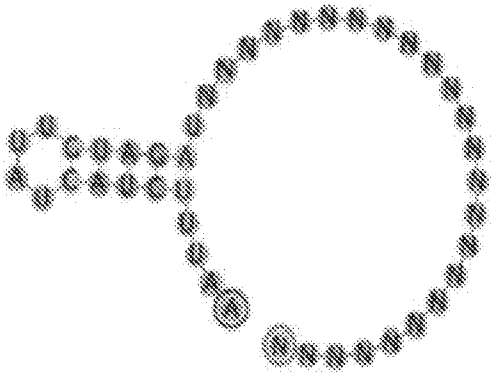


FIG. 87B

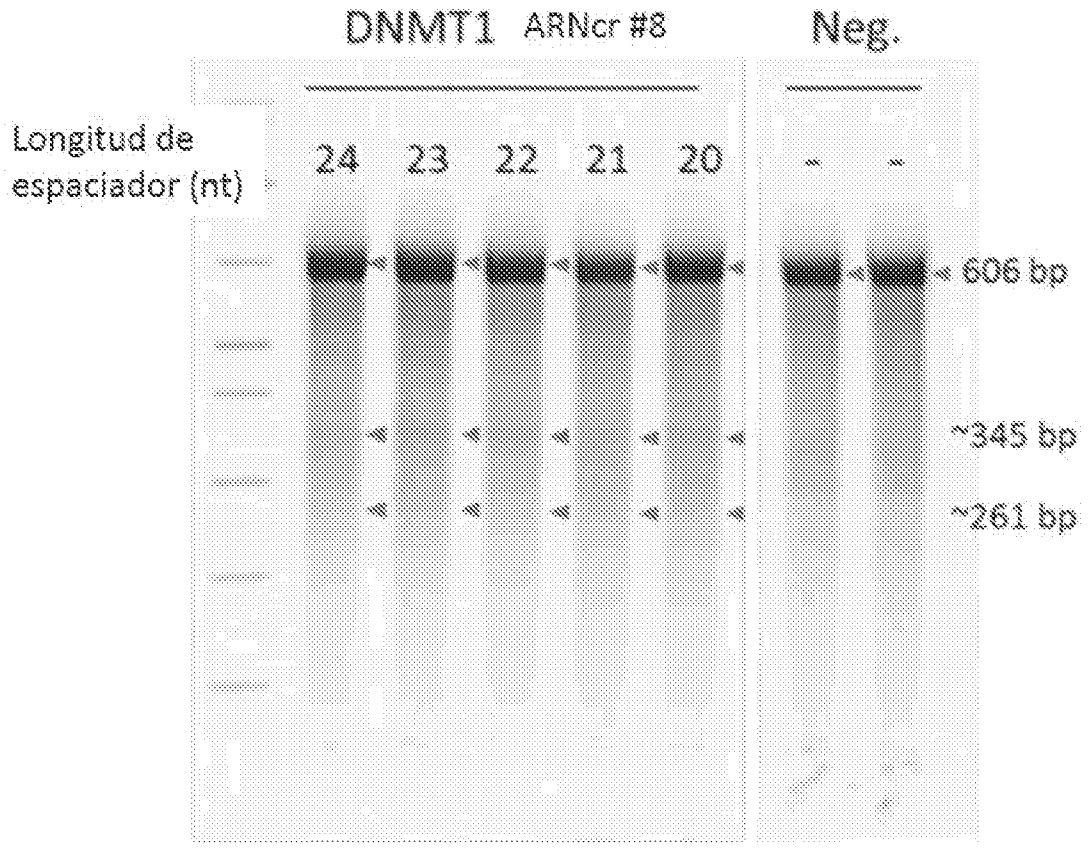

DNMT1 ARNcr #8
Neg.
Longitud de espaciador (nt)
24
23
22
21
20
-
-
606 bp
~345 bp
~261 bp

FIG. 88

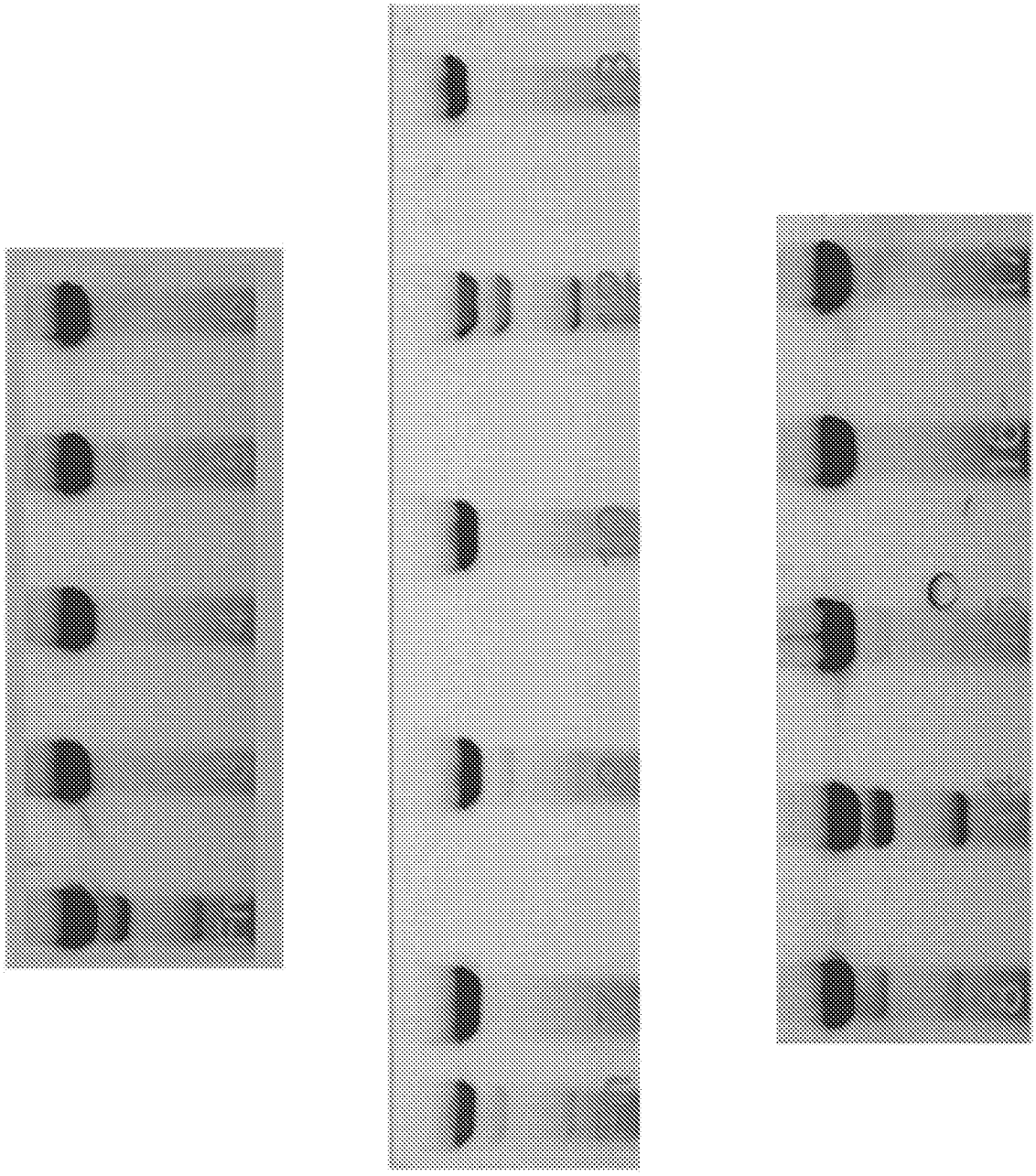


FIG. 69

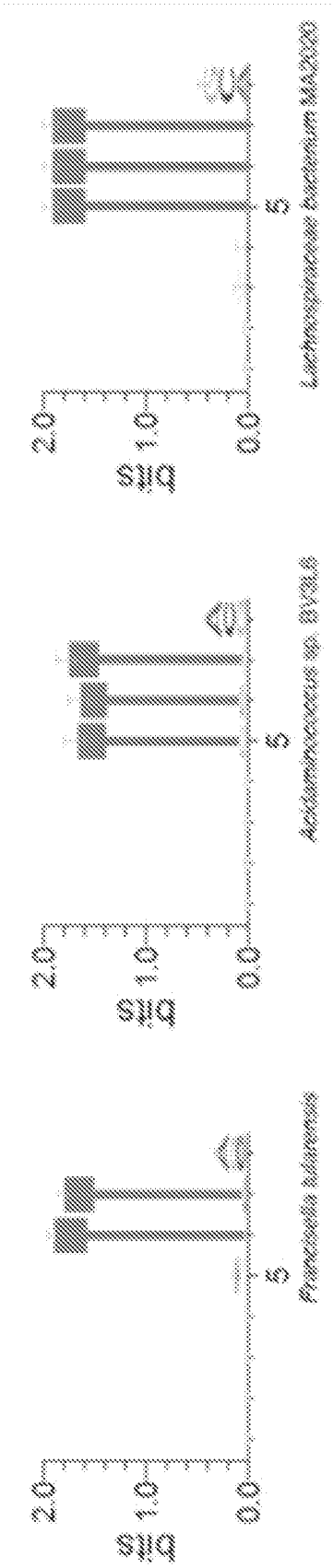

bits
2.0
1.0
0.0
5
Francisella tularensis
bits
2.0
1.0
0.0
5
Actinomyces sp. oral
bits
2.0
1.0
0.0
5
Leptospirillum ferrooxidans

FIG. 90A

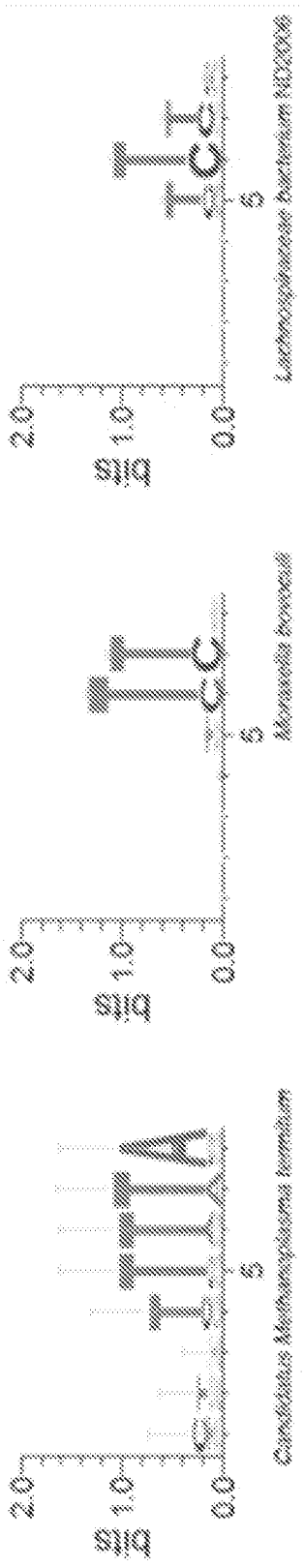

bits
2.0
1.0
0.0
5
Candidatus Methanoplasma termitum
bits
2.0
1.0
0.0
5
Mycoplasma bovoculi
bits
2.0
1.0
0.0
5
Leptospirillum ferriphilum ML-04

FIG. 90B

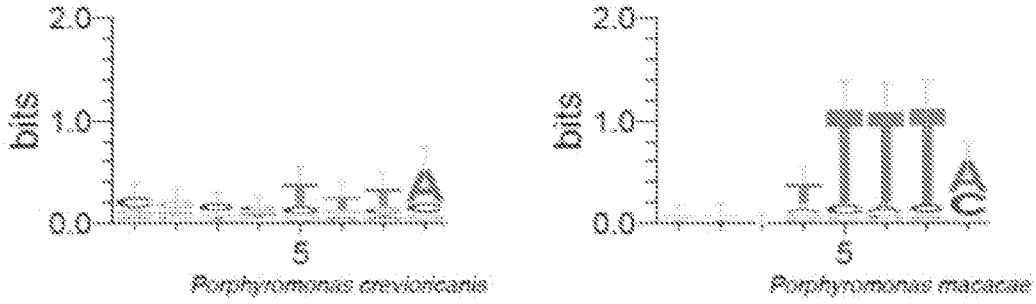

2.0
1.0
0.0
bits
5
Porphyromonas crevioricanis
2.0
1.0
0.0
bits
5
Porphyromonas macacae

FIG. 90C

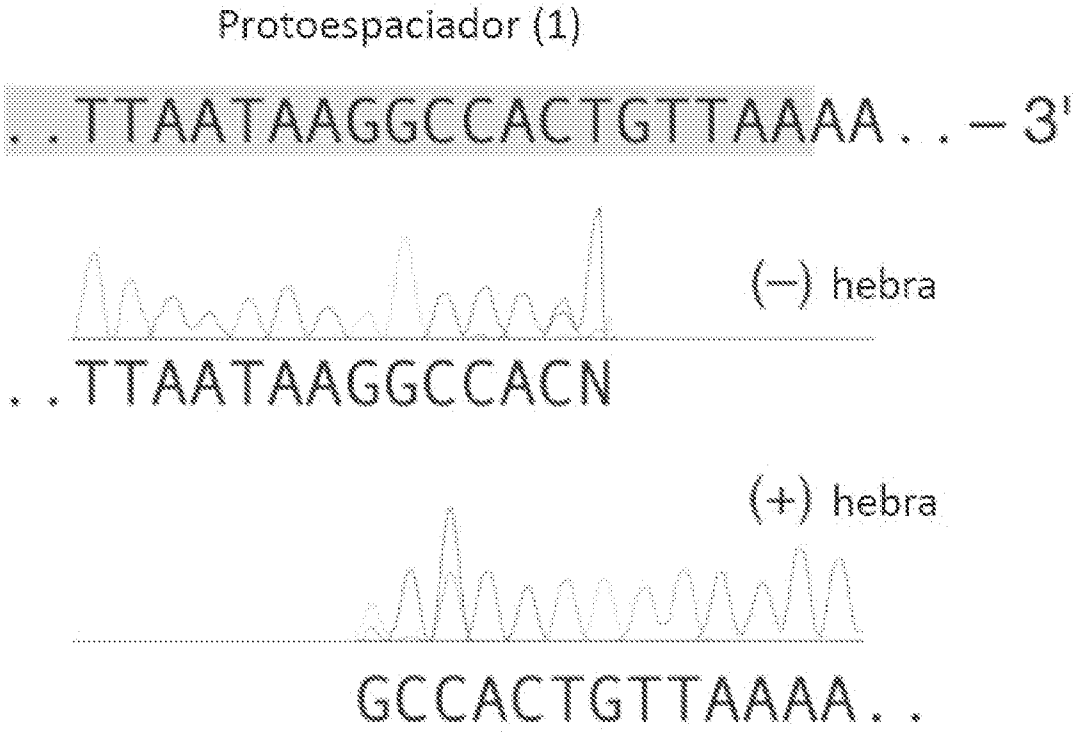

Protoespaciador (1)
. . TTAATAAGGCCACTGTTAAAA . . – 3'
(–) hebra
. . TTAATAAGGCCACN
(+) hebra
GCCACTGTTAAAA . .

FIG. 91

ARNcr
UCAUCUUUAA–5'
GUUUUUUU
UCUUAGAUUAAAGAAGUCAUUUAAUUAAGGCCACU–3'
ARNcr
guías truncadas
1
2
3
4
5
6
7
GAGAAGUCAUUUAAUUAAGGCCA
GAGAAGUCAUUUAAUUAAGGC
GAGAAGUCAUUUAAUUAAG
GAGAAGUCAUUUAAUUAA
GAGAAGUCAUUUAAUUA
GAGAAGUCAUUUAAU
GAGAAGUCAUUUAA
sin cortar
corte
ARNcr
1
2
3
4
5
6
7
ninguno

FIG. 92

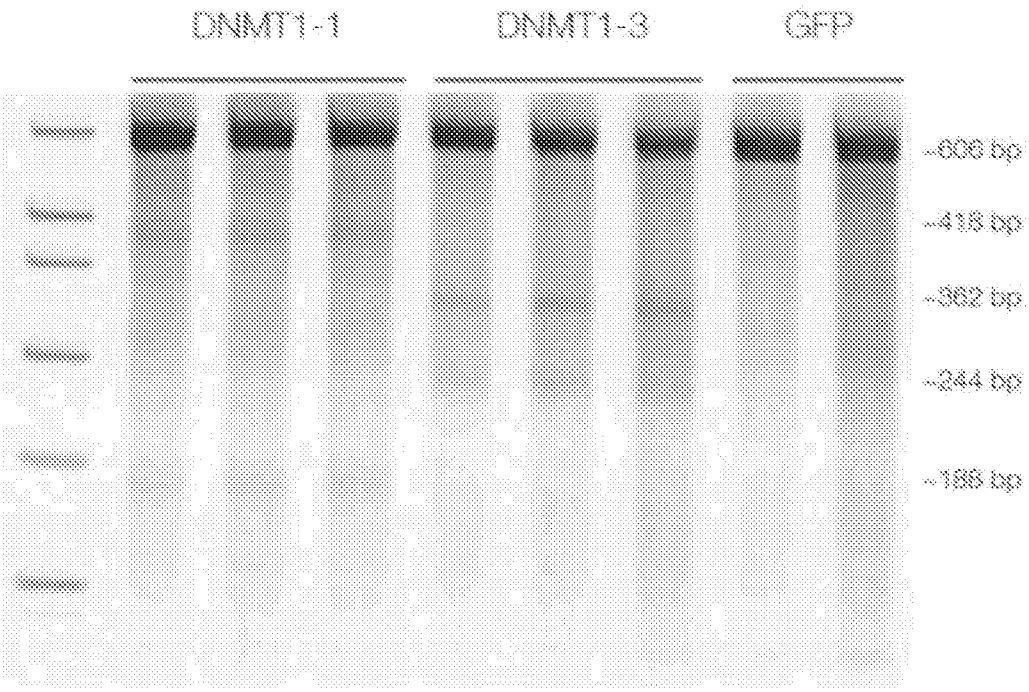


FIG. 93



FIG. 94A

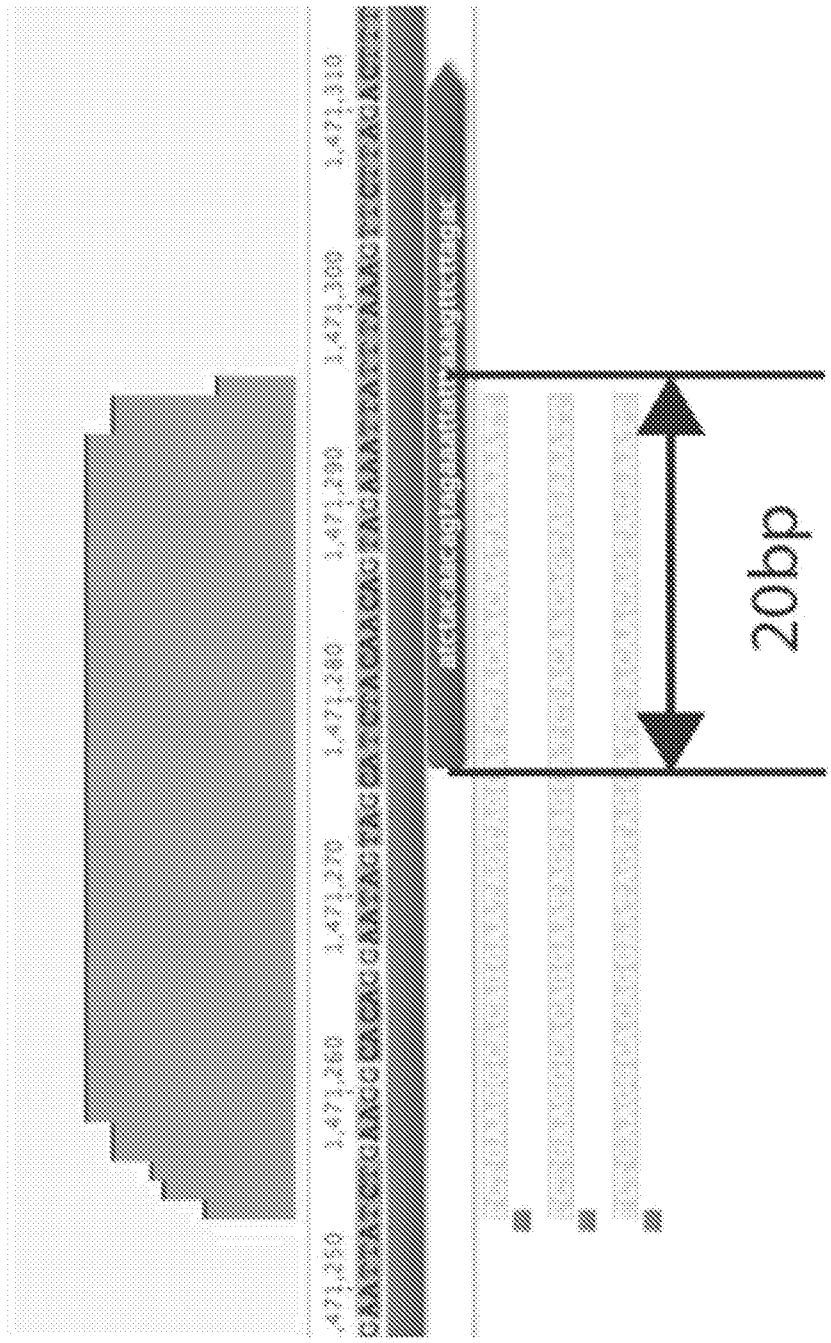


FIG. 94B

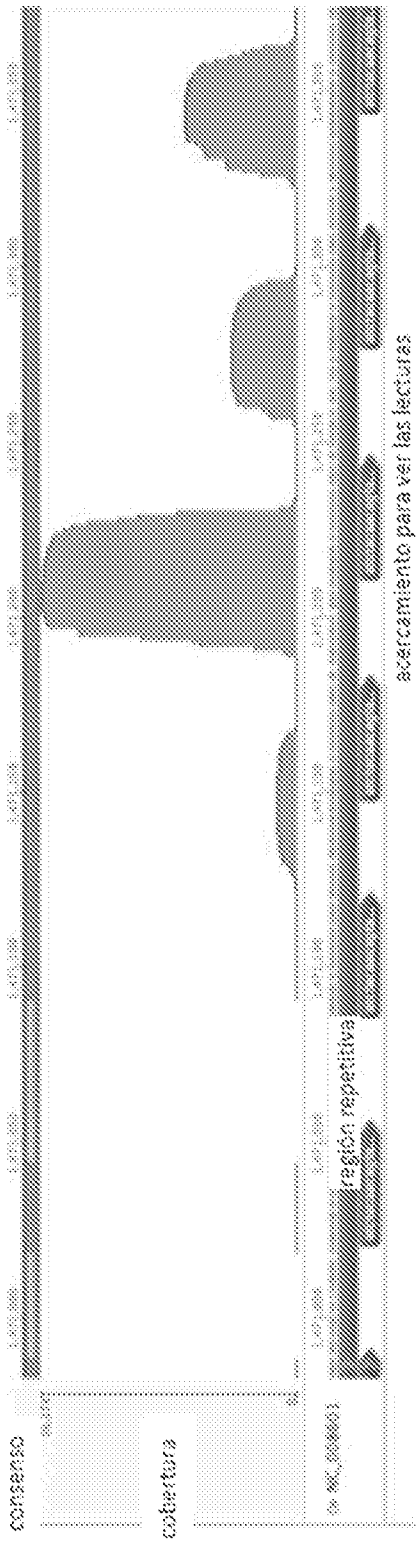

consenso
cobertura
región repetitiva
acercamiento para ver las lecturas

FIG. 94C

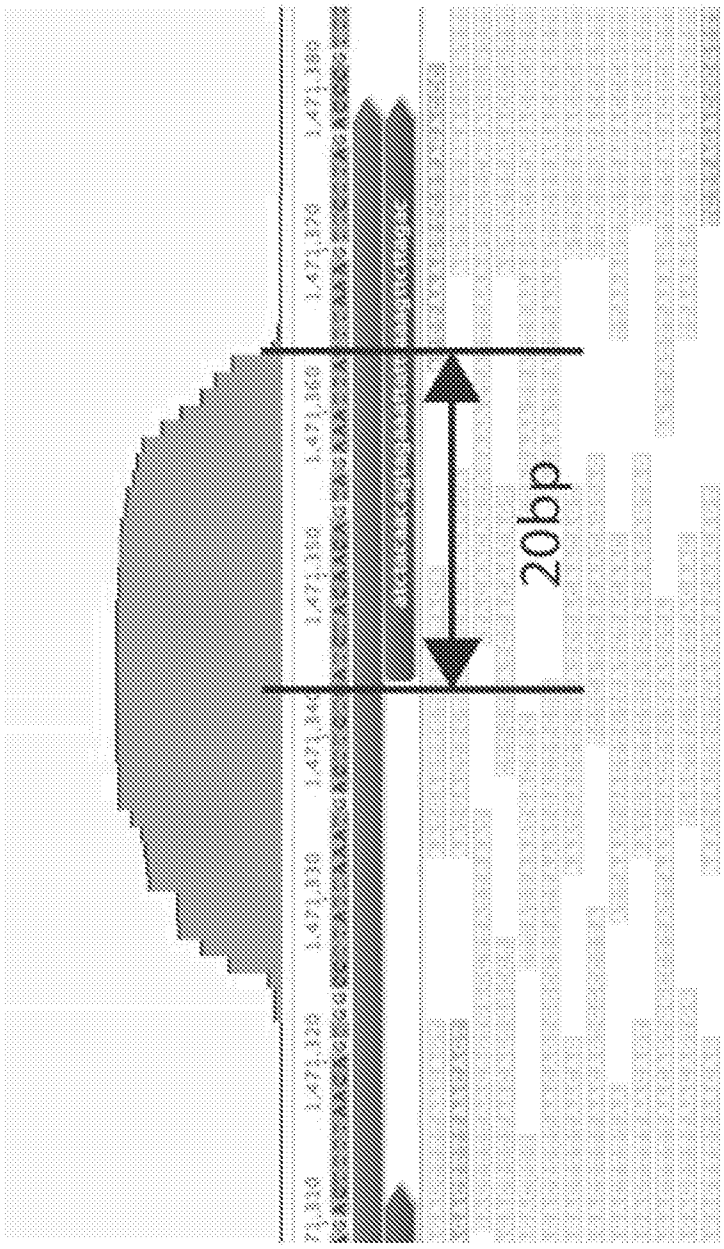


FIG. 84D

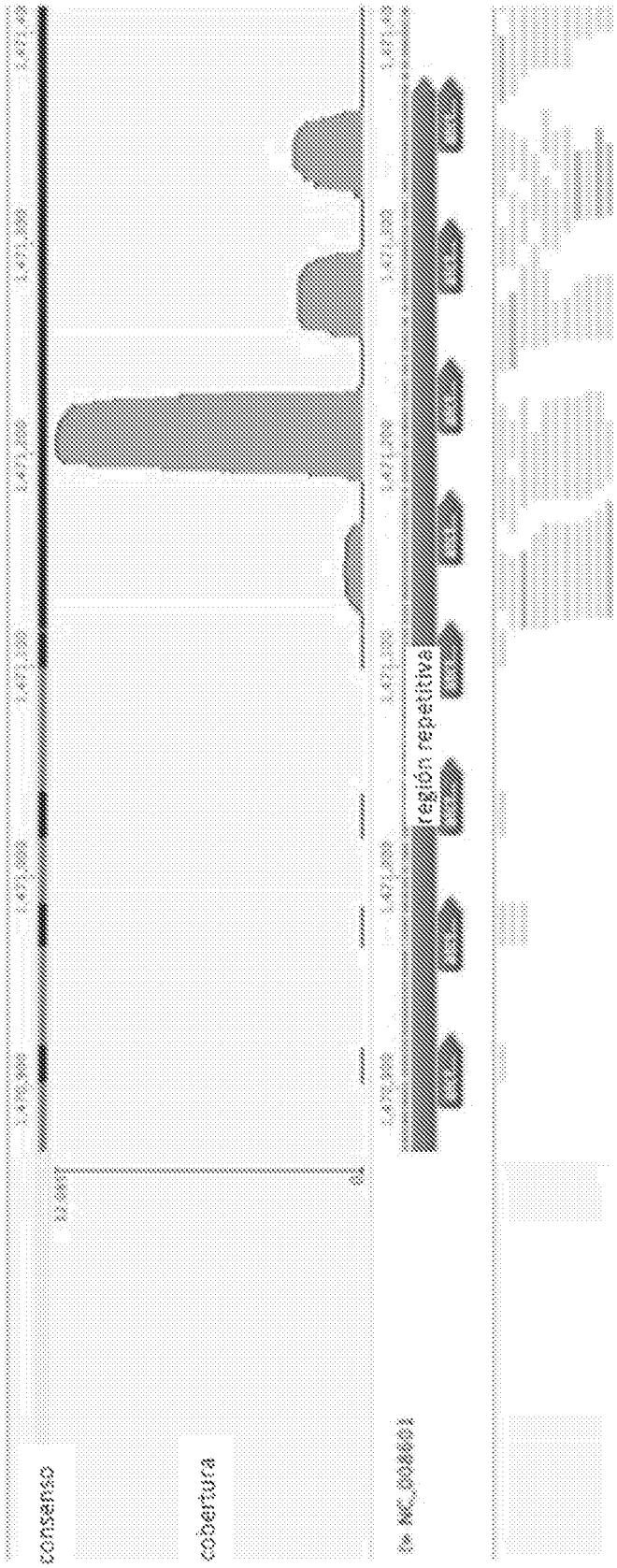
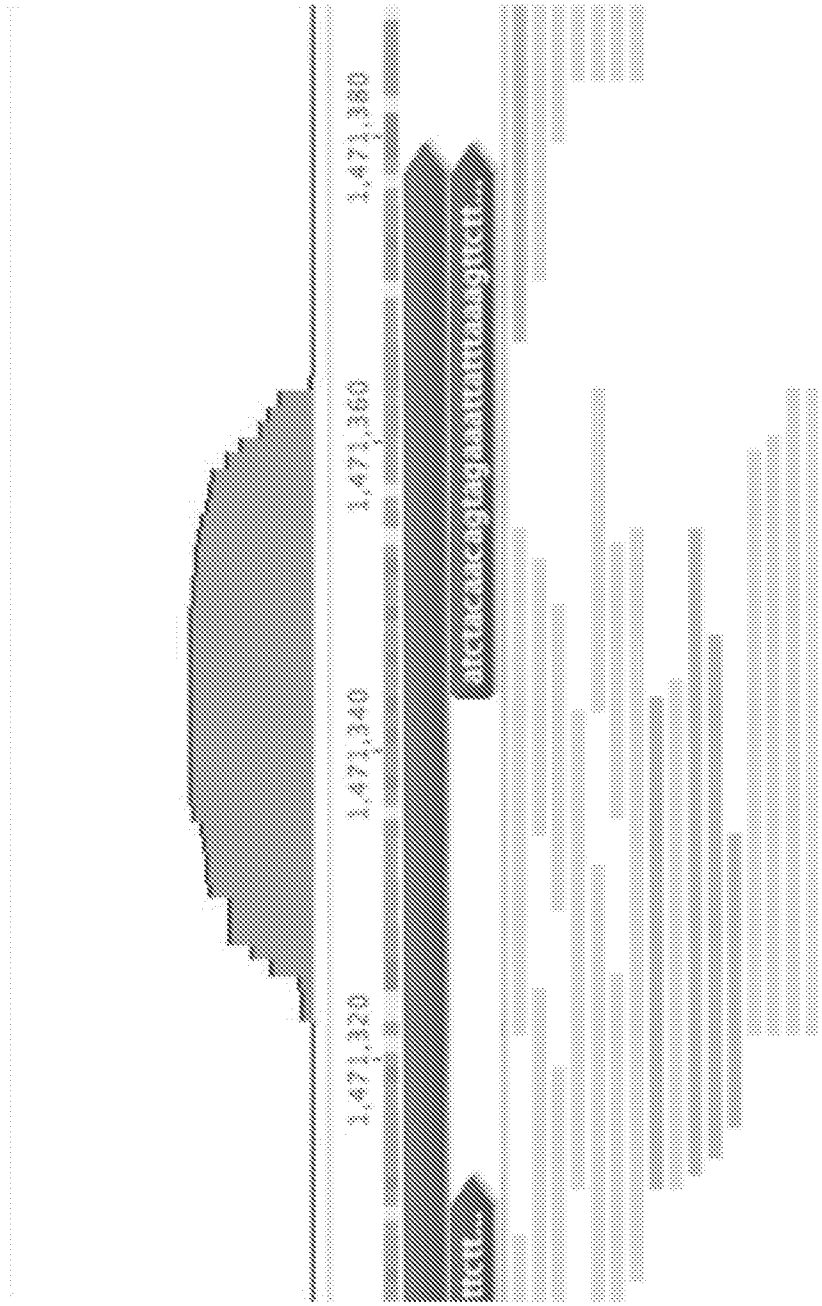


FIG. 94E


1,471,320
1,471,340
1,471,360
1,471,380
HAT1
GTCTTCACCTTT

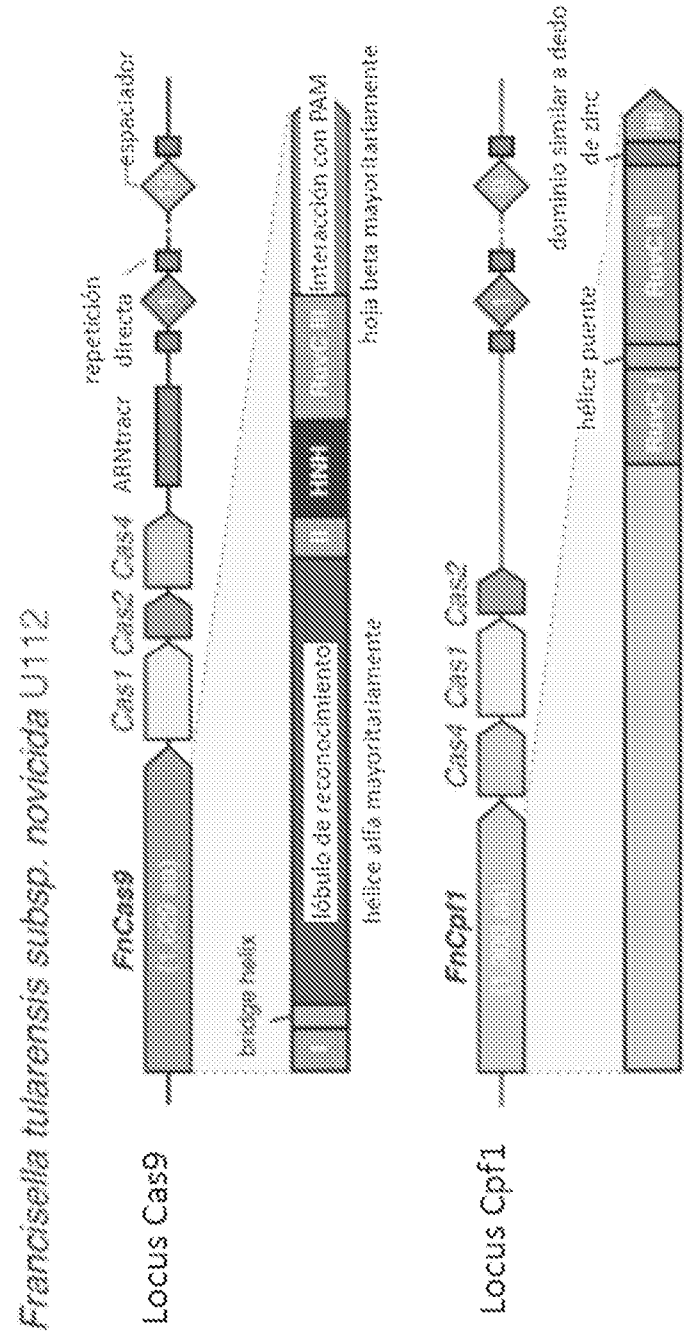

Francisella tularensis subsp. novicida U112
Locus Cas9
FnCas9
Cas1
Cas2
Cas4
ARNtracr
repetición directa
espaciador
bridge helix
lóbulo de reconocimiento
hélice alfa mayoritariamente
interacción con PAM
hoja beta mayoritariamente
Locus Cpf1
FnCpf1
Cas4
Cas1
Cas2
hélice puente
dominio similar a dedo de zinc
alfa/beta mixta

FIG. 95A

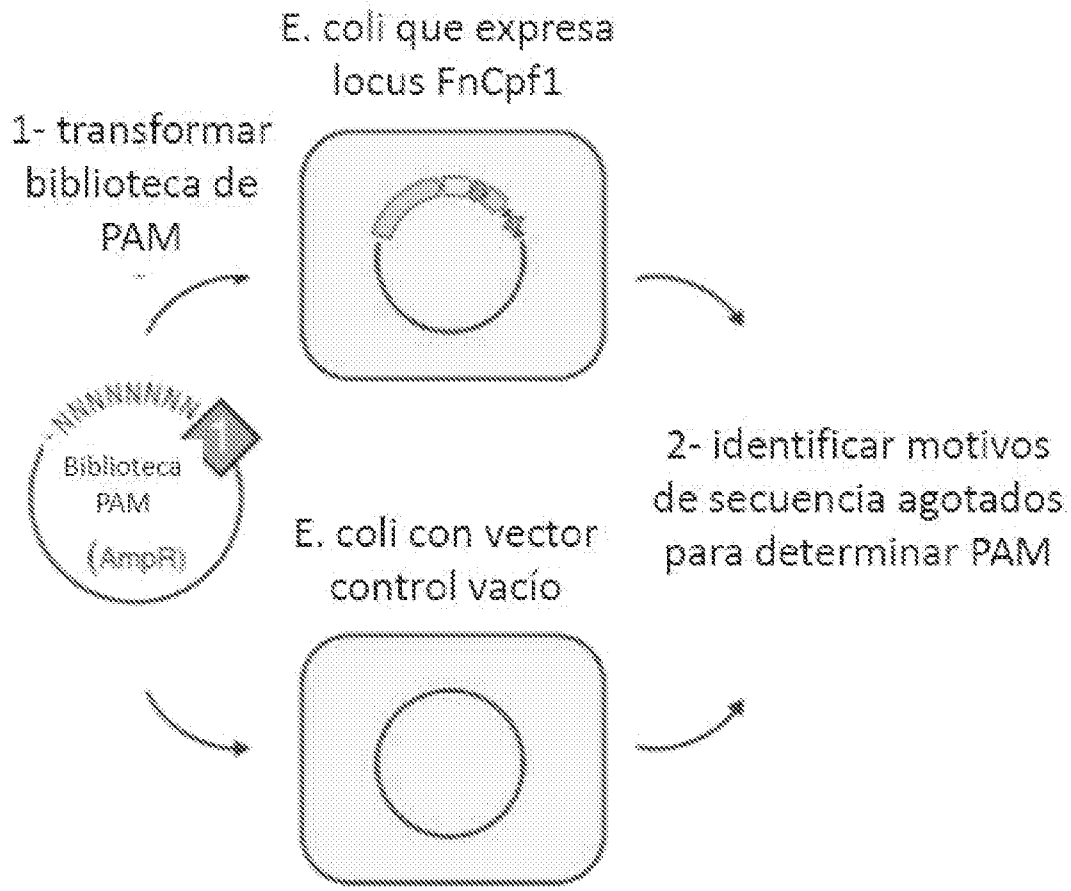

E. coli que expresa locus FnCpf1
1- transformar biblioteca de PAM
NNNNNNNN
Biblioteca PAM
(AmpR)
2- identificar motivos de secuencia agotados para determinar PAM
E. coli con vector control vacío

FIG. 95B

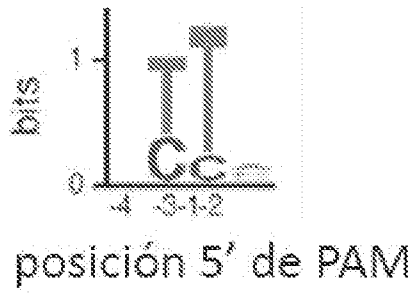


FIG. 95C



FIG. 95D

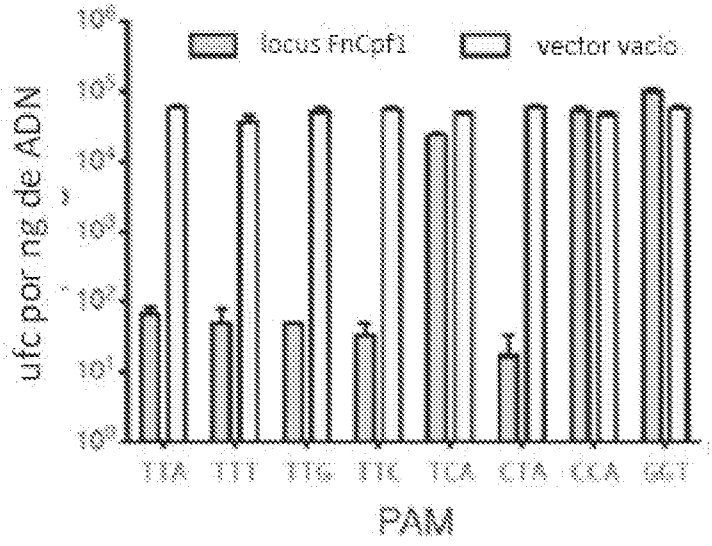


FIG. 95E

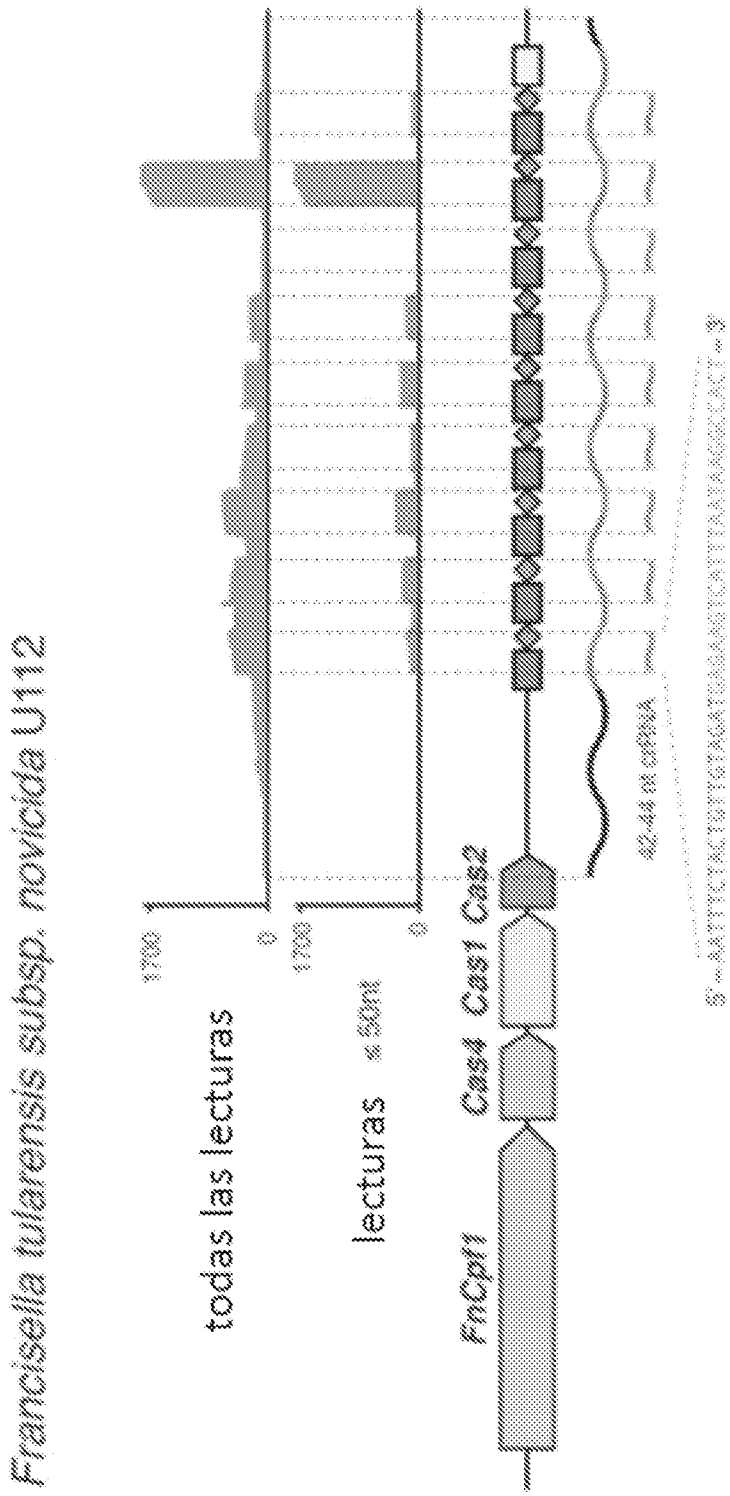

Francisella tularensis subsp. novicida U112
todas las lecturas
1700
1
0
lecturas
1700
1
0
FnCpf1
Cas4
Cas1
Cas2
42.44 m cr#NA
5'-AATTCTACTGTTGTAGATGAAACATTAAAGAAACCACT-3'

FIG. 98A

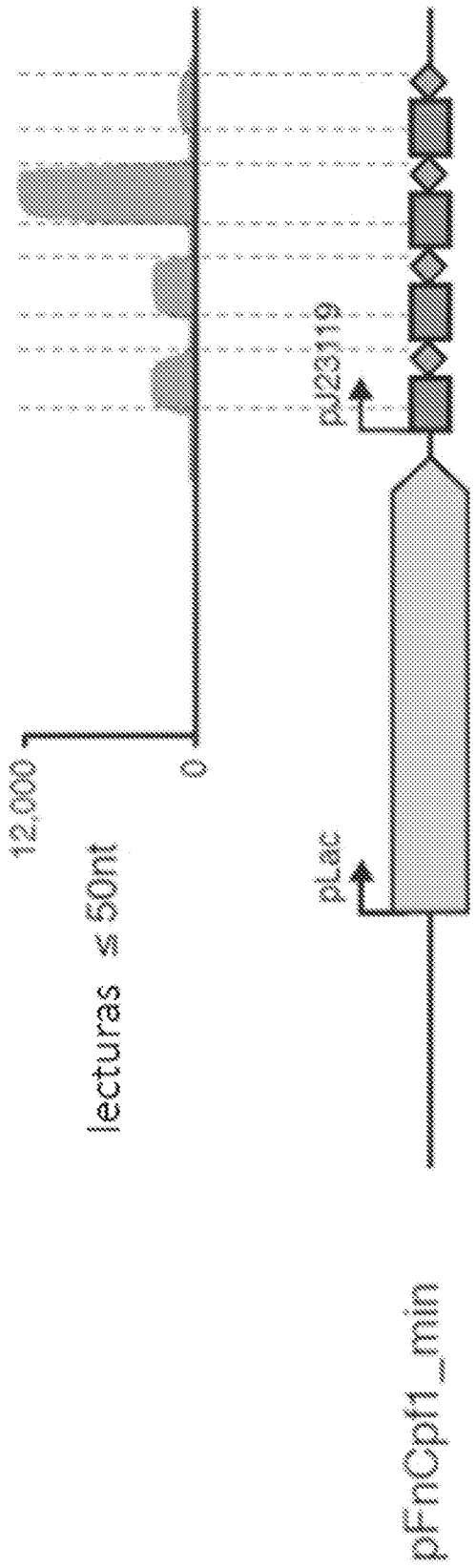


FIG. 96B

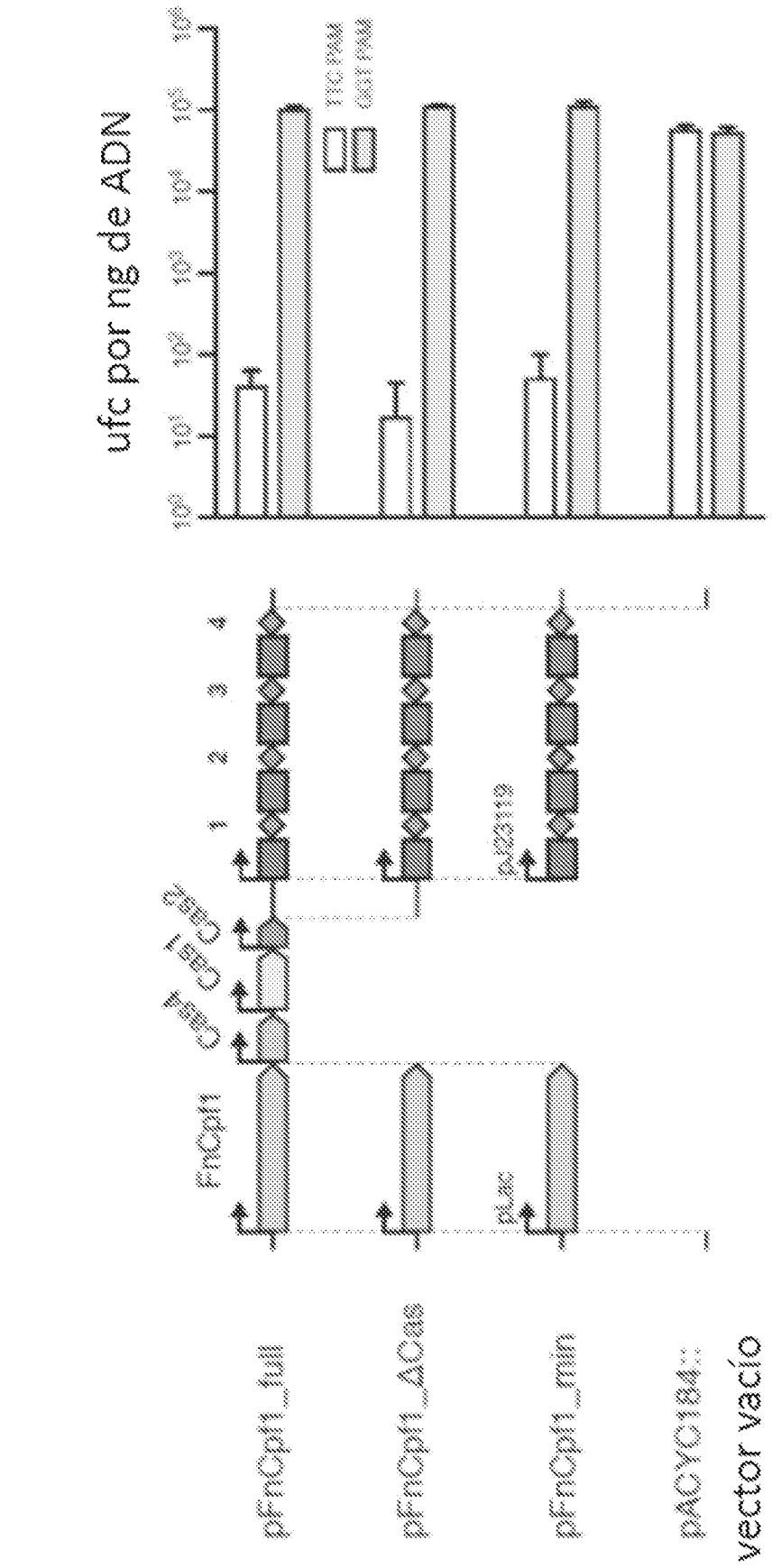

ufc por ng de ADN
TTC PAM
CAT PAM
pFnCpf1_full
pFnCpf1_ΔCas
pFnCpf1_min
pACYC184::
vector vacío
FnCpf1
pLac
ΔR3119
1
2
3
4

FIG. 96C

ADN blanco
PAM
protoespaciador (1)
Cpf1
5' – . . GGTACCCGGGGATCCTTTAGAAGTCATTTAATAGCCACTGTTAAAAAGCTTGGCGTAATCA. . – 3'
3' – . . CCATGGGCCCCTAGGAAATCTCTTCAGTAAATTATTCCGGTGACAATTTTTCGAACCGCATTAGT. . – 5'
ARNcr
U
G
UGUAGAUGAAGUCAUUUAAUAGCCACU – 3'
UCAUCUUUAA – 5'

FIG. 97A

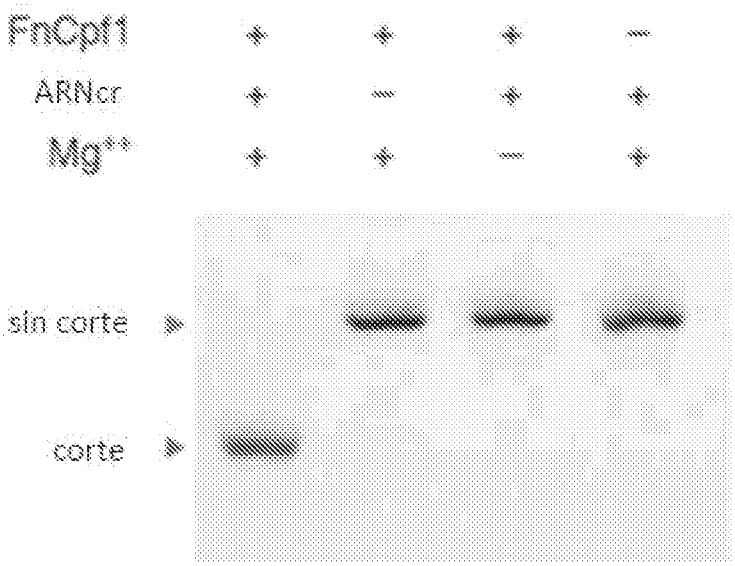


FIG. 97B

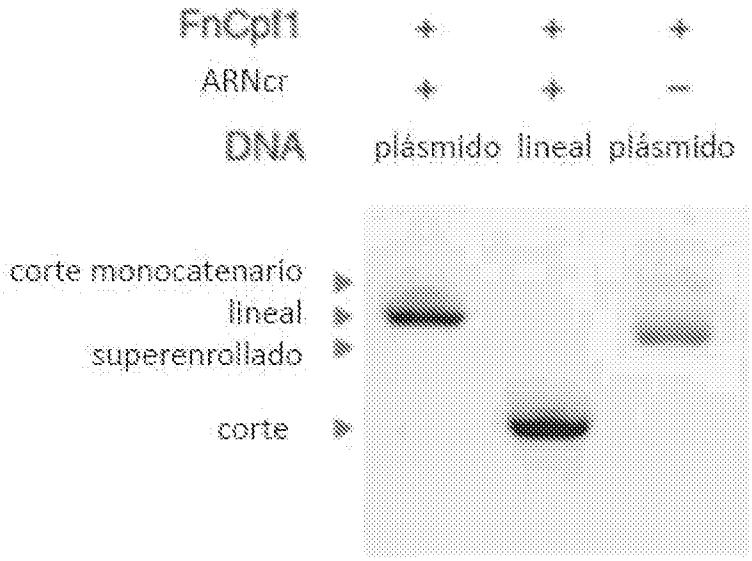


FIG. 97C

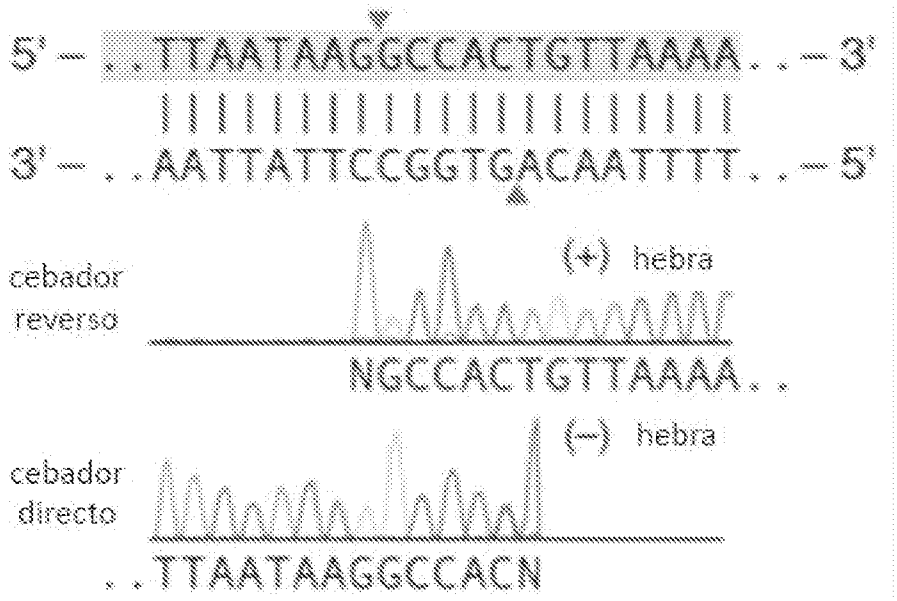

5' – . . TTAATAAGGCCACTGTTAAAA . . – 3'
3' – . . AATTATTCCGGTGACAATTTT . . – 5'
cebador reverso
(+) hebra
NGCCACTGTTAAAA . .
cebador directo
(–) hebra
. . TTAATAAGGCCACN

FIG. 97D

PAM variante
ARNcr
1
2
3
4
5
6
1
2
3
4
5
6
sin corte
corte
1
2
3
4
5'
3'
TTa
AAt
5'
3'
TTa
CCt
AAa
I
AAt
AAa
TTt
CCa
CCt

FIG. 97E

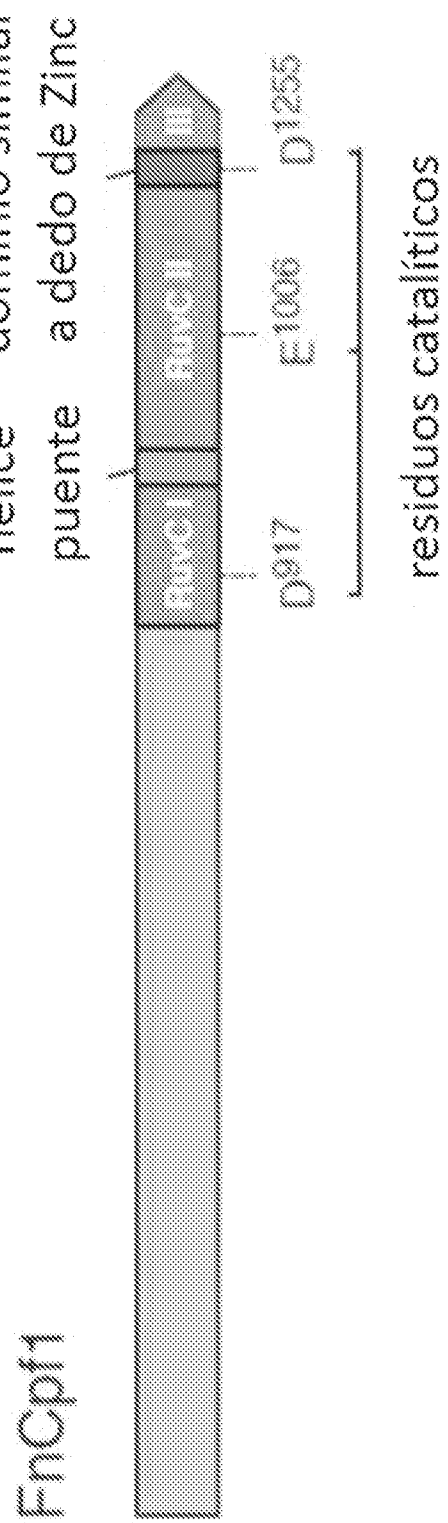

FnCpf1
hélice
puente
dominio similar
a dedo de Zinc
nuc-I
nuc-II
D917
E1006
D1255
residuos catalíticos

FIG. 98A

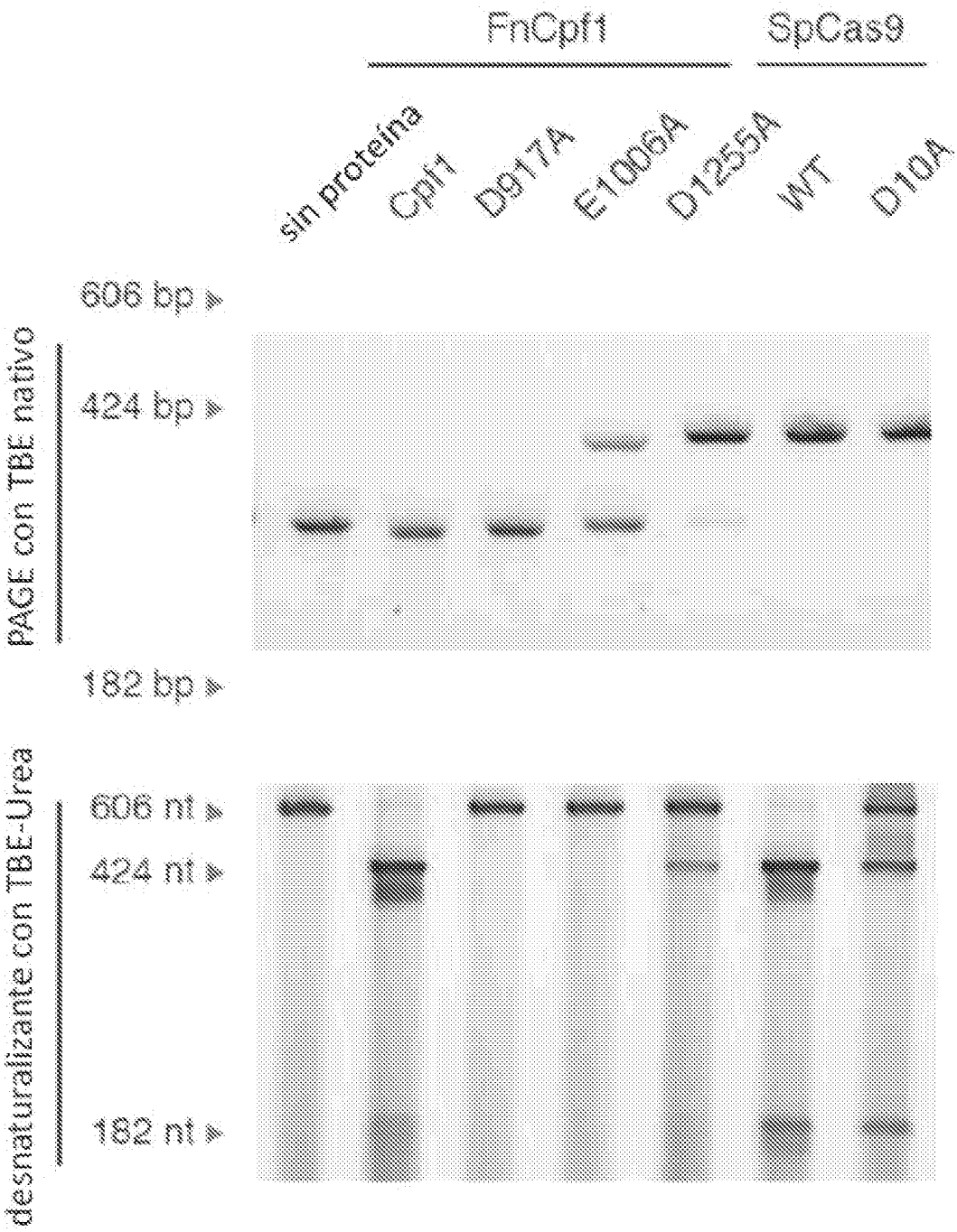
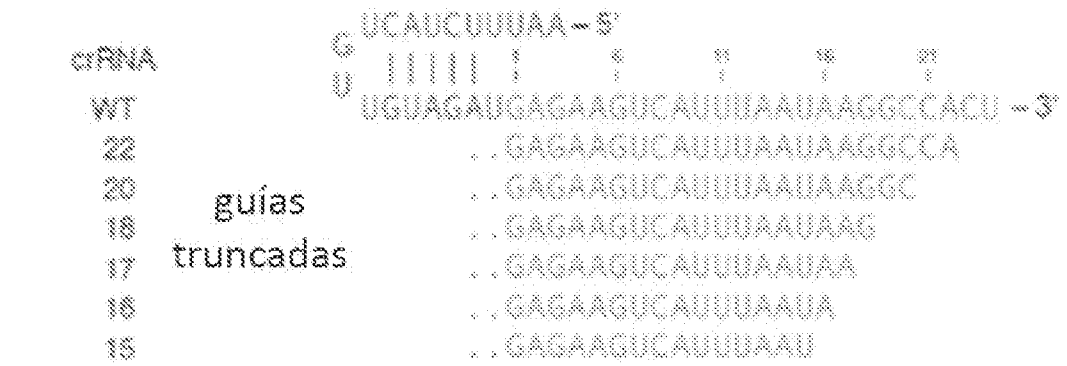

FnCpf1
SpCas9
sin proteína
Cpf1
D917A
E1006A
D1255A
WT
D10A
606 bp
424 bp
182 bp
PAGE con TBE nativo
606 nt
424 nt
182 nt
desnaturalizante con TBE-Urea

FIG. 98B


crRNA
WT
22
20
18
17
16
15
guías
truncadas
UCAUCUUUAA-5'
G
U
UGUAGAUGAGAAGUCAUUUAAUAAGGCCACU-3'
..GAGAAGUCAUUUAAUAAGGCCA
..GAGAAGUCAUUUAAUAAGGC
..GAGAAGUCAUUUAAUAAG
..GAGAAGUCAUUUAAUAA
..GAGAAGUCAUUUAAUA
..GAGAAGUCAUUUAAU

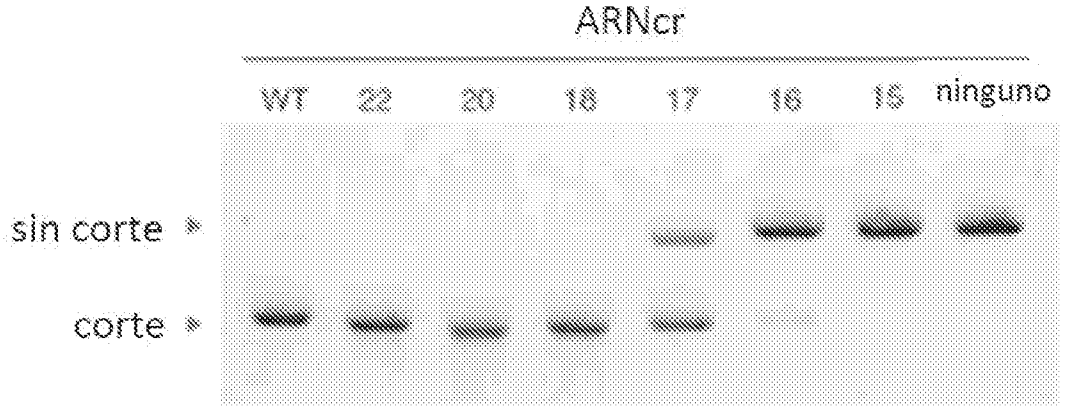
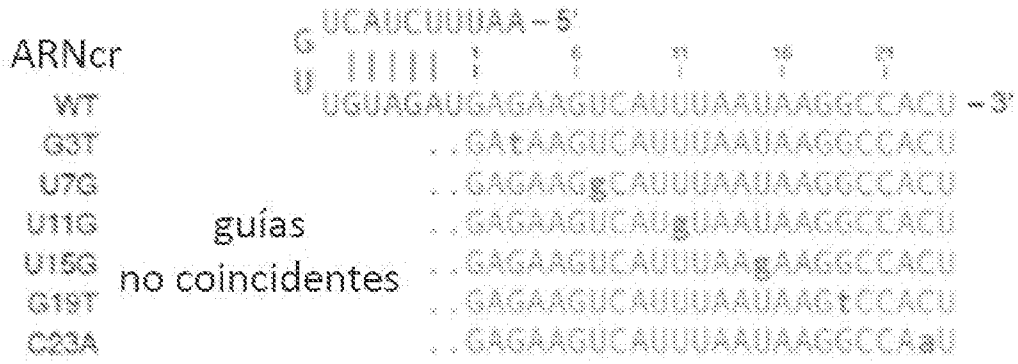

ARNcr
WT
22
20
18
17
16
15
ninguno
sin corte
corte

FIG. 99A


ARNcr
WT
G3T
U7G
U11G
U15G
G19T
C23A
guías
no coincidentes
5' UCAUCUUUAA – 3'
3' UGUAGAUGAGAAGUCAUUUAAUAAGGCCACU – 5'
..GAtAAGUCAUUUAAUAAGGCCACU
..GAGAAGgCAUUUAAUAAGGCCACU
..GAGAAGUCAUgUAAUAAGGCCACU
..GAGAAGUCAUUUAAgAAGGCCACU
..GAGAAGUCAUUUAAUAAGtCCACU
..GAGAAGUCAUUUAAUAAGCCCAgU

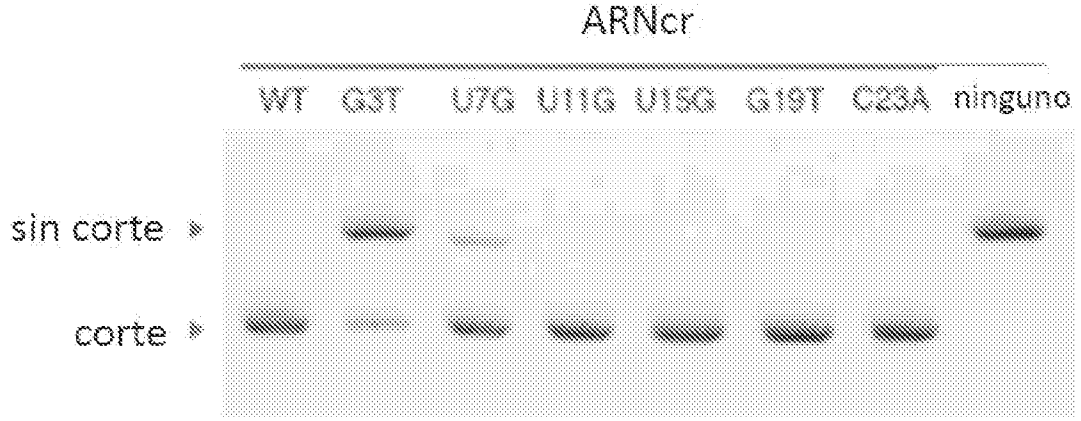
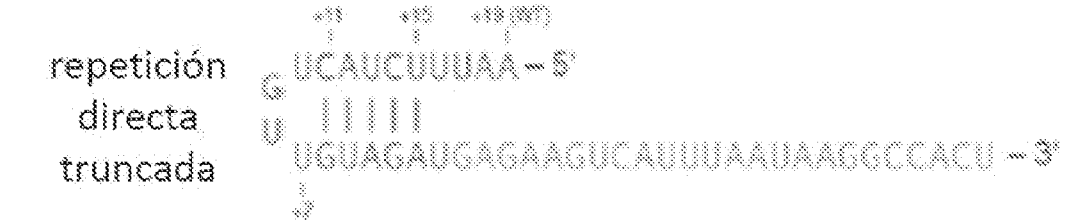

ARNcr
WT
G3T
U7G
U11G
U15G
G19T
C23A
ninguno
sin corte
corte

FIG. 99B


repetición
directa
truncada
UCAUCUUUAA – 5'
UGUAGAUGAGAAGUCAUUUAAUAAGGCCACU – 3'

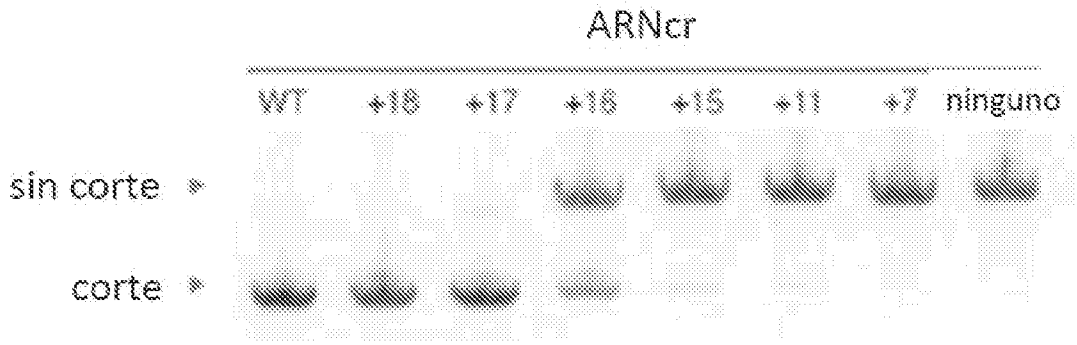
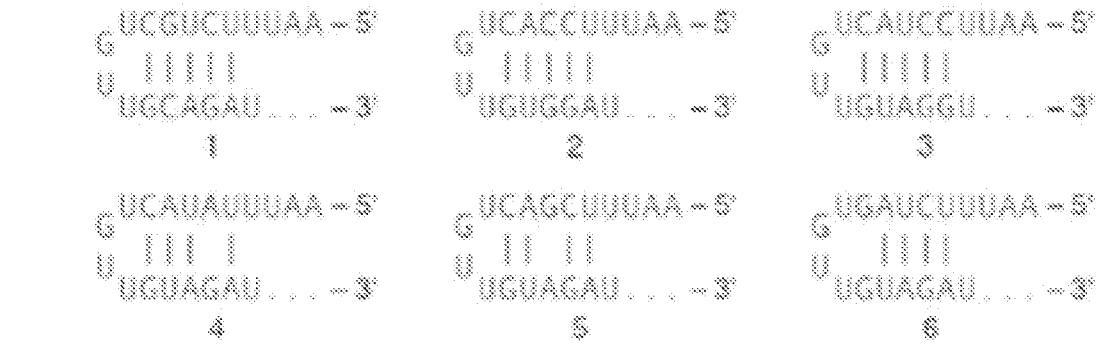

ARNcr
WT
+18
+17
+16
+15
+11
+7
ninguno
sin corte
corte

FIG. 99C


G UCGUCUUUAA – 5'
U UGCAGAU . . . – 3'
1
G UCACCUUUAA – 5'
U UGUGGAU . . . – 3'
2
G UCAUCCUUAA – 5'
U UGUAGGU . . . – 3'
3
G UCAUAUUUAA – 5'
U UGUAGAU . . . – 3'
4
G UCAGCUUUAA – 5'
U UGUAGAU . . . – 3'
5
G UGAUCUUUAA – 5'
U UGUAGAU . . . – 3'
6

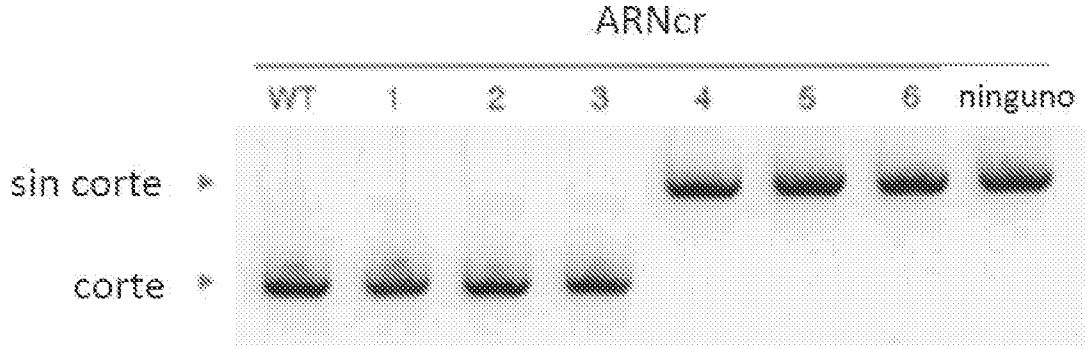

ARNcr
WT
1
2
3
4
5
6
ninguno
sin corte
corte

FIG. 99D

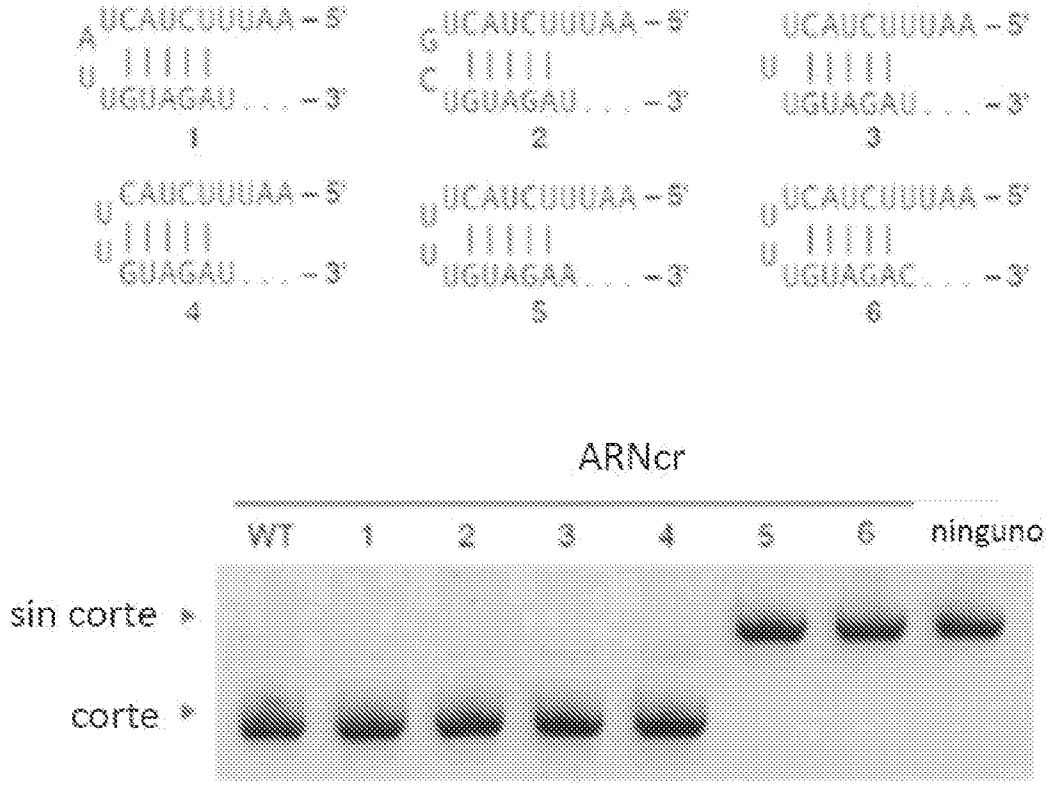

A UCAUCUUUAA – 5'
U | | | | |
UGUAGAU . . . – 3'
1
G UCAUCUUUAA – 5'
C | | | | |
UGUAGAU . . . – 3'
2
UCAUCUUUAA – 5'
U | | | | |
UGUAGAU . . . – 3'
3
U CAUCUUUAA – 5'
U | | | | |
GUAGAU . . . – 3'
4
U UCAUCUUUAA – 5'
U | | | | |
UGUAGAA . . . – 3'
5
U UCAUCUUUAA – 5'
U | | | | |
UGUAGAC . . . – 3'
6
ARNcr
WT
1
2
3
4
5
6
ninguno
sin corte
corte

FIG. 99E

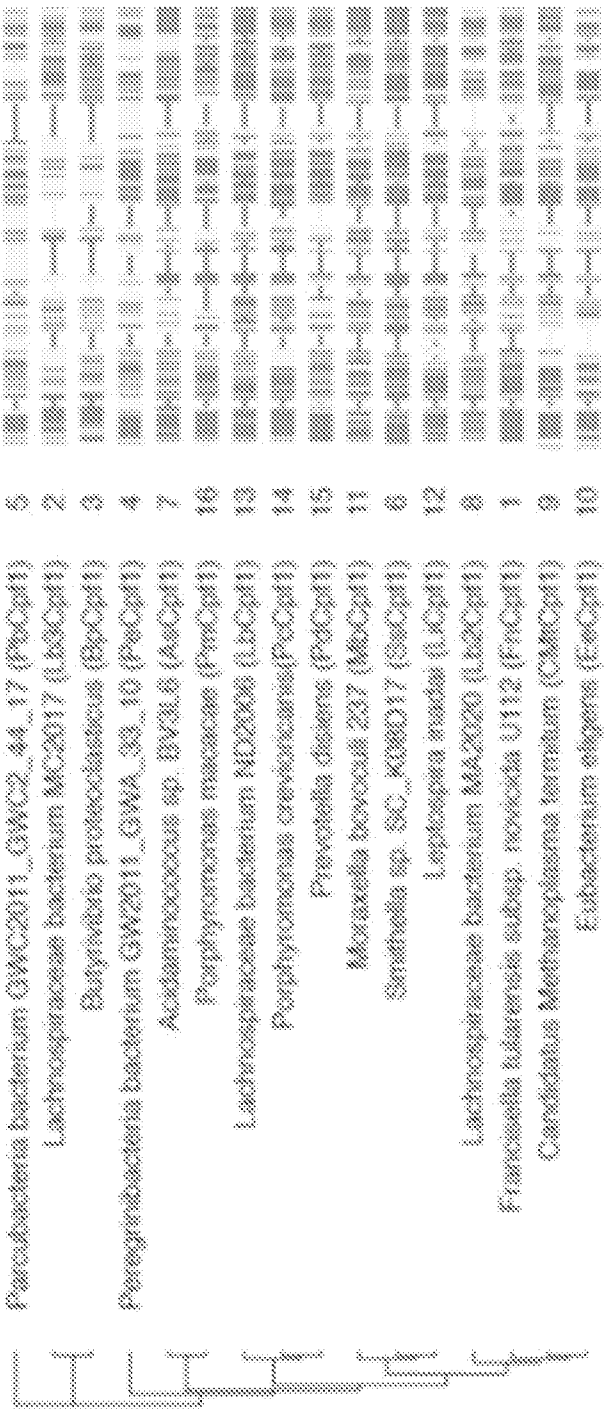

Anomal
ID
Paludisphaera borealis GWC2011_GWC2_44_17 (PboCg41)
Lacipirellulaceae bacterium MC2017 (LbCg41)
Blastopirellula marina (BmCg41)
Pirellulaceae bacterium GWA2_38_10 (PbCg41)
Actinomycetospora sp. DV08 (AsCg41)
Pseudomonas mosselii (PmCg41)
Lacipirellulaceae bacterium MO2000 (LbCg41)
Porphyrobacter cryptocandens(PcCg41)
Planctella desiva (PdCg41)
Monosiga brevicollis 237 (MbCg41)
Sinorhizobium sp. SC_KAM017 (SsCg41)
Leptospira inadai (LiCg41)
Lacipirellulaceae bacterium MA2020 (LbCg41)
Francisella tularensis subsp. novicida U112 (FtCg41)
Candidatus Methanoplasma termitum (CMtCg41)
Eubacterium eligens (EeCg41)
6
2
3
4
7
16
13
14
15
11
8
12
8
1
9
10

FIG. 100A

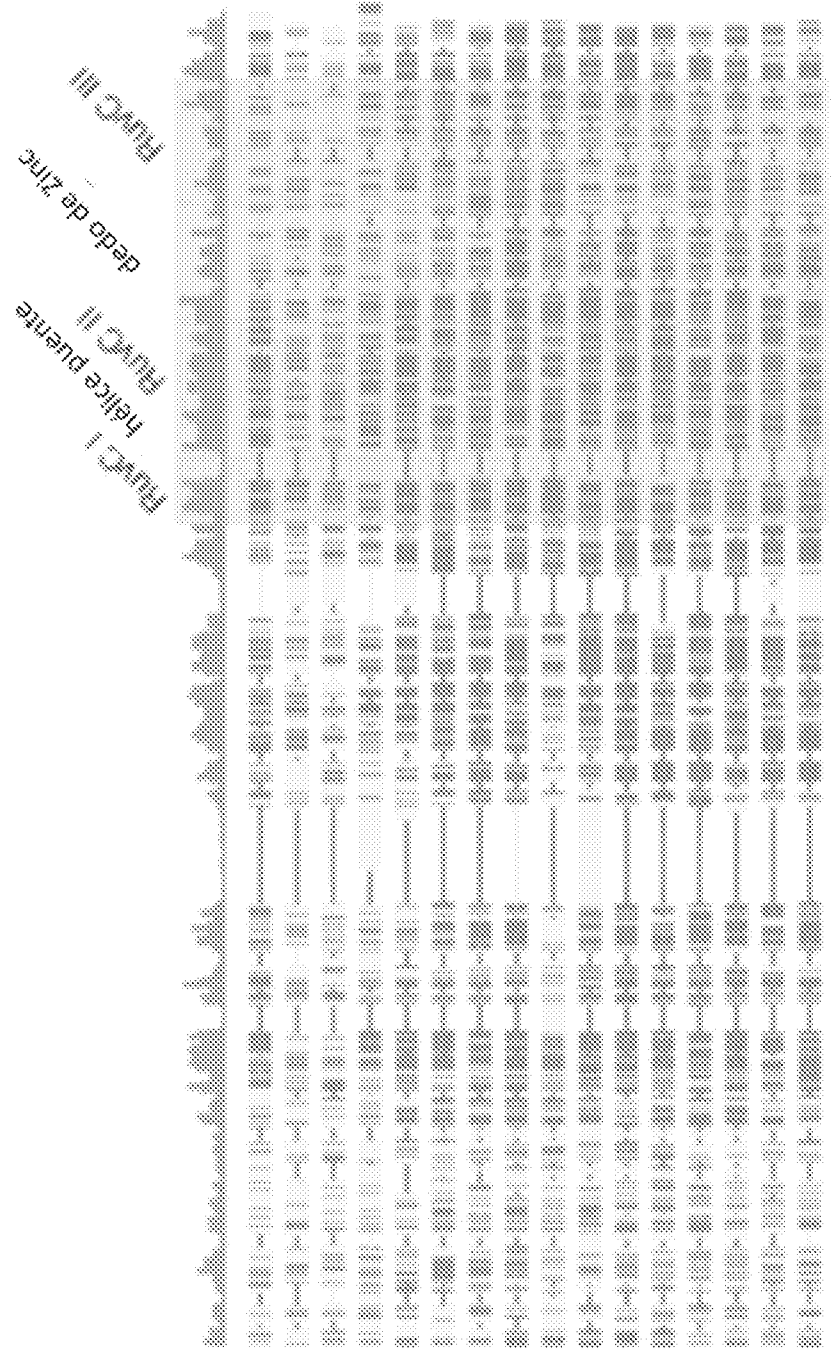


FIG. 100B

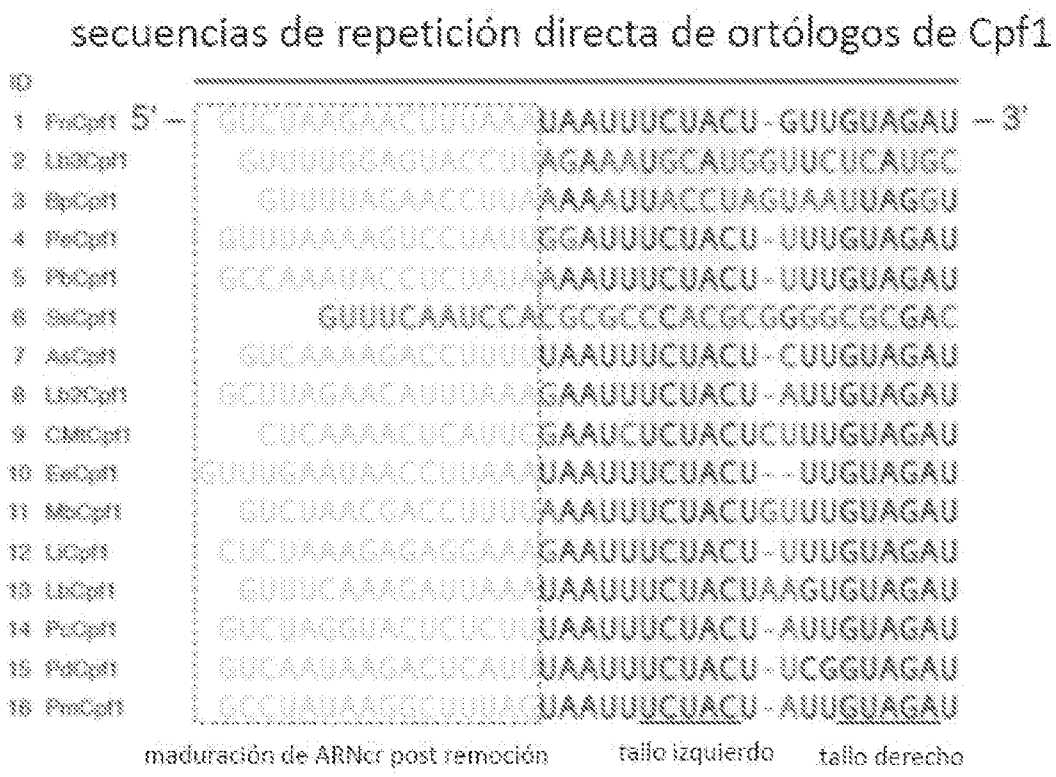

secuencias de repetición directa de ortólogos de Cpf1
ID
1 PreCpf1 5' – GUCUAAGAACUUUAA UAAUUUCUACU-GUUGUAGAU – 3'
2 LbaCpf1 GUUUUGGAGUACCUU AGAAAUGCAUGGUUCUCAUGC
3 BpCpf1 GUUUUAGAACCUUA AAAAUUACCUAGUAAUUAGGU
4 PeCpf1 GUUUAAAAGUCCUAUU GGAUUUCUACU-UUUGUAGAU
5 PbCpf1 GCCAAAUACCUCUAUA AAAUUUCUACU-UUUGUAGAU
6 SsCpf1 GUUUCAAUCCAC GCGCCCACGCGGGGCGCGAC
7 AsCpf1 GUCAAAAGACCUUUU UAAUUUCUACU-CUUGUAGAU
8 LbCpf1 GCUUAGAACAUUUAAA GAAUUUCUACU-AUUGUAGAU
9 CMeCpf1 CUCAAAACUCAUUG GAAUCUCUACUCUUUGUAGAU
10 EeCpf1 GUUUGAAUAACCUUAAA UAAUUUCUACU--UUGUAGAU
11 MbCpf1 GUCUAACGACCUUUU AAAUUUCUACUGUUUGUAGAU
12 UCpf1 CUCUAAAGAGAGGAAA GAAUUUCUACU-UUUGUAGAU
13 LbCpf1 GUUUCAAAGAUUAAA UAAUUUCUACUAAGUGUAGAU
14 PcCpf1 GUCUAGGUACUCUCUU UAAUUUCUACU-AUUGUAGAU
15 PdCpf1 GUCAAUAAGACUCAUU UAAUUUCUACU-UCGGUAGAU
16 PreCpf1 GCCUAUAAGGCUUUAC UAAUUUCUACU-AUUGUAGAU
maduración de ARNcr post remoción
tallo izquierdo
tallo derecho

FIG. 100C

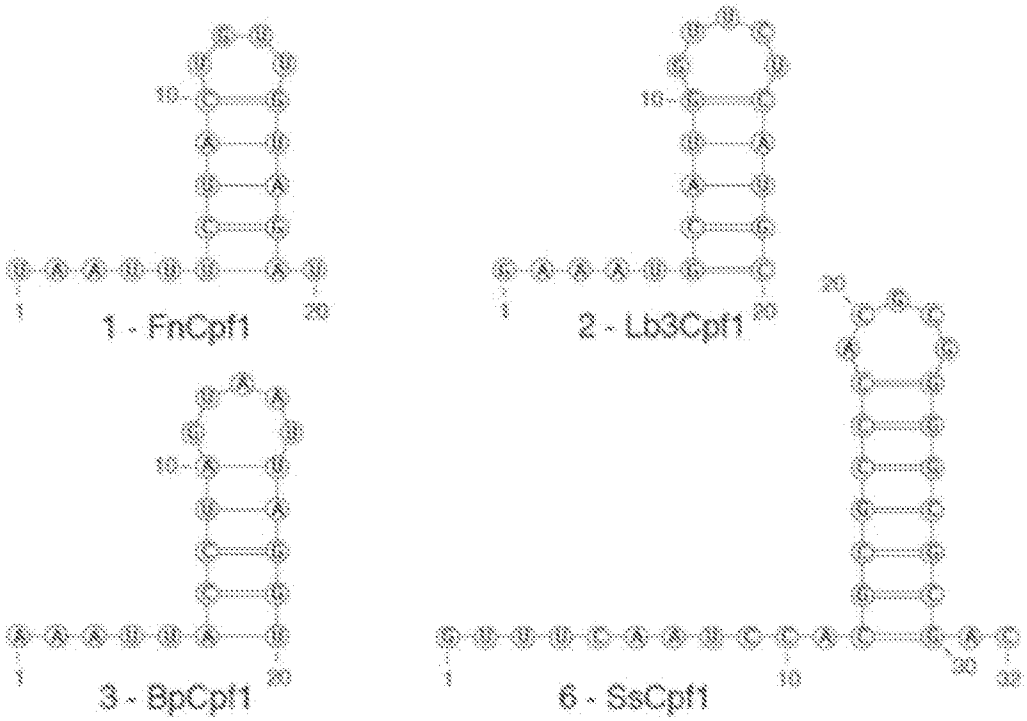


FIG. 100D

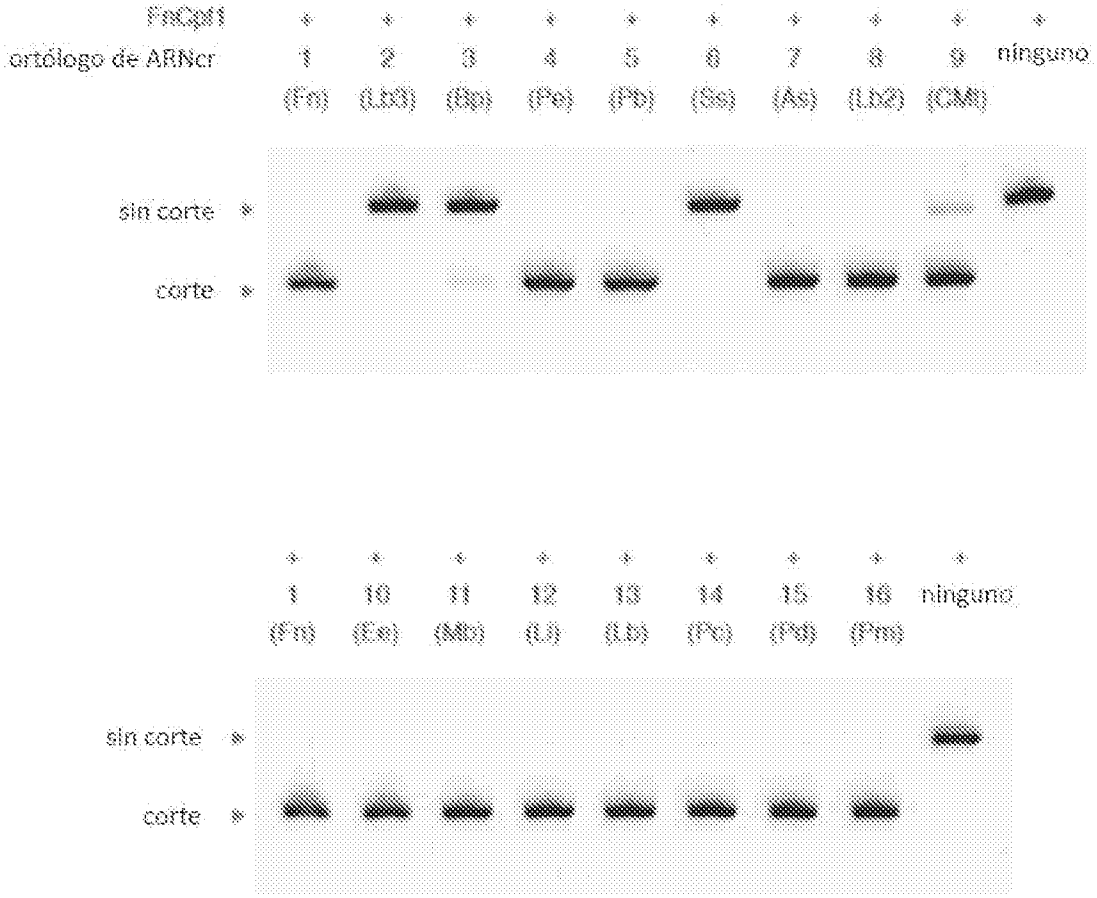

FnCpf1
+ + + + + + + + + +
ortólogo de ARNcr
1 2 3 4 5 6 7 8 9 ninguno
(Fn) (Lb3) (Bp) (Pe) (Pb) (Se) (As) (Lb2) (CMe)
sin corte »
corte »
+ + + + + + + + + +
1 10 11 12 13 14 15 16 ninguno
(Fn) (Ee) (Mb) (Li) (Ld) (Pc) (Pd) (Pm)
sin corte »
corte »

FIG. 100E

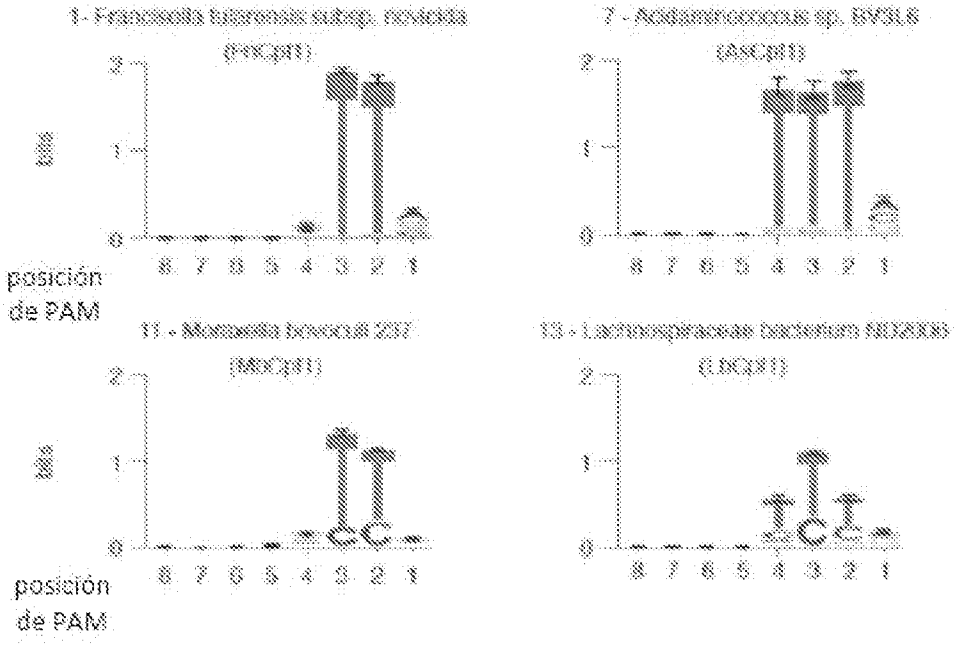
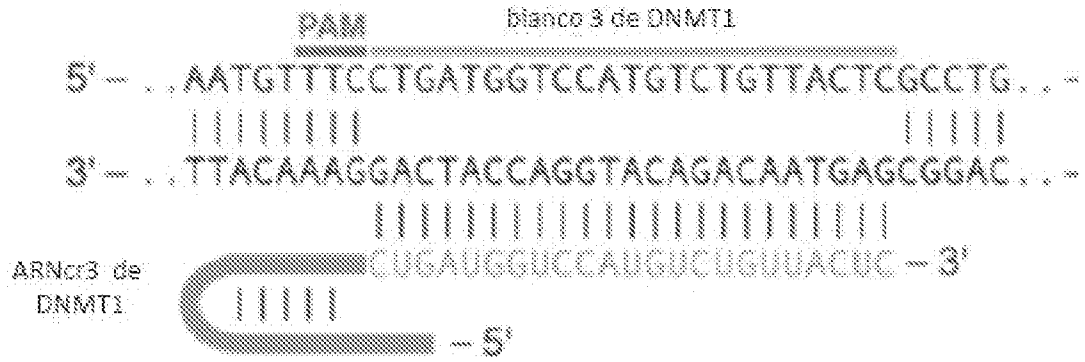


FIG. 100F



FIG. 101A



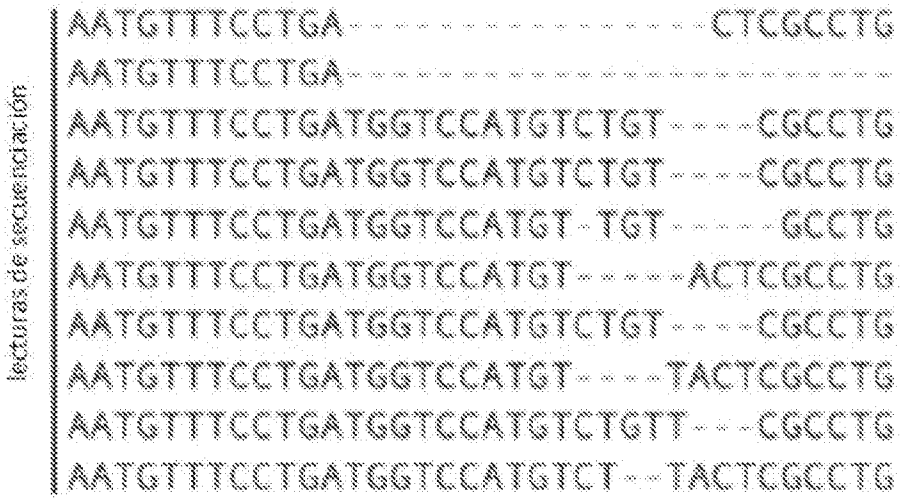


FIG. 101B

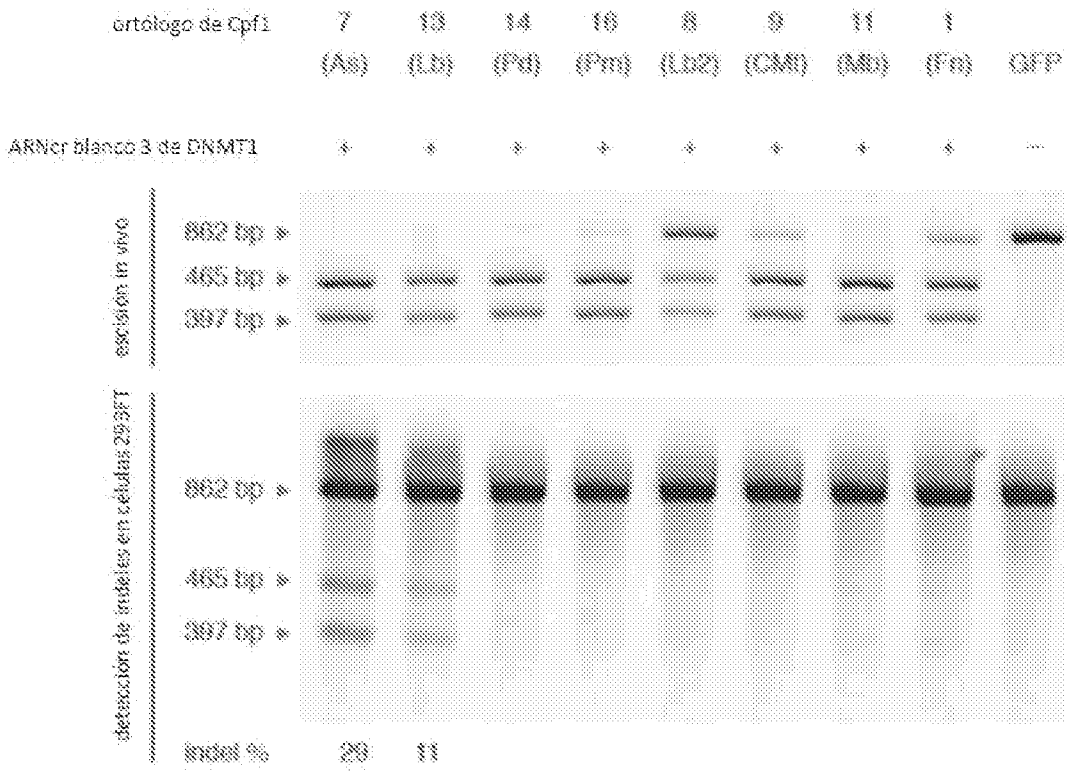

ortólogo de Cpf1
7 (As)
13 (Lb)
14 (Pd)
16 (Pm)
8 (Lb2)
9 (CMt)
11 (Mb)
1 (Fn)
GFP
ARNcr blanco 3 de DNMT1
+
+
+
+
+
+
+
+
—
escisión in vivo
862 bp
465 bp
397 bp
detección de indels en células 293FT
862 bp
465 bp
397 bp
Indel %
29
11

FIG. 101C

		PAM	secuencia guía
sitios blanco de DNMT1 Cpf1	1	TTTC	CCTCACTCCTGCTCGGTGAATTT
	2	TTTG	AGGAGTGTTCAGTCTCCGTGAAC
	3	TTTC	CTGATGGTCCATGTCTGTTACTC
	4	TTTA	TTTCCCTTCAGCTAAAATAAAGG

		secuencia guía	PAM
sitios blanco de DNMT1 SpCas9	1	TCACTCCTGCTCGGTGAATTT	GG
	2	AACCCTCTGGGGACCGTTTG	AGG
	3	AGTACGTTAATGTTTCCTGA	TGG
	4	TTTCCCTTCAGCTAAAATAA	AGG

FIG. 101D

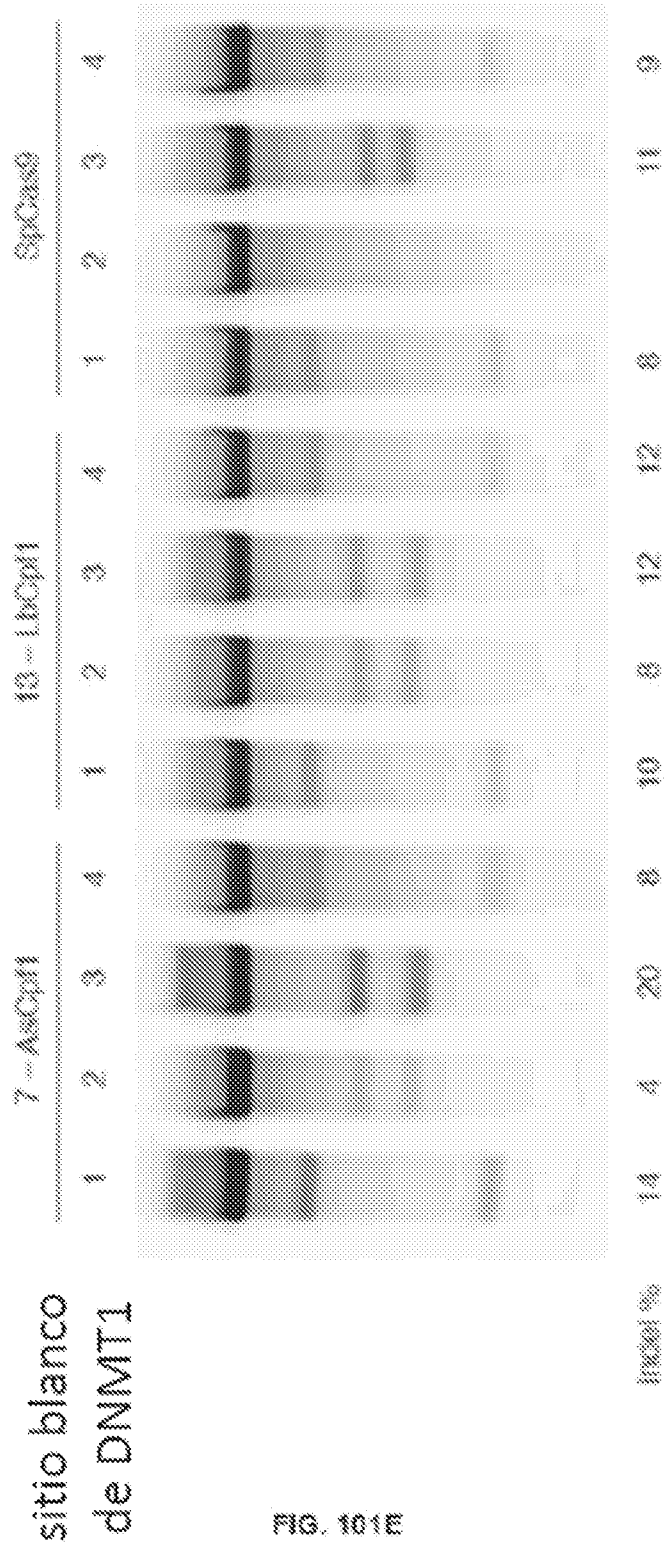

sitio blanco de DNMT1
7 – AsCpf1
1
2
3
4
14
4
20
8
10
8
12
4
8
11
9
18 – LbCpf1
1
2
3
4
12
8
12
12
8
11
9
SpCas9
1
2
3
4
Indel %

FIG. 101E

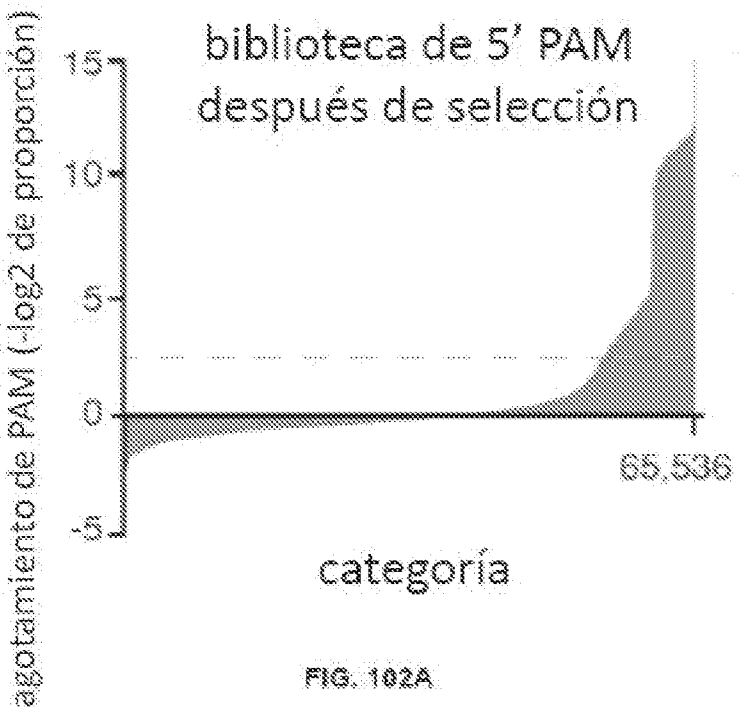


FIG. 102A

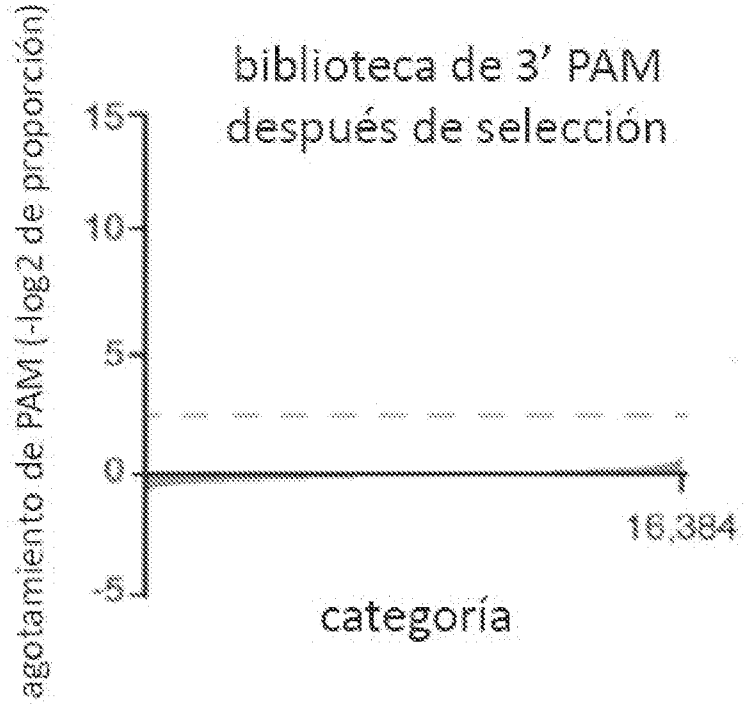


FIG. 102B

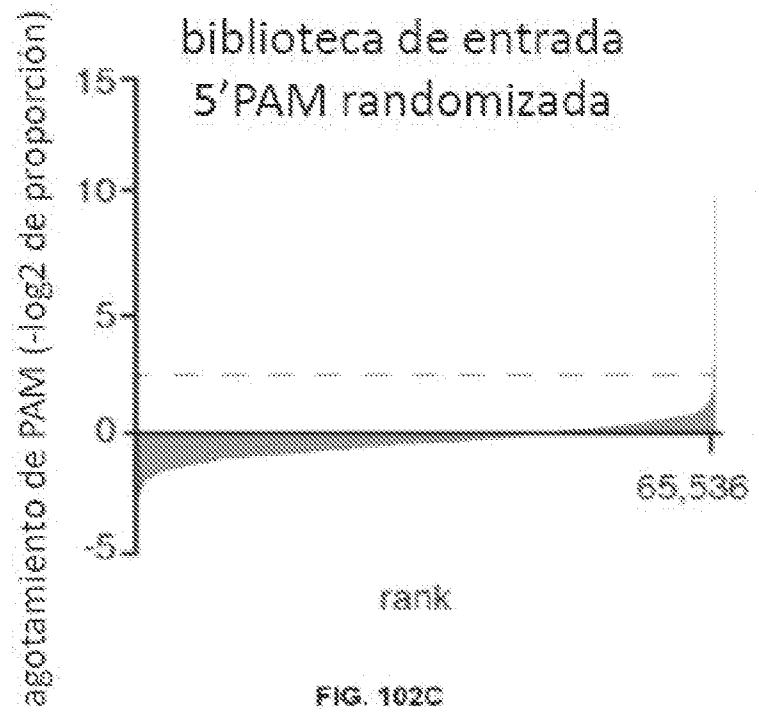


FIG. 102C

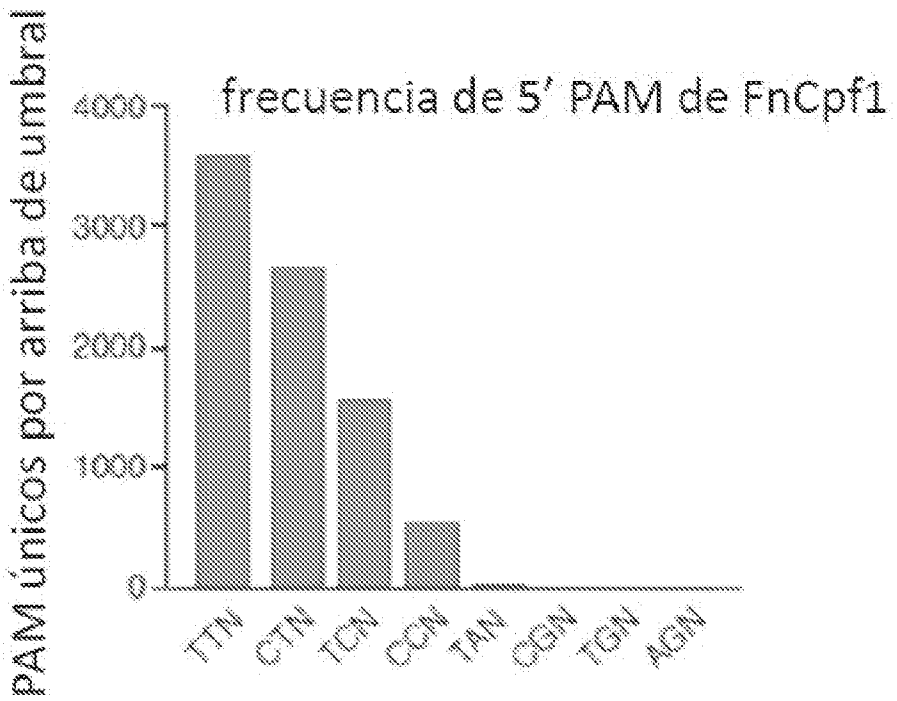


FIG. 102D

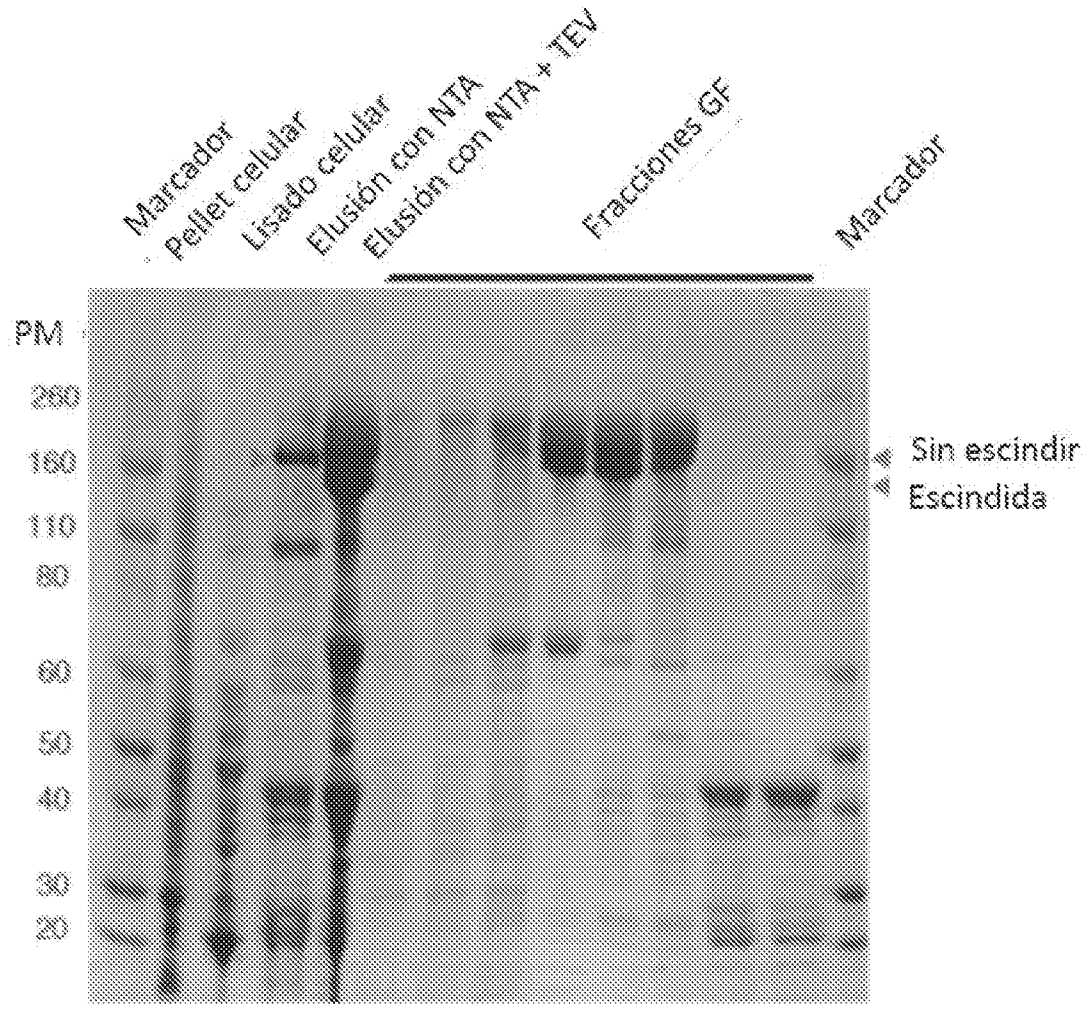

Marcador
Pellet celular
Lisado celular
Elusión con NTA
Elusión con NTA + TEV
Fracciones GF
Marcador
PM
260
160
110
80
60
50
40
30
20
Sin escindir
Escindida

FIG. 103A

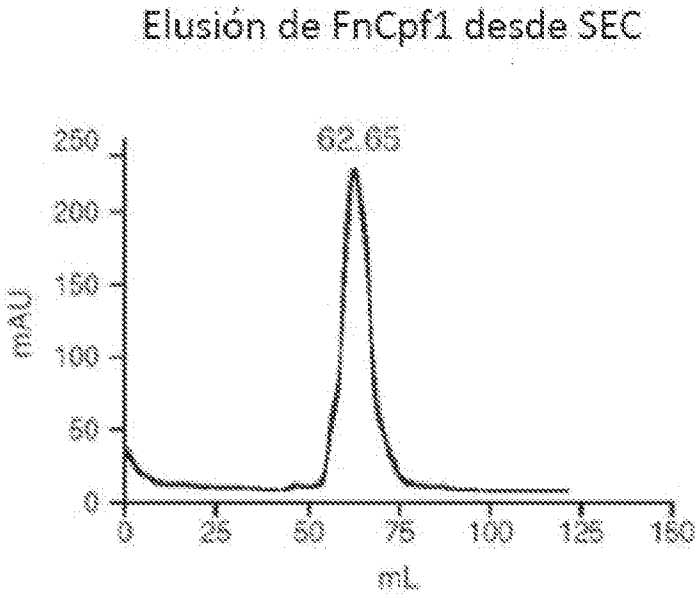


FIG. 103B

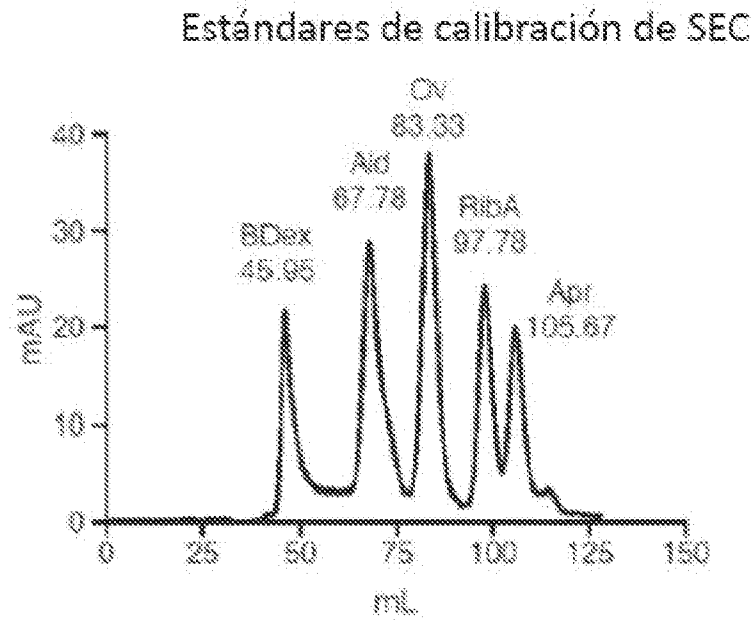


FIG. 103C

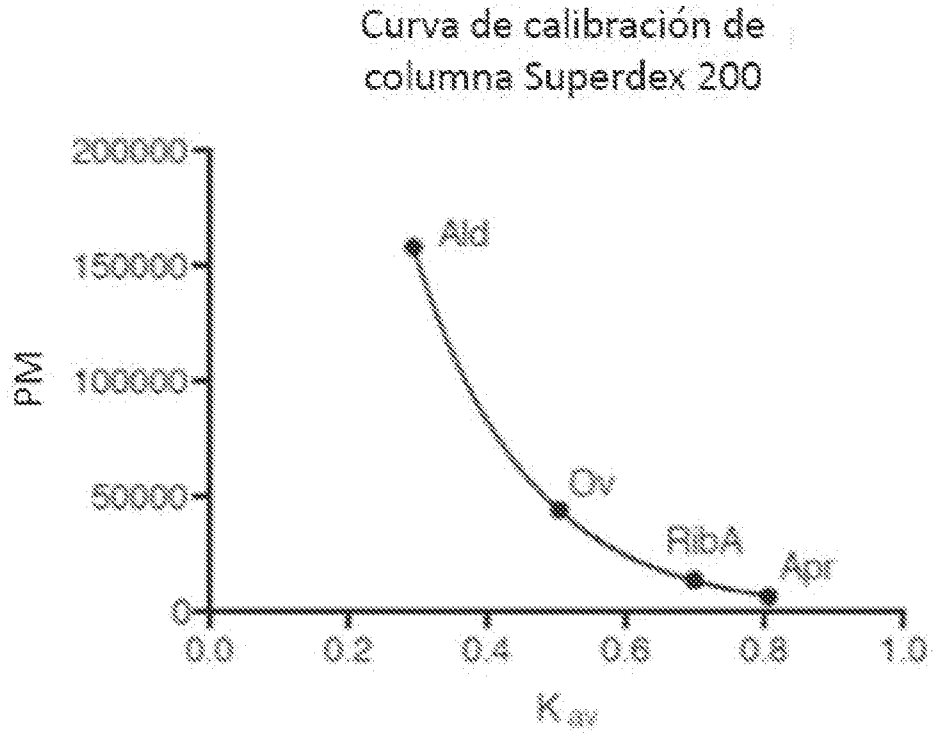

Curva de calibración de columna Superdex 200
PM
200000
150000
100000
50000
0
Ald
Ov
RibA
Apr
0.0
0.2
0.4
0.6
0.8
1.0
K_{av}

FIG. 103D

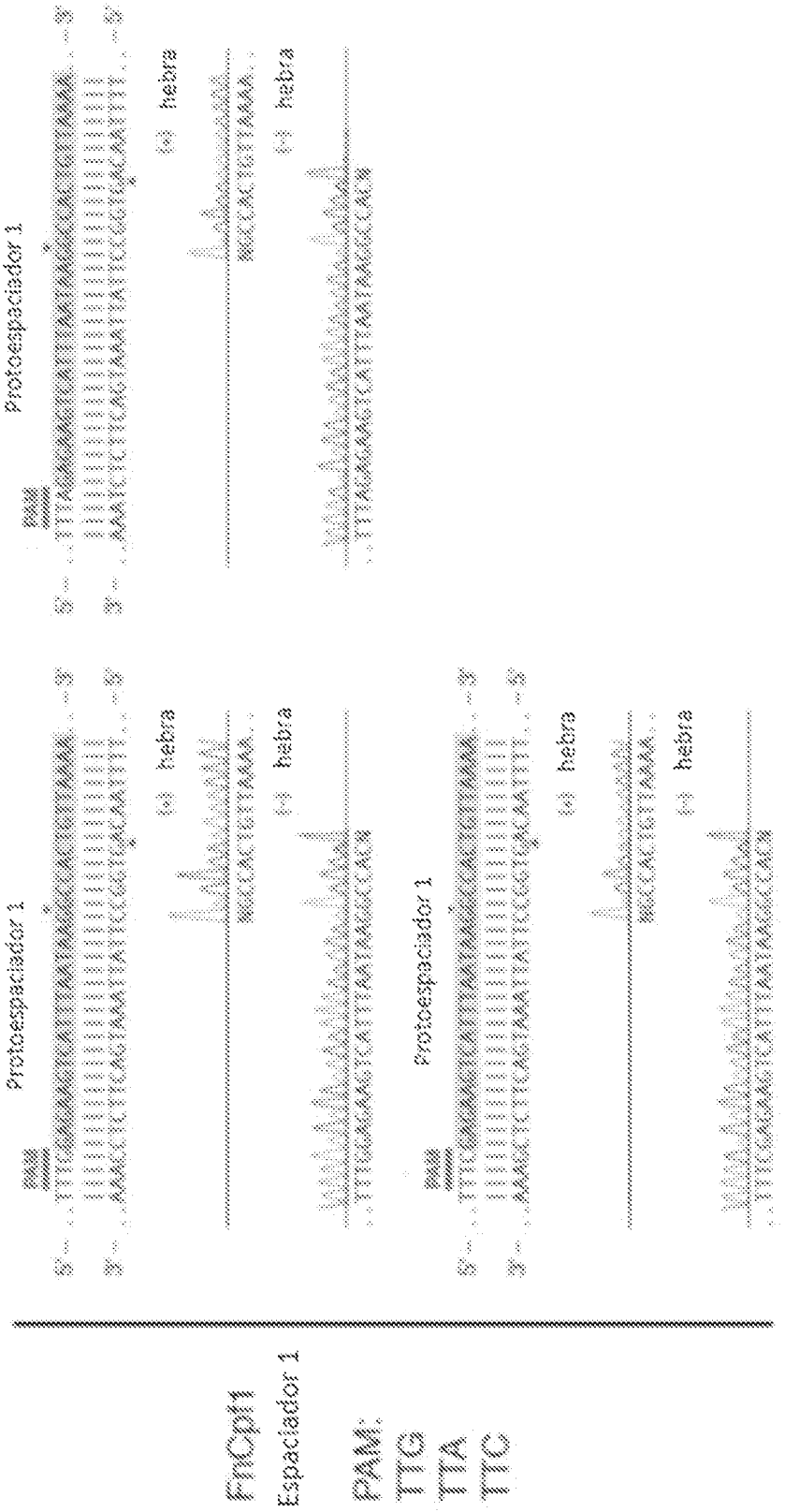

FnCpf1
Espaciador 1
PAM:
TTG
TTA
TTC
Protoespaciador 1
PAM
5' –
3' –
(+) hebra
(–) hebra
Protoespaciador 1
PAM
5' –
3' –
(+) hebra
(–) hebra

FIG. 104A

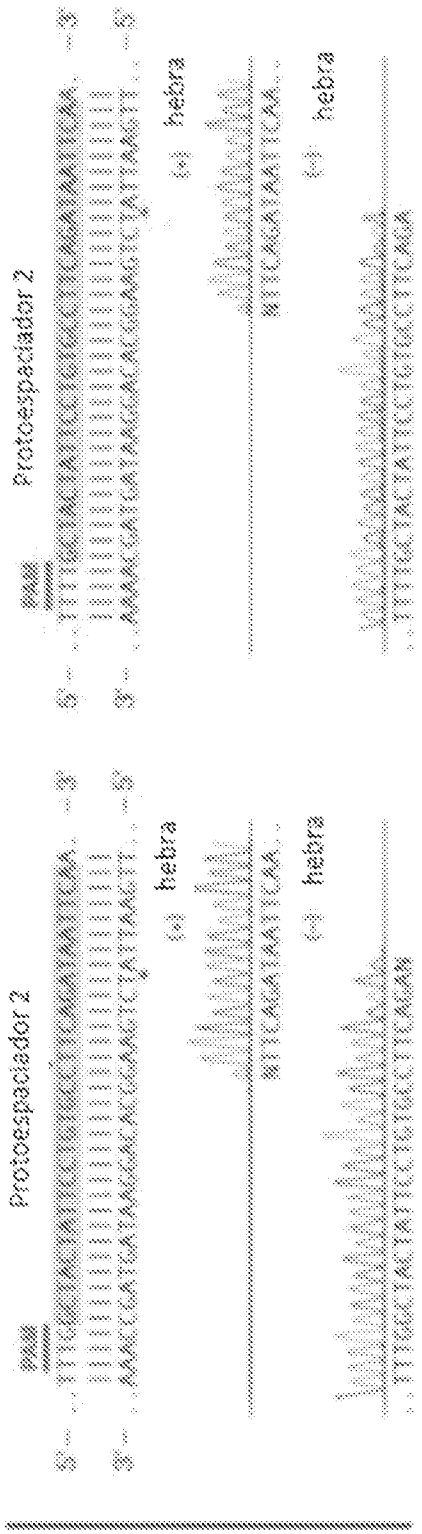

FnCpf1
Espaciador 2
PAM:
TTG
TTT
Protoespaciador 2
PAM
5' - ...TTGcaatactcaaaaaaatctcaaaaaatttcaa... - 3'
3' - ...AAACGATAAGAGTTTTTAGAGTTTTTAAAGT... - 5'
(+) hebra
...TTTCAATAATTCAA...
(-) hebra
...TTTGCTATTCTCTTTTTCAGAGTTTTTCAAA...
Protoespaciador 2
PAM
5' - ...TTGcaatactcaaaaaaatctcaaaaaatttcaa... - 3'
3' - ...AAACGATAAGAGTTTTTAGAGTTTTTAAAGT... - 5'
(+) hebra
...TTTCAATAATTCAA...
(-) hebra
...TTTGCTATTCTCTTTTTCAGAGTTTTTCAAA...

FIG. 104B

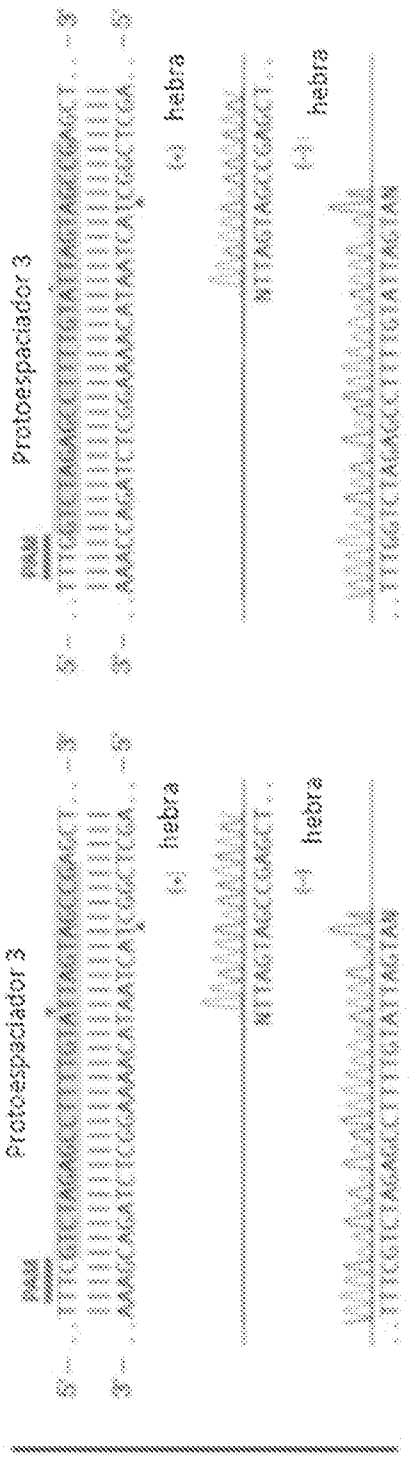

FnCpf1
Espaciador 3
PAM:
TTC
TTG
Protoespaciador 3
PAM
5'
3'
(+) hebra
(-) hebra

FIG. 104C

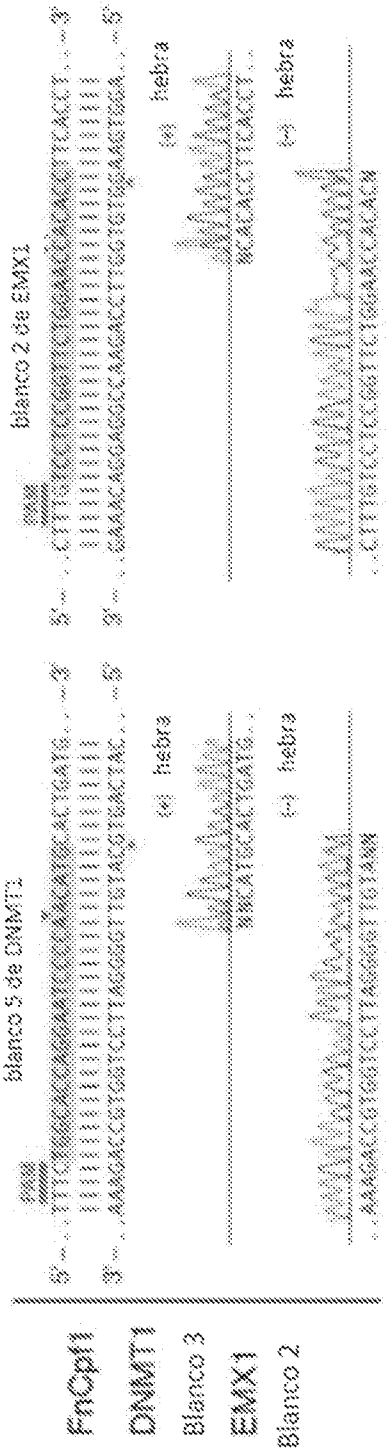

FnCpf1
DNMT1
Blanco 3
EMX1
Blanco 2
PAM
Blanco 5 de DNMT1
5'-
3'-
(+) hebra
(-) hebra
PAM
Blanco 2 de EMX1
5'-
3'-
(+) hebra
(-) hebra

FIG. 104D

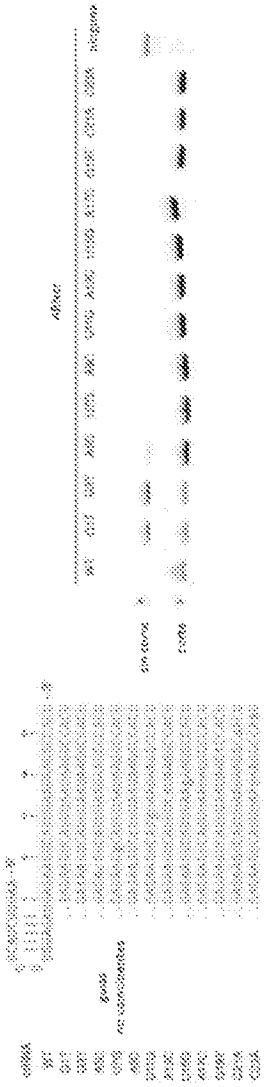


FIG. 19AK

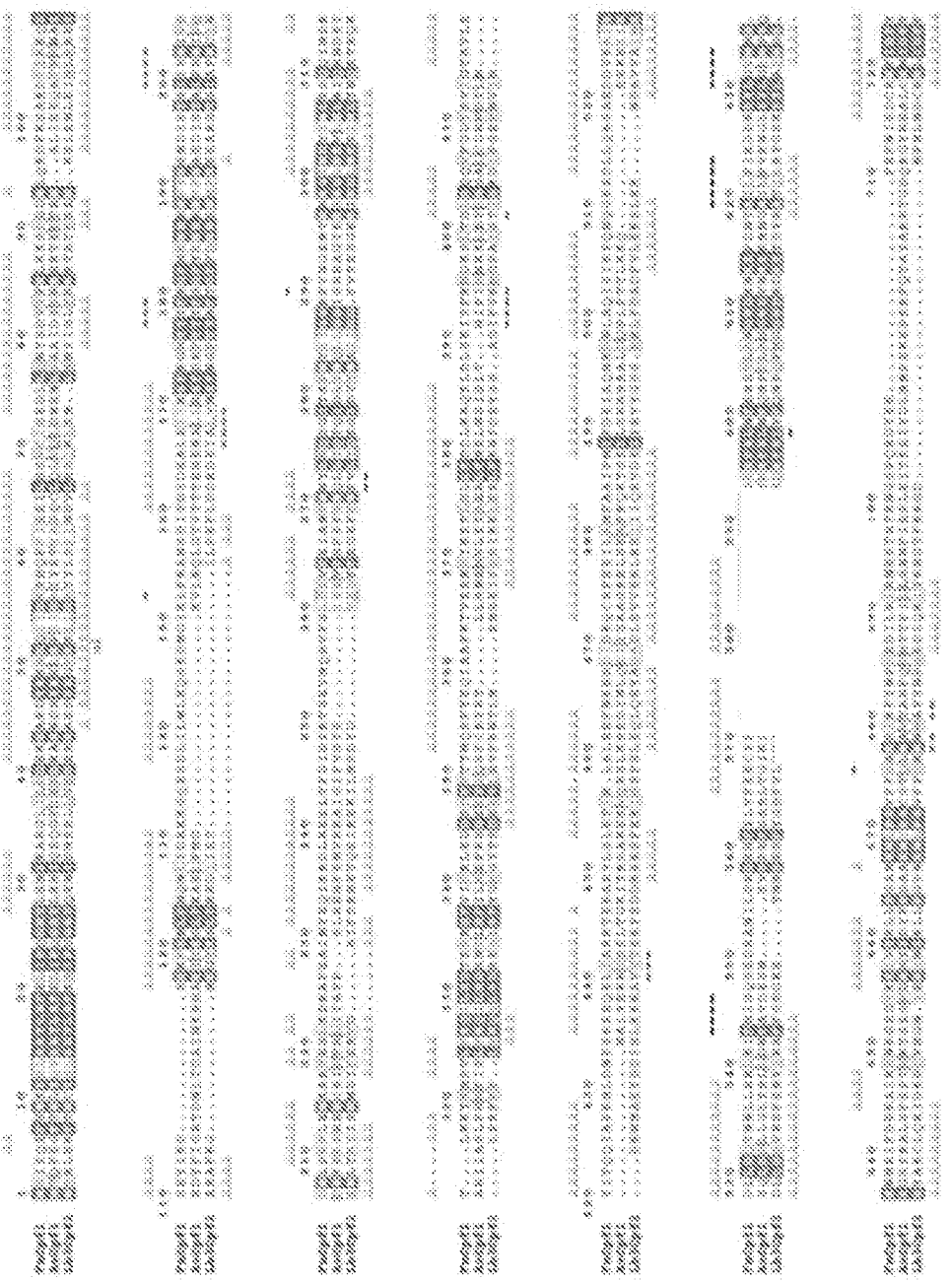


FIG. 105A

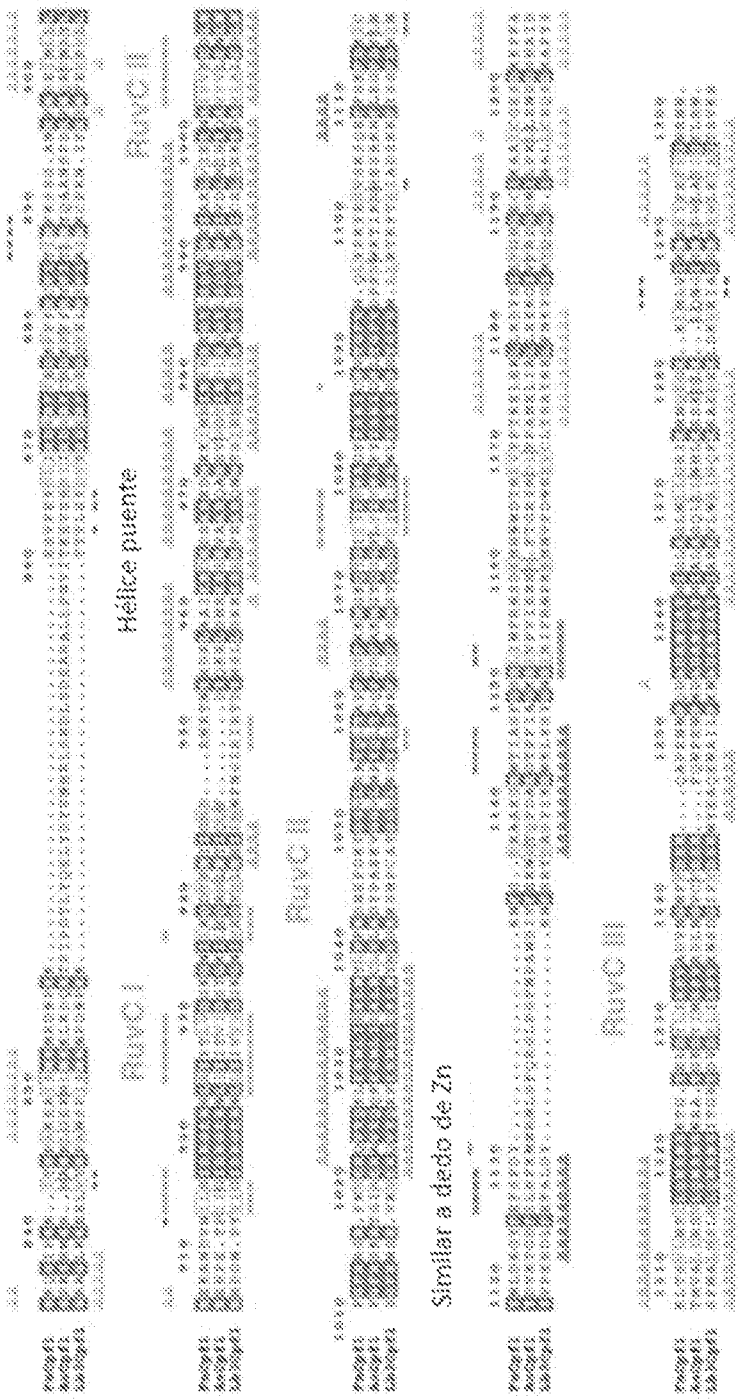

Hélice puente
RuvC I
RuvC II
RuvC III
RuvC II
Similar a dedo de Zn
RuvC III

FIG. 105B

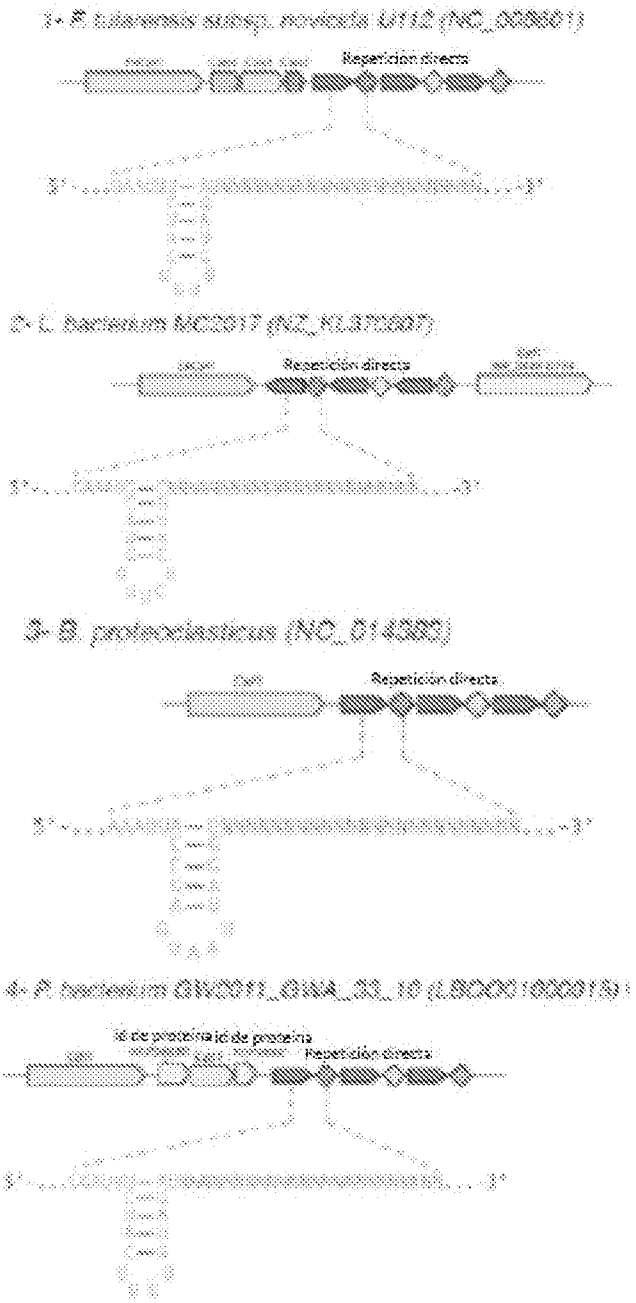

Repetición directa
Repetición directa
Repetición directa
Repetición directa
Repetición directa

FIG. 106A

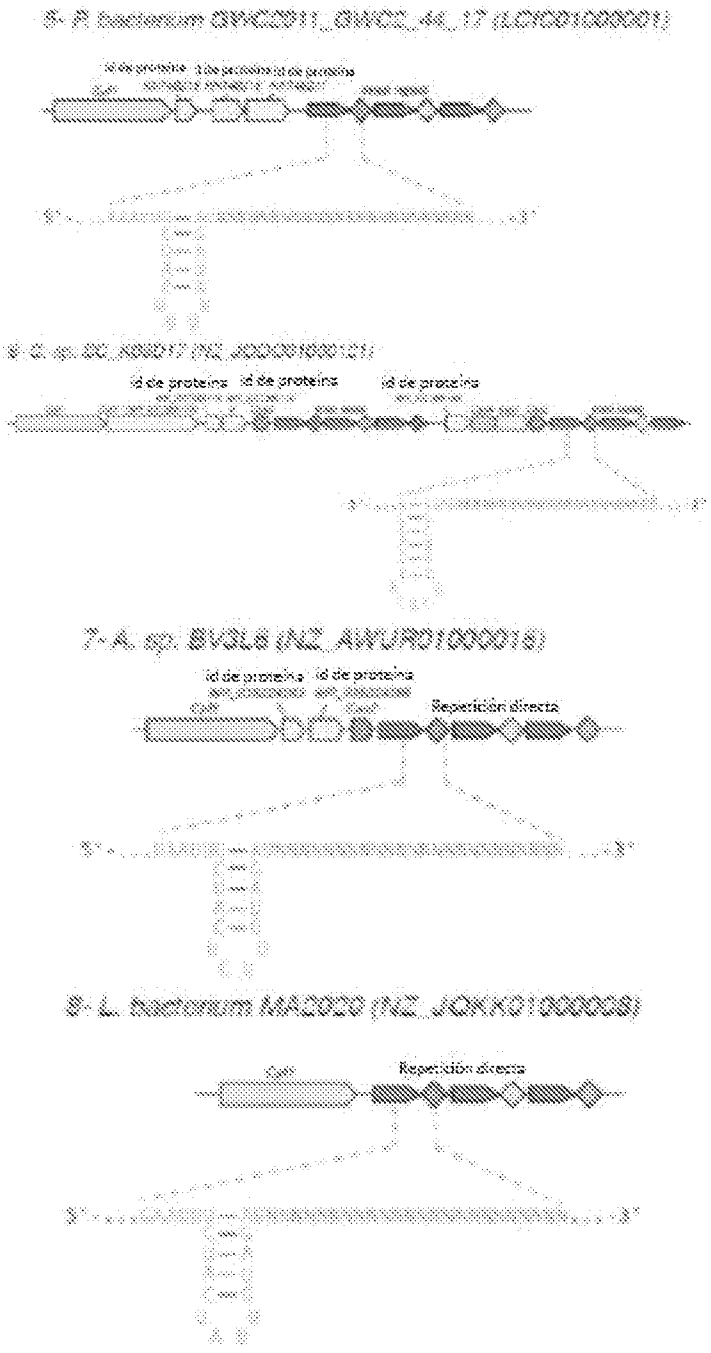

Id de proteína
Id de proteína
Id de proteína
Repetición directa
Repetición directa

FIG. 106B

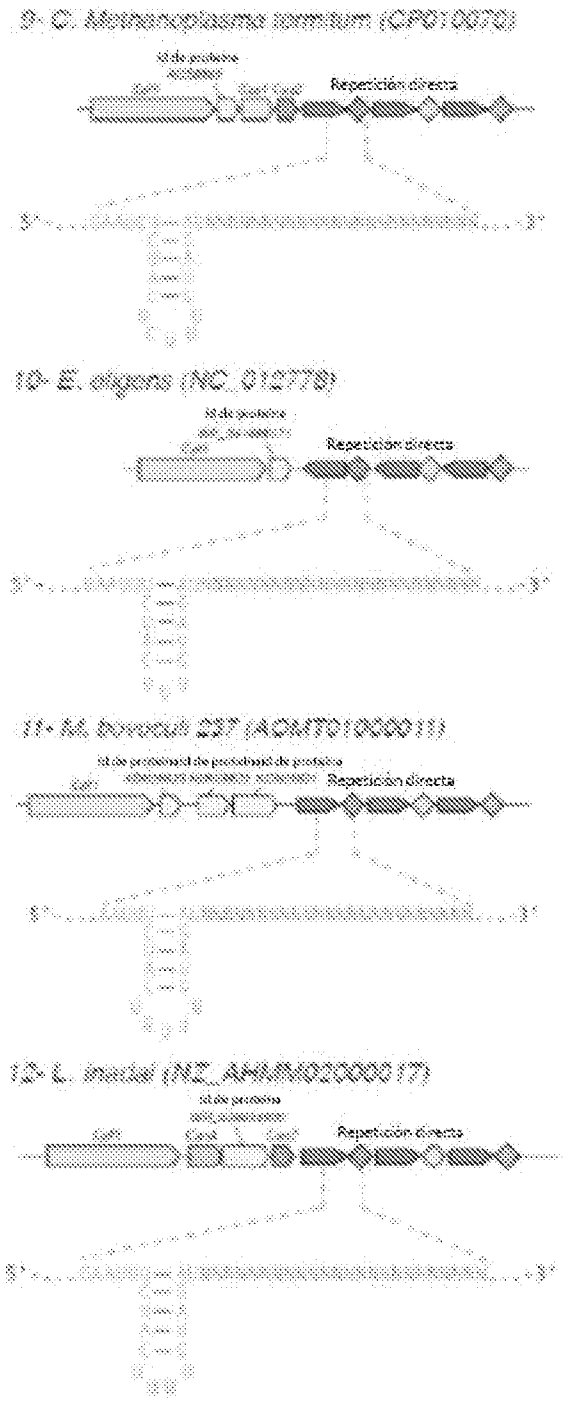

9- C. Methanoplasma termitum (CP010078)
Id de proteína
Repetición directa
5'
3'
10- E. eligens (NC_012778)
Id de proteína
Repetición directa
5'
3'
11- M. bovoculi 237 (AQMT01000011)
Id de proteína
Repetición directa
5'
3'
12- L. innocua (NZ_AMJN02000017)
Id de proteína
Repetición directa
5'
3'

FIG. 10C

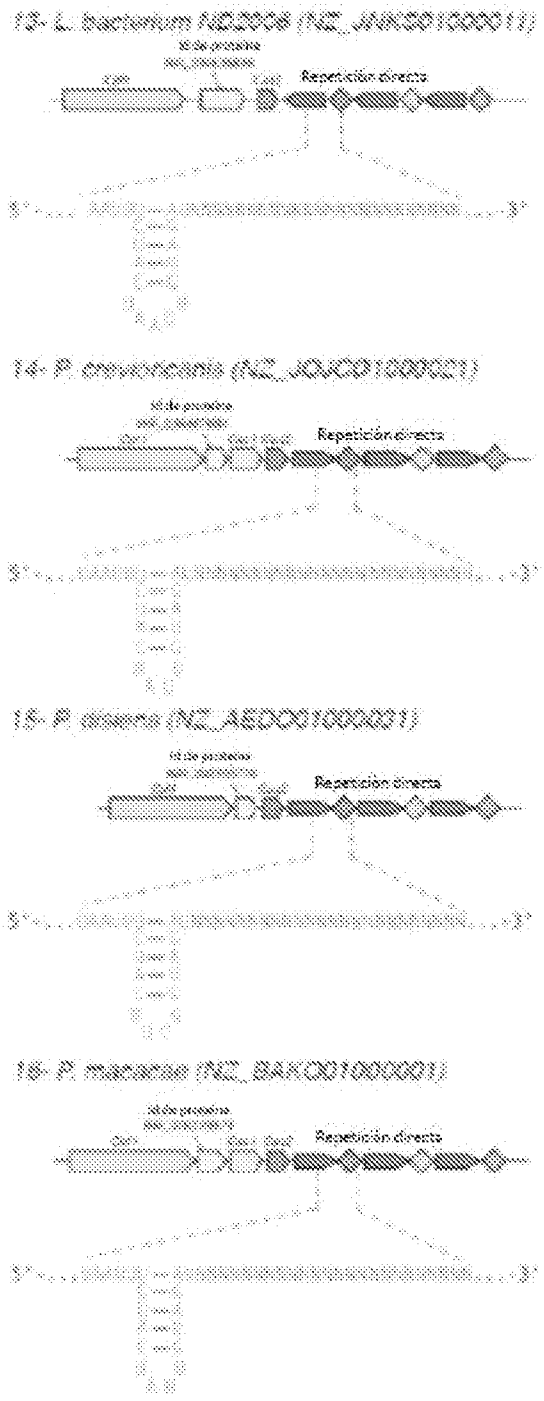

Repetición directa
Repetición directa
Repetición directa
Repetición directa

FIG. 106D

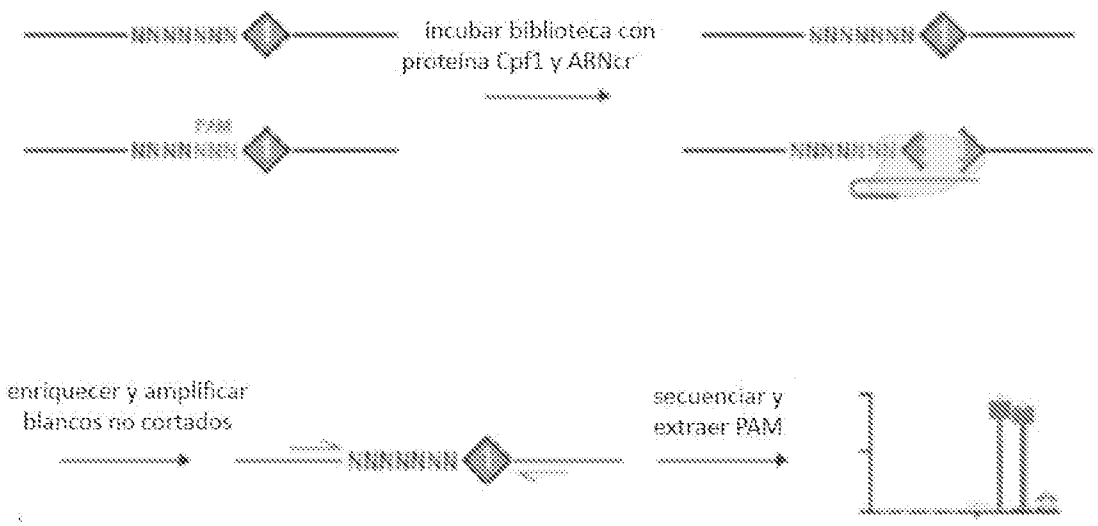

incubar biblioteca con
proteína Cpf1 y ARNcr
PAM
enriquecer y amplificar
blancos no cortados
secuenciar y
extraer PAM

FIG. 107A

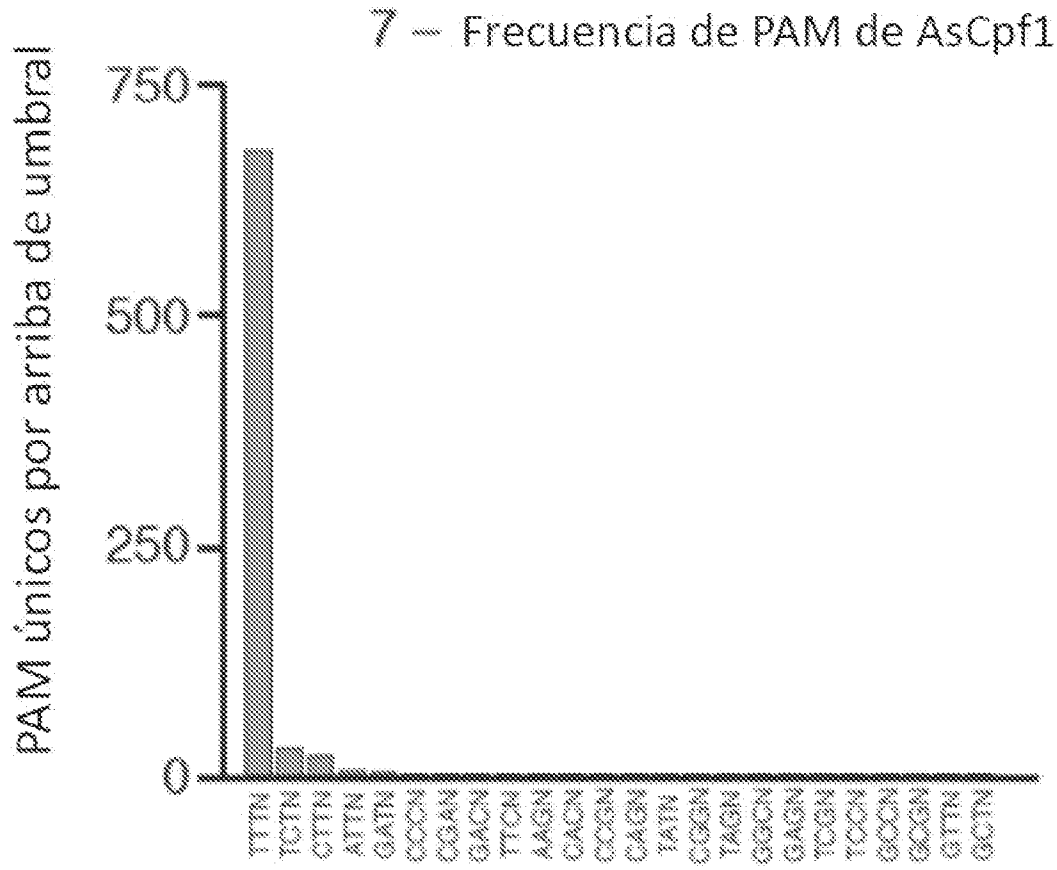

7 – Frecuencia de PAM de AsCpf1
PAM únicos por arriba de umbral
750
500
250
0
TTTN
TCTN
CTTN
ATTN
GATN
CCCN
CGAN
CACN
TTCN
AAGN
CAON
CCGN
CAGN
TATN
CCGN
TAGN
GCCN
GAGN
TCGN
TCCN
GCGN
GCGN
GTTN
GCTN

FIG. 107B

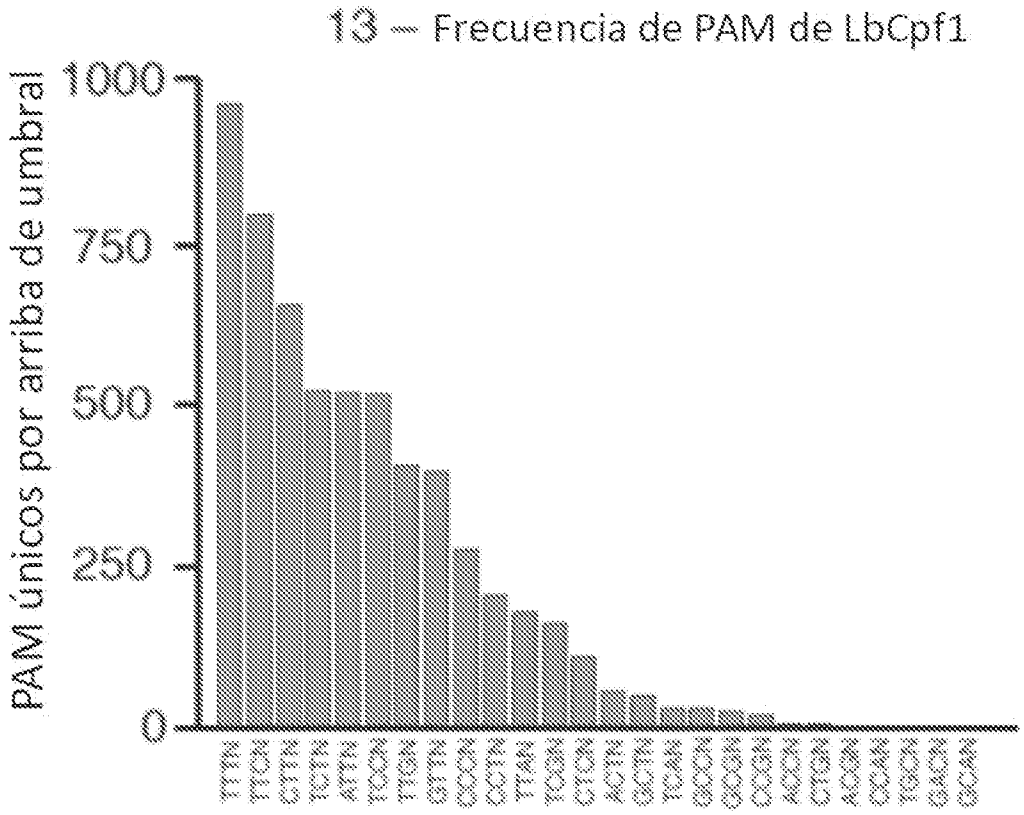

13 — Frecuencia de PAM de LbCpf1
PAM únicos por arriba de umbral
1000
750
500
250
0
TTTN
TTCN
GTTN
TGTN
ATTN
TCCN
TTGN
GTCN
CCTN
CCCN
TTAN
TCGN
CTCN
ACTN
GCTN
TCAN
GCCN
GCGN
CCGN
ACCN
CTGN
ACAN
CCAN
TGCN
GACN
GCAN

FIG. 107C

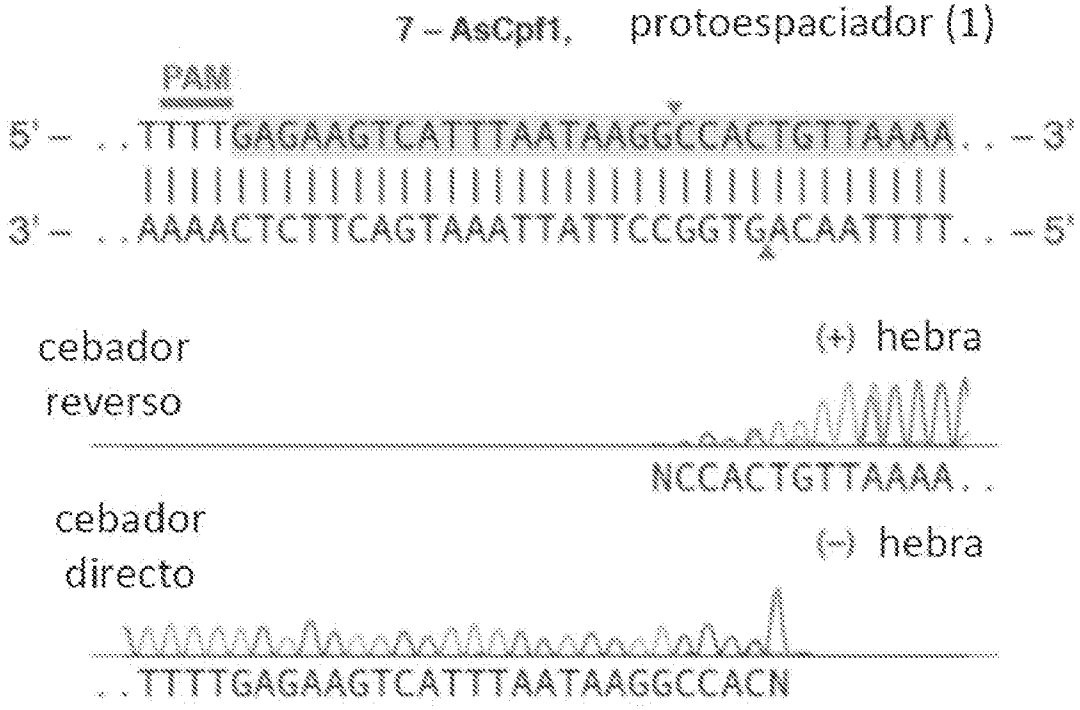

7 – AsCpf1, protoespaciador (1)
PAM
5' – . . TTTTGAGAAGTCATTTAATAAGGCCACTGTTAAAA. . – 3'
3' – . . AAAACTCTTCAGTAAATTATTCCGGTGACAATTTT. . – 5'
cebador reverso
(+) hebra
NCCACTGTTAAAA. .
cebador directo
(–) hebra
. . TTTTGAGAAGTCATTTAATAAGGCCACN

FIG. 107D

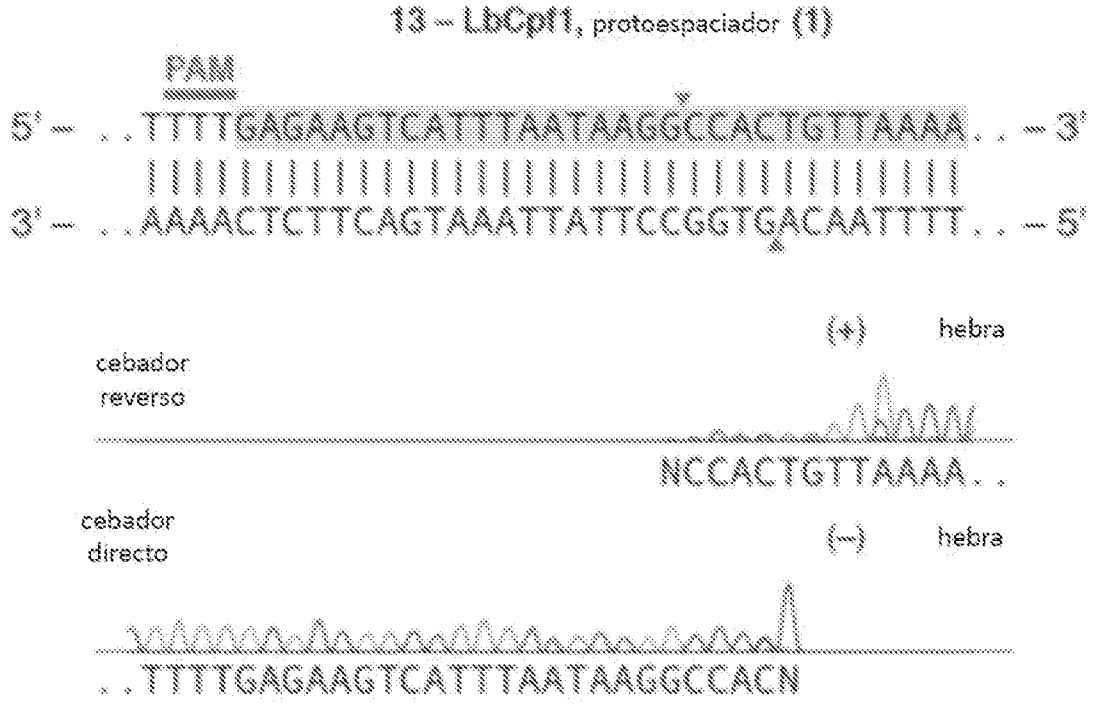

13 – LbCpf1, protoespaciador (1)
PAM
5' – . . TTTTGAGAAGTCATTTAATAAGGCCACTGTTAAAA. . – 3'
3' – . . AAAACTCTTCAGTAAATTATTCCGGTGACAATTTT. . – 5'
(+) hebra
cebador reverso
NCCACTGTTAAAA. .
cebador directo
(–) hebra
. . TTTTGAGAAGTCATTTAATAAGGCCACN

FIG. 107E

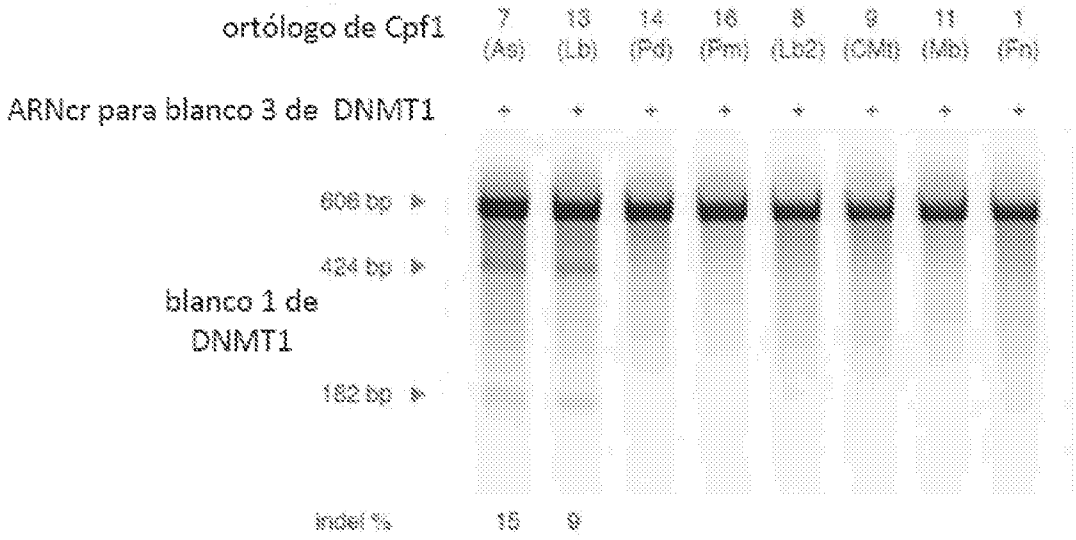


FIG. 108A

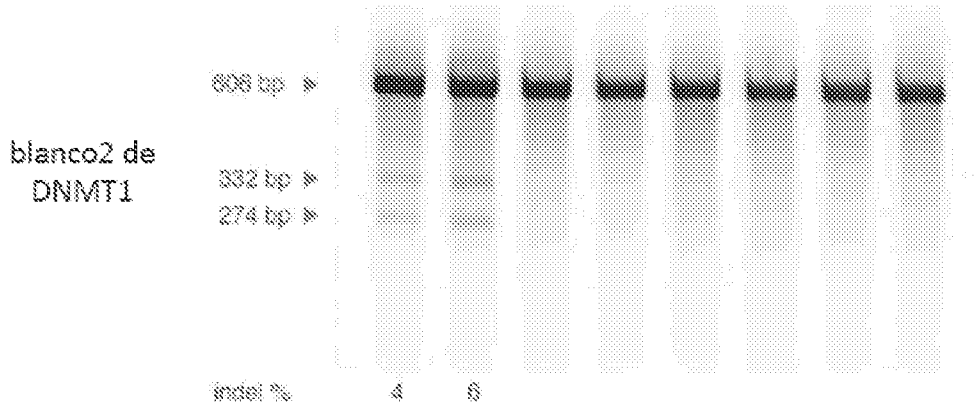


FIG. 108B

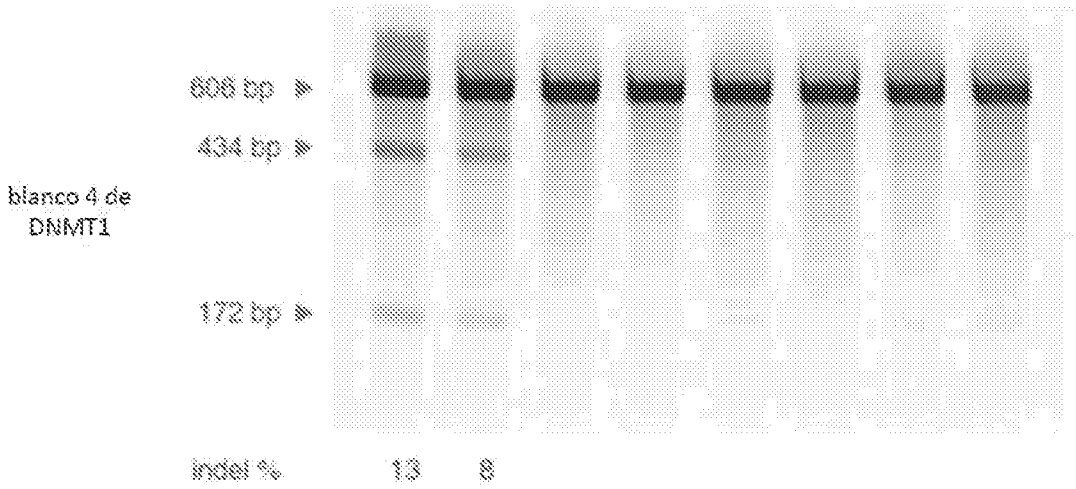


FIG. 108C

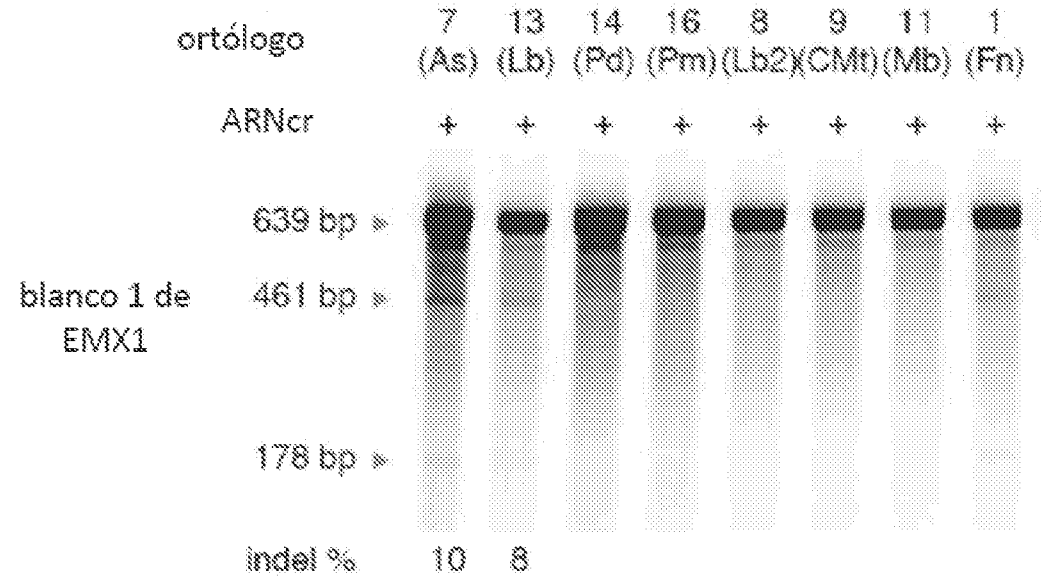


FIG 108D

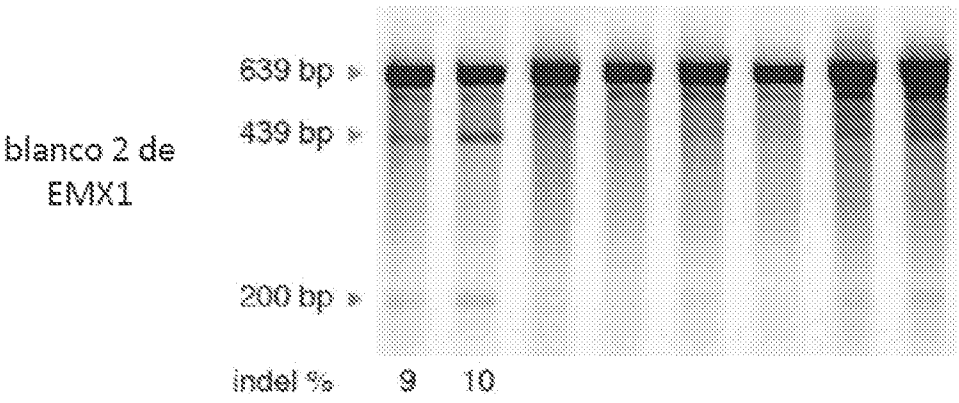


FIG 108E

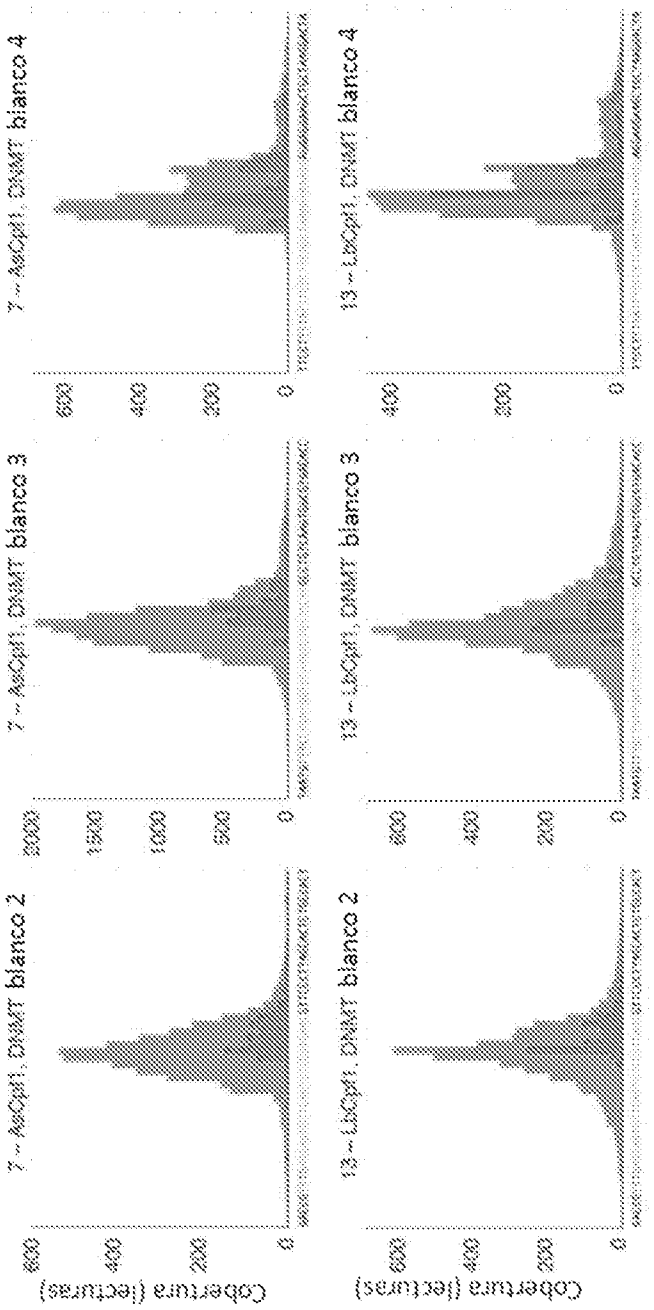

7 – AsCpf1, DNMT blanco 2
7 – AsCpf1, DNMT blanco 3
7 – AsCpf1, DNMT blanco 4
13 – LbCpf1, DNMT blanco 2
13 – LbCpf1, DNMT blanco 3
13 – LbCpf1, DNMT blanco 4
Cobertura (lecturas)
0
200
400
600
800
1000
1200
1400
2000

FIG. 108F

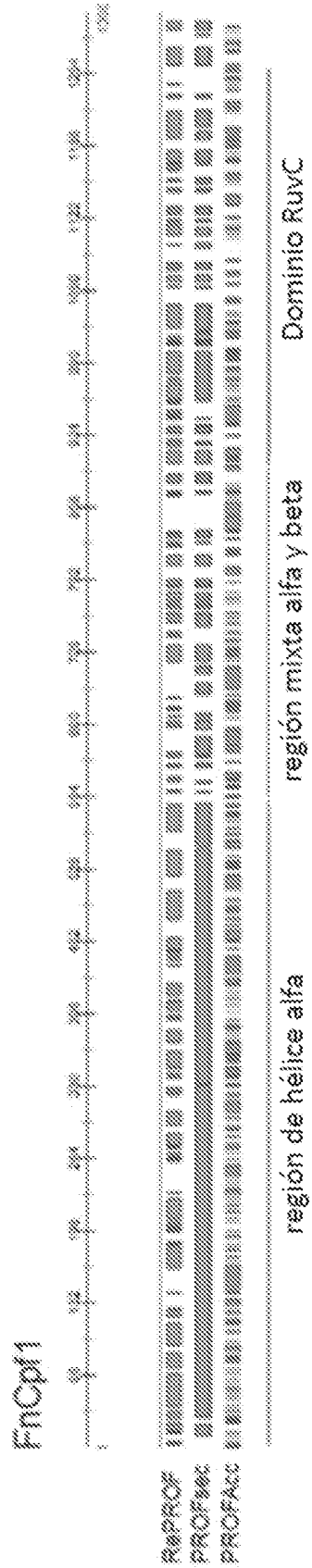

FnCpf1
RxPROF
PROFfunc
PROFAcc
región de hélice alfa
región mixta alfa y beta
Dominio RuvC

FIG. 109A

AsCpf1
RxPROF
PROFsec
PROFAcc
región de hélice alfa
región mixta alfa y beta
Dominio RuvC

FIG. 109B

LbCpf1
RelPROF
PROFsec
PROFAcc
región de hélice alfa
región mixta alfa y beta
Dominio RuvC

FIG. 109C

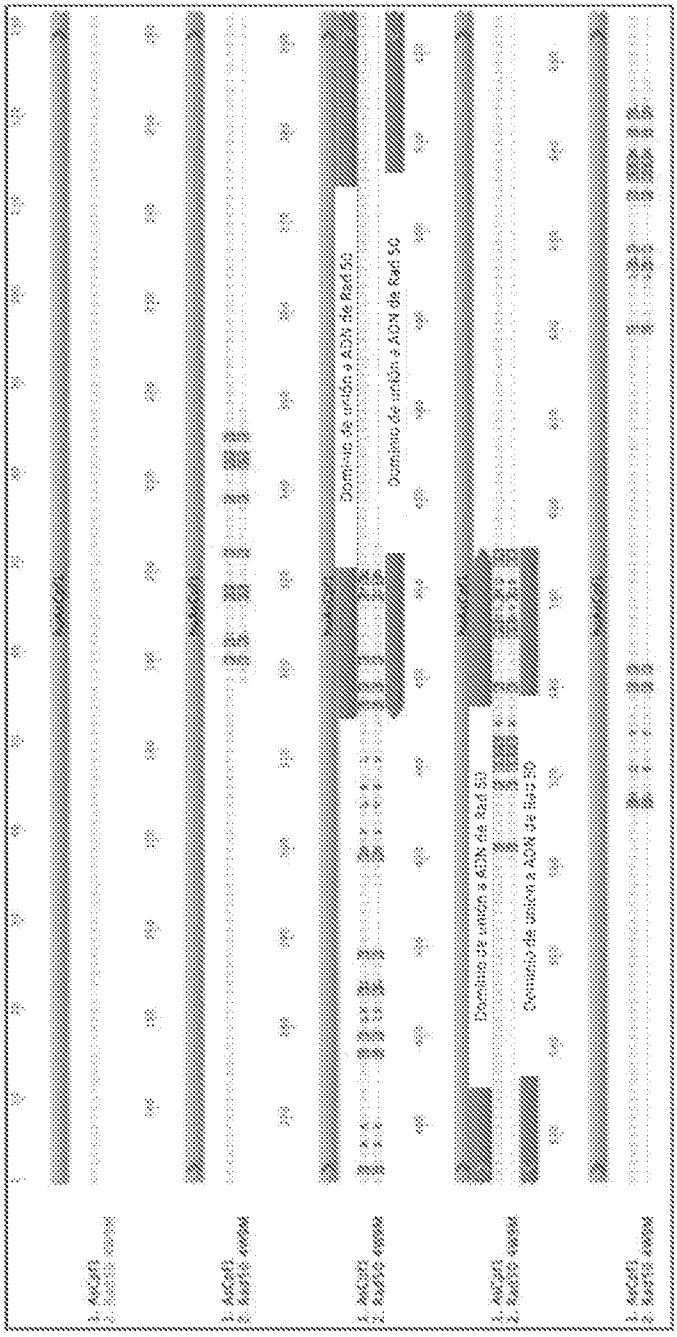

Dominio de unión a ADN de Raf 50
Dominio de unión a ADN de Raf 50
Dominio de unión a ADN de Raf 50
Dominio de unión a ADN de Raf 50

FIG. 110A

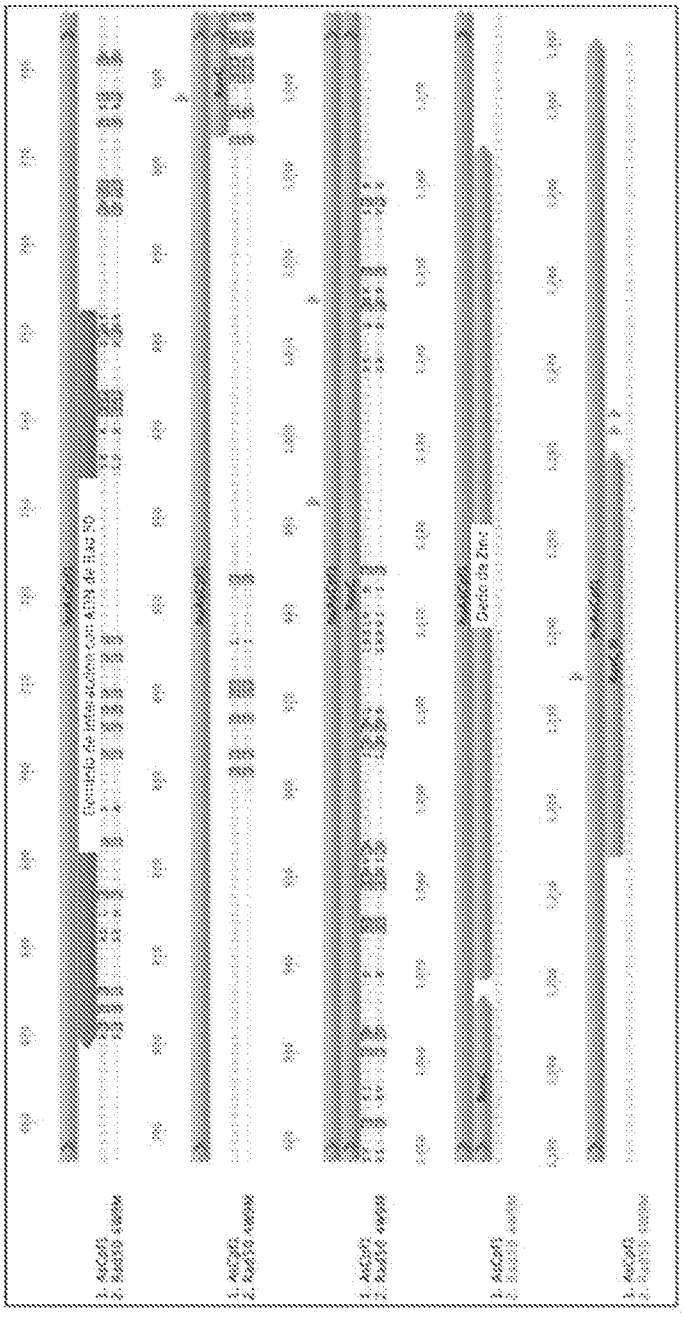


FIG. 110B



FIG. 110C

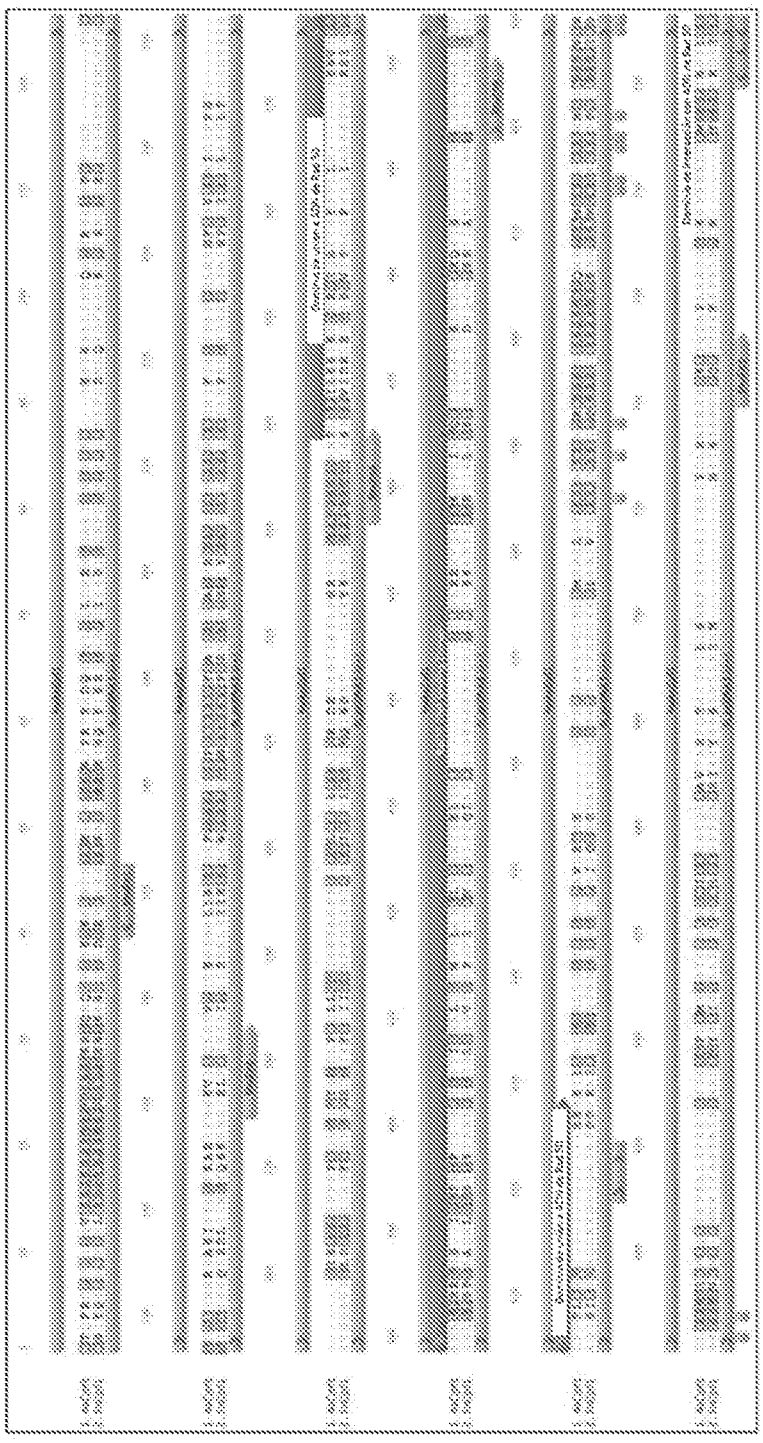


FIG. 110D

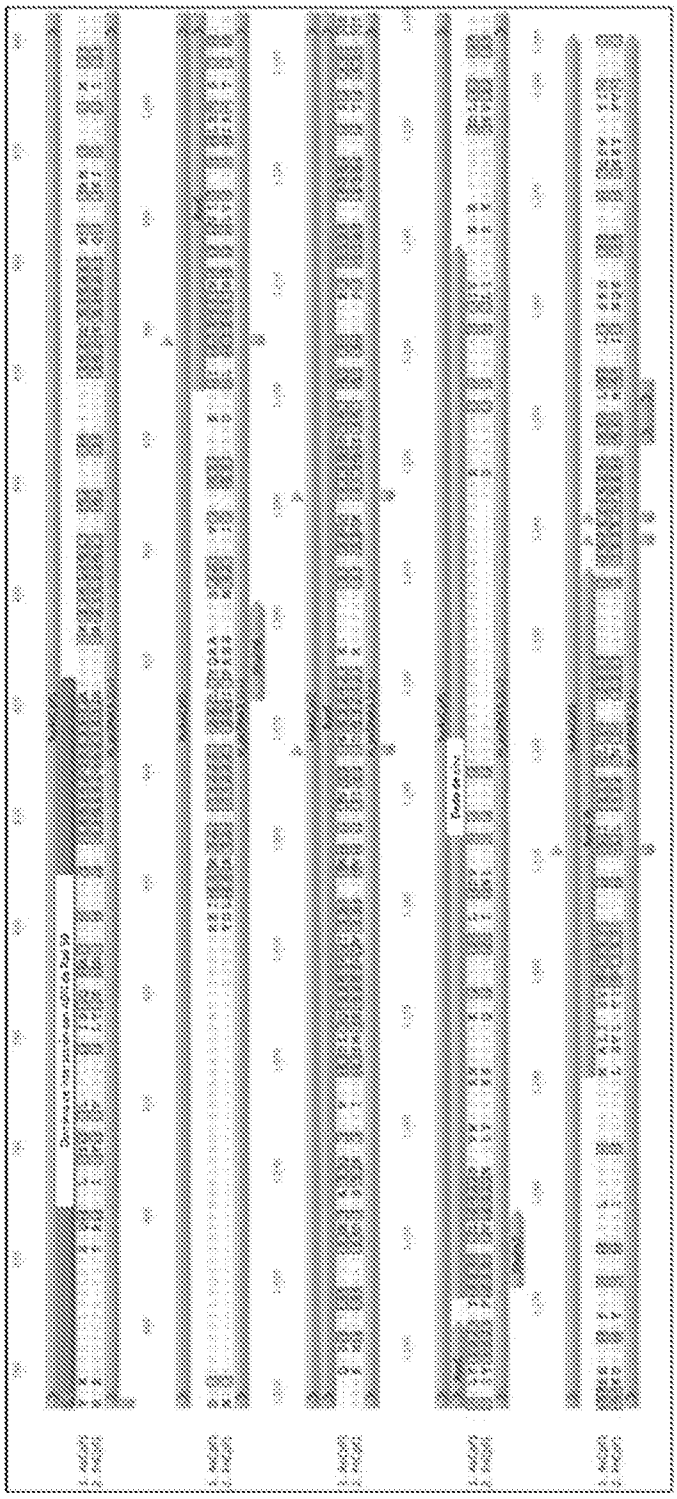


FIG. 110E

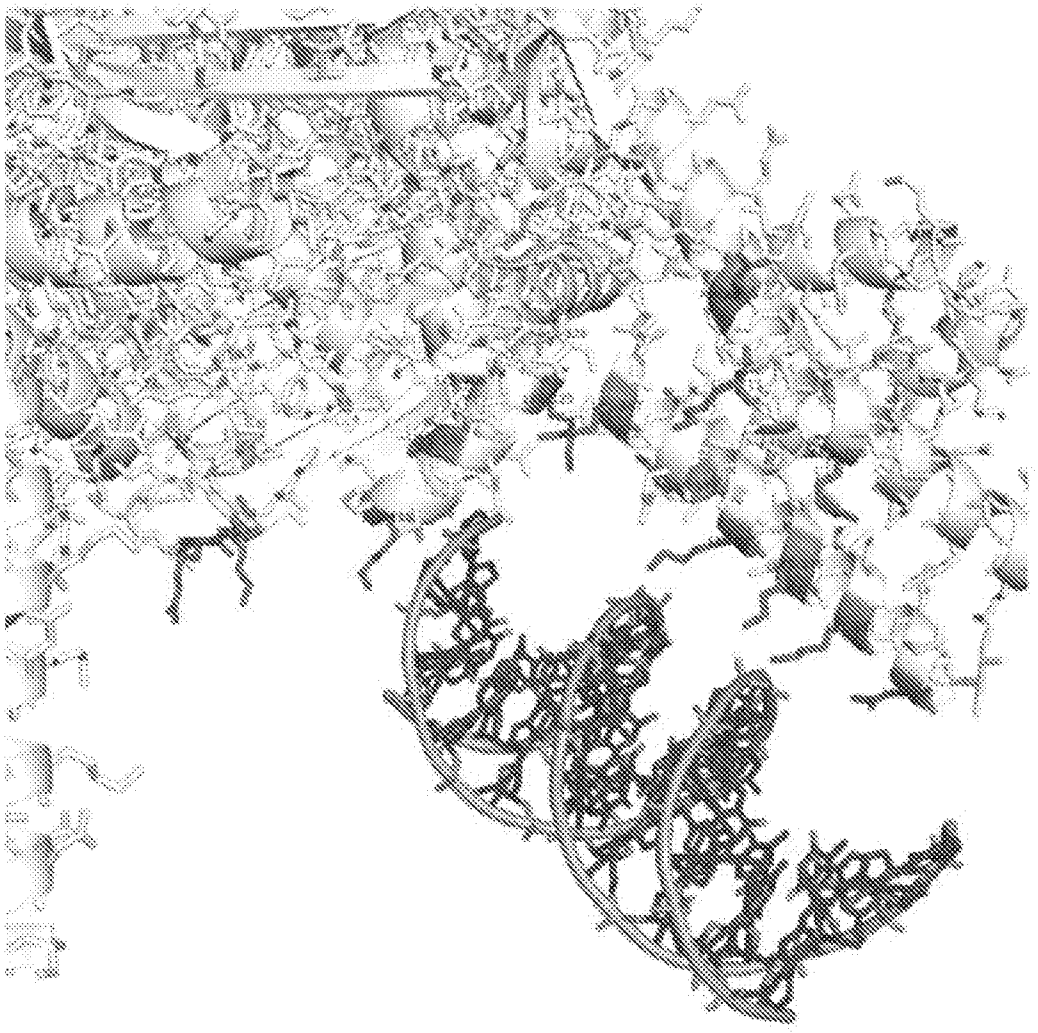


FIG. 111

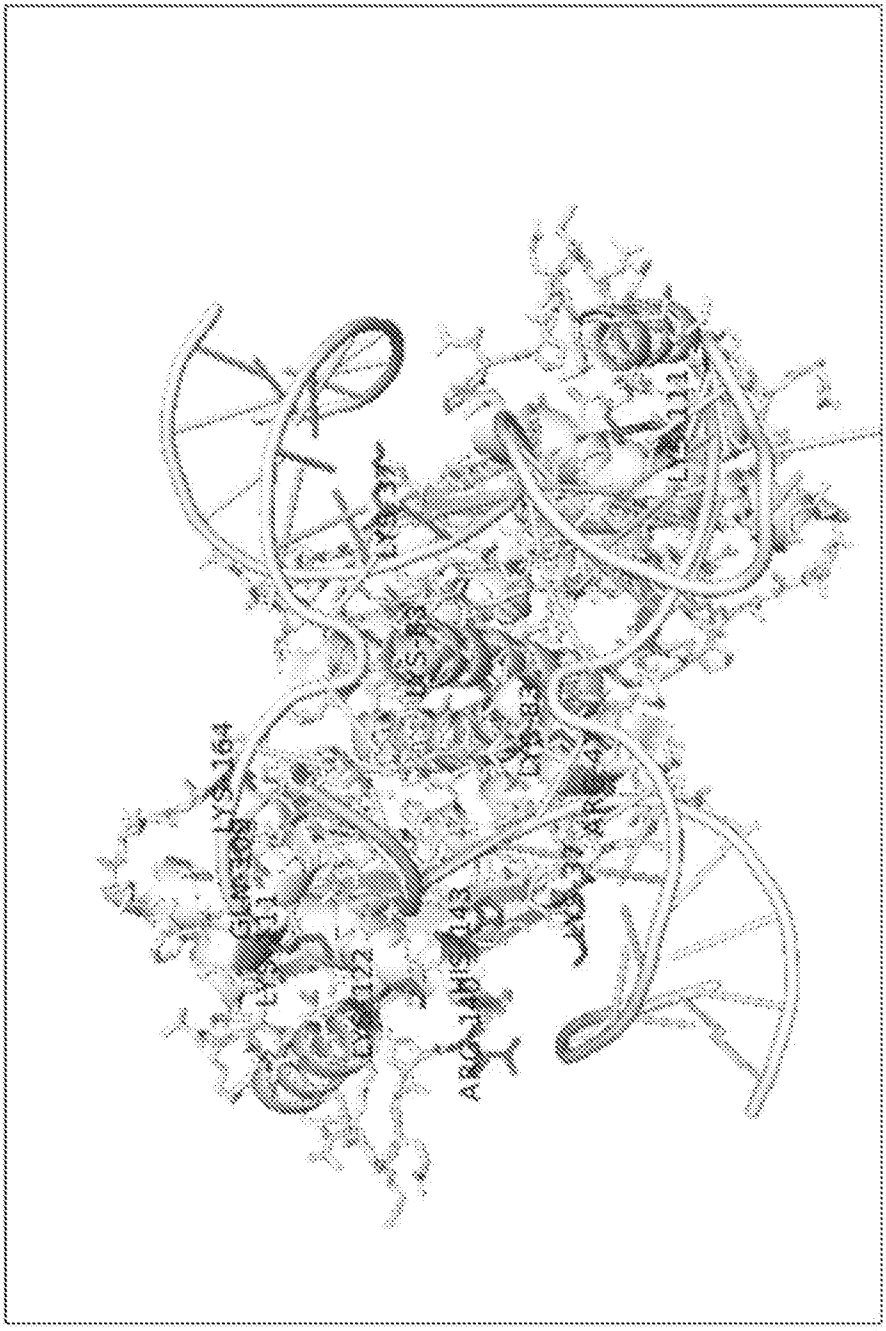


FIG. 112

Recombinasa específica de sitio *XerD [Roseburia intestinalis M50/1]*

ID de secuencia: emb|CBL10059.1| Longitud: 416 Número de coincidencias: 1

Siguiente coincidencia

Siguiente coincidencia previa

Rango 1: 179 to 292 Gráficos Gráficos

Puntuación	Esperado	Método	Identidades	Positivos	Brechas
46.2 bits(108)	0.14	Ajust. a matriz composicional	42/127(33%)	59/127(46%)	17/127(13%)

Query	750	FAKGMMGKF----MLMTLMTMGLFSPEMLAKTSIKLMGQAELTYRFKSRMMAMRLGEK	805
		F +G KP LMWTG+ E LA + ++ +L + R R+G K	
Sbjct	175	FRGVMGKFLSYICFEVLMTGMRGELLALSPADIDIDMKLI------SIKRTYQRIGGK	229
Query	806	MLMKLLKDQKT----PIFDTLYQELYDYVNRRLSDLSDEARALLEMVITKEVSHIIKD	861
		+ K +K+ PIFD L QEL DY+ R D +DE L P V +SE+I+	
Sbjct	230	DVTTSPKTRKSKRTIPIFDFLCQELSDYIQSRYMLD-ADE---RLFP-VTKSYLSEMIRG	285
Query	862	KRTTSDK 868	
		+ T K	
Sbjct	286	CKITGAK 292	

FIG. 113

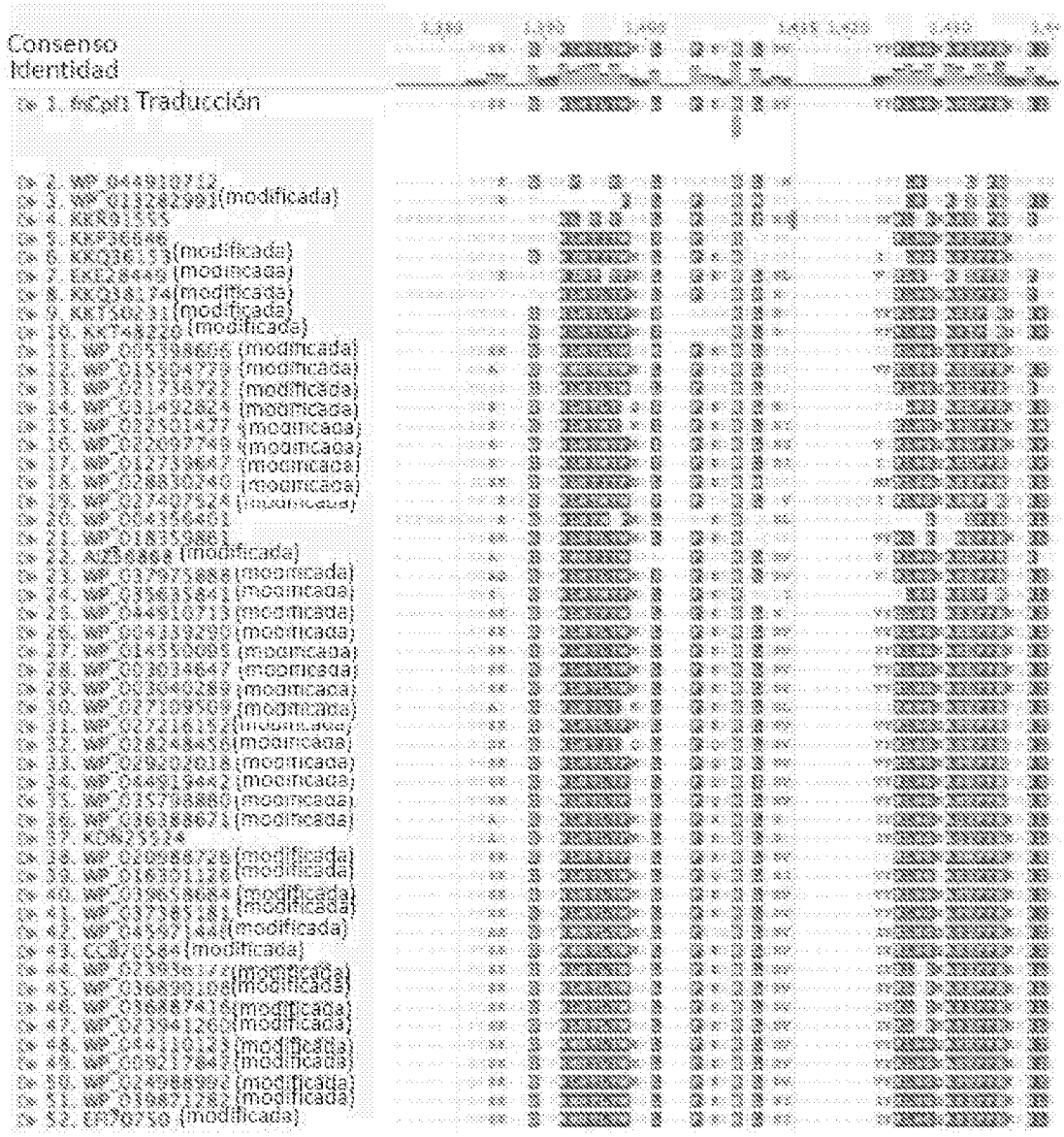

Consenso
Identidad
1. 6vCas Traducción
2. WP_044918712
3. WP_013282393 (modificada)
4. KKK01555
5. KKP36646
6. KKQ36153 (modificada)
7. EKE28448 (modificada)
8. KKQ38174 (modificada)
9. KKT50231 (modificada)
10. KKT48220 (modificada)
11. WP_015788606 (modificada)
12. WP_015584779 (modificada)
13. WP_021736782 (modificada)
14. WP_031492824 (modificada)
15. WP_028501477 (modificada)
16. WP_022097748 (modificada)
17. WP_012739867 (modificada)
18. WP_028830240 (modificada)
19. WP_027407524 (modificada)
20. WP_044356401
21. WP_018359861
22. KZB6868 (modificada)
23. WP_017975888 (modificada)
24. WP_035615841 (modificada)
25. WP_044910713 (modificada)
26. WP_004139280 (modificada)
27. WP_014550088 (modificada)
28. WP_053034647 (modificada)
29. WP_053040288 (modificada)
30. WP_027109509 (modificada)
31. WP_027216182 (modificada)
32. WP_018248456 (modificada)
33. WP_028202018 (modificada)
34. WP_044919442 (modificada)
35. WP_015798880 (modificada)
36. WP_016188671 (modificada)
37. KDN25524
38. WP_028988728 (modificada)
39. WP_016391128 (modificada)
40. WP_031918684 (modificada)
41. WP_037585281 (modificada)
42. WP_045677444 (modificada)
43. CCB70584 (modificada)
44. WP_023893172 (modificada)
45. WP_016890108 (modificada)
46. WP_028887416 (modificada)
47. WP_023941280 (modificada)
48. WP_044110283 (modificada)
49. WP_052217845 (modificada)
50. WP_024988392 (modificada)
51. WP_019871282 (modificada)
52. EFB70750 (modificada)

FIG. 114

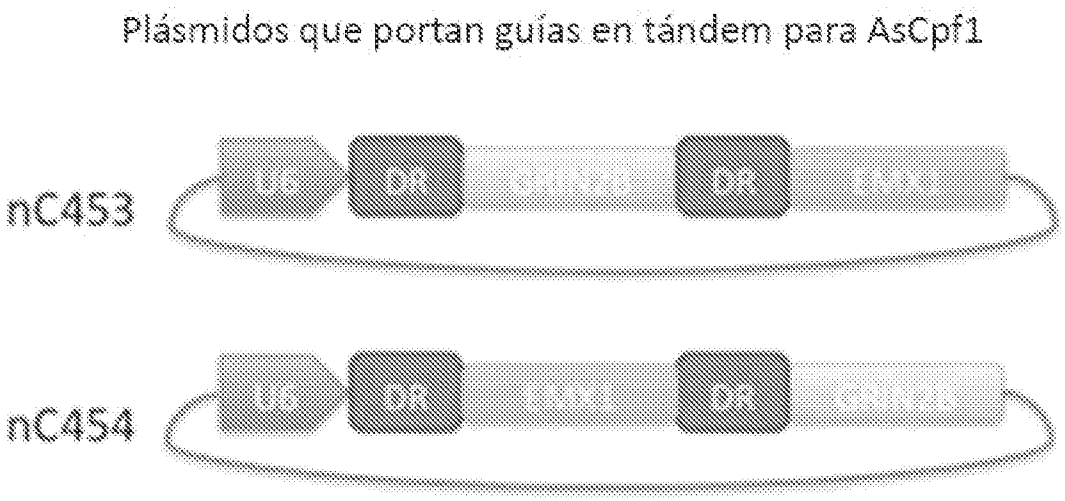

Plásmidos que portan guías en tándem para AsCpf1
nC453
nC454

FIG. 115A

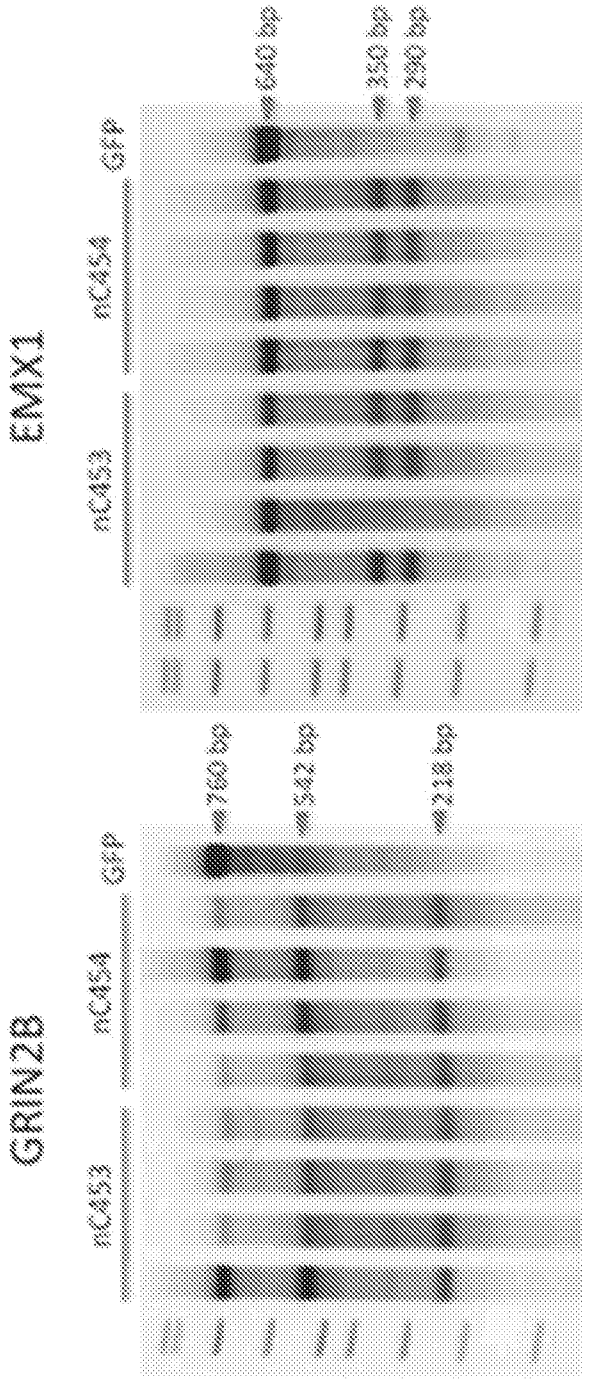

EMX1
nC453
nC454
GFP
640 bp
350 bp
280 bp
GRIN2B
760 bp
342 bp
218 bp

FIG. 115B

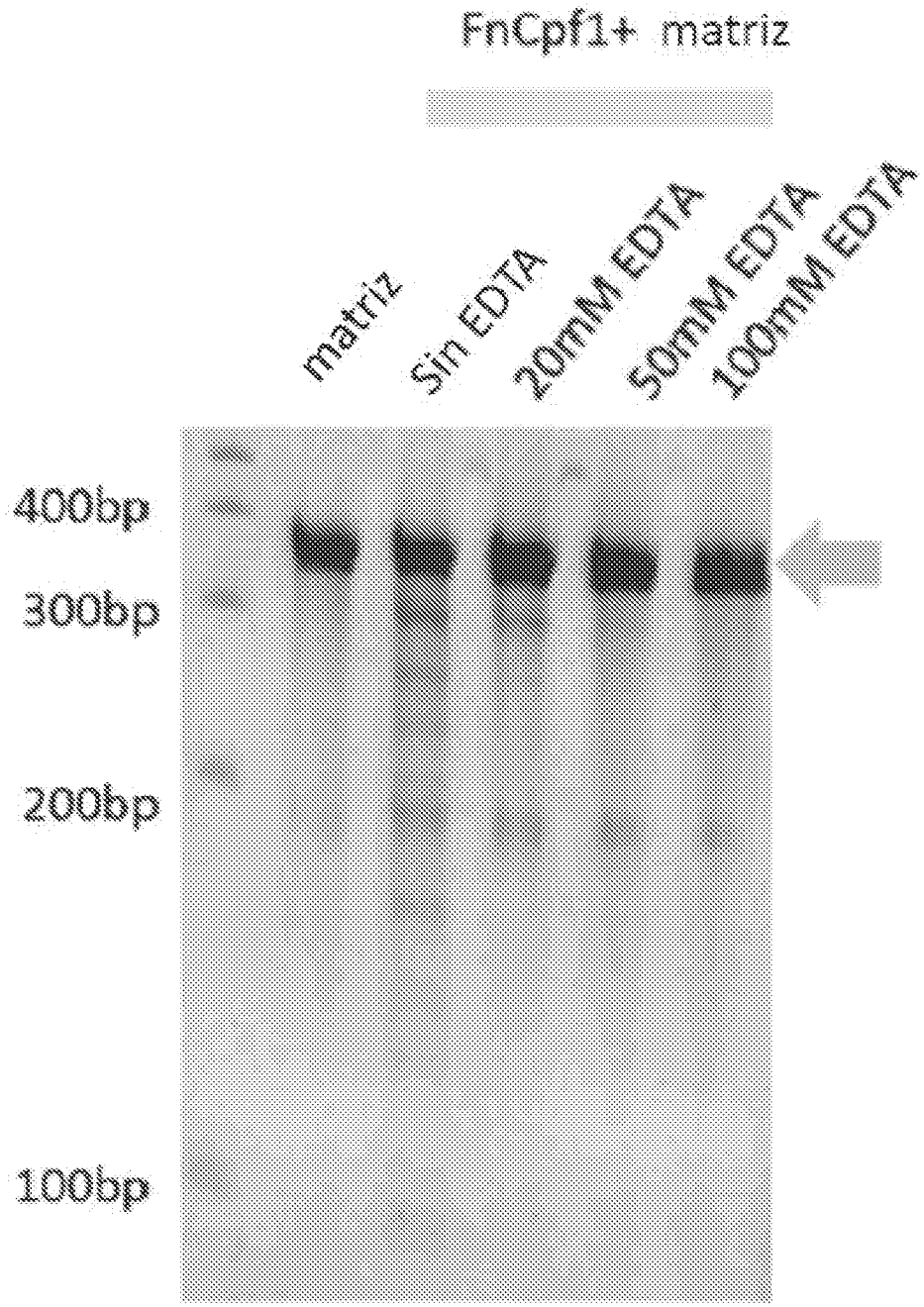

FnCpf1+ matriz
matriz
Sin EDTA
20mM EDTA
50mM EDTA
100mM EDTA
400bp
300bp
200bp
100bp

FIG. 116