

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 974 229**

51 Int. Cl.:

C12N 15/10	(2006.01)
C12N 15/11	(2006.01)
C12N 15/90	(2006.01)
C40B 40/08	(2006.01)
C40B 50/06	(2006.01)
C12N 9/22	(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.02.2015** E 20161652 (1)

97 Fecha y número de publicación de la concesión europea: **10.01.2024** EP 3690044

54 Título: **Ingeniería del genoma multiplex mediante CRISPR**

30 Prioridad:

11.02.2014 US 201461938608 P

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

26.06.2024

73 Titular/es:

**THE REGENTS OF THE UNIVERSITY OF
COLORADO, A BODY CORPORATE (100.0%)
1800 Grant Street, 8th Floor
Denver, CO 80203, US**

72 Inventor/es:

**GILL, RYAN, T. y
GARST, ANDREW**

74 Agente/Representante:

SÁEZ MAESO, Ana

Observaciones:

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

ES 2 974 229 T3

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

Ingeniería del genoma multiplex mediante CRISPR

5 Antecedentes de la invención

La manipulación racional de grandes construcciones de ADN es un desafío central para los esfuerzos actuales de la biología sintética y la ingeniería del genoma. En los años recientes, se han desarrollado una variedad de tecnologías para abordar este desafío y aumentar la especificidad y la velocidad con la que pueden generarse mutaciones. Adicionalmente, las mutaciones adaptativas son un conductor central de la evolución, pero su abundancia y contribución relativa a los fenotipos celulares son poco conocidos incluso en los organismos mejores estudiados. Esto puede atribuirse en gran parte a los desafíos técnicos asociados con la observación y la reconstrucción de estos genotipos y la correlación de su presencia con el fenotipo de interés. Por ejemplo, los métodos de edición del genoma que se basan en mutagénesis aleatoria conducen a genotipos complejos que consisten de muchas mutaciones, cuya contribución relativa de cada uno de ellos es difícil de desconvolucionar. Además, las interacciones epistáticas entre alelos son difíciles de asignar debido a la falta de información referente a las mutaciones individuales. Jiang et al. (NATURE BIOTECHNOLOGY, vol. 31, no.3, 2013, páginas 233-239) divulgan constructos genéticos para la introducción de cambios específicos en el genoma de una célula en las que los componentes del sistema CRISPR/Cas y la plantilla de edición se transfectan como entidades separadas en células.

20 Resumen de la invención

La invención proporciona una biblioteca que comprende una población de células que comprende:

25 un vector que comprende:

- (i) un casete de edición que incluye una región que es homóloga a una región objetivo de un ácido nucleico en una célula e incluye una mutación de al menos un nucleótido con respecto a la región objetivo y un motivo adyacente protoespaciador (PAM) mutado que no es reconocido por un sistema CRISPR;
- 30 (ii) un promotor; y
- (iii) al menos un ARN guía (ARNg), comprendiendo el ARNg una región complementaria a una porción de la región objetivo, y una región que recluta una endonucleasa, y en el que el promotor y el ácido nucleico que codifica el ARNg están operativamente unidos.

35 Las Repeticiones Palindrómicas Cortas Agrupadas Regularmente (CRISPR) existen en muchos genomas bacterianos y se ha descubierto que juegan un papel importante en la inmunidad bacteriana adaptativa. La transcripción de estas series da lugar a ARN CRISPR que dirigen la unión específica de la secuencia de los complejos CRISPR/cas a los objetivos de ADN en las células para la represión génica o la escisión del ADN. La especificidad de estos complejos permite nuevas aplicaciones in vivo para la ingeniería de cepas.

40 En la presente descripción se describen métodos de manipulación racional y multiplexada de cromosomas dentro de marcos de lectura abiertos (por ejemplo, para generar bibliotecas de proteínas) o dentro de múltiples genes en cualquier segmento de un cromosoma, en el que se usan diversos sistemas CRISPR. Estos métodos proporcionan una ingeniería genómica combinatoria más eficiente que las disponibles previamente.

45 Ampliar las capacidades de multiplexación de CRISPR presenta un desafío tecnológico corriente y permitiría usar estos sistemas para generar bibliotecas racionales en formato de alto rendimiento. Tales avances tienen implicaciones de amplio alcance para los campos de la ingeniería metabólica y proteica que buscan refactorizar redes genéticas complejas para una producción óptima.

50 Los métodos comprenden la introducción de los componentes del sistema CRISPR, que incluyen la nucleasa Cas9 asociada a CRISPR y un ARN guía específico de secuencia (ARNg) en las células, que resulta en roturas de doble hebras dirigidas a la secuencia mediante el uso de la capacidad del sistema CRISPR para inducir tales rupturas. Los componentes del sistema CRISPR, que incluyen la nucleasa Cas9 asociada a CRISPR y un ARN guía específico de secuencia (ARNg), pueden introducirse en las células codificadas en uno o más vectores, como un plásmido. Los casetes de recombinación de ADN u oligonucleótidos de edición pueden diseñarse racionalmente para incluir una mutación deseada dentro de un locus objetivo y una mutación, en una ubicación común fuera del locus objetivo que puede reconocerse por el sistema CRISPR. Los métodos descritos pueden usarse para muchas aplicaciones, que incluyen la alteración de una trayectoria de interés.

60 En una realización, el método es un método de ingeniería del genoma, que comprende: (a) introducir en las células un vector que codifica: (i) un casete de edición que incluye una región que es homóloga a la región objetivo del ácido nucleico en la célula e incluye una mutación (referida a una mutación deseada) de al menos un nucleótido con relación a la región objetivo, como una mutación de al menos un nucleótido en al menos un codón con relación a la región objetivo, y una mutación de motivo adyacente protoespaciador (PAM); (ii) un promotor; y (iii) al menos un ARN guía (ARNg), el ARNg comprende: (a) una región (ARN) complementaria a una porción de la región objetivo; y (b) una

- región (ARN) que recluta una nucleasa Cas9, produciendo de esta manera células que comprenden el vector; (b) mantener células que comprenden el vector bajo condiciones bajo las cuales se expresa Cas9, en donde la nucleasa Cas9 se codifica en el vector, codificada en un segundo vector o codificada en el genoma de las células, que resulta en la producción de las células que comprenden el vector y no comprende la mutación PAM y las células que comprenden el vector y la mutación PAM; (c) cultivar el producto (b) bajo condiciones apropiadas para la viabilidad celular, produciendo de esta manera células viables; (d) obtener células viables producidas en (c); y (e) secuenciar el oligonucleótido de edición del vector de al menos una célula viable obtenida en (d) e identificar la mutación de al menos un codón.
- En otra realización, el método es un método de ingeniería del genoma mediante la recombinación enriquecida CRISPR rastreable, que comprende: (a) introducir en una primera población de células un vector que codifica: (i) al menos un casete de edición que comprende: (a) una región homóloga a una región objetivo de un ácido nucleico y que comprende una mutación de al menos un nucleótido con relación a la región objetivo, tal como una mutación de al menos un nucleótido en al menos un codón relativo a la región objetivo, y (b) una mutación de motivo adyacente protoespaciador (PAM); (ii) al menos un promotor; y (iii) al menos un ARN guía (ARNg) que comprende: (a) una región (ARN) complementaria a una porción de la región objetivo y (b) una región (ARN) que recluta una nucleasa Cas9, produciendo de esta manera una segunda población de células que comprenden el vector; (b) mantener la segunda población de células en condiciones en cuyo caso se expresa la nucleasa Cas9, en donde la nucleasa Cas9 se codifica en el vector, un segundo vector o en el genoma de las células de la segunda población de células, que resulta en la escisión del ADN en las células que no comprenden la mutación PAM y la muerte de tales células; (c) obtener células viables producidas en (b); y (d) identificar la mutación de al menos un codón secuenciando el oligonucleótido de edición del vector de al menos una célula de la segunda población de células.
- Cualquiera de las realizaciones anteriores puede comprender además sintetizar y/u obtener una población de oligonucleótidos de edición. Cualquiera de las realizaciones puede comprender además amplificar la población de oligonucleótidos de edición. En cualquiera de las realizaciones, el vector puede comprender además un separador, al menos dos sitios de cebadores o tanto un separador como al menos dos sitios de cebadores. En algunas realizaciones, el casete de edición comprende una región objetivo que comprende una mutación de al menos un codón dentro de los 100 nucleótidos de la mutación PAM.
- Se describe además un vector que comprende:
- (i) un casete de edición que incluye una región que es homóloga a una región objetivo de un ácido nucleico en una célula e incluye una mutación (referida a una mutación deseada) de al menos un nucleótido con relación a la región objetivo, y una mutación con motivo adyacente protoespaciador (PAM);
- (ii) un promotor; y
- (iii) al menos un ARN guía (ARNg) que comprende: (a) una región (ARN) complementaria a una porción de la región objetivo; y (b) una región (ARN) que recluta una nucleasa Cas9.
- Una realización adicional es un vector que comprende:
- (i) un casete de edición que incluye una región que es homóloga a una región objetivo de un ácido nucleico en una célula e incluye una mutación (referida a una mutación deseada) de al menos un nucleótido en al menos un codón con relación a la región objetivo, y una mutación de motivo adyacente protoespaciador (PAM);
- (ii) un promotor; y
- (iii) al menos un ARN guía (ARNg) que comprende: (a) una región (ARN) complementaria a una porción de la región objetivo; y (b) una región (ARN) que recluta una nucleasa Cas9.
- Una realización adicional es un vector que comprende:
- (i) al menos un casete de edición que comprende: (a) una región homóloga a una región objetivo de un ácido nucleico que comprende una mutación de al menos un nucleótido con relación a la región objetivo y (b) una mutación de motivo adyacente protoespaciador (PAM);
- (ii) al menos un promotor; y
- (iii) al menos un ARN guía (ARNg) que comprende: (a) una región (ARN) complementaria a una porción de la región objetivo y (b) una región (ARN) que recluta una nucleasa Cas9.
- Otra realización del vector es un vector que comprende:
- (i) al menos un casete de edición que comprende: (a) una región homóloga a una región objetivo de un ácido nucleico que comprende una mutación de al menos un nucleótido en al menos un codón con relación a la región objetivo y (b) una mutación de motivo adyacente protoespaciador (PAM);
- (ii) al menos un promotor; y
- (iii) al menos un ARN guía (ARNg) que comprende: (a) una región (ARN) complementaria a una porción de la región objetivo y (b) una región (ARN) que recluta una nucleasa Cas9.

En cualquiera de las realizaciones, el vector puede comprender además un separador; al menos dos sitios de cebadores; o un separador y al menos dos sitios de cebadores. En aquellos vectores en los que la mutación es de al menos un nucleótido en al menos un codón, el casete de edición de la mutación puede estar, por ejemplo, dentro de los 100 nucleótidos de la mutación PAM.

Se describe además una biblioteca que comprende una población de células producidas por los métodos descritos en la presente descripción. Una biblioteca de una población de células puede comprender células que tengan cualquiera de los vectores descritos en la presente descripción. Por ejemplo, una población de células puede comprender un vector que comprende:

- (i) un casete de edición que incluye una región que es homóloga a una región objetivo de un ácido nucleico en una célula e incluye una mutación (referida a una mutación deseada) de al menos un nucleótido con relación a la región objetivo, y una mutación con motivo adyacente protoespaciador (PAM);
- (ii) un promotor; y
- (iii) al menos un ARN guía (ARNg) que comprende: (a) una región (ARN) complementaria a una porción de la región objetivo; y (b) una región (ARN) que recluta una nucleasa Cas9.

En una realización adicional, una población de células puede comprender un vector que comprende:

- (i) un casete de edición que incluye una región que es homóloga a una región objetivo de un ácido nucleico en una célula e incluye una mutación (referida a una mutación deseada) de al menos un nucleótido en al menos un codón con relación a la región objetivo, y una mutación de motivo adyacente protoespaciador (PAM);
- (ii) un promotor; y
- (iii) al menos un ARN guía (ARNg) que comprende: (a) una región (ARN) complementaria a una porción de la región objetivo; y (b) una región (ARN) que recluta una nucleasa Cas9.

En una realización adicional, el método es un método de ingeniería racional de proteínas asistida por CRISPR (ingeniería genómica combinatoria), que comprende:

- (a) construir una biblioteca de donantes, que comprende ADN recombinante, como cromosomas recombinantes o ADN recombinante en plásmidos, mediante la introducción, como por co-transformación, una población de primeras células (i) uno o más oligonucleótidos de edición, como oligonucleótidos diseñados racionalmente, que eliminan la pareja de un primer motivo adyacente protoespaciador único (PAM) con mutación en al menos un codón en un gen adyacente al PAM (el gen adyacente) y (b) un ARN guía (ARNg) que se dirige a una secuencia nucleotídica 5' del marco de lectura abierto de un cromosoma, produciendo de esta manera una biblioteca de donantes que comprende una población de primeras células que comprenden cromosomas recombinantes que tienen mutaciones codónicas dirigidas;
- (b) amplificar la biblioteca de donantes construida en (a), como por amplificación por PCR, de cromosomas recombinantes que usa una característica sintética de los oligonucleótidos de edición e incorpora simultáneamente una segunda delección PAM (delección PAM de destino) en el extremo 3' del gen, de esta manera el acoplamiento, como el acoplamiento covalente, mutaciones de codón dirigidas directamente a la eliminación del PAM de destino y producir una biblioteca de donantes recuperada que lleva la eliminación del PAM de destino y mutaciones de codón de destino; e
- (c) introducir (por ejemplo, co-transformar) la biblioteca donante que lleva la delección PAM de destino y las mutaciones codónicas dirigidas y un plásmido de ARNg de destino en una población de segundas células, que típicamente son una población de células vírgenes, produciendo de esta manera una biblioteca de destino que comprende mutaciones codónicas dirigidas.

La población de primeras células y la población de segundas células (por ejemplo, una población de células vírgenes) son típicamente una población en la que todas las células son del mismo tipo y pueden ser procariontas o eucariotas, como por ejemplo, pero no se limitan a, bacterias, células de mamíferos, células vegetales, células de insectos.

En algunas realizaciones, el método comprende además mantener la biblioteca de destino en condiciones bajo las cuales se produce la proteína.

En algunas realizaciones, la primera célula expresa un polipéptido con actividad nucleasa Cas9. En algunas realizaciones, el polipéptido con actividad de nucleasa Cas9 se expresa bajo el control de un promotor inducible.

En algunas realizaciones, los oligonucleótidos de edición son complementarios a un ácido nucleico dirigido (uno, uno o más, al menos uno) presente en la primera célula. En algunas realizaciones, los oligonucleótidos de edición se dirigen a más de un sitio o locus objetivo en la primera célula. En algunas realizaciones, la secuencia del ácido nucleico de los oligonucleótidos de edición [codón deseado] comprende una o más sustituciones, deleciones, inserciones o cualquier combinación de sustituciones, deleciones e inserciones con relación al ácido nucleico objetivo. En algunas realizaciones, los oligonucleótidos de edición se diseñan racionalmente; en realizaciones adicionales, se producen por mutagénesis aleatoria o mediante el uso de oligonucleótidos cebadores degenerados. En algunas realizaciones, los oligonucleótidos de edición se derivan de una colección de ácidos nucleicos (biblioteca).

En algunas realizaciones, el ARNg se codifica en un plásmido. En algunas realizaciones, el oligonucleótido de edición y el ARNg se introducen en la primera célula mediante transformación, mediante la co-transformación del oligonucleótido de edición y el ARNg guía. En algunas realizaciones, el oligonucleótido de edición y el ARNg se introducen secuencialmente en la primera célula. En otras realizaciones, el oligonucleótido de edición y el ARNg se introducen simultáneamente en la primera célula.

En algunas realizaciones, la recuperación de la biblioteca de donantes comprende además (a) seleccionar células para la incorporación del oligonucleótido de edición y (b) seleccionar células que se confirma que han incorporado el oligonucleótido de edición. En algunas realizaciones, la recuperación de la biblioteca de donantes comprende además el procesamiento de la biblioteca de donantes recuperada.

En algunas realizaciones, la célula de destino/célula virgen expresa un polipéptido con actividad de nucleasa Cas9. En algunas realizaciones, el polipéptido con actividad de nucleasa Cas9 se expresa bajo el control de un promotor inducible.

Se describe además un método de ingeniería racional de proteínas asistida por CRISPR, que comprende:

(a) introducir (por ejemplo, co-transformar) (i) casetes de edición de ADNds sintéticos que comprenden oligonucleótidos de edición y (ii) un vector que expresa un ARN guía (ARNg) que se dirige a la secuencia genómica justo aguas arriba de un gen de interés en una población de primeras células, en condiciones bajo las cuales ocurre la recombinación multiplexada y el enriquecimiento selectivo mediante el ARNg de los oligonucleótidos de edición, produciendo de esta manera una biblioteca de donantes;

(b) amplificar la biblioteca de donantes con un oligonucleótido que elimina un motivo adyacente protoespaciador (PAM) adyacente al extremo 3' del gen de interés (PAM de destino), produciendo de esta manera una biblioteca de donantes amplificada que comprende, casetes de edición de ADNds de los cuales el PAM de destino se ha eliminado (con una delección PAM 3'), mutaciones de codones racionales y un sitio P1;

(c) procesar la biblioteca de donantes amplificada con una enzima, como una enzima de restricción (por ejemplo, BsaI), para eliminar el sitio P1; y

(d) co-transformar una población de células vírgenes con la biblioteca de donantes amplificada procesada en (c) y el ARNg de destino, produciendo de esta manera una población de células co-transformadas que comprenden, casetes de edición de ADNds de los cuales se ha eliminado el PAM de destino (con una eliminación del PAM del extremo 3'), las mutaciones de codones racionales y el ARNg de destino.

En todas las realizaciones descritas, una mutación puede ser de cualquier tipo deseado, como una o más inserciones, deleciones, sustituciones o cualquier combinación de dos o tres de las anteriores (por ejemplo, inserción y eliminación; inserción y sustitución; eliminación y sustitución; sustitución e inserción; inserción, eliminación y sustitución). Las inserciones, deleciones y sustituciones pueden ser de cualquier número de nucleótidos. Pueden estar en codones (regiones codificantes) y/o en regiones no codificantes.

Breve descripción de las figuras

Las figuras 1A y 1B presentan una visión general de la ingeniería de proteínas racional asistida por CRISPR (CARPE). La figura 1A muestra un esquema de la construcción de la biblioteca de donantes. Los casetes de edición de ADNds sintéticos se co-transformaron con un vector que expresa un ARN guía (ARNg) dirigido a la secuencia genómica aguas arriba del gen de interés. La co-transformación generó una biblioteca de donantes a través de la recombinación multiplexada de los oligonucleótidos de edición, que se enriquecieron selectivamente por el ARNg. La biblioteca de donantes se amplificó después mediante el uso de un oligonucleótido que muta (elimina) un PAM adyacente al extremo 3' del gen (PAM de destino). La figura 1B muestra un esquema de la generación de la biblioteca de proteínas final. La biblioteca de donantes se procesó con BsaI para eliminar el sitio de P1, y la biblioteca de casetes de ADNds con la delección 3'PAM y las mutaciones de codones racionales se co-transformaron con el ARNg de destino para generar la biblioteca de proteínas final.

La figura 2 presenta la secuencia de ADN de clones de la construcción de la biblioteca de donantes *galK* que confirma la incorporación de la característica P1 del oligonucleótido de edición con alta eficiencia, así como también la mutación en la posición del codón objetivo (subrayado). La secuencia de P1 se proporciona mediante la SQ ID NO: 1.

La figura 3A muestra el diseño del cebador. La figura 3B muestra la densidad esperada con relación al número de cebadores.

La figura 4A presenta los resultados del enlazador y la construcción. La figura 4B muestra 10 ediciones relacionadas con el seguimiento en base al PCR de emulsión.

La figura 5 es un esquema de la edición de proteínas racional para la ingeniería metabólica.

La figura 6 es un esquema de la generación de bibliotecas de proteínas racionales enriquecidas CRISPR.

La figura 7 es un esquema de la configuración y demostración de CARPE.

La figura 8 muestra las estrategias para la co-selección CRISPR iterativa.

La figura 9 presenta una estrategia para la ingeniería de proteínas multiplexada mediante el uso de CARPE.

La figura 10 muestra la construcción de una biblioteca donante *galK* mediante el uso de CARPE.

La figura 11A muestra un esquema de edición en base a CRISPR multiplex mediante el uso de CARPE.

La figura 11B muestra un esquema de edición en base a CRISPR multiplex mediante el uso de la ingeniería genómica mediante la recombinación enriquecida CRISPR rastreada (GEN-TraCER).

La figura 12 muestra un vector GEN-TraCER representativo (construcción) que incluye un casete de edición para editar el codón 24 de *galK*, un promotor y separador.

La figura 13 muestra los resultados de una edición de *galK* mediante el uso de GEN-TraCER. Los paneles superiores muestran los resultados de la secuenciación del ADN del cromosoma y el vector (plásmido) a partir de las células que se habían transformado con el codón 24 *galK* que edita el vector GEN-TraCER, que indica que el casete de edición (oligonucleótido) en el vector puede secuenciarse como un "código de barras trans" que permite el seguimiento de alta eficiencia de la edición genómica deseada (mutación). Los paneles de la parte inferior muestran cromatografías de la secuenciación del ADN de las células que exhiben el fenotipo de tipo salvaje no editado (rojo). El método permite la identificación de las células con cromosomas múltiples que portan tanto el alelo no editado de tipo salvaje como el alelo editado/mutado.

Las figuras 14A-14C muestran esquemas del GEN-TraCER. La figura 14A muestra una visión general de los componentes del diseño. Los casetes GEN-TraCER contienen secuencias de ARN guía (ARNg) para dirigir a un sitio específico en el genoma celular y causar la escisión del ADNds. Una región de homología complementaria a la región objetivo muta el PAM y otros sitios deseados cercanos. Las células que se someten a la recombinación se enriquecen selectivamente a gran abundancia. La secuenciación del casete de edición GEN-TraCER en el vector permite el seguimiento de las ediciones/mutaciones genómicas. La figura 14B muestra un ejemplo de diseño del casete de edición para el gen de *E. coli galK* en el codón 145. El PAM se elimina con la mutación PAM disponible más cercana que puede realizarse para el cambio sinónimo en la posición PAM disponible más cercana. Esto permite la mutagénesis con una "cicatriz silenciosa" de 1-2 nucleótidos en el sitio de delección PAM. La figura 14C muestra que los casetes GEN-TraCER pueden sintetizarse mediante el uso de los métodos de síntesis en base a matrices, lo que permite entonces la síntesis paralela de al menos 10^4 - 10^6 casetes para la orientación sistemática y la evaluación simultánea de la aptitud para miles de mutaciones en una escala de todo el genoma.

La figura 15A muestra una visión general de los vectores GEN-TraCER. La figura 15B muestra una porción de un GEN-TraCER representativo para la generación de una mutación Y145* en el gen *galK* de *E. coli* en el que la mutación PAM y el codón que está mutado están separados por 17 nucleótidos. La secuencia de ácido nucleico de la porción del GEN-TraCER representativo se proporciona mediante la SEQ ID NO: 28 y el complemento inverso se proporciona mediante la SEQ ID NO: 33.

Las figuras 16A-16C presentan controles para el diseño del GEN-TraCER. La figura 16A muestra el efecto del tamaño del casete de edición en la eficiencia del método. La figura 16B muestra el efecto de la distancia entre la mutación/delección del PAM y la mutación deseada sobre la eficiencia del método. La figura 16C muestra el efecto de la presencia o ausencia del sistema MutS en la eficiencia del método.

Descripción detallada

Los sistemas CRISPR de las bacterias y Archaea han surgido como herramientas nuevas poderosas para la edición precisa del genoma. El sistema CRISPR tipo II de *Streptococcus pyogenes* (*S. pyogenes*) ha sido particularmente bien caracterizado *in vitro*, y se han establecido reglas de diseño simples para reprogramar su actividad de unión de ADN bicatenario (ADNds) Jinek et al. *Science* (2012) 337 (6096): 816-821). El uso de los métodos de edición del genoma mediados por CRISPR se ha acumulado rápidamente en la literatura en una amplia variedad de organismos, que incluyen las bacterias (Cong et al. *Science* (2013) 339 (6121): 819-823), *Saccharomyces cerevisiae* (DiCarlo et al. *Nucleic Acids Res.* (2013) 41:4336-4343), *Caenorhabditis elegans* (Waaaijers et al. *Genetics* (2013) 195: 1187-1191) y varias líneas celulares de mamíferos (Cong et al. *Science* (2013) 339 (6121): 819-823; Wang et al. *Cell* (2013) 153:910-918). Al igual que otras tecnologías de edición del genoma en base a endonucleasas, como las nucleasas de dedo de zinc (ZFN), las nucleasas de referencia y TALENS, la capacidad de los sistemas CRISPR para mediar la edición precisa del genoma proviene de la naturaleza altamente específica del reconocimiento de objetivos. Por ejemplo, el sistema CRISPR tipo I de *Escherichia coli* y el sistema de *S. pyogenes* requiere una complementariedad perfecta entre el ARN CRISPR (ARNcr) y un objetivo de reconocimiento de 14-15 pares de bases, lo que sugiere que las funciones

inmunes de los sistemas CRISPR se emplean naturalmente (Jinek et al. Science (2012) 337(6096): 816-821; Brouns et al. Science (2008) 321:960-964; Semenova et al. PNAS (2011) 108: 10098-10103).

5 En la presente descripción se describen métodos para la edición del genoma que emplean una endonucleasa, como la nucleasa Cas9 codificada por un gen *cas9*, para realizar la evolución del genoma dirigida/producir cambios (deleciones, sustituciones, adiciones) en el ADN, como el ADN genómico. El gen *cas9* puede obtenerse de cualquier fuente, como una bacteria, como la bacteria *S. pyogenes*. La secuencia de ácido nucleico de la *cas9* y/o la secuencia de aminoácidos de Cas9 puede estar mutada, con relación a la secuencia de una ocurrencia natural *cas9* y/o Cas9; las mutaciones pueden ser, por ejemplo, una o más inserciones, deleciones, sustituciones o cualquier combinación de dos o tres de las anteriores. En tales realizaciones, la Cas9 mutada resultante puede tener actividad nucleasa mejorada o reducida con relación a la Cas9 natural.

15 Las figuras 1A, 1B y 11A presentan un método de edición del genoma mediado por CRISPR referido como CRISPR asistido racional proteína ingeniería (CARPE). CARPE es un proceso de construcción de dos etapas que se basa en la generación de bibliotecas "donantes" y "de destino" que incorporan mutaciones directas de casetes de edición de ADN monocatenario (ADNss) o ADN bicatenario (ADNds) directamente en el genoma. En la primera etapa de la construcción del donante (Figura 1A), los oligos de edición diseñados racionalmente se cotransforman en células con un ARN guía (ARNg) que se hibrida con/se dirige a una secuencia de ADN objetivo, como una secuencia 5' de un marco de lectura abierto u otra secuencia de interés. Una innovación clave de CARPE está en el diseño de los oligonucleótidos de edición que combinan la delección o mutación de un único motivo adyacente protoespaciador (PAM) con la mutación de uno o más codones deseados en el gen adyacente, lo que permite de esta manera, la generación de toda la biblioteca de donantes en una sola transformación. La biblioteca de donantes se recupera después mediante la amplificación de los cromosomas recombinantes, *por ejemplo* mediante una reacción de PCR, mediante el uso de una característica sintética del oligonucleótido de edición; una segunda delección o mutación PAM se incorpora simultáneamente en el extremo 3' del gen. De este modo, este enfoque acopla de forma covalente las mutaciones dirigidas al codón directamente a una delección PAM. En la segunda etapa de CARPE (Figura 1B), las bibliotecas de donantes amplificadas por PCR que llevan la eliminación/mutación PAM de destino y las mutaciones dirigidas (mutación(es) deseada(s) de uno o más nucleótidos, como uno o más nucleótidos en uno o más codones) se cotransforman en las células vírgenes con un vector de ARNg de destino para generar una población de células que expresan una biblioteca de proteínas diseñada racionalmente.

35 En el sistema CRISPR, la trans-activación CRISPR (tracrARN) y el ARN separador (crARN) guían la selección de una región objetivo. Como se usa en la presente, una región objetivo se refiere a cualquier locus en el ácido nucleico de una célula o una población de células en la que se desea una mutación de al menos un nucleótido, como una mutación de al menos un nucleótido en al menos un codón (uno o más codones). La región objetivo puede ser, por ejemplo, un locus genómico (secuencia genómica objetivo) o un locus extracromosómico. El tracrARN y el crARN pueden expresarse como una molécula de ARN quimérica única, referida como ARN de guía única, ARN guía o ARNg. La secuencia de ácido nucleico del ARNg comprende una primera secuencia de ácido nucleico, referida además a una primera región, que es complementaria a una región de la región objetivo y una segunda secuencia de ácido nucleico, referida además a una segunda región, que forma una estructura de lazo de tallo y funciones para reclutar Cas9 en la región objetivo. En algunas realizaciones, la primera región del ARNg es complementaria a una región aguas arriba de la secuencia genómica objetivo. En algunas realizaciones, la primera región del ARNg es complementaria al menos a una porción de la región objetivo. La primera región del ARNg puede ser completamente complementaria (100 % complementaria) a la secuencia genómica objetivo o incluir uno o más desajustes, proporcionando que sea suficientemente complementaria a la secuencia genómica objetivo para hibridar/guiar específicamente y reclutar Cas9. En algunas realizaciones, la primera región del ARNg tiene al menos 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o al menos 30 nucleótidos de longitud. En algunas realizaciones, la primera región del ARNg tiene al menos 20 nucleótidos de longitud. En algunas realizaciones, la estructura de lazo del tallo que se forma por la segunda secuencia del ácido nucleico es de al menos 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, o 100 nucleótidos de longitud. En realizaciones específicas, la estructura de lazo del tallo tiene una longitud de 80 a 90 u 82 a 85 nucleótidos y, en realizaciones específicas adicionales, la segunda región del ARNg que forma una estructura de lazo del tallo tiene 83 nucleótidos de longitud.

55 En algunas realizaciones, la secuencia del ARNg (de la biblioteca de donantes) que se introduce en la primera célula mediante el uso del método CARPE es la misma que la secuencia del ARNg (de la biblioteca de destino) que se introduce en la segunda célula/virgen. En algunas realizaciones, se introduce más de un ARNg en la población de las primeras células y/o la población de segundas células. En algunas realizaciones, las más de una molécula de ARNg comprenden primeras secuencias de ácido nucleico que son complementarias de más de una región objetivo.

60 En el método CARPE, los casetes de ADN bicatenarios, referenciados además como oligonucleótidos de edición, para usar en los métodos descritos pueden obtenerse o derivarse de muchas fuentes. Por ejemplo, en algunas realizaciones, los casetes de ADNds se derivan de una biblioteca de ácido nucleico que se ha diversificado por recombinación aleatoria no homóloga (NRR); dicha biblioteca se referencia como biblioteca NRR. En algunas realizaciones, los oligonucleótidos de edición se sintetizan, por ejemplo, mediante síntesis en base a matrices. La longitud del oligonucleótido de edición puede depender del método usado para obtener el oligonucleótido de edición.

En algunas realizaciones, el oligonucleótido de edición tiene aproximadamente 50-200 nucleótidos, 75-150 nucleótidos, o entre 80-120 nucleótidos de longitud.

Un oligonucleótido de edición incluye (a) una región que es homóloga a una región objetivo del ácido nucleico de la célula e incluye una mutación (referida a una mutación deseada) de al menos un codón con relación a la región objetivo, y (b) una mutación del motivo adyacente protoespaciador (PAM). La mutación PAM puede ser cualquier inserción, delección o sustitución de uno o más nucleótidos que muta la secuencia PAM de manera que ya no se reconoce por el sistema CRISPR. Puede decirse que una célula que comprende dicha mutación PAM es "inmune" a la muerte mediada por CRISPR. La mutación deseada con relación a la secuencia de la región objetivo puede ser una inserción, delección y/o sustitución de uno o más nucleótidos en al menos un codón de la región objetivo.

El método CARPE se describe más abajo con referencia a un gen bacteriano solo con fines ilustrativos. Los métodos pueden aplicarse a cualquier gen(es) de interés, que incluyen los genes de cualquier procarionte, que incluyen las bacterias y las Archea, o cualquier eucariota, que incluyen los genes de las levaduras y los mamíferos (que incluyen los humanos). El método CARPE se realizó en el gen *galK* en el genoma de *E. coli*, en parte debido a la disponibilidad de los ensayos de la actividad para este gen. El método se llevó a cabo mediante el uso de las cepas parentales BW23115 y el vector pSIM5 (Datta et al. *Gene* (2008) 379: 109-115) para mediar la recombinación. El gen *cas9* se clonó en la cadena principal pBTBx-2 bajo el control de un promotor pBAD para permitir el control de la actividad de escisión mediante la adición de la arabinosa. La evaluación de la capacidad de incorporar selectivamente casetes de DNAds sintéticos (127 pb) se llevó a cabo mediante el uso de casetes de DNAds de bibliotecas NNK que se construyeron a partir de cebadores degenerados y/o oligonucleótidos (oligos) diseñados racionalmente sintetizados como parte de una biblioteca de 27 000 miembros a través de la tecnología de micromatriz. En ambos casos, los oligonucleótidos se diseñaron para mutar los residuos del sitio activo del producto genético *galK*. La recuperación altamente eficiente de las bibliotecas de cepas de donantes se verificó en base a los cambios en los tamaños del amplicón obtenidos con cebadores dirigidos al locus *galK*. La secuenciación de estos productos de PCR de colonias de las bibliotecas NRR indicó que el sitio de cebado sintético (P1) de los casetes de DNAds se incorporó con aproximadamente 90-100 % de eficiencia. Esto indicó que estas bibliotecas pueden generarse con alta eficiencia sin depender de los errores propensos de las cepas knockout *mutS* que típicamente se han usado en otros enfoques de edición en base a la recombinación (Costantino et al. *PNAS* (2003) 100:15748-15753; Wang et al. *Nature* (2009) 460:894-898). Hubo una caída en la eficiencia de las mutaciones del codón (alrededor del 20 %), lo que puede deberse a las correcciones *mutS* durante el reemplazo alélico. La evaluación preliminar de los clones en las bibliotecas de destino indicó que la eficiencia de edición final de los codones fue de aproximadamente 10 % cuando ambas fases de construcción se llevan a cabo en el antecedente *mutS*⁺.

Se realizó una comparación con otros protocolos publicados recientemente para la edición co-seleccionable, mediante el uso de los protocolos alternativos que no enlazan de manera covalente las mutaciones de los codones y el PAM, sino que dependen de la cercanía entre sí durante la replicación (Wang et al. *Nat. Métodos* (2012) 9:591-593). En estos experimentos no covalentes, se usaron los mismos oligos de edición que los anteriores y se hicieron esfuerzos para co-seleccionar su inserción mediante el uso de los oligos de ADNss que se dirigen a los mismos sitios PAM de donante/destino. La selección de las colonias de los mutantes resultantes revela una alta eficiencia en la recuperación de los mutantes PAM. Sin embargo, no parece haber una co-selección fuerte para la inserción de los casetes de edición de DNAds. Esto puede deberse a grandes diferencias en las eficiencias relativas de la recombinación de los oligonucleótidos de delección PAM y los casetes de edición que generan delecciones cromosómicas considerables.

Puede evaluarse la capacidad de mejorar las eficiencias de edición finales del método CARPE, tal como llevando a cabo la construcción de donantes en las cepas deficientes *demutS* antes de transferir a una cepa donante de tipo salvaje en un esfuerzo por evitar la pérdida de mutaciones durante la fase de construcción del donante. Además, puede evaluarse la generalidad del método CARPE, como la utilización de CARPE en un número de genes esenciales, que incluyen *dxs*, *metA*, y *folA*. Los genes esenciales se han dirigido efectivamente mediante el uso de las estrategias de diseño de ARNg descritas. Los resultados indican además que a pesar de la interrupción del gen que ocurre durante la creación de la biblioteca de donantes, las bibliotecas de donantes pueden construirse y recuperarse efectivamente dentro de 1-3 horas después de la recombinación.

Se proporcionan además en la presente descripción, los métodos para la edición del genoma de precisión rastreable mediante el uso de un sistema mediado por CRISPR referido Genoma Ingeniería mediante Rastreable CRISPR Enriquecido Recombinación (GEN-TraCER). Los métodos GEN-TraCER logran una edición/mutación de alta eficiencia mediante el uso de un solo vector que codifica tanto el casete de edición como el ARNg. Cuando se usa con síntesis de ADN paralela, como la síntesis de ADN en base a matrices, el GEN-TraCER proporciona la generación en una sola etapa de miles de ediciones/mutaciones de precisión y hace posible mapear la mutación mediante la secuenciación del casete de edición en el vector, en lugar de secuenciar del genoma de la célula (ADN genómico). Los métodos tienen una amplia utilidad en las aplicaciones de ingeniería de proteínas y genomas, así como también para la reconstrucción de mutaciones, como las mutaciones identificadas en los experimentos de evolución de laboratorio.

Los métodos y vectores GEN-TraCER combinan un casete de edición, que incluye una mutación deseada y una mutación PAM, con un gen que codifica un ARNg en un solo vector, lo que hace posible generar una biblioteca de mutaciones en una sola reacción. Como se muestra en la figura 11B, el método implica introducir un vector que

comprende un casete de edición que incluye la mutación deseada y la mutación PAM en una célula o una población de células. En algunas realizaciones, las células en las que se introduce el vector codifica además Cas9. En algunas realizaciones, un gen que codifica Cas9 se introduce subsecuentemente en la célula o la población de células. Se activa la expresión del sistema CRISPR, que incluye Cas9 y el ARNg, en la célula o la población de células; el ARNg recluta Cas9 a la región objetivo, donde se produce la escisión del DNAd. Sin desear limitarse a ninguna teoría particular, la región homóloga del casete de edición complementaria a la región objetivo, muta el PAM y uno o más codones de la región objetivo. Las células de la población de células que no integraron la mutación PAM sufren una muerte celular no editada debido a la escisión del DNAd mediado por Cas9. Las células de la población de células que integran la mutación PAM no sufren muerte celular; permanecen viables y se enriquecen selectivamente a gran abundancia. Se obtienen células viables y proporcionan una biblioteca de mutaciones dirigidas.

El método de edición del genoma rastreable mediante el uso del GEn-TraCER comprende: (a) introducir un vector que codifica al menos un casete de edición, un promotor y al menos un ARNg en una célula o una población de células, lo que produce de esta manera una célula o una población de células que comprende el vector (una segunda población de células); (b) mantener la segunda población de células bajo condiciones en las que se expresa Cas9, en donde la nucleasa de Cas9 se codifica en el vector, un segundo vector o en el genoma de las células de la segunda población de células, resulta en la escisión del ADN y la muerte de células de la segunda población de células que no comprenden la mutación PAM, mientras que las células de la segunda población de células que comprenden la mutación PAM son viables; (c) obtener células viables; y (d) secuenciar el casete de edición del vector en al menos una célula de la segunda población de células para identificar la mutación de al menos un codón.

En algunas realizaciones, una codificación vectorial separada *cas9* se introduce además en la célula o la población de células. La introducción de un vector en una célula o una población de células puede realizarse mediante el uso de cualquier método o técnica conocida en la técnica. Por ejemplo, los vectores pueden introducirse mediante protocolos estándar, como la transformación, que incluyen la transformación química y la electroporación, la transducción y el bombardeo de partículas.

Un casete de edición incluye (a) una región, que reconoce (hibrida con) una región objetivo de un ácido nucleico en una célula o una población de células, es homóloga a la región objetivo del ácido nucleico de la célula e incluye una mutación (referida a una mutación deseada) de al menos un nucleótido en al menos un codón con relación a la región objetivo, y (b) una mutación de motivo adyacente protoespaciador (PAM). La mutación PAM puede ser cualquier inserción, deleción o sustitución de uno o más nucleótidos que muta la secuencia de PAM de manera que el PAM mutado no reconoce el sistema CRISPR. Puede decirse que una célula que comprende, una mutación PAM, es "inmune" a la muerte mediada por CRISPR. La mutación deseada con relación a la secuencia de la región objetivo puede ser una inserción, deleción y/o sustitución de uno o más nucleótidos en al menos un codón de la región objetivo. En algunas realizaciones, la distancia entre la mutación PAM y la mutación deseada es al menos 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 nucleótidos en el casete de edición. En algunas realizaciones, la mutación PAM se localiza al menos 9 nucleótidos a partir del final del casete de edición. En algunas realizaciones, la mutación deseada se localiza al menos 9 nucleótidos a partir del final del casete de edición.

En algunas realizaciones, la mutación deseada con relación a la secuencia de la región objetivo es una inserción de una secuencia de ácido nucleico. La secuencia del ácido nucleico insertada en la región objetivo puede ser de cualquier longitud. En algunas realizaciones, la secuencia del ácido nucleico insertada es al menos 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900 o al menos 2000 nucleótidos de longitud. En las realizaciones en las que se inserta una secuencia del ácido nucleico en la región objetivo, el casete de edición comprende una región que es al menos 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, o al menos 60 nucleótidos de longitud y homólogos a la región objetivo.

El término "casete GEn-TraCER" puede usarse para referirse a un casete de edición, promotor, secuencia separadora y al menos una porción de un gen que codifica un ARNg. En algunas realizaciones, la porción del gen que codifica el ARNg en el casete GEn-TraCER codifica la porción del ARNg que es complementaria a la región objetivo. En algunas realizaciones, la porción del ARNg que es complementaria a la región objetivo es de al menos 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, o al menos 30 nucleótidos de longitud. En algunas realizaciones, la porción del ARNg que es complementaria a la región objetivo es de 24 nucleótidos de longitud. En algunas realizaciones, el casete GEn-TraCER comprende además al menos dos sitios de cebado. En algunas realizaciones, los sitios de cebado pueden usarse para amplificar el casete GEn-TraCER, por ejemplo mediante PCR. En algunas realizaciones, la porción del ARNg que es complementaria a la región objetivo se usa como un sitio de cebado.

En el método GEn-TraCER, los casetes de edición y los casetes GEn-TraCER para usar en los métodos descritos pueden obtenerse o derivarse de muchas fuentes. Por ejemplo, en algunas realizaciones, el casete de edición se sintetiza, por ejemplo, mediante síntesis en base a matrices. En algunas realizaciones, el casete GEn-TraCER se sintetiza, por ejemplo, mediante síntesis en base a matrices. La longitud del casete de edición y/o del casete GEn-TraCER puede depender del método usado para obtener el casete de edición y/o el casete GEn-TraCER. En algunas realizaciones, el casete de edición tiene aproximadamente 50-300 nucleótidos, 75-200 nucleótidos, o entre 80-120

nucleótidos de longitud. En algunas realizaciones, el casete GEN-TraCER tiene aproximadamente 50-300 nucleótidos, 75-200 nucleótidos, o entre 80-120 nucleótidos de longitud.

5 En algunas realizaciones, el método implica además obtener casetes GEN-TraCER, por ejemplo mediante síntesis en base a matrices, y construir el vector. Los métodos de construcción de un vector serán conocidos por un experto en la técnica y pueden implicar ligar el casete GEN-TraCER en un vector. En algunas realizaciones, los casetes GEN-TraCER o un subconjunto (conjunto) de los casetes GEN-TraCER se amplifican antes de la construcción del vector, por ejemplo mediante PCR.

10 La célula o la población de células que comprende el vector y que codifica además Cas9 se mantienen o cultivan bajo condiciones en las que se expresa Cas9. La expresión de Cas9 puede controlarse. Los métodos descritos en la presente descripción implican el mantenimiento de las células bajo condiciones en las que se activa la expresión de Cas9, lo que resulta en la producción de Cas9. Las condiciones específicas bajo las cuales se expresa Cas9 dependerán de factores, como la naturaleza del promotor usado para regular la expresión de Cas9. En algunas realizaciones, la expresión de Cas9 se induce en presencia de una molécula inductora, como la arabinosa. Cuando la célula o la población de células que comprende el ADN que codifica Cas9 están en presencia de la molécula inductora, se produce la expresión de Cas9. En algunas realizaciones, la expresión de Cas9 se reprime en presencia de una molécula represora. Cuando la célula o la población de células que comprende el ADN que codifica Cas9 están en ausencia de la molécula represora de la expresión de Cas9, ocurre la expresión de Cas9.

15 Las células de la población de células que permanecen viables se obtienen o se separan de las células que sufren muerte celular no editada como resultado de la muerte mediada por Cas9; esto puede hacerse, por ejemplo, diseminando la población de células en la superficie de cultivo, permitiendo el crecimiento de las células viables, que después están disponibles para su evaluación.

20 La mutación deseada acoplada a la mutación PAM puede rastrearse mediante el uso del método GEN-TraCER secuenciando el casete de edición en el vector en células viables (células que integran la mutación PAM) de la población. Esto permite la identificación fácil de la mutación sin necesidad de secuenciar el genoma de la célula. Los métodos implican la secuenciación del casete de edición para identificar la mutación de uno o más codones. Puede realizarse la secuenciación del casete de edición como un componente del vector o después de su separación del vector y, opcionalmente, la amplificación. La secuenciación puede realizarse mediante el uso de cualquier método de secuenciación conocido en la técnica, como la secuenciación de Sanger.

25 Los métodos descritos en la presente descripción pueden llevarse a cabo en cualquier tipo de célula en la que pueda funcionar el sistema CRISPR (*por ejemplo*, marcar y escindir el ADN), que incluyen las células procariotas y eucariotas. En algunas realizaciones, la célula es una célula bacteriana, como *Escherichia* spp. (*por ejemplo*, *E. coli*). En otras realizaciones, la célula es una célula fúngica, como una célula de levadura, *por ejemplo*, *Saccharomyces* spp. En otras realizaciones, la célula es una célula de algas, una célula vegetal, una célula de insecto o una célula de mamífero, que incluye una célula humana.

30 Un "vector" es cualquiera de una variedad de ácidos nucleicos que comprenden una secuencia o secuencias deseadas para ser suministradas o expresadas en una célula. La secuencia o secuencias deseadas pueden incluirse en un vector, tal como por restricción y ligadura o por recombinación. Los vectores están compuestos típicamente de ADN, aunque los vectores de ARN están disponibles además. Los vectores incluyen, pero no se limitan a: plásmidos, fásmidos, fagémidos, genomas de virus y cromosomas artificiales.

35 Los vectores útiles en el método GEN-TraCER comprenden al menos un casete de edición como se describió en la presente descripción, un promotor y al menos un gen que codifica un ARNg. En algunas realizaciones, se incluyen más de un casete de edición (por ejemplo 2, 3, 4, 5, 6, 7, 8, 9, 10 o más casetes de edición) en el vector. En algunas realizaciones, los más de un casete de edición son homólogos con diferentes regiones objetivo (por ejemplo, hay diferentes casetes de edición, cada uno de los cuales es homólogo con una región objetivo diferente). Alternativamente o además, el vector puede incluir más de un gen que codifica más de un ARNg (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más ARNg). En algunas realizaciones, los más de un ARNg contienen regiones que son complementarias a una porción de diferentes regiones objetivo (por ejemplo, hay diferentes ARNg, cada uno de los cuales es complementario a una porción de una región objetivo diferente).

40 En algunas realizaciones, un casete GEN-TraCER que comprende al menos un casete de edición, un promotor y un gen que codifica una porción de un ARNg se unen en un vector que codifica otra porción de un ARNg. Tras la unión, la porción del ARNg del casete GEN-TraCER y la otra porción del ARNg se unen y forman un ARNg funcional.

45 El promotor y el gen que codifica el ARNg están operablemente unidos. En algunas realizaciones, los métodos implican la introducción de un segundo vector que codifica Cas9. En tales realizaciones, el vector puede comprender además uno o más promotores unidos operablemente a un gen que codifica Cas9. Como se usa en la presente descripción, "operablemente" unido significa que el promotor afecta o regula la transcripción del ADN que codifica un gen, como el gen que codifica el ARNg o el gen que codifica Cas9. El promotor puede ser un promotor nativo (un promotor presente en la célula en la que se introduce el vector). En algunas realizaciones, el promotor es un promotor inducible o

reprimible (el promotor está regulado permitiendo la transcripción inducible o reprimible de un gen, como el gen que codifica el ARNg o el gen que codifica Cas9), como los promotores que están regulados por la presencia o la ausencia de una molécula (*por ejemplo*, un inductor o un represor). La naturaleza del promotor necesaria para la expresión del ARNg puede variar en base a la especie o el tipo de célula y será reconocida por un experto en la técnica.

En algunas realizaciones, el método comprende introducir un vector separado que codifica Cas9 en la célula o la población de células antes o al mismo tiempo que la introducción del vector que comprende al menos un casete de edición como se describió en la presente descripción, un promotor y al menos un ARNg. En algunas realizaciones, el gen que codifica Cas9 está integrado en el genoma de la célula o la población de células. El ADN que codifica Cas9 puede integrarse en el genoma celular antes de la introducción del vector que comprende al menos, un casete de edición como se describió en la presente descripción, un promotor y al menos un ARNg o después de la introducción del vector que comprende al menos un casete de edición como se describió en la presente descripción, un promotor y al menos un ARNg. Alternativamente, una molécula de ácido nucleico, como Cas9 que codifica el ADN, puede expresarse a partir del ADN integrado en el genoma. En algunas realizaciones, el gen que codifica Cas9 está integrado en el genoma de la célula.

Los vectores útiles en los métodos GEn-TraCER descritos en la presente descripción pueden comprender además una secuencia separadora, dos o más sitios de cebado o, tanto una secuencia espaciadora como dos o más sitios de cebado. En algunas realizaciones, la presencia de los sitios de cebado que bordean el casete GEn-TraCER permiten la amplificación de las secuencias del casete de edición, promotor y ácido nucleico de ARNg.

Ejemplos

Ejemplo 1: Uso del método CARPE para editar *galK*

El enfoque CARPE se llevó a cabo en el gen de la galactoquinasa, *galK*, en el genoma de *E. coli*, hay muchos ensayos disponibles para evaluar la actividad del producto génico. Los experimentos se llevaron a cabo mediante el uso de la cepa parental de *E. coli* BW23115 y el vector pSIM5 (Datta et al. Gene (2008) 379:109-115) para mediar la recombinación. El gen que codifica Cas9 se clonó en la cadena principal pBTBx-2 bajo el control de un promotor pBAD para permitir el control de la actividad de escisión de Cas9 mediante la adición de la arabinosa al medio de cultivo.

Primero, se probó la capacidad de incorporar selectivamente los casetes de ADNds sintéticos (127 pb). Los casetes de DNAds sintéticos se derivaron de las bibliotecas NNR que se construyeron a partir de los cebadores degenerados o de oligos diseñados racionalmente sintetizados como parte de una biblioteca de 27,000 miembros a través de la tecnología de micromatriz. En ambos casos, los oligonucleótidos se diseñaron para mutar los residuos del sitio activo del producto genético *galK*, así como también para contener el sitio de cebado sintético, P1 (SEQ ID NO: 1). La recuperación altamente eficiente de las bibliotecas de las cepas de donantes se verificó en base a los cambios en los tamaños del amplicón obtenido por PCR de colonias mediante el uso de cebadores dirigidos al locus *galK*. La secuenciación de los productos de PCR de colonias de las bibliotecas NNR indicó que el sitio de cebado sintético (P1) de los casetes de ADNds se incorporó con aproximadamente 90-100 % de eficiencia (Figura 2). Este resultado sorprendente e inesperado sugiere que las bibliotecas pueden generarse con alta eficiencia sin depender de las cepas propensas a errores deficientes de mutS que se han usado típicamente en otros enfoques de edición en base a la recombinación (Constantino, et al, PNAS (2003) 100:15748-15753; Wang et al Nature (2009) 460: 894-898). Sin embargo, hubo una caída en la eficiencia de las mutaciones del codón (alrededor del 20 %), lo que puede deberse a la corrección mediante MutS durante el reemplazo alélico. En este trabajo, la eficiencia de edición final de los codones fue de aproximadamente el 10 % cuando ambas fases de construcción se llevaron a cabo en el antecedente mutS+.

Para mejorar las eficiencias de edición final y la generalidad del método CARPE, la construcción del donante puede realizarse en las cepas deficientes de mutS antes de transferir a una cepa de donante mutS+ en un esfuerzo por evitar la pérdida de las mutaciones durante la fase de construcción del donante.

Ejemplo 2: Uso del método CARPE para dirigir genes esenciales

Para probar la generalidad del enfoque CARPE, el método se usó, como se describió anteriormente, en varios genes esenciales, que incluyen *dxs*, *metA*, y *folA*. Los genes esenciales pueden dirigirse mediante el uso de las estrategias del diseño del ARNg (Figura 3).

Los datos de los experimentos CARPE dirigidos al gen *dxs* sugieren además que a pesar de la interrupción del gen que ocurre durante la creación de la biblioteca de donantes, es posible construir y recuperar efectivamente las bibliotecas de donantes dentro de 1-3 horas después de la recombinación.

Ejemplo 3: Uso del método CARPE para modular la producción de isopentenol

La búsqueda de mejores biocombustibles para la fabricación industrial a través de la producción bacteriana requiere la capacidad de realizar el diseño, la ingeniería y la detección del genoma más modernos para el producto deseado. Anteriormente, demostramos la capacidad de modificar individualmente los niveles de expresión de cada gen en el

genoma de *E. coli* Warner et al. Nat. Biotechnol (2010) 28:856-862). Este método, denominado recombinación múltiple rastreable (TRMR), produjo una biblioteca de aproximadamente 8000 células genómicamente modificadas (-4000 genes sobre expresados y -4000 genes noqueados). Posteriormente, esta biblioteca se examinó en diferentes condiciones, lo que permitió una comprensión más profunda de las actividades de los productos génicos y resultó en cepas de mejor rendimiento bajo estas selecciones. El TRMR permitió la modificación de la expresión de proteínas para dos niveles (sobre expresado y noqueado) pero no permitió la modificación del marco de lectura abierto (ORF). Aquí, nuestro objetivo es producir grandes bibliotecas de modificaciones del ORF y diseñar trayectorias metabólicas completas para la producción óptima de biocombustibles.

Una dificultad importante en la producción de tales bibliotecas, que están diseñadas racionalmente (por el contrario con la mutagénesis aleatoria), es la eficiencia de la inserción de las mutaciones deseadas en las células objetivos. La recombinación, el método canónico para modificaciones genómicas en *E. coli*, usa genes recombinantes del fago Lambda para facilitar la inserción del ADN extraño en el genoma del huésped. Sin embargo, este proceso adolece de bajas eficiencias y puede superarse agregando un gen de resistencia a los antibióticos seguido de la selección (como en TRMR) o induciendo recursivamente eventos de recombinación (es decir, mediante MAGE (Wang et al. Nature (2008) 460:894-898). El método CARPE descrito en la presente descripción aumenta la eficiencia de la recombinación que implica el uso del sistema CRISPR para eliminar todas las células no recombinantes de la población. CRISPR es un mecanismo de defensa adaptativo en base al ARN de bacterias y Archaea recientemente descubierto contra fagos y plásmidos invasores (Bhaya et al Anm. Rev. of Genetics (2011) 45:273-297). Este sistema se sometió a una ingeniería masiva para permitir roturas de doble cadena dirigidas por la secuencia mediante el uso de dos plásmidos; un plásmido que codifica la nucleasa Cas9 asociada a CRISPR y el segundo plásmido que codifica el ARN guía específico de la secuencia (ARNg) que guía a Cas9 a su ubicación única (Qi et al Cell (2013) 45:273-297). El método CARPE utiliza la capacidad del sistema CRISPR para inducir roturas del ADN y, en consecuencia, la muerte celular, de una manera dependiente de la secuencia. Produjimos casetes de recombinación del ADN que, además de la mutación deseada dentro del ORF, incluyen una mutación en una ubicación común fuera del marco de lectura abierto del gen que es el objetivo de la maquinaria CRISPR. Este enfoque de enlace/acoplamiento de las mutaciones deseadas con la prevención de la muerte mediada por CRISPR, debido a la mutación/delección de PAM, permite un enriquecimiento dramático de las células modificadas dentro de la población total de células.

El método se demuestra además mediante el uso de la trayectoria DXS. La trayectoria DSX resulta en la producción de pirofosfato de isopentenilo (IPP) que resulta en la biosíntesis de terpenos y terpenoides. Es interesante, que el IP puede ser además, precursor del licopeno o isopentenol, dada la adición de los genes requeridos. Mientras que el licopeno hace que las colonias bacterianas se vuelvan rojas y, por lo tanto, es fácilmente detectable, el isopentenol se considera un biocombustible de 'segunda generación' con mayor densidad de energía y menor miscibilidad del agua que del etanol. Se seleccionaron tres proteínas para la ingeniería: 1) La DSX, la primera y la enzima limitante de la velocidad de la trayectoria, 2) la IspB, que desvía el flujo metabólico de la trayectoria DXS, y 3) la NudF, que se ha demostrado que convierte el IPP en isopentenol en ambos *E. coli* y *B. subtilis* (Withers et al. App. Environ. Microbiol (2007) 73: 6277-6283; Zheng et al Biotechnol. for biofuels (2013) 6:57). Las mutaciones en los genes que codifican la DXS e IspB se evaluarán para aumentar la producción del licopeno con una nueva herramienta de análisis de imágenes desarrollada para la cuantificación del color de las colonias. La actividad de la NudF se analizará directamente midiendo los niveles del isopentenol por GC/MS e indirectamente por las células auxotróficas de isopentenol que servirán como biosensores. Este método proporciona la capacidad de diseñar racionalmente grandes bibliotecas mutacionales en el genoma de *E. coli* con alta precisión y eficiencia y una cepa que produce un alto rendimiento de isopentenol.

Ejemplo 4: Uso del método GEn-TraCER para editar *galk*

El método GEn-TraCER se usó para editar el gen *galk*, que ha servido como un sistema modelo para la recombinación en *E. coli* (Yu et al, 2000). Los primeros casetes GEn-TraCER construidos se diseñaron para introducir un codón de parada en lugar de un PAM dentro del marco en el codón 24 de *galk*, referido *galk_Q24* (Figura 12). Las construcciones y los vectores se diseñaron mediante el uso de un script de Python personalizado para generar las mutaciones requeridas de alto rendimiento.

Los casetes de control se clonaron en el vector de ARNg descrito por Qi y col. Cell (2013) mediante el uso del método de clonación de la polimerasa circular (CPEC). La cadena principal se linealizó con los siguientes cebadores: CCAGAAATCATCCTTAGCGAAAGCTAAGGAT (SEQ ID NO: 29) y GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT (SEQ ID NO: 30).

Los casetes GenTRACER se ordenaron como gblocks y se amplificaron mediante el uso de los siguientes cebadores:

ATCACGAGGCAGAATTTTCAGATAAAAAAATCCTTAGCTTTTCGCTAAGGATGATT
TCTGG (SEQ ID NO: 31),

ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTATTTCTAGCTCTA

AAAC (SEQ ID NO: 32).

5 Los componentes se unieron mediante el uso de CPEC y se transformaron en *E. coli* para generar los vectores. Este procedimiento se realizará en multiplex mediante el uso de las bibliotecas de oligonucleótidos agrupados con eficiencias de clonación del orden de 10^4 - 10^5 UFC/ μ g.

10 Las células MG1655 de *E. coli* que llevan pSIM5 (plásmido lambda-RED) y el plásmido X2-cas9 se cultivó hasta la mitad de la fase logarítmica (0,4-0,7 DO) a 30 °C en LB con 50 μ g/mL de kanamicina y 34 μ g/mL de cloranfenicol. Las funciones recombinantes del vector pSIM5 se indujeron a 42 °C durante 15 minutos y después se colocaron en hielo durante 10 minutos. Después, las células se hicieron electrocompetentes granulando y lavando 2X con 10 mL de H₂O enfriado. Las células se transformaron con 100 ng de un plásmido GEn-TraCER (que además codifica la resistencia a la carbenicilina) y se recuperaron durante 3 horas a 37 °C. Se colocaron en placa 50 -100 μ L de células en los medios apropiados que contenían 50 μ g/mL de kanamicina y 100 μ g/mL de carbenicilina para enriquecer selectivamente las cepas editadas con CRISPR. Edición de las eficiencias para el gen *galK* se calculó mediante el uso de la detección rojo/blanco en agar MacConkey suplementado con galactosa.

15 En base a la detección de la eficiencia de edición de agar MacConkey de ~ 100 % se observó con el diseño *galK_Q24**. Es interesante, a diferencia de los métodos de recombinación mediada por oligo que requieren reparaciones de desajuste para lograr una alta eficiencia (Li et al. 2003; Sawitzke et al. 2011; Wang et al. 2011), no hubo efecto en las cepas con o sin la maquinaria de reparación de desajuste intacto.

20 Las secuencias cromosómicas y vectoriales se verificaron después mediante la secuenciación de Sanger.

25 Como se anticipó, la mutación diseñada en el vector se reflejó en el cromosoma (Figura 13), lo que indica que la mutación estaba presente en ambas ubicaciones y que el plásmido sirve como un código de barras de transacción (código de barras trans) o registro de la edición del genoma.

30 El diseño se adaptó para la mutagénesis racional de los marcos de codificación de las proteínas a escala del genoma mediante la generación de "cicatrices seleccionables silenciosas" que consisten en la mutación sinónima de PAM (Figura 14B, Δ PAM) para "inmunizar" la célula contra la escisión mediada por Cas9 pero dejar el producto de traducción impávido. Razonamos que las cicatrices silenciosas pueden permitir la co-selección para las ediciones cercanas en un codón u otra característica de interés con alta eficiencia. Los efectos de la longitud del brazo de homología y la distancia entre la mutación/delección PAM y la mutación deseada en *galK* fueron evaluados y las eficiencias comparadas (Figura 16B). Un aumento significativo en la eficiencia mutacional en la posición 145 de *galK* se observó cuando la longitud del brazo de homología se extendió de 80 a 100 nucleótidos (-5 % y 45 %, respectivamente) con ediciones PAM idénticas.

35 Ejemplo 5: Uso del método GEn-TraCER para reconstruir mutaciones

40 El enfoque GEn-TraCER se extendió a una escala genómica mediante el uso de un programa de diseño automatizado personalizado que permite la dirección de sitios alrededor del genoma con una simple definición de entrada del usuario. El enfoque se probó mediante la reconstrucción de todas las mutaciones puntuales no sinónimas de un estudio recientemente reportado de adaptación térmica en *E. coli* (Tenailon et al. 2012). Este estudio caracterizó el conjunto completo de mutaciones que ocurrieron en 115 aislamientos de cepas propagadas independientemente. Este conjunto de datos proporciona una fuente diversa de mutaciones cuyos efectos de aptitud individual arrojan más luz sobre los fundamentos mecanicistas de este fenotipo complejo. Cada una de estas mutaciones se reconstruyó con una redundancia de 2 veces en el uso del codón y Δ PAM, cuando fue posible, para permitir la corrección estadística tanto de PAM como de las mutaciones del codón objetivo en el análisis de aptitud abajo.

45 Ejemplo 6: Uso del método GEn-TraCER para modular las interacciones genéticas

50 Se genera una biblioteca de cableado del promotor integrando un promotor que está regulado dinámicamente por una señal ambiental (nivel de oxígeno, fuente de carbono, estrés) aguas arriba de cada gen en el genoma de *E. coli*. Mediante el uso del método GEn-TraCER, las cepas se generan con genotipos recableados que pueden ser beneficiosos, por ejemplo, para la tolerancia a los productos químicos de interés para la producción.

REIVINDICACIONES

1. Una biblioteca que comprende una población de células que comprende:
un vector que comprende:
 - 5 (i) un casete de edición que incluye una región que es homóloga a una región objetivo de un ácido nucleico en una célula e incluye una mutación de al menos un nucleótido con respecto a la región objetivo y un motivo adyacente protoespaciador (PAM) mutado que no es reconocido por un sistema CRISPR;
 - (ii) un promotor; y
 - 10 (iii) un ácido nucleico que codifica al menos un ARN guía (ARNg), comprendiendo el ARNg una región complementaria a una porción de la región objetivo, y una región que recluta una endonucleasa, y en el que el promotor y el ácido nucleico que codifica el ARNg están operativamente unidos.
2. La biblioteca de la reivindicación 1, en la que el promotor es un promotor inducible.
3. La biblioteca de la reivindicación 1, en la que el promotor es un promotor reprimible.
4. La biblioteca de la reivindicación 1, en la que dicha célula es una célula procariótica.
5. La biblioteca de la reivindicación 1, en la que dicha célula es una célula eucariótica.
- 15 6. La biblioteca de la reivindicación 1, en la que el vector comprende más de un casete de edición.
7. La biblioteca de la reivindicación 1, en la que el vector codifica al menos dos ARNg que son complementarios a una porción de diferentes regiones objetivo.
8. La biblioteca de la reivindicación 1, en la que el vector comprende además al menos dos sitios de cebadores.
9. La biblioteca de la reivindicación 1, en la que la endonucleasa es Cas9.
- 20 10. La biblioteca de la reivindicación 1, en la que el vector es un vector de ADN.

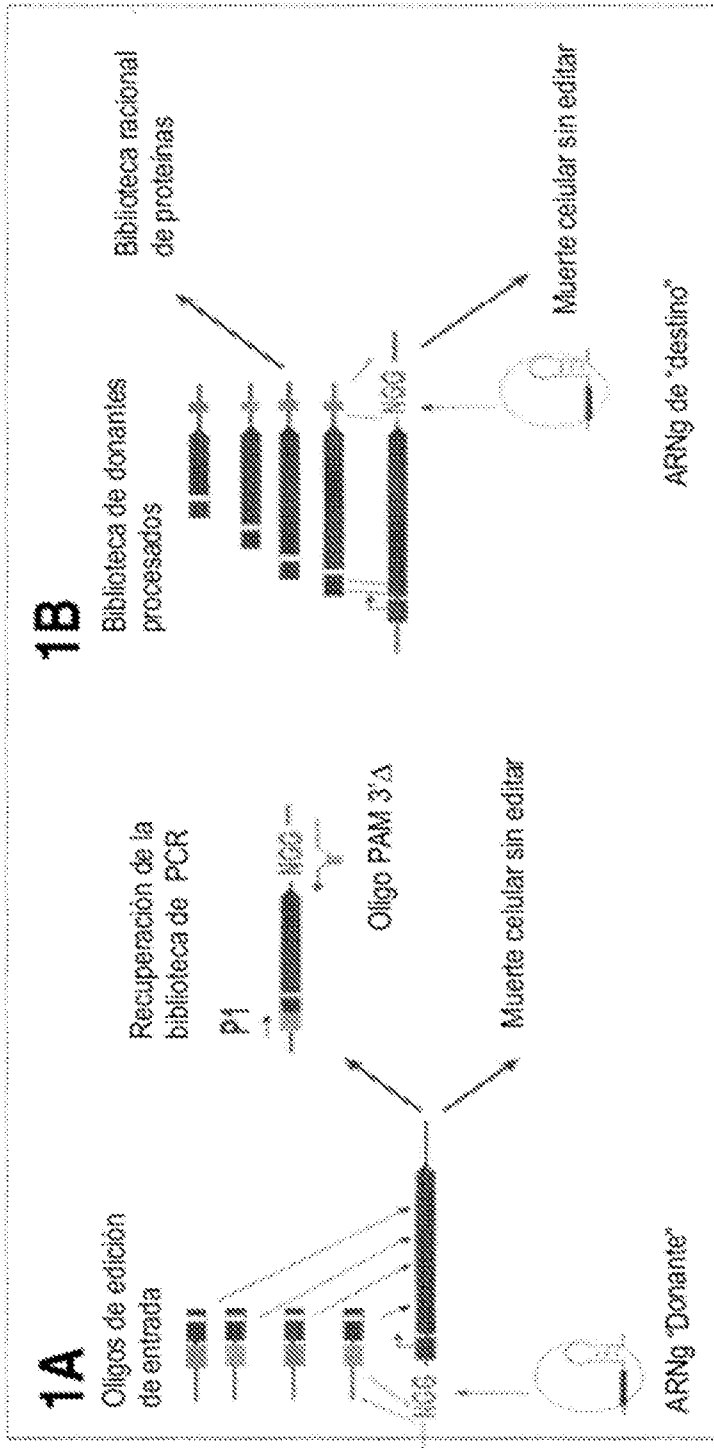


Figura 1

Clon	#1	Codón	SEQ ID NO:
F_001	...GCCCGC...	TAA	SEQ ID NO: 2
F_002	...AGCAAA	ACA	SEQ ID NO: 3
F_003	...AACAGG	TAT	SEQ ID NO: 4
F_004	...GCCCGC...	CGC	SEQ ID NO: 5
F_005	Mala calidad de lectura		
F_006	...GCCCGC...	CGC	SEQ ID NO: 5
F_007	...CCAGTT	CAA	SEQ ID NO: 6
F_008	...CCAGTT	GTA	SEQ ID NO: 7
F_009	Mala calidad de lectura		
F_010	...CCAGTT	GTA	SEQ ID NO: 7
F_001	...CTCCAA	CGT	SEQ ID NO: 8
F_002	...CCAGTT	GTA	SEQ ID NO: 7
F_003	...GCCCGC...	CGC	SEQ ID NO: 5
F_004	...CCAGTT	GTA	SEQ ID NO: 7
F_005	...CTCCAA	CGT	SEQ ID NO: 8
F_006	...GCCCGC...	CGC	SEQ ID NO: 5
F_007	...AGCCAA	ATT	SEQ ID NO: 9
F_008	...GCCCGC...	CGC	SEQ ID NO: 10
F_009	...CTCCAA	CGG	SEQ ID NO: 11
F_010	...CTCCAA	ACA	SEQ ID NO: 12

Eficiencia de la mutación 90 %

20 %

Figura 2

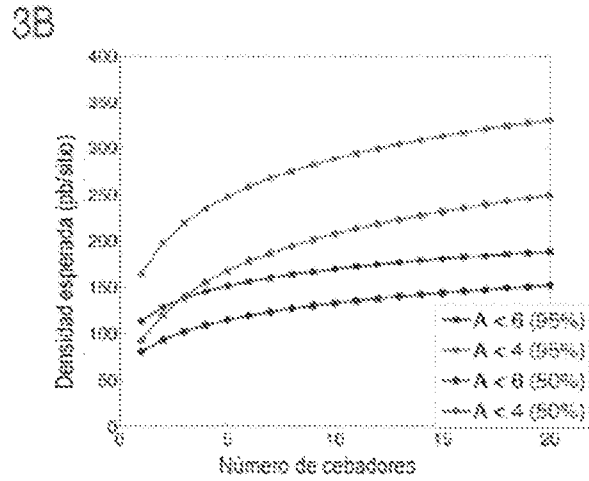
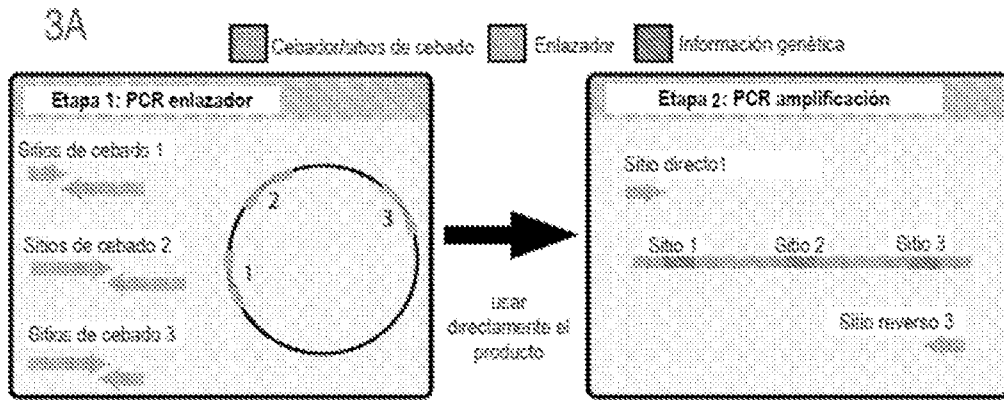


Figura 3

4A

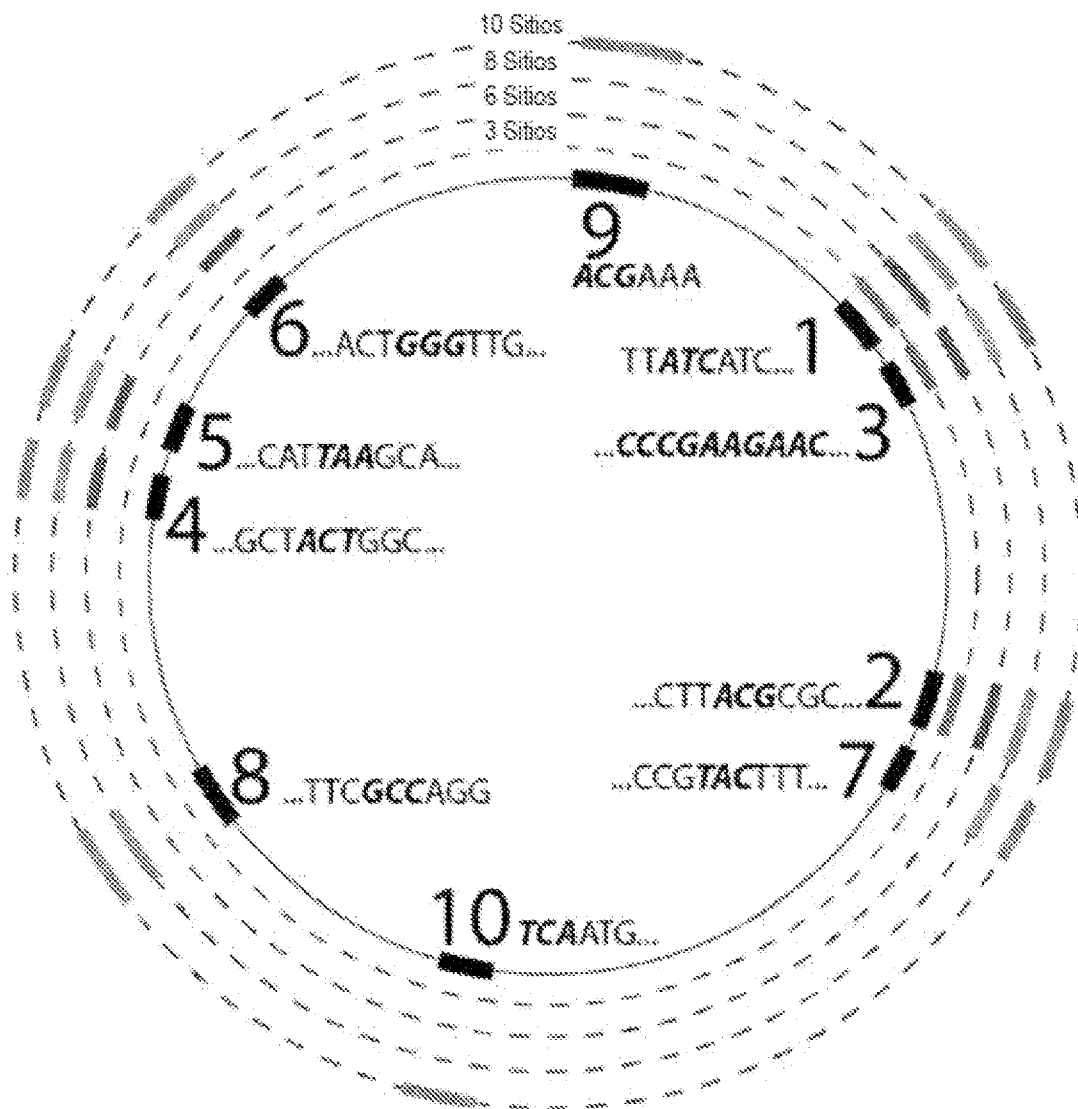
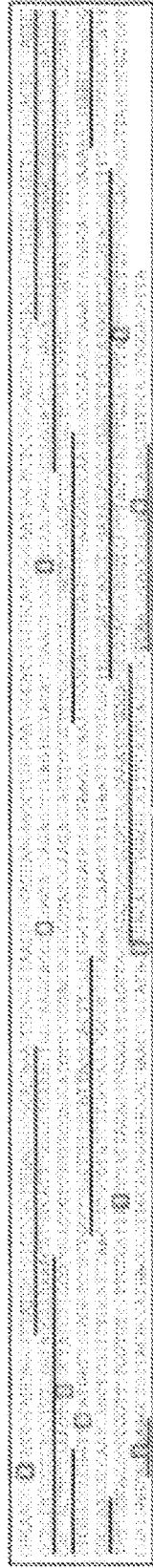


Figura 4

4B

10 Sílices



SEQ ID NO: 13

Figura 4, continuación

Figura 4

Edición de proteínas racional para la ingeniería metabólica

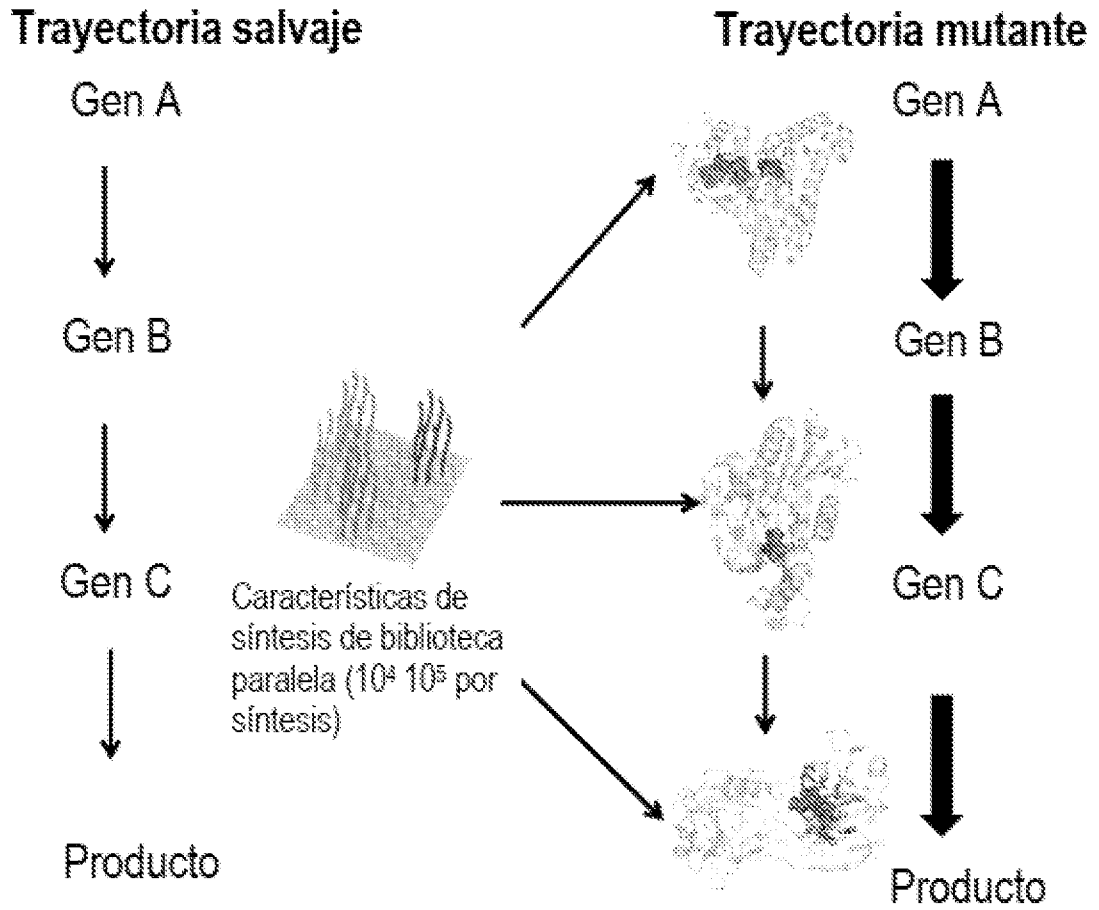
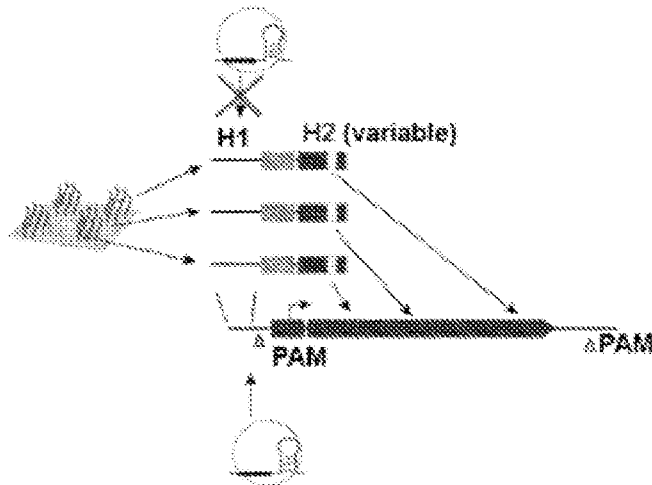


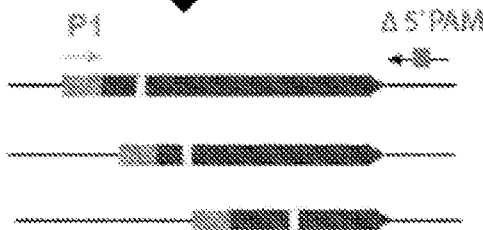
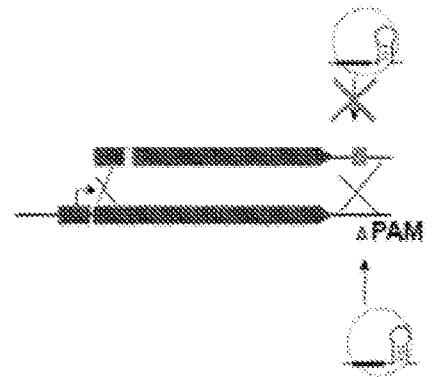
Figura 5

CRISPR biblioteca de proteínas racional enriquecida

Etapa 1: Recombinación multiplex



Etapa 3: Recombinación multiplex para acoplar PAM a la biblioteca de mutaciones



CRIPS permite la selección rápida de genotipos recombinantes para enriquecer las bibliotecas de proteínas de la placa

Etapa 2: Recuperación de la biblioteca por PCR

Figura 6

Configuración experimental y prueba de concepto

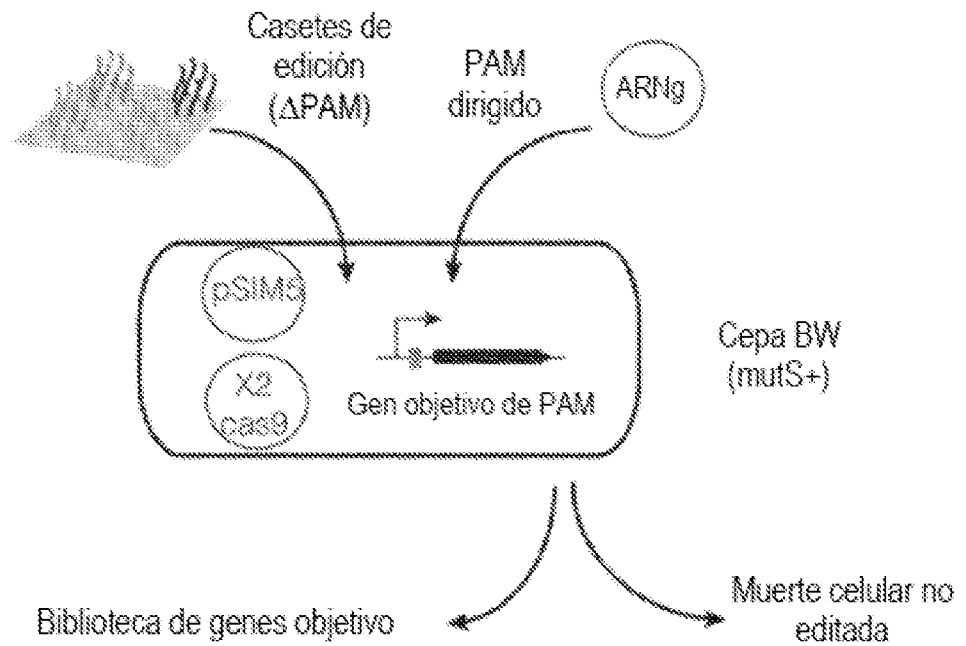
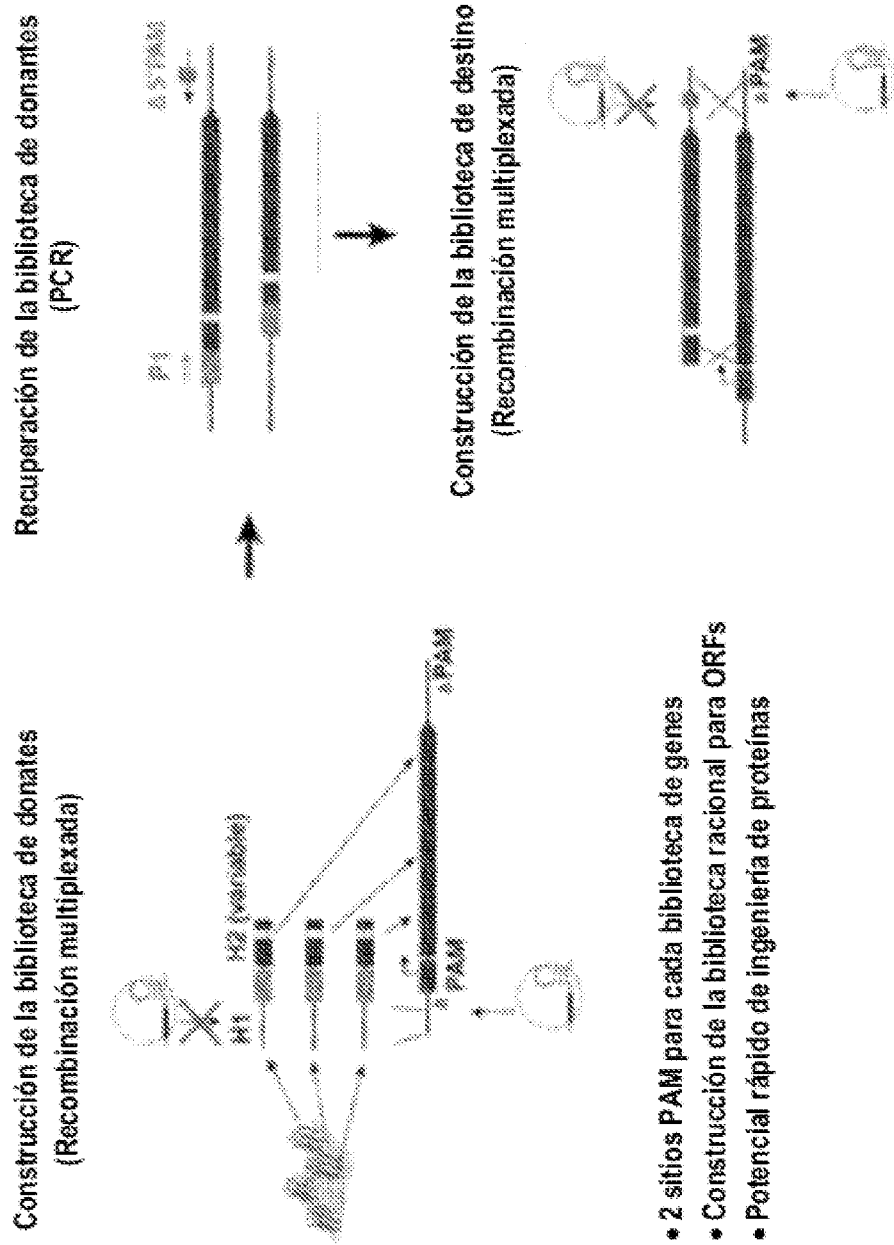


Figura 7

Estrategia para la ingeniería de proteínas multiplexadas



- 2 sitios PAM para cada biblioteca de genes
- Construcción de la biblioteca racional para ORFs
- Potencial rápido de ingeniería de proteínas

Figura 9

Prueba de concepto de la construcción de la biblioteca donante (galk)

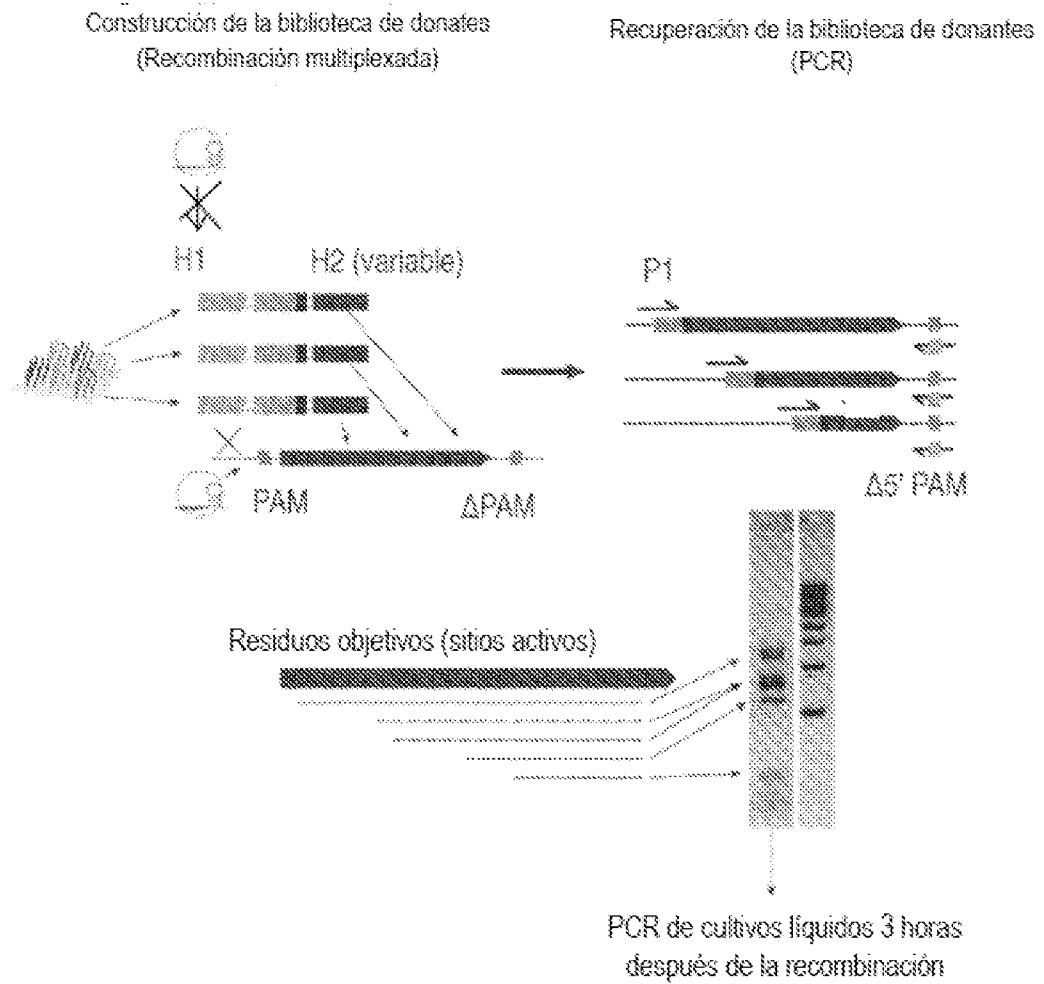


Figura 10

11A

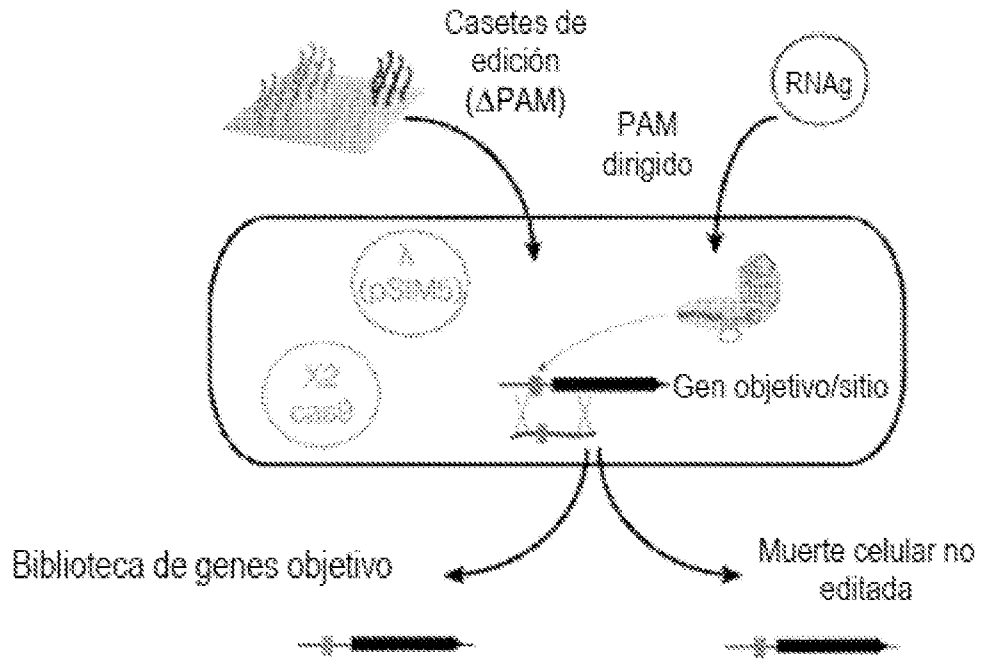


Figura 11

11B

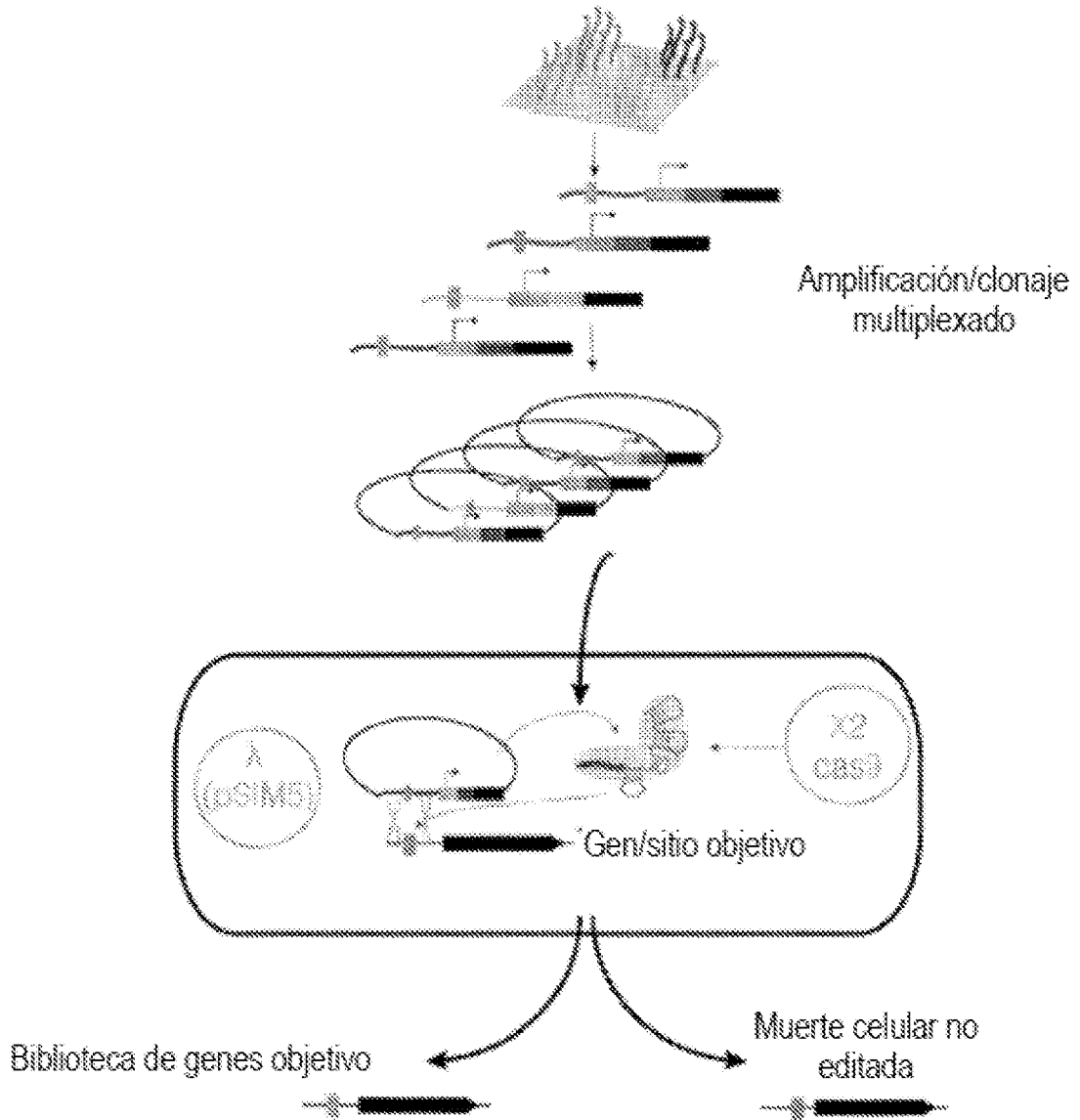


Figura 11, continuación

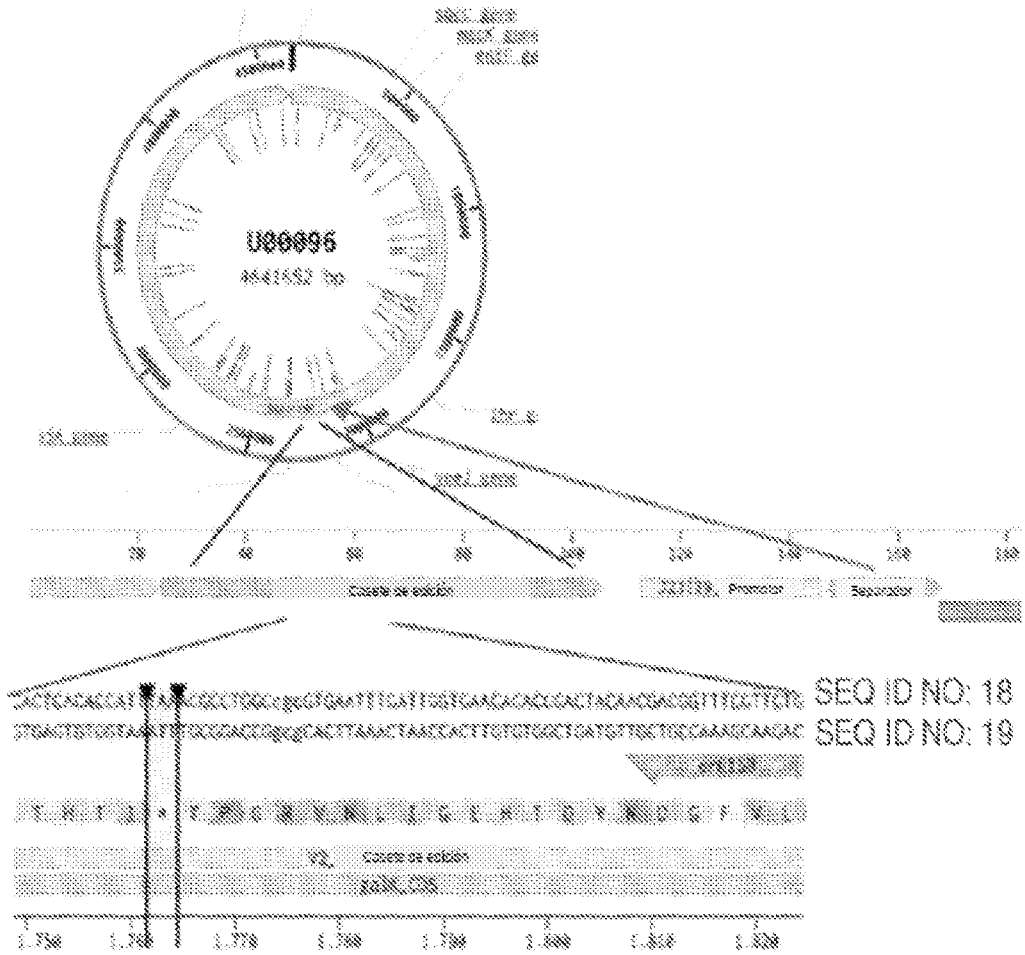


Figura 12

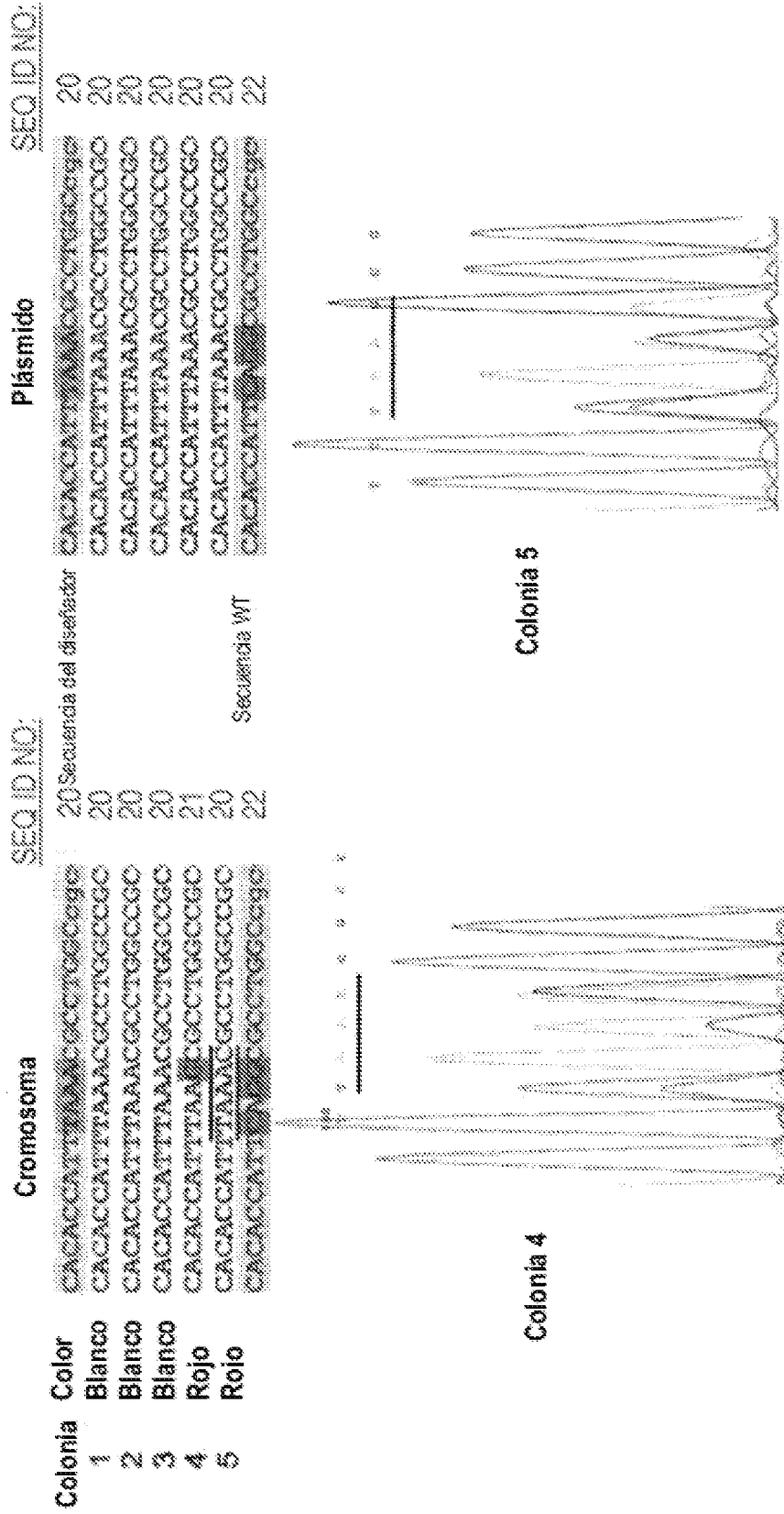


Figura 13

14A

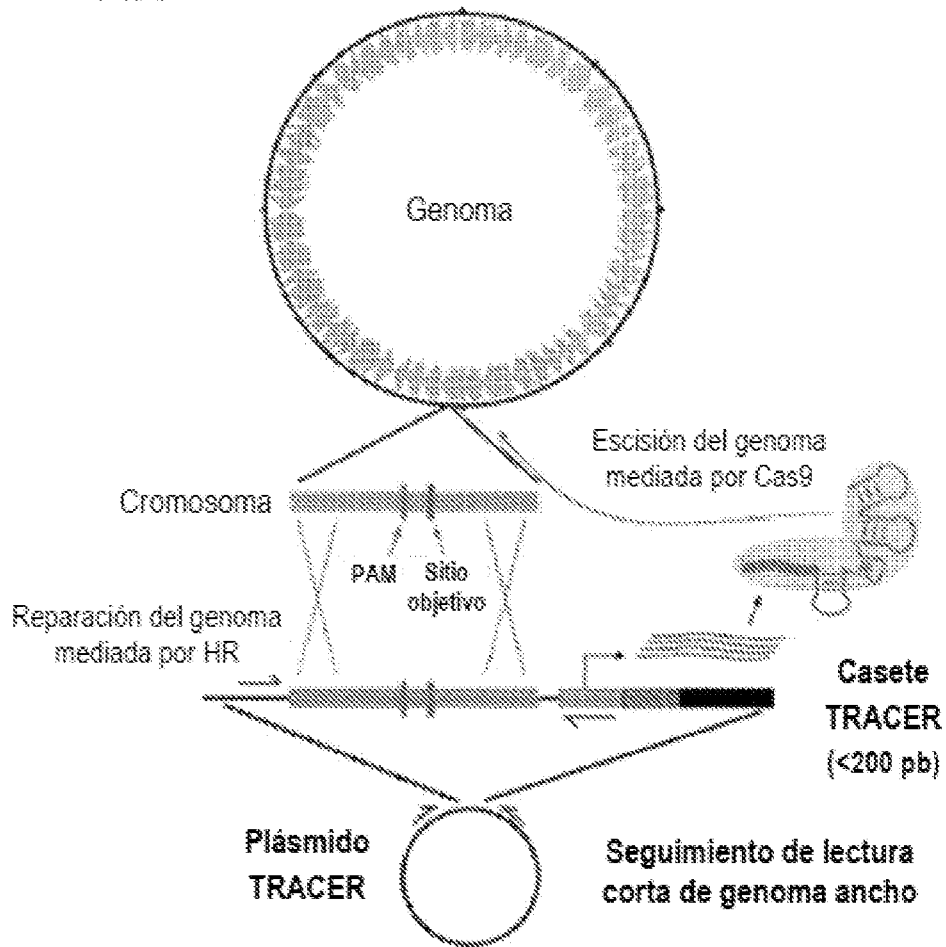


Figura 14

14B

ΔPAM		Separador		Codón		objetivo									
GGG	ACC	GTA	TTC	CAG	CAG	CTT	TAT	CAT	CTG	CCG	CTG	GAC	GGC	GCA	SEQ ID NO: 23
G	T	V	L	Q	Q	L	T	H	L	P	L	D	S	A	SEQ ID NO: 24

ΔPAM		Separador		Codón		objetivo									
GGG	ACC	GTA	TTC	CAG	CAG	CTT	TAT	CAT	CTG	CCG	CTG	GAC	GGC	GCA	SEQ ID NO: 25
G	T	V	L	Q	Q	L	T	H	L	P	L	D	S	A	SEQ ID NO: 26
															SEQ ID NO: 27

14C

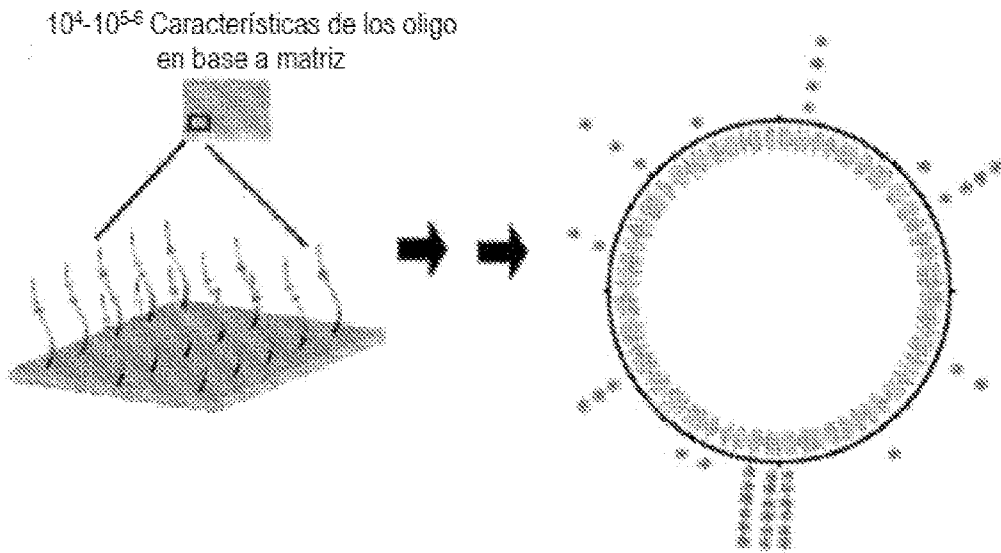
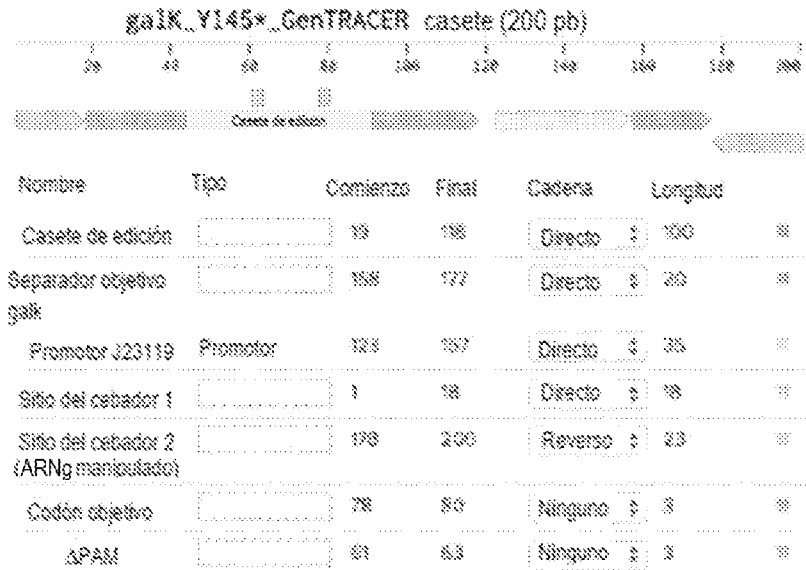


Figura 14, continuación

15A



15B

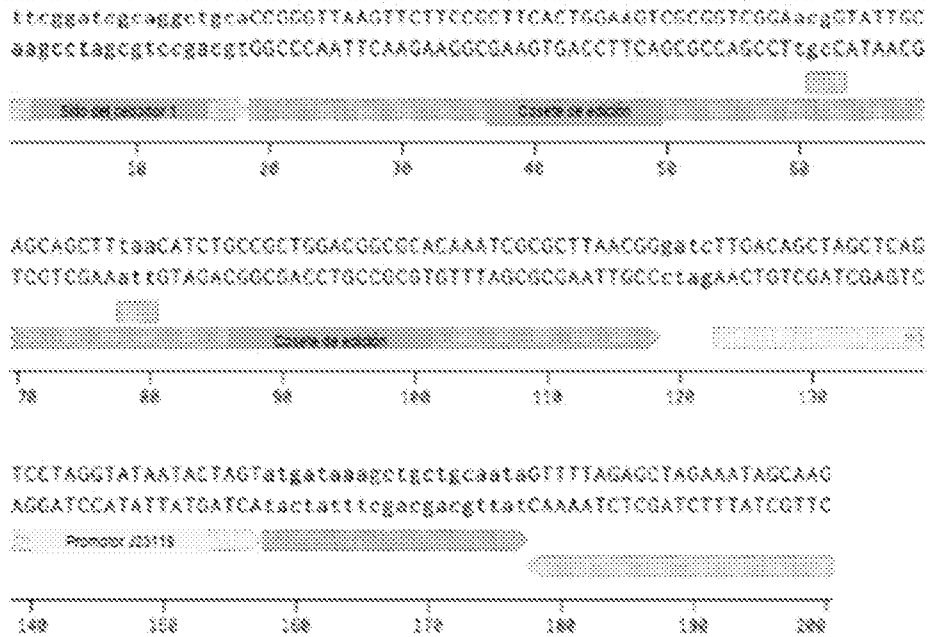


Figura 15

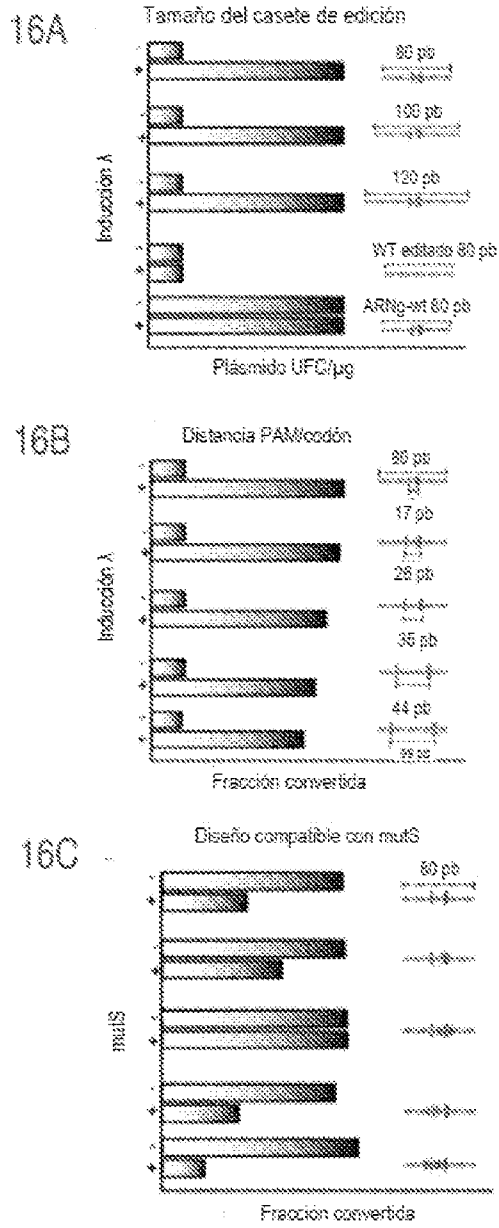


Figura 16