

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 014 020**

51 Int. Cl.:

**C12N 15/113** (2010.01)

**C07K 14/26** (2006.01)

**C12N 9/22** (2006.01)

**C12N 9/78** (2006.01)

**C12N 15/62** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **31.08.2023** **PCT/EP2023/073881**

87 Fecha y número de publicación internacional: **07.03.2024** **WO24047151**

96 Fecha de presentación y número de la solicitud europea: **31.08.2023** **E 23765455 (3)**

97 Fecha y número de publicación de la concesión europea: **26.02.2025** **EP 4392563**

54 Título: **Un nuevo tipo de sistema CRISPR/Cas**

30 Prioridad:

**31.08.2022 GB 202212688**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**16.04.2025**

73 Titular/es:

**SNIPR BIOME APS (100.00%)**  
**Lersø Parkallé 44, 5. floor**  
**2100 Copenhagen, DK**

72 Inventor/es:

**SEMSEY, SZABOLCS;**  
**MOUGIAKOS, IOANNIS;**  
**SØNDBERG, EMILIE y**  
**CLUBE, JASPER**

74 Agente/Representante:

**ELZABURU, S.L.P**

Observaciones:

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

ES 3 014 020 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

Un nuevo tipo de sistema CRISPR/Cas

### Campo técnico

5 La tecnología descrita en la presente memoria se relaciona con sistemas CRISPR/Cas nuevos y sintéticos y componentes de los mismos, vectores, proteínas, complejos de ribonucleoproteínas y métodos de utilización del nuevo sistema y los componentes.

Llamamos al nuevo sistema un sistema CRISPR/Cas "Tipo S".

### Antecedentes

Actualmente, se conocen los sistemas CRISPR/Cas Tipos I-VI.

10 Sería deseable proporcionar un nuevo sistema y componentes del mismo que puedan aumentar la caja de herramientas y las utilidades para el direccionamiento CRISPR/Cas en general y que puedan aportar nuevos tipos de funcionalidades para estos sistemas útiles.

### Sumario

15 La invención satisface estas necesidades al proporcionar un nuevo sistema, CRISPR/Cas Tipo-S. Los componentes básicos de este sistema no se basan en los genes característicos de ninguno de los sistemas Tipo I-VI que ya se han dilucidado. El Tipo-S no utiliza una Cas3 (la proteína característica de los sistemas Tipo I), una Cas9 (la proteína característica de los sistemas Tipo II), una Cas10 (la proteína característica de los sistemas Tipo III), una Cas12 (la proteína característica de los sistemas Tipo V) o una Cas13 (la proteína característica de los sistemas Tipo VI). Hemos descrito el nuevo sistema como Tipo-S.

20 Además, identificamos que los componentes básicos (Cas-S1, Cas-S2, Cas-S3, Cas-S4 y Cas-S5) no comprenden una nucleasa de ADN guiada por ARN. Sin embargo, descubrimos sorprendentemente que un complejo de ribonucleoproteína de estos componentes con un ARNcr es capaz de dirigirse de manera eficaz y específica a un sitio predeterminado en el ADN. Usando este sistema, también fue sorprendentemente posible modificar una célula sin inducir roturas de ADNds. Sorprendentemente, también pudimos usar el sistema para inhibir el crecimiento de células bacterianas. Fusionamos varios componentes básicos del sistema a un editor de bases y descubrimos que era capaz de realizar la edición de bases de manera eficaz. También fusionamos varios componentes básicos del sistema a una nucleasa de ADN y descubrimos que era capaz de realizar el corte del ADN de manera eficaz. Armados con este conocimiento, podemos ver muchas aplicaciones del nuevo sistema. Es posible, por ejemplo, proporcionar una o más proteínas o dominios efectores adicionales (por ejemplo, una nucleasa, un editor de bases o un editor principal) junto con uno o más de los componentes del sistema para producir sistemas sintéticos que puedan modificar células y ácidos nucleicos de una manera dirigida por el ARN. La unión al ADN del sistema Tipo-S no requiere la presencia de todos los componentes básicos (Cas-S1, Cas-S2, Cas-S3, Cas-S4 y Cas-S5). Cuando se fusiona con un editor de bases o una nucleasa de ADN, de nuevo, sorprendentemente, el sistema sintético no requiere la presencia de todos los componentes básicos (Cas-S1, Cas-S2, Cas-S3, Cas-S4 y Cas-S5) del sistema Tipo-S para realizar de forma eficaz las funciones de los dominios efectores.

Por lo tanto, el sistema y sus componentes proporcionan nuevos medios para llevar a cabo la edición de ácidos nucleicos y el direccionamiento celular en entornos ambientales, médicos y de otro tipo.

Becker *et al.* (Genome Announcements, vol. 3, no. 4, 9 de julio de 2015) se refiere a la secuencia genómica completa de una cepa de brote de *Klebsiella pneumoniae* productora de CTX-M-15 de secuencia multilocus tipo 514.

40 Schlüter *et al.* (Antimicrobial Agents and Chemotherapy, vol. 58, no. 7, 21 de abril de 2014, páginas 3768-3773) se refiere a un plásmido de tipo incH que alberga los genes *bla*<sub>CTX-M-15</sub>, *bla*<sub>DHA-1</sub> y *qnrB4* recuperados de aislados animales.

Young *et al.* (Communications Biology, vol. 2, 383, 18 de octubre de 2019, páginas 1-7) se refiere a la reutilización de la Cascada de CRISPR tipo IE para la activación génica en plantas.

45 Medina-Aparicio *et al.* (Pathogens and Disease, vol. 76, no. 1, 9 de enero de 2018) se refiere al sistema CRISPR-Cas en *Enterobacteriaceae*.

Zheng *et al.* (Frontiers in Bioengineering and Biotechnology, vol. 8, 4 de marzo de 2020) se refiere al CRISPR-Cas tipo I endógeno: de la defensa del ADN extraño a la ingeniería procariota.

### Declaraciones de la invención:

50 La invención es como se reivindica en las reivindicaciones 1 a 17. La invención reivindicada requiere la presencia obligatoria de los ácidos nucleicos como se recita en la reivindicación 3(a) y la reivindicación 3(b) y la reivindicación 3(c) juntas en combinación con un ARNcr como se reivindica; o los polipéptidos como se recita en la reivindicación

5(a) y la reivindicación 5(b) y la reivindicación 5(c) juntas en combinación con un ARNcr como se reivindica. Cualquier parte de la divulgación relacionada con estos ácidos nucleicos y proteínas específicos individualmente no forma parte de la invención per se, sino que sirve para la comprensión de la invención reivindicada.

#### Breve descripción de los dibujos

- 5 **Figura 1: Componentes de clonación de un sistema CRISPR Tipo S.** Se construyó un plásmido (p1624) que contenía un origen de replicación pSC101 y un marcador de resistencia a la tetraciclina, así como secuencias de ADN que codificaban una endonucleasa guiada por ARN (*rge*), dos ORF en 3' del gen *rge*, una matriz CRISPR cognada y cuatro ORF inmediatamente en 5' de la matriz CRISPR. Se predijo que los ORF 140 y 423 codificarían una ADN polimerasa propensa a errores, mientras que se predijo que los ORF 624 y 188 codificarían una helicasa y una ARN endonucleasa, respectivamente. El plásmido diana p1631 se construyó creando primero un fragmento de ADN sintético que contenía secuencias similares a espaciadores del sistema CRISPR/Cas Tipo S separadas por la secuencia PAM predicha (AAG), y luego insertando esta secuencia en un plásmido que contenía un origen de replicación p15A, un marcador de resistencia al cloranfenicol (CmR) y el gen *amiCP* produciendo un color morado.
- 10 **Figura 2: El sistema CRISPR-Cas inhibe el crecimiento de colonias transformadas con el plásmido diana p1631.** La cepa bSNP3127 se transformó con p1760 (derecha, véase la Figura 3 para ver el contenido), o p1624 (centro, véase la Figura 1 para ver el contenido) o un control sin plásmido (izquierda). Cada una de las cepas resultantes se transformó con una mezcla 1:1 de plásmidos diana (morado) y no diana (blanco) y se sembró en placas LB suplementadas con los antibióticos adecuados para la selección. No se observaron colonias moradas en presencia de p1760.
- 15 **Figura 3: Representación esquemática de la biblioteca de plásmidos** construida después de la delección de una serie de genes en p1760 (véase la Figura 3 para ver el contenido). Los recuadros indican las delecciones de genes. Las construcciones que mostraron inhibición del plásmido se indican con \*.
- 20 **Figura 4: Cinco genes y una matriz CRISPR comprenden el sistema CRISPR-Cas Tipo S.** Las cepas bSNP3127 que portan algunos de los plásmidos que se muestran en la Figura 3 se transforman con una mezcla 1:1 de plásmidos p1631 diana (morado) y no diana (incolores). Las colonias moradas con los plásmidos ensayados se observaron solo en ausencia de la matriz CRISPR.
- 25 **Figura 5: El sistema CRISPR-Cas utiliza los cinco genes (S1-S5) y una matriz CRISPR para la inhibición de plásmidos.** Las cepas b5700 ( $\Delta$ Cas-S1), b5702 ( $\Delta$ Cas-S3), b5703 ( $\Delta$ Cas-S4), b5704 ( $\Delta$ Cas-S5) portaban plásmidos con delecciones de un solo gen de los genes CRISPR-Cas Tipo-S. Estas cepas, junto con las cepas b4816 (control negativo, sin sistema CRISPR-Cas Tipo-S) y b5408 (control positivo, sistema CRISPR-Cas Tipo-S completo presente) se transformaron con el plásmido diana p1631 o el plásmido no diana p144. Las veces de reducción en las eficiencias de transformación de cada cepa se presentan en esta representación gráfica.
- 30 **Figura 6: La inducción controlada del sistema CRISPR-Cas Tipo S dificulta el crecimiento de colonias con el plásmido diana p1631.**
- 35 **Figura 7: Ensayos de aclaramiento de plásmidos** para la evaluación de la actividad de direccionamiento de cada espaciador en la matriz CRISPR Tipo S. Los cuatro espaciadores ensayados muestran una eficiencia de direccionamiento similar.
- Figura 8: El sistema CRISPR-Cas Tipo S no induce roturas letales de ADNds en el cromosoma de *E. coli*.**
- 40 **Figura 9: Representación esquemática de las posiciones de los protoespaciadores empleados para los ensayos de unión CRISPR-Cas Tipo S.**
- Figura 10: El CRISPR-Cas Tipo S se une al ADN cromosómico.** Curvas de crecimiento de las cepas derivadas de bSNP5810 en presencia de concentraciones crecientes de cloranfenicol.
- 45 **Figura 11: Representación gráfica de las preferencias PAM de CRISPR-Cas Tipo S en forma de 'rueda PAM'** (véase Leenay *et al.*, Technology, 62(1), 137-147, 2016, doi: <https://doi.org/10.1016/j.molcel.2016.02.031>). El anillo interior corresponde a la posición del 3<sup>er</sup> nucleótido aguas arriba (5') del protoespaciador, el anillo medio corresponde a la posición del 2<sup>o</sup> nucleótido aguas arriba (5') del protoespaciador y el anillo exterior corresponde a la posición del 1<sup>er</sup> nucleótido aguas arriba (5') del protoespaciador. En la rueda, cada secuencia ocupa un sector, con un área proporcional al enriquecimiento relativo de la secuencia.
- 50 **Figura 12: Representación gráfica de la eficiencia del direccionamiento de CRISPR/Cas Tipo S en los PAM más preferidos.** El eje x indica los PAM de los miembros de la biblioteca p2259 más diana, según los datos NGS generados. El eje y representa las veces de reducción de lecturas de secuenciación para cada PAM, al comparar los datos NGS generados por la cepa b5408 (direccionamiento de Cas-S) con los datos NGS generados por la cepa b6259 (control Cas-S).
- Figura 13: Representación gráfica de la eficiencia de transformación de b5408 (Tipo-S wt) con cada miembro de la**

biblioteca de plásmidos p1935 en comparación con su eficiencia de transformación con el plásmido de control negativo p2361 (sin PAM). El protoespaciador de tipo salvaje se representa en la parte superior de cada gráfico. Las secciones espaciador-protoespaciador complementarias para cada protoespaciador ensayado se indican con puntos. Las secciones de protoespaciador mutado se indican con los códigos de nucleótidos de una sola letra correspondientes. Su protoespaciador ensayado tiene un código único que se indica en el lado izquierdo de los gráficos.

(A) Errores de emparejamiento proximales de PAM (cadenas); (B) Errores de emparejamiento distales de PAM (cadenas); (C) Errores de emparejamiento simples proximales de PAM; (D) Errores de emparejamiento dobles proximales de PAM; (E) Errores de emparejamiento triples proximales de PAM; (F) Errores de emparejamiento cuádruples proximales de PAM.

Nota: los símbolos de "cruz(es)" indican la presencia de un número bajo (cruz simple) o alto (cruz doble) de "microcolonias" (que no se pueden contar) en las placas de transformación correspondientes. Las microcolonias posiblemente indican un direccionamiento ineficiente de CRISPR/Cas Tipo S que no detiene por completo la replicación del plásmido. Como resultado, las colonias crecen a una tasa significativamente más lenta en el medio de selección.

**Figura 14: Representación gráfica de los plásmidos utilizados en los Ejemplos 4 y 5.** La variante de Tipo-S que cada plásmido expresa se indica en el lado izquierdo de la figura. El nombre de cada plásmido se indica en el lado inferior izquierdo de cada plásmido. Los genes de Tipo-S se presentan como recuadros con rellenos sólidos y los nombres de los genes correspondientes se indican encima de cada recuadro. Las líneas punteadas corresponden a los genes de Tipo-S delecionados. La expresión de los genes de Tipo-S estaba dirigida por el promotor pB<sub>ol</sub>A, que se presenta como una flecha que apunta a la dirección de la transcripción. La expresión del módulo ARNcr estaba dirigida por la secuencia 'Líder', presentada como un recuadro lleno con rayas diagonales y está localizado justo aguas arriba (5') del módulo de expresión de ARNcr. Las repeticiones CRISPR (SEQ ID No:6) del módulo de expresión de ARNcr se presentan como triángulos y el espaciador se presenta como un círculo. Las localizaciones de los sitios de reconocimiento de BsmBI se indican con signos de asterisco (\*). El gen del marcador de resistencia a la tetraciclina, el origen de replicación SC101 y el gen de la proteína de replicación repA101 se presentan como recuadros con rellenos de cuadros, ladrillos y cuadrículas respectivamente.

**Figura 15: Representación esquemática de las posiciones de los protoespaciadores empleados en los ensayos CRISPRi de Tipo-S para la regulación a la baja de la expresión de GFP de la cepa b5815 que contiene el gen gfp integrado cromosómicamente.** Cada posición de protoespaciador se indica con un triángulo y el ID de cada protoespaciador se indica sobre el triángulo.

**Figura 16: Representación gráfica de la actividad CRISPRi cromosómica del sistema Tipo-S de tipo salvaje sobre la expresión de GFP de la cepa b5815 durante un periodo de 24 horas.** Los protoespaciadores seleccionados estaban localizados A) al comienzo del orf de gfp y en la región promotora p70a (Protoespaciadores indicados entre paréntesis: S1F, S1R, S2F, S2R, S3F, S3R), B) al final del orf de gfp y las regiones de 200pb en 5' y en 3' de gfp (Protoespaciadores indicados entre paréntesis S4F, S4R, S5F, S6F, S6R) C) 1 kb en 5' (Protoespaciadores indicados entre paréntesis: U1kF, U1kR) y 2 kb en 5' (Protoespaciadores indicados entre paréntesis: U2kF, U2kR) del orf de gfp.

**Figura 17: Representación gráfica de la actividad CRISPRi cromosómica del sistema ΔcasS3 Tipo-S sobre la expresión de GFP de la cepa b5815 durante un periodo de 24 horas.** Los protoespaciadores seleccionados estaban localizados A) al comienzo del orf de gfp y en la región promotora p70a (Protoespaciadores S1F, S1R, S2F, S2R, S3F, S3R), B) al final del orf de gfp y las regiones de 200 pb en 5' y en 3' de gfp (Protoespaciadores S4F, S4R, S5F, S6F, S6R).

**Figura 18: Representación gráfica de la actividad CRISPRi cromosómica del A) sistema ΔcasS3 Tipo-S, B) sistema ΔcasS1ΔcasS3 Tipo-S y C) sistema ΔcasS1 Tipo-S sobre la expresión de GFP de la cepa b5815 durante un periodo de 24 horas.** Los protoespaciadores seleccionados estaban localizados al comienzo del orf de gfp y en la región promotora p70a (Protoespaciadores S1F, S1R, S2F, S2R, S3F, S3R). Los plásmidos empleados para la transformación de la cepa b5815 se indican con (:).

**Figura 19: Representación gráfica de la actividad CRISPRi en el plásmido del A) sistema ΔcasS3 Tipo-S, B) sistema ΔcasS4 Tipo-S, C) sistema ΔcasS4 Tipo-S y D) sistema ΔcasS1ΔcasS3 Tipo-S sobre la expresión de GFP de la cepa b6386 durante un periodo de 24 horas.** Los protoespaciadores seleccionados estaban localizados al comienzo del orf de gfp y en la región promotora p70a (Protoespaciadores S1F, S1R, S2F, S2R, S3F, S3R).

**Figura 20: Representación esquemática de las combinaciones de plásmidos empleadas para los ensayos de edición de bases A) CasS1, B) CasS1 (ΔCasS3), C) CasS3 y D) CasS4.** El asterisco (\*) indica que el espaciador está dirigido contra uno de: un control no dirigido, S1F, S1R, S2F, S2R, S3F o S3R.

**Figura 21: Visualización basada en mapas de calor de resultados de secuenciación representativos de los ensayos de edición de bases de CRISPR/Cas Tipo-S.** La citidina desaminasa PmCDA1 se fusiona al extremo C de A) la subunidad Cas-S1 del sistema CRISPR/Cas Tipo-S de tipo salvaje, B) la subunidad Cas-S1 del sistema CRISPR/Cas Tipo-S ΔCas-S3, C) la subunidad Cas-S3 del sistema CRISPR/Cas Tipo-S de tipo salvaje, D) la

subunidad Cas-S2 del sistema CRISPR/Cas Tipo-S de tipo salvaje. En la parte superior de cada recuadro se indican i) el código del protoespaciador ensayado y ii) la secuencia de nucleótidos del protoespaciador (resaltada con una flecha cuadrada que además muestra la dirección del protoespaciador) junto con la secuencia de las regiones genómicas inmediatamente vecinas donde se detectaron modificaciones de edición de bases. En el lado inferior de cada recuadro se indican i) los códigos de resultados de secuenciación de Sanger únicos (cada uno de una colonia distinta) y ii) todas las modificaciones de C a T, o modificaciones de G a A cuando se tomó como diana y se modificó la cadena inversa, que se detectaron en los resultados de la secuenciación. La intensidad del color de cada celda del mapa de calor representa la abundancia relativa de C a T (o G a A) para cada posición, según la relación de altura de los picos C y T correspondientes de los resultados de la secuenciación de Sanger. Por ejemplo, los recuadros negros representan colonias con un genotipo mutado de C a T del 100 % para las posiciones correspondientes, mientras que los recuadros blanquecinos corresponden a genotipos mutantes mixtos de C a T del 95 % al 5 % (límite confiable de detección para la secuenciación de Sanger) para las posiciones correspondientes.

**Figura 22: Representación esquemática de las combinaciones de plásmidos empleadas para los ensayos de direccionamiento del plásmido TevCasS.**

**Figura 23: Representación gráfica de los resultados de los ensayos de direccionamiento del plásmido TevCasS.** El eje x corresponde a la caída en la eficiencia de transformación de una cepa MG1655 que expresa TevCasS1 y está dirigida a la secuencia diana en un plásmido diana TevCasS en relación con la eficiencia de transformación de una cepa MG1655 que expresa TevCasS1 y no está dirigida a la secuencia diana en un plásmido diana TevCasS. El eje y corresponde al tamaño de la secuencia espaciadora I-TevI en la secuencia diana de un plásmido diana TevCasS.

**Figura 24: Representación gráfica de los resultados de los ensayos de direccionamiento cromosómico de TevCasS.** El eje x corresponde a la caída en la eficiencia de transformación de una cepa MG1655 que expresa TevCasS1 y está dirigida a una secuencia diana cromosómica de TevCasS en relación con la eficiencia de transformación de una cepa MG1655 que expresa TevCasS1 y no está dirigida a una secuencia diana cromosómica de TevCasS. El eje y corresponde al tamaño de la secuencia espaciadora I-TevI en la secuencia diana de una secuencia diana cromosómica de TevCasS.

**Figura 25: Representación gráfica de las preferencias del sitio de escisión diana de TevCasS en forma de 'gráfico de Krona'.** La secuencia del sitio de escisión diana de TevCasS es 5'CN<sub>1</sub>N<sub>2</sub>N<sub>3</sub>G-3'. El anillo interior corresponde a N<sub>1</sub>, el anillo medio corresponde al nucleótido N<sub>2</sub> y el anillo exterior corresponde a N<sub>3</sub>. En la rueda, cada secuencia ocupa un sector, con un área proporcional al enriquecimiento relativo de la secuencia.

#### Descripción detallada

La tecnología descrita en la presente memoria se refiere a sistemas CRISPR/Cas nuevos y sintéticos y componentes de los mismos, vectores, proteínas, complejos de ribonucleoproteínas y métodos de uso del nuevo sistema y los componentes.

Llamamos al nuevo sistema "**CRISPR-S**" y nos referimos a él como un sistema "**CRISPR/Cas Tipo-S**".

Los componentes básicos de este sistema no se basan en los genes característicos de ninguno de los sistemas de Tipo I-III, o V o VI que ya se han dilucidado. El Tipo-S no utiliza una Cas3 (la proteína característica de los sistemas de Tipo I), una Cas9 (la proteína característica de los sistemas de Tipo II), una Cas10 (la proteína característica de los sistemas de Tipo III), una Cas12 (la proteína característica de los sistemas de Tipo V) o una Cas13 (la proteína característica de los sistemas de Tipo VI). Los componentes básicos no comprenden una proteína IscB, IsrB o IshB (las IscB y las IsrB comparten una historia evolutiva común con Cas9). Nos referimos al nuevo sistema como Tipo-S. Por lo tanto, en un ejemplo, el o los vectores de la presente memoria no codifican una Cas3, 9, 10, 12 y/o 13. El ácido nucleico de la presente memoria puede no codificar una Cas3, 9, 10, 12 y/o 13. El complejo de la presente memoria puede no comprender una Cas3, 9, 10, 12 y/o 13. El método de la presente memoria puede no utilizar una Cas3, 9, 10, 12 y/o 13. En un ejemplo, el o los vectores o ácidos nucleicos de la presente memoria no codifican una TnpB. En un ejemplo, los complejos descritos en la presente memoria no comprenden una TnpB. En un ejemplo, el o los vectores o ácidos nucleicos de la presente memoria no codifican un dominio de nucleasa HD. En un ejemplo, los complejos de la presente memoria no comprenden un dominio de nucleasa HD. En un ejemplo, el o los vectores o ácidos nucleicos de la presente memoria codifican una Cas8. En un ejemplo, los complejos de la presente memoria comprenden una Cas8.

Como se mencionó anteriormente, la invención reivindicada requiere la presencia obligatoria de los ácidos nucleicos que se indican en la reivindicación 3(a) y la reivindicación 3(b) y la reivindicación 3(c) juntos en combinación con un ARNcr como se reivindica; o los polipéptidos que se indican en la reivindicación 5(a) y la reivindicación 5(b) y la reivindicación 5(c) juntos en combinación con un ARNcr como se reivindica. Cualquier parte de la siguiente divulgación relacionada con estos ácidos nucleicos y proteínas específicos individualmente no forma parte de la invención *per se*, pero sirve para la comprensión de la invención reivindicada.

Los complejos pueden comprender una proteína para la identificación de PAM, una proteína para la interacción con una Cas8, una proteína para procesar pre-ARNcr y una proteína de soporte, en donde los complejos comprenden una

proteína Cas-S1 a Cas-S5. Los complejos pueden comprender una proteína para la identificación de PAM, una proteína para la interacción con una Cas8 y una proteína de soporte, en donde el complejo comprende una proteína Cas-S1, Cas-S2, Cas-S4 y Cas-S5.

En un ejemplo, el o los vectores o ácidos nucleicos comprenden una matriz CRISPR para producir el ARNcr. La matriz puede comprender una o más secuencias repetidas y una o más secuencias espaciadoras. Las secuencias espaciadoras pueden ser complementarias (como se define en otra parte de la presente memoria) a una primera secuencia protoespaciadora. En algunos ejemplos, las secuencias espaciadoras de la matriz pueden ser complementarias a un segundo, tercer, cuarto... etc. protoespaciador en una secuencia diana, por ejemplo, si se desea una edición o modificación multiplex. Por ejemplo, la secuencia repetida es la SEQ ID NO:6 o SEQ ID No:70 o una secuencia que es idéntica excepto por 1, 2, 3, 4 o 5 cambios, en particular 1 o 2, por ejemplo, 1 cambio.

El o los vectores o ácidos nucleicos de la presente memoria pueden comprender una matriz de Tipo-S, es decir, una matriz que codifica un ARNcr que es operable con al menos Cas-S1, 2, 3, 4 y 5 para tomar como diana un primer protoespaciador (o, por ejemplo, un primer y un segundo protoespaciador) de un ácido nucleico o una secuencia diana. El o los vectores o ácidos nucleicos de la presente memoria pueden comprender una matriz de Tipo-S, es decir, una matriz que codifica un ARNcr que es operable con al menos Cas-S1, 2, 4 y 5 para tomar como diana un primer protoespaciador (o, por ejemplo, un primer y un segundo protoespaciador) de un ácido nucleico o una secuencia diana. El o los vectores o ácidos nucleicos de la presente memoria pueden comprender una matriz de Tipo-S, es decir, una matriz que codifica un ARNcr que es operable con al menos Cas-S2, 3, 4 y 5 para tomar como diana un primer protoespaciador (o, por ejemplo, un primer y un segundo protoespaciador) de un ácido nucleico o una secuencia diana. El o los vectores o ácidos nucleicos de la presente memoria pueden comprender una matriz de Tipo-S, es decir, una matriz que codifica un ARNcr que es operable con al menos Cas-S2, 4 y 5 para tomar como diana un primer protoespaciador (o, por ejemplo, un primer y un segundo protoespaciador) de un ácido nucleico o una secuencia diana.

El protoespaciador se encuentra en 3' de un PAM 5'-AAG-3' en el ácido nucleico o secuencia diana. Alternativamente,

- el PAM puede ser AAN, ANG, NAG; o
- el PAM puede ser AAG sin cambio de nucleótido o con un solo cambio de nucleótido.

El PAM no es CGG.

Alternativamente, el PAM puede ser AHN, KAG, AGG, GAC o GTG.

Alternativamente, el PAM se selecciona de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3', 5'-ACA-3', 5'-ACT-3', 5'-ATC-3', 5'-ATA-3', 5'-GAG-3', 5'-TAG-3', 5'-ACC-3', 5'-AGG-3', 5'-ATT-3', 5'-GAC-3' y 5'-GTG-3'. Por ejemplo, el PAM se selecciona de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3', 5'-ACA-3', 5'-ACT-3', 5'-ATC-3', 5'-ATA-3', 5'-GAG-3' y 5'-TAG-3'. Por ejemplo, el PAM se selecciona de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3' y 5'-ACA-3', por ejemplo, 5'-AAC-3' y 5'-ATG-3'. En particular, el PAM es 5'-AAC-3'.

Significativamente, identificamos que los componentes del sistema Tipo-S (Cas-S1, Cas-S2, Cas-S3, Cas-S4 y Cas-S5) no comprenden una nucleasa de ADN guiada por ARN. Sin embargo, descubrimos sorprendentemente que un complejo de ribonucleoproteína de estos componentes con un ARNcr es capaz de dirigirse de manera efectiva y específica a un sitio predeterminado en el ADN. Usando este sistema, también fue sorprendentemente posible modificar una célula sin inducir roturas de ADNds, véase el Ejemplo 1.3.4 a continuación en la presente memoria. Armados con este conocimiento, podemos ver muchas aplicaciones del nuevo sistema. Es posible, por ejemplo, proporcionar una o más proteínas o dominios con actividad efectora (por ejemplo, nucleasa (véase el Ejemplo 5 a continuación en la presente memoria), edición de bases (véase el Ejemplo 4 a continuación en la presente memoria) o actividad de *prime editing*) junto con uno o más de los componentes del sistema para producir sistemas sintéticos que pueden modificar células y ácidos nucleicos de una manera dirigida por ARN o dirigida por ADN.

Por lo tanto, el sistema Tipo-S y sus componentes proporcionan nuevos medios para llevar a cabo la edición de ácidos nucleicos y el direccionamiento celular en entornos ambientales, médicos y de otro tipo.

En un primer aspecto de la descripción detallada, se proporcionan los siguientes aspectos del nuevo sistema Tipo-S:

Una proteína Cas-S1 o un ácido nucleico que codifica una Cas-S1.

Una proteína Cas-S2 o un ácido nucleico que codifica una Cas-S2.

Una proteína Cas-S3 o un ácido nucleico que codifica una Cas-S3.

Una proteína Cas-S4 o un ácido nucleico que codifica una Cas-S4.

Una proteína Cas-S5 o un ácido nucleico que codifica una Cas-S5.

Uno o más ácidos nucleicos que codifican una proteína Cas-S1, S2, S3, S4 y S5.

Uno o más ácidos nucleicos que codifican una proteína Cas-S4 y S5.

Uno o más ácidos nucleicos que codifican una proteína Cas-S3, S4 y S5.

Uno o más ácidos nucleicos que codifican una proteína Cas-S2, S3, S4 y S5.

Uno o más ácidos nucleicos que codifican una proteína Cas-S1, S2, S4 y S5.

5 Uno o más ácidos nucleicos que codifican una proteína Cas-S2, S3, S4 y S5.

Uno o más ácidos nucleicos que codifican una proteína Cas-S2, S4 y S5.

10 Un vector o ácido nucleico que codifica una proteína Cas-S1, S2, S3, S4 y S5, en donde el vector o ácido nucleico comprende la SEQ ID NO: 12 o una secuencia de ADN que es al menos (aproximadamente) un 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 % idéntica a la SEQ ID NO: 12.

Un vector de ácido nucleico o un ácido nucleico que comprende la SEQ ID NO:12 o una secuencia de ADN que es al menos (aproximadamente) un 80 % (o al menos aproximadamente un 80 %, 85 %, 90 %, 91, 92, 93, 94, 95, 96, 97, 98 o al menos aproximadamente un 99 %) idéntica a la SEQ ID NO:12.

15 En un ejemplo, cada una de dichas secuencias está conectada operativamente a un promotor heterólogo (es decir, un promotor que no está conectado operativamente de forma natural a la secuencia). En un ejemplo, cada una de dichas secuencias está conectada operativamente a un promotor que es un promotor eucariota o un promotor de virus (por ejemplo, fago). El promotor puede ser un promotor constitutivo. El promotor puede ser un promotor inducible. Los promotores pueden ser como se describe en cualquier opción o característica descrita en la presente memoria.

20 Opcionalmente, ninguno del o de los ácidos nucleicos o vectores codifica una nucleasa Cas. El o los ácidos nucleicos pueden codificar una proteína o dominio efector que no es Cas, como una nucleasa que no sea una nucleasa Cas.

Se proporciona, en un segundo aspecto:

Un complejo de proteína o ribonucleoproteína que comprende 1, 2, 3, 4 o la totalidad de una proteína Cas-S1, S2, S3, S4 y S5.

Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S1, S2, S3, S4 y S5.

25 Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S4 y S5.

Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S3, S4 y S5.

Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S2, S3, S4 y S5.

Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S1, S2, S4 y S5.

Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S2, S3, S4 y S5.

30 Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S2, S4 y S5.

Opcionalmente, el complejo no comprende una nucleasa Cas. El complejo puede comprender una proteína o dominio efector que no es Cas, como una nucleasa que no sea una nucleasa Cas.

Las proteínas de Tipo-S son capaces de realizar modificaciones dirigidas por ARN de una secuencia de ADN, opcionalmente sin crear una ruptura en el ADN. Tal como se utiliza en la presente memoria:

35 **Proteínas Cas-S1:** Cas-S1 es una proteína Cas que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO: 1. Preferiblemente, la identidad es al menos (aproximadamente) un 80 %. Aún más preferiblemente, la identidad es al menos (aproximadamente) un 90 o 95 %. En otro ejemplo, la identidad es al menos aproximadamente un 96 %. En otro ejemplo, la identidad es al menos aproximadamente un 97 %. En otro ejemplo, la identidad es al menos aproximadamente un 98 %. En otro ejemplo, la identidad es al menos aproximadamente un 99 %.

40 Preferiblemente, la Cas-S1 es Cas-S1.1. Cas-S1.1 es una proteína Cas que comprende el aminoácido de la SEQ ID NO: 1.

**Proteínas Cas-S2:** Cas-S2 es una proteína Cas que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO: 2. Preferiblemente, la identidad es al menos (aproximadamente) un 80 %. Aún más preferiblemente, la identidad es al menos (aproximadamente) un 90 o 95 %. En otro ejemplo, la identidad es al menos aproximadamente un 96 %. En otro ejemplo, la identidad es al menos aproximadamente un 97 %. En otro ejemplo, la identidad es al menos aproximadamente un 98 %. En otro ejemplo, la identidad es de al menos aproximadamente un

99 %. Preferentemente, la Cas-S2 es Cas-S2.1. Cas-S2.1 es una proteína Cas que comprende el aminoácido de la SEQ ID NO:2.

**Proteínas Cas-S3:** Cas-S3 es una proteína Cas que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO: 3. Preferiblemente, la identidad es al menos (aproximadamente) un 80 %. Aún más preferiblemente, la identidad es al menos (aproximadamente) un 90 o 95 %. En otro ejemplo, la identidad es al menos aproximadamente un 96 %. En otro ejemplo, la identidad es al menos aproximadamente un 97 %. En otro ejemplo, la identidad es al menos aproximadamente un 98 %. En otro ejemplo, la identidad es de al menos aproximadamente un 99 %. Preferentemente, la Cas-S3 es Cas-S3.1. Cas-S3.1 es una proteína Cas que comprende el aminoácido de la SEQ ID NO:3.

**Proteínas Cas-S4:** Cas-S4 es una proteína Cas que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO: 4. Preferiblemente, la identidad es al menos (aproximadamente) un 80 %. Aún más preferiblemente, la identidad es al menos (aproximadamente) un 90 o 95 %. En otro ejemplo, la identidad es al menos aproximadamente un 96 %. En otro ejemplo, la identidad es al menos aproximadamente un 97 %. En otro ejemplo, la identidad es al menos aproximadamente un 98 %. En otro ejemplo, la identidad es de al menos aproximadamente un 99 %. Preferentemente, la Cas-S4 es Cas-S4.1. Cas-S4.1 es una proteína Cas que comprende el aminoácido de la SEQ ID NO:4.

**Proteínas Cas-S5:** Cas-S5 es una proteína Cas que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO: 5. Preferiblemente, la identidad es al menos (aproximadamente) un 80 %. Aún más preferiblemente, la identidad es al menos (aproximadamente) un 90 o 95 %. En otro ejemplo, la identidad es al menos aproximadamente un 96 %. En otro ejemplo, la identidad es al menos aproximadamente un 97 %. En otro ejemplo, la identidad es al menos aproximadamente un 98 %. En otro ejemplo, la identidad es de al menos aproximadamente un 99 %. Preferentemente, la Cas-S5 es Cas-S5.1. Cas-S5.1 es una proteína Cas que comprende el aminoácido de la SEQ ID NO:5.

Por ejemplo, cualquier porcentaje de identidad en la presente memoria, a menos que se especifique lo contrario, es al menos (aproximadamente) un 70 %. Por ejemplo, cualquier porcentaje de identidad en la presente memoria, a menos que se especifique lo contrario, es al menos (aproximadamente) un 80 %. Por ejemplo, cualquier porcentaje de identidad en la presente memoria, a menos que se especifique lo contrario, es al menos (aproximadamente) un 90 %. Por ejemplo, cualquier porcentaje de identidad en la presente memoria es al menos un 95 %. Por ejemplo, cualquier porcentaje de identidad en la presente memoria es al menos aproximadamente un 96 %. Por ejemplo, cualquier porcentaje de identidad en la presente memoria es al menos aproximadamente un 97 %. Por ejemplo, cualquier porcentaje de identidad es al menos aproximadamente un 98 %. Por ejemplo, cualquier porcentaje de identidad en la presente memoria es al menos aproximadamente un 99 %.

El porcentaje de identidad de las secuencias de aminoácidos se determina utilizando el algoritmo blastP con los siguientes parámetros:

Los parámetros por defecto se ajustan para secuencias de entrada cortas, el umbral esperado se establece en 0,05 y la longitud de la secuencia semilla que inicia una alineación se establece en 6. Las regiones de baja complejidad compositiva se enmascaran. La matriz de puntuación empleada es 'BLOSUM62', siendo los costos de puntuación para crear y extender un hueco de 11 y 1 respectivamente. Se emplea un ajuste condicional de la matriz de puntuación compositiva para compensar la composición de aminoácidos de las secuencias comparadas.

El porcentaje de identidad de las secuencias de nucleótidos se determina utilizando el algoritmo blastn con los siguientes parámetros:

Los parámetros por defecto se ajustan automáticamente para secuencias de entrada cortas, el umbral esperado se establece en 0,05 y la longitud de la secuencia de semillas que inicia una alineación se establece en 28. Las regiones de baja complejidad compositiva se enmascaran. La secuencia de consulta se enmascara mientras se producen secuencias semilla que se utilizan para escanear bases de datos, pero no se enmascara para extensiones. Las coincidencias se puntúan como +1 y las discordancias se puntúan como -2.

Por ejemplo, una secuencia de aminoácidos descrita en la presente memoria es idéntica a la SEQ ID NO de referencia excepto por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 cambios de aminoácidos, en particular de 1 a 5, por ejemplo, de 1 a 3, tal como 1 o 2, por ejemplo, 1 cambio de aminoácido. Por ejemplo, una secuencia de ácido nucleico descrita en la presente memoria es idéntica a la SEQ ID NO de referencia excepto por 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 cambios de nucleótidos. Por ejemplo, una secuencia de aminoácidos del polipéptido es idéntica a la SEQ ID NO: 1 excepto por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 cambios de aminoácidos, en particular de 1 a 5, por ejemplo, de 1 a 3, tal como 1 o 2, por ejemplo, 1 cambio de aminoácido. Por ejemplo, una secuencia de aminoácidos del polipéptido es idéntica a la SEQ ID NO: 2 excepto por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 cambios de aminoácidos, en particular de 1 a 5, por ejemplo, de 1 a 3, tal como 1 o 2, por ejemplo, 1 cambio de

aminoácido. Por ejemplo, una secuencia de aminoácidos del polipéptido es idéntica a la SEQ ID NO: 3 excepto por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 cambios de aminoácidos, en particular de 1 a 5, por ejemplo, de 1 a 3, tal como 1 o 2, por ejemplo, 1 cambio de aminoácido. Por ejemplo, una secuencia de aminoácidos del polipéptido es idéntica a la SEQ ID NO: 4 excepto por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 cambios de aminoácidos, en particular de 1 a 5, por ejemplo, de 1 a 3, tal como 1 o 2, por ejemplo, 1 cambio de aminoácido. Por ejemplo, una secuencia de aminoácidos del polipéptido es idéntica a la SEQ ID NO: 5 excepto por 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 cambios de aminoácidos, en particular de 1 a 5, por ejemplo, de 1 a 3, tal como 1 o 2, por ejemplo, 1 cambio de aminoácido.

Por ejemplo, una secuencia de aminoácidos descrita en la presente memoria es idéntica a la SEQ ID NO de referencia, excepto por un número total de cambios de aminoácidos en donde el número total no es más de (aproximadamente) el 5, 10, 15, 20, 25 o 30 % del número de aminoácidos en la secuencia de referencia. Por ejemplo, una secuencia de nucleótidos descrita en la presente memoria es idéntica a la SEQ ID NO de referencia, excepto por un número total de cambios de nucleótidos en donde el número total no es más de (aproximadamente) el 5, 10, 20, 25 o 30 % del número de nucleótidos en la secuencia de referencia. Así, en un ejemplo, la secuencia de aminoácidos del polipéptido es la SEQ ID NO:1 o una secuencia de aminoácidos que es idéntica a la SEQ ID NO:1 excepto por un número total de cambios de aminoácidos en donde el número total no es más de (aproximadamente) el 5, 10, 15, 20, 25 o 30 % del número de aminoácidos en la SEQ ID NO:1, en particular no más de (aproximadamente) el 20 %, tal como no más de (aproximadamente) el 10 % del número de aminoácidos en la SEQ ID NO:1. Así, en un ejemplo, la secuencia de aminoácidos del polipéptido es la SEQ ID NO:2 o una secuencia de aminoácidos que es idéntica a la SEQ ID NO:2 excepto por un número total de cambios de aminoácidos en donde el número total no es más de (aproximadamente) el 5, 10, 15, 20, 25 o 30 % del número de aminoácidos en la SEQ ID NO:2, en particular no más de (aproximadamente) el 20 %, tal como no más de (aproximadamente) el 10 % del número de aminoácidos en la SEQ ID NO:2. Así, en un ejemplo, la secuencia de aminoácidos del polipéptido es la SEQ ID NO:3 o una secuencia de aminoácidos que es idéntica a la SEQ ID NO:3 excepto por un número total de cambios de aminoácidos en donde el número total no es más de (aproximadamente) el 5, 10, 15, 20, 25 o 30 % del número de aminoácidos en la SEQ ID NO:3, en particular no más de (aproximadamente) el 20 %, tal como no más de (aproximadamente) el 10 % del número de aminoácidos en la SEQ ID NO:3. Así, en un ejemplo, la secuencia de aminoácidos del polipéptido es la SEQ ID NO:4 o una secuencia de aminoácidos que es idéntica a la SEQ ID NO:4 excepto por un número total de cambios de aminoácidos en donde el número total no es más de (aproximadamente) el 5, 10, 15, 20, 25 o 30 % del número de aminoácidos en la SEQ ID NO:4, en particular no más de (aproximadamente) el 20 %, tal como no más de (aproximadamente) el 10 % del número de aminoácidos en la SEQ ID NO:4. Así, en un ejemplo, la secuencia de aminoácidos del polipéptido es la SEQ ID NO:5 o una secuencia de aminoácidos que es idéntica a la SEQ ID NO:5 excepto por un número total de cambios de aminoácidos en donde el número total no es más de (aproximadamente) el 5, 10, 15, 20, 25 o 30 % del número de aminoácidos en la SEQ ID NO:5, en particular no más de (aproximadamente) el 20 %, tal como no más de (aproximadamente) el 10 % del número de aminoácidos en la SEQ ID NO:5.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1, 2, 3, 4 y 5.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S4 y 5.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S3, 4 y 5.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S2, 3, 4 y 5.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1 y 2.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1, 2 y 3.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1, 2, 3 y 4.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1, 2, 4 y 5.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S2, 4 y 5. En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1.1, 2.1, 3.1, 4.1 y 5.1.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S4.1 y 5.1.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S3.1, 4.1 y 5.1.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S2.1, 3.1, 4.1 y 5.1.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1.1 y 2.1.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1.1, 2.1 y 3.1.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1.1, 2.1, 3.2 y 4.1.

En un ejemplo, el o los vectores de la presente memoria codifican Cas-S1.1, 2.1, 4.1 y 5.1.

- En un ejemplo, el o los vectores de la presente memoria codifican Cas-S2.1, 4.1 y 5.1.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1, 2, 3, 4 y 5.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S4 y 5.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S3, 4 y 5.
- 5 En un ejemplo, el complejo de la presente memoria comprende Cas-S2, 3, 4 y 5.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1 y 2.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1, 2 y 3.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1, 2, 3 y 4.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1, 2, 4 y 5.
- 10 En un ejemplo, el complejo de la presente memoria comprende Cas-S2, 4 y 5.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1.1, 2.1, 3.1, 4.1 y 5.1.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S4.1 y 5.1.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S3.1, 4.1 y 5.1.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S2.1, 3.1, 4.1 y 5.1.
- 15 En un ejemplo, el complejo de la presente memoria comprende Cas-S1.1 y 2.1.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1.1, 2.1 y 3.1.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1.1, 2.1, 3.1 y 4.1.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S1.1, 2.1, 4.1 y 5.1.
- En un ejemplo, el complejo de la presente memoria comprende Cas-S2.1, 4.1 y 5.1.
- 20 En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S1.1, 2.1, 3.1, 4.1 o 5.1.
- En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S1.1, 2.1, 3.1, 4.1 y 5.1.
- En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S4.1 y 5.1.
- En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S3.1, 4.1 y 5.1.
- En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S2.1, 3.1, 4.1 y 5.1.
- 25 En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S1.1 y 2.1.
- En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S1.1, 2.1 y 3.1.
- En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S1.1, 2.1, 3.1 y 4.1.
- En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S1.1, 2.1, 4.1 y 5.1.
- En un ejemplo, la proteína de fusión de la presente memoria comprende Cas-S2.1, 4.1 y 5.1.
- 30 En un ejemplo, el método de la presente memoria utiliza Cas-S1.1, 2.1, 3.1, 4.1 o 5.1.
- En un ejemplo, el método de la presente memoria utiliza Cas-S1.1, 2.1, 3.1, 4.1 y 5.1.
- En un ejemplo, el método de la presente memoria utiliza Cas-S4.1 y 5.1.
- En un ejemplo, el método de la presente memoria utiliza Cas-S3.1, 4.1 y 5.1.
- En un ejemplo, el método de la presente memoria utiliza Cas-S2.1, 3.1, 4.1 y 5.1.
- 35 En un ejemplo, el método de la presente memoria utiliza Cas-S1.1 y 2.1.
- En un ejemplo, el método de la presente memoria utiliza Cas-S1.1, 2.1 y 3.1.
- En un ejemplo, el método de la presente memoria utiliza Cas-S1.1, 2.1, 3.1 y 4.1.

En un ejemplo, el método de la presente memoria utiliza Cas-S1.1, 2.1, 4.1 y 5.1.

En un ejemplo, el método de la presente memoria utiliza Cas-S2.1, 4.1 y 5.1.

En un aspecto, se proporciona uno o más vectores y ácidos nucleicos que comprenden una o más secuencias de nucleótidos seleccionadas de las SEQ ID NO:7-11.

5 Se proporciona:-

Uno o más ácidos nucleicos que comprenden las SEQ ID NO:7-11.

Uno o más ácidos nucleicos que comprenden las SEQ ID NO:8-11.

Uno o más ácidos nucleicos que comprenden las SEQ ID NO:9-11.

Uno o más ácidos nucleicos que comprenden las SEQ ID NO:10 y 11.

10 Uno o más ácidos nucleicos que comprenden las SEQ ID NO:7, 8, 10 y 11.

Uno o más ácidos nucleicos que comprenden las SEQ ID NO:8, 9, 10 y 11.

Uno o más ácidos nucleicos que comprenden las SEQ ID NO:8, 10 y 11.

Un ácido nucleico que comprende la SEQ ID NO:7.

Un ácido nucleico que comprende la SEQ ID NO:8.

15 Un ácido nucleico que comprende la SEQ ID NO: 9.

Un ácido nucleico que comprende la SEQ ID NO: 10.

Un ácido nucleico que comprende la SEQ ID NO: 11.

Un ácido nucleico que comprende las SEQ ID NO:7-11.

Un ácido nucleico que comprende las SEQ ID NO:7, 8, 10 y 11.

20 Un ácido nucleico que comprende las SEQ ID NO:8, 9, 10 y 11.

Un ácido nucleico que comprende las SEQ ID NO:8, 10 y 11.

Uno o más vectores de ácidos nucleicos que comprenden las SEQ ID NO:7-11.

Uno o más vectores de ácidos nucleicos que comprenden las SEQ ID NO:8-11.

Uno o más vectores de ácidos nucleicos que comprenden las SEQ ID NO:9-11.

25 Uno o más vectores de ácidos nucleicos que comprenden las SEQ ID NO:10 y 11.

Uno o más vectores de ácidos nucleicos que comprenden las SEQ ID NO:7, 8, 10 y 11.

Uno o más vectores de ácidos nucleicos que comprenden las SEQ ID NO:8, 9, 10 y 11.

Uno o más vectores de ácidos nucleicos que comprenden las SEQ ID NO:8, 10 y 11.

Un vector de ácido nucleico que comprende la SEQ ID NO:7.

30 Un vector de ácido nucleico que comprende la SEQ ID NO:8.

Un vector de ácido nucleico que comprende la SEQ ID NO:9.

Un vector de ácido nucleico que comprende la SEQ ID NO:10.

Un vector de ácido nucleico que comprende la SEQ ID NO:11.

Un vector de ácido nucleico que comprende las SEQ ID NO:7-11.

35 Un vector de ácido nucleico que comprende las SEQ ID NO: 7, 8, 10 y 11.

Un vector de ácido nucleico que comprende las SEQ ID NO:8, 9, 10 y 11.

Un vector de ácido nucleico que comprende las SEQ ID NO:8, 10 y 11.

Un vector que comprende la SEQ ID NO: 12.

Un ácido nucleico que comprende la SEQ ID NO: 12.

Opcionalmente, dicha secuencia de nucleótidos recitada o al menos una de dichas secuencias de nucleótidos recitadas está conectada operativamente a un promotor que es

- 5           a)     heterólogo a la(s) secuencia(s) de nucleótidos;
- b)     no es un promotor de *E. coli* o *Klebsiella*;
- c)     un promotor de células eucariotas;
- d)     un promotor animal (opcionalmente, un promotor de mamífero o humano);
- e)     un promotor de planta;
- 10          f)     un promotor de hongos (opcionalmente, un promotor de levadura);
- g)     un promotor de insecto;
- h)     un promotor de virus (opcionalmente, un promotor de virus, AAV o lentivirus); o
- i)     un promotor sintético.

15           Cualquier promotor descrito en la presente memoria se puede utilizar con los vectores y ácidos nucleicos descritos en la presente memoria.

En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula que no es una célula de *E. coli* o *Klebsiella*. (por ejemplo, no en una de *K. pneumoniae*). En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula que no es una célula de *Pseudomonas*. En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula que no es una célula de *E. coli* o *Klebsiella* que comprende una secuencia de nucleótidos endógena que comprende las SEQ ID NO: 7-11. En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula (por ejemplo, una célula bacteriana) que no comprende una secuencia de nucleótidos endógena que comprende las SEQ ID NO: 7-11. En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula eucariota. En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula humana. En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula animal. En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula vegetal. En un ejemplo, el o los vectores o ácidos nucleicos están comprendidos por una célula fúngica.

También se proporciona:

Uno o más vectores de ácido nucleico o uno o más ácidos nucleicos que comprenden al menos una secuencia de nucleótidos seleccionada de las SEQ ID NO: 7-11, en donde una de dicha secuencia de nucleótidos está

- 30           a)     conectada operativamente a un promotor heterólogo, promotor sintético, promotor eucariota o promotor no bacteriano; y/o
- b)     comprendida por una célula que es una célula eucariota, una célula no bacteriana o una célula que no es una célula bacteriana (por ejemplo, una célula de *E. coli*, *Pseudomonas* o *Klebsiella*) que comprende una secuencia de nucleótidos endógena que comprende las SEQ ID NO: 7-11.

35           Uno o más vectores de ácidos nucleicos o uno o más ácidos nucleicos que comprenden al menos una secuencia de nucleótidos seleccionada de las SEQ ID NO: 7-11, en donde una de dicha secuencia de nucleótidos está comprendida por una célula (por ejemplo, una célula bacteriana) que no comprende una secuencia de nucleótidos endógena que comprende las SEQ ID NO: 7-11.

40           Opcionalmente, cualquier vector o ácido nucleico o secuencia de nucleótidos de la presente memoria está optimizada por codones para su uso en una célula humana, animal (por ejemplo, de mamífero, roedor, ratón o rata), vegetal o fúngica. Opcionalmente, cualquier vector o ácido nucleico o secuencia de nucleótidos de la presente memoria está optimizada por codones para su uso en una célula eucariota. Los vectores y secuencias de ácido nucleico para su uso en células eucariotas pueden comprender secuencias de localización nuclear (NLS). Las NLS facilitan la entrada de las proteínas y polipéptidos descritos en la presente memoria en las células eucariotas.

45           Opcionalmente, cualquier vector o ácido nucleico o secuencia de nucleótidos de la presente memoria está optimizada por codones para su uso en una célula procariota, por ejemplo, una célula bacteriana o de arquea. Por ejemplo, la célula bacteriana no es una célula de *E. coli*. Por ejemplo, la célula bacteriana no es una célula de *Klebsiella* (por ejemplo, *K. pneumoniae*). Por ejemplo, la célula bacteriana no es una célula de *Pseudomonas*. En un ejemplo, la célula comprende un ADN o polinucleótido que se modifica o se selecciona como diana mediante los métodos descritos en la presente memoria.

En un ejemplo, la célula de la presente memoria es una célula humana, animal (por ejemplo, de mamífero, roedor, ratón o rata), vegetal o fúngica. En un ejemplo, la célula de la presente memoria es una célula de *Homo sapiens*, *Drosophila melanogaster*, *Mus musculus*, *Rattus norvegicus*, *Caenorhabditis elegans* o *Arabidopsis thaliana*. En un ejemplo, la célula de la presente memoria es una célula procariota, por ejemplo, una célula bacteriana o de arquea. Por ejemplo, la célula bacteriana no es una célula de *E. coli*. Por ejemplo, la célula bacteriana no es una célula de *Klebsiella*. Por ejemplo, la célula bacteriana no es una célula de *Pseudomonas*.

En un ejemplo, un animal o mamífero de la presente memoria es un vertebrado.

Opcionalmente, el ARNcr divulgado en la presente memoria (véase, por ejemplo, el Concepto 10 a continuación) tiene una longitud de 15-100 nucleótidos y comprende una secuencia (por ejemplo, una secuencia espaciadora) de al menos 10 nucleótidos contiguos que es complementaria a la secuencia diana. Por ejemplo, el ARNcr tiene una longitud de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 o 50 nucleótidos. Por ejemplo, el ARNcr comprende una secuencia (por ejemplo, una secuencia espaciadora) de 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 o 40 nucleótidos contiguos que es complementaria a la secuencia diana.

En un ejemplo, el ARNcr es como se describe en otra parte de la presente memoria. El ARNcr puede comprender dos secuencias repetidas. El ARNcr puede comprender al menos una secuencia repetida que tiene la secuencia de nucleótidos de la SEQ ID No:6. El ARNcr puede comprender dos secuencias repetidas (por ejemplo, que tienen la secuencia de nucleótidos de la SEQ ID No:6) y una secuencia espaciadora que se hibrida con una primera secuencia protoespaciadora en una secuencia diana. El ARNcr puede comprender al menos una secuencia repetida que tiene la secuencia de nucleótidos de la SEQ ID No:70. El ARNcr puede comprender dos secuencias repetidas (por ejemplo, que tienen la secuencia de nucleótidos de la SEQ ID No:70) y una secuencia espaciadora que se hibrida con una primera secuencia protoespaciadora en una secuencia diana.

La secuencia espaciadora puede tener una longitud de aproximadamente 25 a 39 nucleótidos. La secuencia espaciadora puede tener una longitud de aproximadamente 29 a 35 (por ejemplo, de aproximadamente 30 a 34, o de aproximadamente 31 a 33) nucleótidos. La secuencia espaciadora puede tener una longitud de aproximadamente 32 nucleótidos. La secuencia espaciadora puede tener una longitud de 32 nucleótidos. El espaciador puede ser (aproximadamente) un 70 % (tal como (aproximadamente) un 80 %, o (aproximadamente) un 90 %, o (aproximadamente) un 95 %) complementario a la secuencia protoespaciadora en la secuencia diana. El espaciador puede ser 100 % complementario a la secuencia protoespaciadora en la secuencia diana. El espaciador y/o protoespaciador pueden ser como se describe en el Concepto 10 en la presente memoria.

La secuencia diana puede ser un ARN. La secuencia diana puede ser un ADN monocatenario.

En un ejemplo particular, la secuencia diana puede ser una secuencia de un ADNds. La secuencia diana puede ser una secuencia en el genoma de un mamífero, por ejemplo, un ser humano. Opcionalmente, la secuencia diana comprende una secuencia asociada con una enfermedad o trastorno en un sujeto humano o animal.

Opcionalmente, la secuencia diana comprende una mutación puntual asociada con una enfermedad o trastorno y el complejo edita una mutación puntual en la secuencia diana. Alternativamente, la secuencia diana está compuesta por un gen cuya expresión es errática (por ejemplo, está presente o sobreexpresada) en una enfermedad o trastorno, y el complejo introduce una mutación puntual (o introduce una o más mutaciones en una región de la secuencia diana) que disrumen o impiden la expresión del gen. La (una o más) mutación(es) puntual(es) pueden estar ubicadas entre aproximadamente 10 y aproximadamente 20 nucleótidos aguas arriba (5') de un PAM 5'-AAG-3' en la secuencia diana. La (una o más) mutación(es) puntual(es) pueden estar ubicadas entre aproximadamente 10 y aproximadamente 400 nucleótidos, o entre aproximadamente 10 y 350 nucleótidos, o entre aproximadamente 10 y 300 nucleótidos, o entre aproximadamente 10 y 250 nucleótidos, o entre aproximadamente 10 y 200 nucleótidos, o entre aproximadamente 10 y 150 nucleótidos aguas arriba (5') o aguas abajo (3') (por ejemplo, aguas abajo) de un PAM en la secuencia diana. La (una o más) mutación(es) puntual(es) pueden estar ubicadas entre aproximadamente 30 y 120 nucleótidos (por ejemplo, entre aproximadamente 50 y 100 nucleótidos) aguas arriba (5') o aguas abajo (3') (por ejemplo, aguas abajo) de un PAM en la secuencia diana. Alternativamente, el

— el PAM puede ser AAN, ANG, NAG; o

— el PAM puede ser AAG sin cambio de nucleótido o con un solo cambio de nucleótido.

Alternativamente, el PAM puede ser AHN, KAG, AGG, GAC o GTG.

Alternativamente, el PAM se selecciona de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3', 5'-ACA-3', 5'-ACT-3', 5'-ATC-3', 5'-ATA-3', 5'-GAG-3', 5'-TAG-3', 5'-ACC-3', 5'-AGG-3', 5'-ATT-3', 5'-GAC-3' y 5'-GTG-3'. Por ejemplo, el PAM se selecciona de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3', 5'-ACA-3', 5'-ACT-3', 5'-ATC-3', 5'-ATA-3', 5'-GAG-3' y 5'-TAG-3'. Por ejemplo, el PAM se selecciona de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3' y 5'-ACA-3', por ejemplo, 5'-AAC-3' y 5'-ATG-3'. En particular, el PAM es 5'-AAC-3'.

Por ejemplo, la mutación es una mutación puntual de C a T (o la una o más mutaciones son mutaciones puntuales de

- C a T), en donde el complejo desamina la mutación puntual de C diana, y en donde la desaminación da como resultado una secuencia que no está asociada con una enfermedad o trastorno. Por ejemplo, la mutación puntual de C diana (o la una o más mutaciones son mutaciones puntuales de C diana) está(n) presente(s) en la cadena de ADN que no es complementaria al ARNcr. Por ejemplo, la (una o más) mutación(es) puntual(es) de C a T se introduce(n) mediante una citidina desaminasa (por ejemplo, una citidina desaminasa PmCDA1 como se describe en otra parte de la presente memoria), que está fusionada (por ejemplo, al extremo C) de una proteína Cas-S1, Cas-S3 o Cas-S4, en particular Cas-S1 o Cas-S3. La proteína de fusión puede comprender todas de Cas-S1, Cas-S2, Cas-S3, Cas-S4 y Cas-S5. Alternativamente, la proteína de fusión que comprende la citidina desaminasa fusionada a una proteína Cas-S1 puede estar desprovista de una proteína Cas-S3. La proteína de fusión puede comprender además una proteína UGI.
- 10 Por tanto, se proporciona un complejo de proteína de fusión que comprende una proteína de fusión en un complejo con otras proteínas, y en donde la proteína de fusión comprende una proteína de la fórmula A-B-C-D-E (en orientación 5' a 3'), en donde:
- A es una proteína CasS seleccionada de una Cas-S1, Cas-S3 o Cas-S4 (en particular, Cas-S1);
- 15 B está presente opcionalmente y, cuando está presente, comprende un conector (por ejemplo, cualquiera de los conectores como se describe en la presente memoria, en particular un conector XTEN, por ejemplo, que tiene la secuencia de aminoácidos de la SEQ ID No:17);
- C es un editor de bases, tal como una citidina desaminasa (por ejemplo, una citidina desaminasa PmCDA1 como se describe en otra parte de la presente memoria, por ejemplo, que tiene la secuencia de aminoácidos de la SEQ ID No:19);
- 20 D está presente opcionalmente y, cuando está presente, comprende un conector (por ejemplo, cualquiera de los conectores como se describe en la presente memoria, en particular el conector de 10 aminoácidos de la SEQ ID No:23);
- E es una proteína inhibidora de la uracil ADN glicosilasa (UGI) (como se describe con más detalle en otra parte de la presente memoria, por ejemplo, que tiene la secuencia de aminoácidos de la SEQ ID No: 21); y
- 25 las demás proteínas en complejo con la proteína de fusión comprenden una proteína Cas-S2 y una proteína Cas-S5 y una proteína Cas-S1 y/o una proteína Cas-S4 según se requiera, de modo que el complejo de proteína de fusión comprende al menos una proteína Cas-S1, Cas-S2, Cas-S4 y Cas-S5.
- En un ejemplo, el complejo de proteína de fusión puede comprender además una proteína Cas-S3. También se proporciona un primer vector que expresa la proteína de fusión de fórmula A-B-C-D-E y un segundo vector que expresa las demás proteínas del complejo de proteína de fusión.
- 30 En un ejemplo, el complejo de proteína de fusión puede comprender además un ARNcr como se describe en otra parte de la presente memoria.
- En otro ejemplo, la mutación es una mutación puntual de A a G, en donde el complejo desamina la mutación puntual A diana y en donde la desaminación da como resultado una secuencia que no está asociada con una enfermedad o trastorno. Por ejemplo, la mutación puntual A diana está presente en la cadena de ADN que no es complementaria al ARNcr.
- 35 En un ejemplo, se proporciona una composición aislada, vector(es), ácido(s) nucleico(s) o proteína(s), tal como se describe en la presente memoria. El aislamiento puede significar, por ejemplo, que la composición, vector(es), ácido(s) nucleico(s) o proteína(s) se proporcionan con exclusión de una célula de *E. coli*.
- 40 Opcionalmente, la secuencia diana está compuesta por un gen cuya expresión es errática (por ejemplo, está presente o sobreexpresada) en una enfermedad o trastorno, y el complejo introduce una rotura bicatenaria (o monocatenaria) en la secuencia diana, disrumpiendo o impidiendo así la expresión del gen. Opcionalmente, la secuencia diana está compuesta por una bacteria patógena que causa una enfermedad o trastorno, y el complejo introduce una rotura bicatenaria (o o una rotura monocatenaria) en el cromosoma de la bacteria, matando así (por ejemplo, matando selectivamente) la bacteria. Opcionalmente, la secuencia diana está compuesta por un gen de resistencia a antibióticos compuesto por una bacteria patógena que causa una enfermedad o trastorno, y el complejo introduce una rotura bicatenaria (o una rotura monocatenaria) en la secuencia diana, disrumpiendo o impidiendo así la expresión del gen del antibiótico y dando como resultado la resensibilización de la bacteria al antibiótico. La rotura bicatenaria (o rotura monocatenaria) puede estar ubicada aguas arriba (5') de una secuencia PAM (por ejemplo, entre aproximadamente 29 y aproximadamente 40 nucleótidos (por ejemplo, entre 30 y 32 nucleótidos) aguas arriba (5') de una secuencia PAM) en la secuencia diana. La secuencia PAM puede ser cualquiera de las descritas en la presente memoria, por ejemplo, seleccionada de AHN, KAG, AGG, GAC y GTG (véase también el Concepto 39 en la presente memoria). Las roturas bicatenarias pueden introducirse mediante proteínas de fusión como se describe en otra parte de la presente memoria en las que Py es una nucleasa (por ejemplo, cualquier nucleasa descrita en la presente memoria). Las roturas monocatenarias en un cromosoma (o en ADN bicatenario) pueden introducirse mediante proteínas de fusión como se describe en otra parte de la presente memoria en las que Py es una nickasa (por ejemplo, cualquier nickasa descrita
- 55

en la presente memoria).

Por ejemplo, la rotura bicatenaria se introduce mediante una nucleasa. Por ejemplo, una rotura monocatenaria se introduce mediante una nickasa. Por ejemplo, la rotura bicatenaria se introduce mediante una nucleasa I-Tev (por ejemplo, una nucleasa I-Tevl como se describe en otra parte de la presente memoria), que se fusiona (por ejemplo, al extremo N) de una proteína Cas-S1 o Cas-S4, en particular Cas-S1. El complejo de proteína de fusión puede comprender todas de Cas-S1, Cas-S2, Cas-S4 y Cas-S5, y opcionalmente puede comprender además una proteína Cas-S3.

Por tanto, se proporciona un complejo de proteína de fusión que comprende una proteína de fusión en un complejo con otras proteínas, y en donde la proteína de fusión comprende una proteína de la fórmula A-B-C (en orientación 5' a 3'), en donde:

A es una nucleasa I-Tev (por ejemplo, cualquiera de las proteínas I-Tevl descritas en la presente memoria, por ejemplo, que tiene la secuencia de aminoácidos de la SEQ ID No:45);

B está presente opcionalmente y, cuando está presente, comprende un conector (por ejemplo, cualquiera de los conectores como se describe en la presente memoria, en particular un conector XTEN, por ejemplo, que tiene la secuencia de aminoácidos de la SEQ ID No:17);

C es una proteína CasS seleccionada de Cas-S1 y Cas-S4; y

las demás proteínas en complejo con la proteína de fusión comprenden una proteína Cas-S2 y Cas-S5 y una proteína Cas-S1 o Cas-S4 según se requiera, de modo que el complejo de proteína de fusión comprende al menos una proteína Cas-S1, Cas-S2, Cas-S4 y Cas-S5.

En un ejemplo, el complejo de proteína de fusión puede comprender además una proteína Cas-S3. También se proporciona un primer vector que expresa la proteína de fusión de fórmula A-B-C y un segundo vector que expresa las demás proteínas del complejo de proteína de fusión.

En un ejemplo, el complejo de proteína de fusión puede comprender además un ARNcr como se describe en otra parte de la presente memoria.

En un ejemplo, el complejo de proteína de fusión que comprende una nucleasa I-Tev reconoce y escinde una secuencia diana que comprende, en orientación 5' a 3':

a) una secuencia de nucleótidos del sitio de escisión de I-Tevl (como se describe en otra parte de la presente memoria, pero en particular en donde la secuencia de nucleótidos del sitio de escisión es 5'-CNNNG-3', tal como en donde la secuencia de nucleótidos del sitio de escisión de I-Tevl tiene la secuencia de nucleótidos de la SEQ ID No:42);

b) una secuencia de nucleótidos espaciadora de I-Tevl (como se describe en otra parte de la presente memoria, pero en particular en donde la secuencia de nucleótidos espaciadora de I-Tevl tiene aproximadamente 31 nucleótidos de longitud, por ejemplo, en donde la secuencia de nucleótidos espaciadora de I-Tevl es al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idéntica a la secuencia de nucleótidos de la SEQ ID No:43 (por ejemplo, es 100 % idéntica a la secuencia de nucleótidos de la SEQ ID No:43));

c) una secuencia PAM (como se describe en otra parte de la presente memoria, pero en particular en donde el PAM se selecciona de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3' y 5'-ACA-3', por ejemplo, 5'-AAC-3' y 5'-ATG-3', por ejemplo, en donde el PAM es 5'-AAC-3'; y

d) una secuencia protoespaciadora (como se describe en otra parte de la presente memoria, pero en particular en donde el protoespaciador es una secuencia espaciadora que tiene una longitud de aproximadamente 32 nucleótidos.

y opcionalmente en donde las secuencias a) a d) son cada una inmediatamente adyacentes entre sí.

En un ejemplo, el complejo de proteína de fusión que comprende una nucleasa I-Tev introduce una rotura bicatenaria dentro de la secuencia de nucleótidos del sitio de escisión. Por ejemplo, cuando la secuencia de nucleótidos del sitio de escisión comprende la secuencia de la SEQ ID No:42, la proteína de fusión introduce una rotura después de la segunda C en la cadena directa de la secuencia diana y después de la primera T en la cadena inversa de la secuencia diana (ya que la secuencia de nucleótidos del sitio de escisión en la cadena inversa comprende la secuencia 5'-CGTTG-3'). La nucleasa I-Tevl corta la cadena directa inmediatamente después de la segunda C, e inmediatamente después de la primera T en la cadena inversa. La doble mella del sitio de escisión da como resultado una rotura bicatenaria escalonada del ADN y cada sitio de la rotura alberga salientes de dinucleótidos monocatenarios; 5'-AC-3' para la cadena directa y 5'-GT-3' para la cadena inversa.

En un ejemplo, el o los vectores, complejos, ácidos nucleicos o proteínas no están comprendidos por una célula, en

donde la célula comprende secuencias de nucleótidos endógenas que codifican las proteínas Cas-S1, S2, S3, S4 y S5. En un ejemplo, el o los vectores, complejos, ácidos nucleicos o proteínas no están comprendidos por una célula de una especie seleccionada de las especies en la Tabla 3. Por ejemplo, el o los vectores, complejos, ácidos nucleicos o proteínas no están comprendidos por una célula de *E. coli*. Por ejemplo, el o los vectores, complejos, ácidos nucleicos o proteínas no están comprendidos por una célula de *Ktedonobacter*, por ejemplo, *Ktedonobacter racemifer*. Por ejemplo, el o los vectores, complejos, ácidos nucleicos o proteínas no están comprendidos por una célula de *Allochromatium*, por ejemplo, *Allochromatium warmingii*. Por ejemplo, el o los vectores, complejos, ácidos nucleicos o proteínas no están comprendidos en una célula de *Ignatius*, por ejemplo, *Ignatius tetrasporus*.

En un ejemplo, el o los vectores, complejos, ácidos nucleicos o proteínas no están comprendidos por una célula, en donde la célula comprende secuencias de nucleótidos endógenas que codifican un polipéptido que comprende la SEQ ID NO: 1, un polipéptido que comprende la SEQ ID NO: 2, un polipéptido que comprende la SEQ ID NO: 3, un polipéptido que comprende la SEQ ID NO: 4 y un polipéptido que comprende la SEQ ID NO: 5.

#### Conceptos:

También se proporcionan los siguientes Conceptos numerados relacionados con el nuevo sistema, sus componentes y usos. Cualquier Concepto puede combinarse con cualquier opción o característica divulgada en la presente memoria.

Concepto 1A. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico que comprenden secuencias de nucleótidos expresables, en donde las secuencias comprenden

- a) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 1;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 2;
- c) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 3;
- d) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 4; y
- e) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 5.

Concepto 1B. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico que comprenden secuencias de nucleótidos expresables, en donde las secuencias comprenden

- a) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 94 % (tal como un 95 %) idéntica a la SEQ ID NO: 1;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 98 % (tal como un 99 %) idéntica a la SEQ ID NO: 2;
- c) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 99 % (tal como un 100 %) idéntica a la SEQ ID NO: 3;
- d) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 98 % (tal como un 99 %) idéntica a la SEQ ID NO: 4; y
- e) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 88 % (tal como un 89 %) idéntica a la SEQ ID NO: 5.

Concepto 1C. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico que comprenden secuencias de nucleótidos expresables, en donde las secuencias comprenden

- a) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 98 % (tal como un 99 %) idéntica a la SEQ ID NO: 2;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 98 % (tal como un 99 %) idéntica a la SEQ ID NO: 4; y

- c) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 88 % (tal como un 89 %) idéntica a la SEQ ID NO: 5.

Concepto 1C-1. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico según el Concepto 1C, en donde el o los vectores comprenden además una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 94 % (tal como un 95 %) idéntica a la SEQ ID NO:1.

Concepto 1C-2. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico según el Concepto 1C o 1C-1, en donde el o los vectores comprenden además una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 99 % (tal como un 100 %) a la SEQ ID NO:3.

Concepto 1D. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico que comprenden secuencias de nucleótidos expresables, en donde las secuencias comprenden

- a) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos aproximadamente un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 2;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos aproximadamente un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 4; y
- c) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos aproximadamente un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 5.

Concepto 1D-1. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico según el Concepto 1D, en donde el o los vectores comprenden además una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos aproximadamente un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO:1.

Concepto 1D-2. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico según el Concepto 1D o 1D-1, en donde el o los vectores comprenden además una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos aproximadamente un 80 % (tal como aproximadamente un 90 %) a la SEQ ID NO:3.

En lugar de vector(es) en estos Conceptos, se proporciona alternativamente uno o más ácidos nucleicos que comprenden los componentes citados. Las divulgaciones relacionadas con vector(es) en la presente memoria también deben leerse como aplicables *mutatis mutandis* a ácido(s) nucleico(s). Por ejemplo, el ácido nucleico puede estar comprendido por un cromosoma de una célula, por ejemplo, una célula procariota. Por ejemplo, el ácido nucleico puede estar comprendido por un plásmido. Por ejemplo, la modificación puede ser la modificación de dicho cromosoma o plásmido, en particular el cromosoma. Por lo tanto, el ARNcr puede comprender un espaciador que es capaz de hibridar con un protoespaciador comprendido por el cromosoma o plásmido, en particular el cromosoma. Por lo tanto, el ARNcr puede comprender un espaciador que es cognado a un protoespaciador comprendido por el cromosoma o plásmido, en particular el cromosoma.

Concepto 2. El o los vectores del Concepto 1, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es

- a) heterólogo para al menos una de las secuencias de nucleótidos;
- b) no es un promotor de *E. coli* o *Klebsiella*;
- c) un promotor de célula eucariota;
- d) un promotor animal (opcionalmente, un promotor de mamífero o humano);
- e) un promotor de planta;
- f) un promotor fúngico (opcionalmente, un promotor de levadura);
- g) un promotor de insecto;
- h) un promotor de virus (opcionalmente, un promotor de virus, AAV o lentivirus); o
- i) un promotor sintético.

En cualquier característica de la presente memoria relacionada con promotores, el promotor puede ser un promotor constitutivo. En cualquier opción o característica de la presente memoria relacionada con promotores, el promotor puede ser un promotor inducible. Los promotores adecuados son bien conocidos por los expertos en la técnica. Los

promotores inducibles pueden ser particularmente útiles cuando se desea expresar la(s) proteína(s) y complejos de Tipo-S solo bajo ciertos parámetros definidos. En otras circunstancias en las que se desea producir constantemente la(s) proteína(s) y complejos de Tipo-S descritos en la presente memoria, se pueden emplear promotores constitutivos.

5 En cualquier característica de la presente memoria relacionada con promotores, el promotor puede ser un promotor BolA, por ejemplo, un promotor pBolA que comprende la secuencia de nucleótidos de la SEQ ID No:14. En cualquier característica de C de la presente memoria relacionada con promotores, el promotor puede ser un promotor p70a, por ejemplo, un promotor p70a que comprende la secuencia de nucleótidos de la SEQ ID No:15. En cualquier opción o característica de la presente memoria relacionada con promotores, el promotor puede ser un promotor inducible por arabinosa, por ejemplo, un promotor pBAD que comprende la secuencia de nucleótidos de la SEQ ID No:24.

10 Concepto 3. El o los vectores del Concepto 1 o del Concepto 2, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor constitutivo.

En un ejemplo, todas las secuencias de nucleótidos están conectadas operativamente a un promotor constitutivo. En un ejemplo, todos los nucleótidos están compuestos por un único operón bajo el control de un único promotor constitutivo.

15 Concepto 4. El o los vectores del Concepto 1 o del Concepto 2, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor inducible.

En un ejemplo, todas las secuencias de nucleótidos están conectadas operativamente a un promotor inducible. En un ejemplo, todos los nucleótidos están compuestos por un único operón bajo el control de un único promotor inducible.

20 Concepto 5. El o los vectores de uno cualquiera de los Conceptos 1 a 4, en donde el o los vectores están desprovistos de la secuencia de nucleótidos que codifica el polipéptido de la parte c) del Concepto 1A o 1B.

Concepto 6. El o los vectores de uno cualquiera de los Conceptos 1 a 5, en donde el o los vectores están desprovistos de la secuencia de nucleótidos que codifica un polipéptido de la parte a) del Concepto 1A o 1B.

Concepto 7A: Un vector de ácido nucleico que comprende una secuencia de nucleótidos expresable que comprende

- 25 a) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 90 % idéntica a la SEQ ID NO:2;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 90 % idéntica a la SEQ ID NO:4; y
- c) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 90 % idéntica a la SEQ ID NO:5.

30 Concepto 7B. Un vector de ácido nucleico que comprende una secuencia de nucleótidos expresable seleccionada de

- a) una secuencia de nucleótidos que codifica un polipéptido, que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 1;
- 35 b) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 2;
- c) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 3;
- 40 d) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 4; y
- e) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (tal como aproximadamente un 90 %) idéntica a la SEQ ID NO: 5.
- 45

Concepto 8A. El vector del Concepto 7A, en donde el vector comprende además una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 90 % idéntica a la SEQ ID NO:1.

50 Concepto 8B. El vector del Concepto 7A o 7B, en donde el vector comprende además una secuencia de nucleótidos que codifica un dominio de proteína Cas que es operativo con los polipéptidos de las partes a), b) y c) para unirse a

una secuencia de nucleótidos seleccionada de ADN bicatenario, ADN monocatenario y ARN.

Dichas proteínas Cas pueden seleccionarse de proteínas Cas de complejos CRISPR/Cas de Tipo I, II, III, IV, V o VI de origen natural y que son bien conocidos por los expertos en la técnica. Los métodos proporcionados en la presente memoria pueden utilizarse para identificar proteínas que sean operativas con las proteínas Cas-S2, Cas-S4 y Cas-S5 para proporcionar actividades adicionales.

Concepto 9. El vector del Concepto 7A o del Concepto 8A u 8B, en donde el vector comprende además una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 90 % idéntica a la SEQ ID NO:3.

Concepto 10. El o los vectores de cualquier Concepto anterior, en donde el o los vectores comprenden una o más secuencias de nucleótidos para producir un ARNcr, en donde el ARNcr comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, opcionalmente en donde el protoespaciador

- a) no se encuentra en *E coli*;
- b) es un protoespaciador de célula eucariota; o
- c) es un protoespaciador de células animales (opcionalmente humanas), vegetales, de insectos o de hongos.

Es posible que el protoespaciador no se encuentre en una bacteria que comprenda secuencias de nucleótidos endógenas que codifiquen los polipéptidos de a) a e), como se describe en el Concepto 1.

El ARNcr comprende opcionalmente un espaciador que tiene al menos 10 secuencias de nucleótidos consecutivas que son complementarias al protoespaciador. El ARNcr comprende opcionalmente un espaciador que tiene al menos 15 secuencias de nucleótidos consecutivas que son complementarias al protoespaciador. El ARNcr comprende opcionalmente un espaciador que tiene al menos 20 secuencias de nucleótidos consecutivas que son complementarias al protoespaciador. El ARNcr comprende opcionalmente un espaciador que tiene al menos 25 secuencias de nucleótidos consecutivas que son complementarias al protoespaciador. El ARNcr comprende opcionalmente un espaciador que tiene al menos 28 secuencias de nucleótidos consecutivas que son complementarias al protoespaciador.

En cualquier opción o característica de la presente memoria relacionada con protoespaciadores y espaciadores, el espaciador y el protoespaciador pueden tener una longitud de aproximadamente 10 a 40 nucleótidos. Por ejemplo, el espaciador y el protoespaciador pueden tener entre aproximadamente 25 y 40 nucleótidos, tal como entre aproximadamente 25 y 38 nucleótidos, tal como entre aproximadamente 28 y 35 nucleótidos. El espaciador y el protoespaciador pueden tener una longitud de aproximadamente 25 a 39 nucleótidos. El espaciador y el protoespaciador pueden tener una longitud de aproximadamente 29 a 35 (por ejemplo, de aproximadamente 30 a 34, o de aproximadamente 31 a 33) nucleótidos. Por ejemplo, el espaciador y el protoespaciador pueden tener una longitud de aproximadamente 32 nucleótidos. Por ejemplo, el espaciador y el protoespaciador pueden tener una longitud de aproximadamente 28 nucleótidos. Por ejemplo, el espaciador y el protoespaciador pueden tener una longitud de 32 nucleótidos. Por ejemplo, el espaciador y el protoespaciador pueden tener una longitud de 28 nucleótidos. El espaciador y/o el protoespaciador pueden tener cualquiera de las características descritas en cualquier opción o característica de la presente memoria.

En cualquier opción o característica de la presente memoria relacionada con protoespaciadores y espaciadores, el espaciador puede ser (aproximadamente) un 90 % complementario al protoespaciador. El espaciador puede ser (aproximadamente) un 70 % complementario al protoespaciador. El espaciador puede ser (aproximadamente) un 80 % complementario al protoespaciador. El espaciador puede ser (aproximadamente) un 95 % complementario al protoespaciador. El espaciador puede ser un 100 % complementario al protoespaciador en toda su longitud.

En un ejemplo, cuando el protoespaciador tiene una longitud de aproximadamente 32 (por ejemplo, 32) nucleótidos (o más), los nucleótidos 1 a 28 del espaciador, que están en 5' de la secuencia PAM, son complementarios a la secuencia del protoespaciador.

En cualquier opción o característica de la presente memoria relacionada con ARNcr, el ARNcr comprende al menos una secuencia repetida. Por ejemplo, el ARNcr comprende dos secuencias repetidas. La secuencia repetida puede ser cualquier secuencia que sea capaz de formar un bucle de horquilla y ser reconocida por el sistema Tipo-S. La secuencia repetida puede tener una longitud de aproximadamente 20 a 25 nucleótidos (por ejemplo, 22 o 23 nucleótidos). La secuencia repetida puede ser una secuencia de nucleótidos que sea al menos aproximadamente un 80 %, por ejemplo, aproximadamente un 90 % (tal como al menos aproximadamente un 95, 96, 97 o 98 %) idéntica a la secuencia de nucleótidos de la SEQ. ID No:6. La secuencia repetida puede ser una secuencia de nucleótidos de la SEQ ID No:6. La secuencia repetida puede ser una secuencia de nucleótidos que sea al menos aproximadamente un 80 %, por ejemplo, aproximadamente un 90 % (tal como al menos aproximadamente un 95, 96, 97 o 98 %) idéntica a la secuencia de nucleótidos de la SEQ ID No:70. La secuencia repetida puede ser una secuencia de nucleótidos de la SEQ ID No:70.

En cualquier ácido nucleico o vector descrito en la presente memoria, el ARNcr puede estar codificado por una matriz CRISPR (por ejemplo, una matriz CRISPR como se describe en otra parte de la presente memoria).

Según cualquier opción o característica de la presente memoria, una planta puede ser monocotiledónea o dicotiledónea. Opcionalmente, la planta se selecciona del grupo que consiste en maíz, soja, algodón, trigo, canola, colza, sorgo, arroz, centeno, cebada, mijo, avena, caña de azúcar, césped, pasto varilla, alfalfa, girasol, tabaco, maní, patata, *Arabidopsis*, cártamo y tomate.

Como sabrá el experto en la técnica, "cognado a" o "cognado con" o "complementario a" se refiere a componentes que pueden funcionar juntos. Por ejemplo, se sabe que un ARNcr funciona con un PAM para guiar a Cas hacia un protoespaciador en un ácido nucleico diana. En este sentido, el ARNcr es cognado al PAM y es cognado al protoespaciador.

Concepto 11. El o los vectores de cualquier Concepto anterior, en donde el o los vectores comprenden

- a) una secuencia de localización nuclear (NLS);
- b) una secuencia de empaquetamiento de fagos (opcionalmente, un sitio *pac* o *cos*);
- c) un origen de replicación de plásmido;
- d) un origen de transferencia de plásmido;
- e) un núcleo de plásmido bacteriano;
- f) un núcleo de plásmido eucariota;
- g) un gen de proteína estructural de un virus humano (opcionalmente, AAV o lentivirus);
- h) una secuencia *rep* y/o *cap* de un virus (opcionalmente, AAV o lentivirus);
- i) un marcador de selección o secuencia que codifica un marcador seleccionable;
- j) un promotor eucariota; o
- k) una secuencia de nucleótidos de un gen humano, animal, vegetal o fúngico.

Cuando la secuencia diana está en una célula eucariota, en cualquier opción o característica, el o los vectores, proteínas o complejos de la presente memoria comprenden además una NLS. Los expertos en la técnica conocen las NLS e incluyen, pero no se limitan a, las NLS monopartitas o bipartitas. La secuencia de NLS puede ser de un antígeno T grande de SV40 (véase Kalderon *et al.*, Cell, 39(3), 499-509, 1984, doi:[https://doi.org/10.1016/0092-8674\(84\)90457-4](https://doi.org/10.1016/0092-8674(84)90457-4)), de nucleoplasmina, importina  $\alpha$ , c-myc, EGL-13, proteína TUS, hnRNP A1 o represor de transcripción de levadura Mata2.

Cualquiera de los vectores descritos en la presente memoria puede administrarse a células bacterianas a través de una partícula de fago. Cualquiera de los vectores descritos en la presente memoria puede administrarse a células bacterianas a través de una partícula de fagémido empaquetada dentro de una partícula de fago. Cualquiera de los vectores descritos en la presente memoria puede administrarse a células bacterianas a través de un plásmido conjugativo.

Concepto 12. El o los vectores de cualquier Concepto anterior, en donde cada vector es

- a) un vector plasmídico (opcionalmente, un plásmido conjugativo)
- b) un vector de transposón (opcionalmente, un transposón conjugativo);
- c) un vector de virus (opcionalmente, un vector de fago, AAV o lentivirus);
- d) un fagémido (opcionalmente, un fagémido empaquetado); o
- e) una nanopartícula (opcionalmente, una nanopartícula lipídica).

Concepto 13A. Uno o más polipéptidos como los enunciados en cualquier Concepto precedente.

Concepto 13B. Uno o más polipéptidos expresados a partir de los vectores enunciados en cualquier Concepto anterior.

Concepto 13C. Uno o más polipéptidos como se enuncia en cualquier reivindicación anterior expresados a partir de los vectores de cualquier Concepto anterior.

Concepto 14A. Una proteína de fusión que comprende un polipéptido (Px), en donde Px

- a) comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % idéntica (por ejemplo, (aproximadamente) un 90 % idéntica) a una secuencia seleccionada de las SEQ ID No: 1-5; y
- b) se fusiona a un polipéptido heterólogo (Py).

El 14<sup>º</sup> Concepto también proporciona:

- 5 Una proteína de fusión que comprende una Cas-S1 fusionada a un polipéptido heterólogo (Py).  
Una proteína de fusión que comprende una Cas-S2 fusionada a un polipéptido heterólogo (Py).  
Una proteína de fusión que comprende una Cas-S3 fusionada a un polipéptido heterólogo (Py).  
Una proteína de fusión que comprende una Cas-S4 fusionada a un polipéptido heterólogo (Py).  
Una proteína de fusión que comprende una Cas-S5 fusionada a un polipéptido heterólogo (Py).
- 10 Concepto 14B. Un complejo de proteína de fusión que comprende:
  - (i) un polipéptido de proteína de fusión (Px), en donde Px
    - a) comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % idéntica (por ejemplo, (aproximadamente) un 90 % idéntica) a una secuencia seleccionada de las SEQ ID No: 1, 3 y 4; y
    - 15 b) se fusiona a un polipéptido heterólogo (Py);
  - (ii) un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % idéntica (por ejemplo, (aproximadamente) un 90 % idéntica) a la SEQ ID No:2;
  - (iii) un polipéptido que comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % idéntica (por ejemplo, (aproximadamente) un 90 % idéntica) a la SEQ ID No: 5;
  - 20 (iv) un ARNcr que comprende una secuencia espaciadora que es cognada a un primer protoespaciador en una secuencia diana; y
  - (v) opcionalmente, un polipéptido que es al menos (aproximadamente) un 80 % idéntico (por ejemplo, (aproximadamente) un 90 % idéntico) a una secuencia seleccionada de las SEQ ID No: 1, 3 y 4 y no se basa en la secuencia de aminoácidos del polipéptido citado en la parte (i)a); y
  - 25 (vi) opcionalmente, un polipéptido que es al menos (aproximadamente) un 80 % idéntico (por ejemplo, (aproximadamente) un 90 % idéntico) a una secuencia seleccionada de las SEQ ID No: 1, 3 y 4 y no se basa en la secuencia de aminoácidos del polipéptido citado en la parte (i)a), y, si está presente, no se basa en la secuencia de aminoácidos del polipéptido citado en la parte (v).

Concepto 14C. Un complejo de proteína de fusión que comprende:

- 30 (i) un polipéptido de proteína de fusión (Px), en donde Px
  - a) comprende una proteína Cas-S1, Cas-S3 o Cas-S4; y
  - b) se fusiona a un polipéptido heterólogo (Py);
- (ii) una proteína Cas-S2;
- (iii) una proteína Cas-S5;
- 35 (iv) un ARNcr que comprende una secuencia espaciadora que es cognada a un primer protoespaciador en una secuencia diana; y
- (v) opcionalmente, una proteína que se selecciona de una proteína Cas-S1, Cas-S3 y Cas-S4, y es de una clase de proteína CasS diferente a la proteína citada en la parte (i)a); y
- 40 (vi) opcionalmente, una proteína que se selecciona de una proteína Cas-S1, Cas-S3 y Cas-S4, y es de una clase de proteína CasS diferente a la proteína citada en la parte (i)a), y, si está presente, es de una clase de proteína CasS diferente a la proteína citada en la parte (v).

En un ejemplo, la proteína de fusión comprende 2 o más de Cas-S1 a S5, por ejemplo, comprende Cas-S1 y S2, o S4 y S5.

En cualquier opción o característica relacionada con las proteínas de fusión, Py puede fusionarse (directa o

indirectamente) al extremo N de la proteína CasS. En cualquier opción o característica relacionada con las proteínas de fusión, Py puede fusionarse (directa o indirectamente) al extremo C de la proteína CasS.

La fusión puede ser directa o indirecta (por ejemplo, a través de un conector peptídico, tal como un conector  $(G_4S)_n$ , en donde  $n = 1, 2, 3, 4, 5, 6, 7, 8, 9$  o  $10$ ). Un conector peptídico puede tener una longitud de aproximadamente 10 a 20 aminoácidos (por ejemplo, aproximadamente 16 aminoácidos). Un conector peptídico puede tener una longitud de aproximadamente 8 a 12 aminoácidos (por ejemplo, aproximadamente 10 aminoácidos). Un conector peptídico puede tener la secuencia de aminoácidos de la SEQ ID No:17. Un conector peptídico puede tener la secuencia de aminoácidos de la SEQ ID No:23. Los conectores pueden ser como se describe en otra parte de la presente memoria (por ejemplo, en los Conceptos 27 o 29 de la presente memoria).

Aquí, "heterólogo" se refiere a un polipéptido que no se encuentra naturalmente fusionado a Px. En un ejemplo, Py comprende una nucleasa I-Tev o proteína mutH.

En cualquier opción o característica, Py es una nucleasa I-Tev, tal como una nucleasa I-TevI, por ejemplo, como se describe en otra parte de la presente memoria. En un ejemplo, la nucleasa I-TevI es una proteína que comprende la secuencia de aminoácidos de la SEQ ID No:45. En un ejemplo, la nucleasa I-TevI está codificada por una secuencia de ácido nucleico que codifica una secuencia de aminoácidos que comprende los aminoácidos de la SEQ ID:45. En un ejemplo, la nucleasa I-TevI está codificada por una secuencia de ácido nucleico de la SEQ ID:46. En un ejemplo, la nucleasa I-TevI es una proteína que tiene una secuencia de aminoácidos que es al menos aproximadamente un 80 % idéntica (por ejemplo, es aproximadamente un 85 %, 90 %, 95 %, 96 %, 97 % o 98 % idéntica) al aminoácido de la SEQ ID No:45 y es capaz de escindir una secuencia de nucleótidos del sitio de escisión de I-TevI que tiene la secuencia de nucleótidos de la SEQ ID No:42 y es capaz de reconocer una secuencia de nucleótidos del espaciador de I-TevI que tiene la secuencia de nucleótidos de la SEQ ID No:43.

Generalmente, cuando un componente es heterólogo respecto a otro componente, los componentes no ocurren naturalmente en la naturaleza, o en la misma célula.

Por ejemplo, la secuencia seleccionada es la SEQ ID NO: 1. Por ejemplo, la secuencia seleccionada es la SEQ ID NO: 2. Por ejemplo, la secuencia seleccionada es la SEQ ID NO: 3. Por ejemplo, la secuencia seleccionada es la SEQ ID NO: 4. Por ejemplo, la secuencia seleccionada es la SEQ ID NO: 5.

Py puede comprender un dominio efector. Por ejemplo, el dominio efector es un dominio que comprende actividad de nucleasa, actividad de nickasa, actividad de recombinasa, transcriptasa inversa, helicasa, actividad de desaminasa, actividad de metiltransferasa, actividad de metilasa, actividad de acetilasa, actividad de acetiltransferasa, actividad de activación transcripcional o actividad de represión transcripcional.

Py puede ser de origen natural. Py puede ser una secuencia sintética. Py puede ser un derivado mutado sintéticamente de una secuencia de origen natural.

Opcionalmente, el dominio efector es un dominio de edición de ácidos nucleicos. Por ejemplo, el dominio de edición de ácidos nucleicos comprende un dominio de desaminasa. El dominio de desaminasa puede ser un dominio de citidina desaminasa. El dominio de citidina desaminasa puede ser una desaminasa de la familia del complejo de edición de ARNm de apolipoproteína B (APOBEC). El dominio de citidina desaminasa puede ser al menos (aproximadamente) un 80 %, al menos (aproximadamente) un 85 %, al menos (aproximadamente) un 90 %, al menos (aproximadamente) un 92 %, al menos (aproximadamente) un 95 %, al menos (aproximadamente) un 96 %, al menos (aproximadamente) un 97 %, al menos (aproximadamente) un 98 %, al menos (aproximadamente) un 99 %, o al menos (aproximadamente) un 99,5 % idéntico al dominio de citidina desaminasa de una cualquiera de las SEQ ID NO:350-389 divulgadas en la publicación estadounidense número US 2022/0307001 A1 o WO2019/168953, secuencias que forman parte de la divulgación.

Opcionalmente, Py comprende un dominio inhibidor de la uracil glicosilasa (UGI). El dominio UGI puede comprender la secuencia de aminoácidos de la SEQ ID NO:500 divulgada en la publicación estadounidense número US 2022/0307001 A1 o WO2019/168953, secuencia que forma parte de la divulgación. El dominio UGI puede tener la secuencia de aminoácidos de la SEQ ID No:21. El dominio UGI puede tener la secuencia de nucleótidos de la SEQ ID No:20.

Opcionalmente, Py comprende la secuencia de aminoácidos de la SEQ ID NO:534. Opcionalmente, Py puede comprender la secuencia de aminoácidos de la SEQ ID NO:538. Estas secuencias se divulgan en la publicación estadounidense número US 2022/0307001 A1 o WO2019/168953, secuencias que forman parte de la divulgación.

Opcionalmente, Py comprende la proteína de fusión que comprende además un segundo dominio UGI. Opcionalmente, Py comprende la secuencia de aminoácidos de la SEQ ID NO:540. Opcionalmente, Py comprende la secuencia de aminoácidos de la SEQ ID NO:541. Estas secuencias se divulgan en la publicación estadounidense número US 2022/0307001 A1 o WO2019/168953, secuencias que forman parte de la divulgación.

Opcionalmente, el dominio de desaminasa es un dominio de adenosina desaminasa. La proteína de fusión puede comprender un segundo dominio de adenosina desaminasa. Por ejemplo, el primer dominio de adenosina desaminasa

y el segundo dominio de adenosina desaminasa comprenden un dominio ecTadA, o una variante del mismo. El primer dominio de adenosina desaminasa y el segundo dominio de adenosina desaminasa pueden comprender la secuencia de aminoácidos de una cualquiera de las SEQ ID NO:400-458. La primera adenosina desaminasa puede comprender la secuencia de aminoácidos de la SEQ ID NO:400. La segunda adenosina desaminasa puede comprender la secuencia de aminoácidos de la SEQ ID NO:458. Opcionalmente, Py comprende la secuencia de aminoácidos de la SEQ ID NO:535 o 539. Estas secuencias se divulgan en la publicación estadounidense número US 2022/0307001 A1 o WO2019/168953, secuencias que forman parte de la divulgación.

Los editores de bases que se pueden utilizar con el sistema Cas-S son bien conocidos en la técnica. Los editores de bases incluyen sistemas en donde una proteína CasS se fusiona con un dominio de citosina o adenosina desaminasa y se dirige a la secuencia diana para realizar la modificación deseada.

En un ejemplo, el editor de bases se selecciona de:

- 1) Editores de bases de citosina (CBE) que convierten C:G en T:A (Komar *et al.*, Nature, 533:420-4, 2016)
- 2) Editores de bases de adenina (ABE) que convierten A:T en G:C (Gaudelli *et al.*, Nature, 551 (7681), 464-471, 2017)
- 3) Editores de bases de guanina citosina (CGBE) que convierten C:G en G:C (Chen *et al.*, Biorxiv, 2020; Kurt *et al.*, Nature Biotechnology, 2020)
- 4) Editores de bases de adenina citosina (CABE) que convierten C:G en A:T (Zhao *et al.*, Nature Biotechnology, 2020)
- 5) Editores de bases de citosina adenina (ACBE) que convierten A:T en C:G (WO2020/181180)
- 6) Editores de bases de timina adenina (ATBE) que convierten A:T en T:A (WO2020/181202)
- 7) Editor de bases de timina adenina (TABE) que convierte T:A en A:T (WO2020/181193 (o US2022/0170013); WO2020/181178; WO2020/181195).

Los editores de bases difieren en las enzimas de modificación de bases. Los CBE se basan en la citidina desaminasa de ADNss, entre las que se encuentran: APOBEC1, rAPOBEC1, APOBEC1 mutante o versión evolucionada (evoAPOBEC1) y homólogos de APOBEC (APOBEC3A (eA3A), Anc689), citidina desaminasa 1 (CDA 1), evoCDA 1, FERNY, evoFERNY.

Los ABE se basan en la actividad desoxiadenosina desaminasa de una fusión en tándem TadA-TadA\*, donde TadA\* es una versión evolucionada de TadA, una enzima adenosina desaminasa de ARNT de *E. coli*, capaz de convertir la adenosina en inosina en ADNss. TadA\* incluye TadA-8a-e y TadA-7.10.

Los ejemplos de proteínas editoras basadas en ADN incluyen, pero no se limitan a BE1, BE2, BE3, BE4, BE4-GAM, HF-BE3, Sniper-BE3, Target-AID, Target-AID-NG, ABE, EE-BE3, YE1-BE3, YE2-BE3, YEE-BE3, BE-PLUS, SaBE3, SaBE4, SaBE4-GAM, Sa(KKH)-BE3, VQR-BE3, VRER-BE3, EQR-BE3, xBE3, Cas12a-BE, Ea3A-BE3, A3A-BE3, TAM, CRISPR-X, ABE7.9, ABE7.10, ABE7.10\*, xABE, ABESa, VQR-ABE, VRER-ABE, Sa(KKH)-ABE, ABE8e, SpRY-ABE, SpRYCBE, SpG-CBE4, SpG-ABE, SpRY-CBE4, SpCas9-NG-ABE, SpCas9-NG-CBE4, enAsBE1.1, enAsBE1.2, enAsBE1.3, enAsBE1.4, AsBE1.1, AsBE1.4, CRISPR-Abest, CRISPR-Cbest, eA3A-BE3 y AncBE4.

En un ejemplo, Py es una citidina desaminasa. La citidina desaminasa puede ser una citidina desaminasa PmCDA1. La citidina desaminasa puede ser de lamprea marina. En un ejemplo, la citidina desaminasa PmCDA1 comprende la secuencia de aminoácidos de la SEQ ID No:19. En un ejemplo, la citidina desaminasa PmCDA1 está codificada por una secuencia de ácido nucleico que codifica una secuencia de aminoácidos que comprende los aminoácidos de la SEQ ID No:19. En un ejemplo, la citidina desaminasa PmCDA1 está codificada por una secuencia de ácido nucleico de la SEQ ID No:18. En un ejemplo, la citidina desaminasa PmCDA1 es una proteína que tiene una secuencia de aminoácidos que es al menos aproximadamente un 80 % idéntica (por ejemplo, es aproximadamente un 85 %, 90 %, 95 %, 96 %, 97 % o 98 % idéntica) al aminoácido de la SEQ ID No:19 y es capaz de realizar una mutación de C a T en el ácido nucleico diana.

Concepto 15. La proteína de fusión del Concepto 14, en donde la proteína de fusión está en forma de un complejo de proteína de fusión que comprende al menos dos polipéptidos adicionales,

en donde los dos polipéptidos adicionales tienen cada uno una secuencia de aminoácidos que es al menos un 90 % idéntica a una secuencia seleccionada de las secuencias de la SEQ ID No:2, SEQ ID No:4 y SEQ ID No:5,

y en donde el complejo de proteína de fusión comprende secuencias de aminoácidos que son al menos un 90 % idénticas a cada una de las secuencias de la SEQ ID No:2, SEQ ID No:4 y SEQ ID No:5.

Px puede comprender una secuencia de aminoácidos que es al menos un 90 % idéntica a la secuencia de la SEQ ID No:2, y los polipéptidos adicionales comprenden una secuencia de aminoácidos que es al menos un 90 % idéntica a

5

10

15

25

30

40

45

en donde la proteína de fusión está en forma de un complejo de proteína de fusión que comprende al menos dos

polipéptidos adicionales,

en donde los dos polipéptidos adicionales tienen cada uno una secuencia de aminoácidos que es al menos un 90 % idéntica a una secuencia seleccionada de las secuencias de la SEQ ID No:2 y SEQ ID No:5, y

- 5 (i) opcionalmente, un polipéptido que es al menos un 90 % idéntico a una secuencia seleccionada de las SEQ ID No: 1, 3 y 4 y no está basado en la secuencia de aminoácidos del polipéptido citado en la parte I; y
- (ii) opcionalmente, un polipéptido que es al menos un 90 % idéntico a una secuencia seleccionada de las SEQ ID No: 1, 3 y 4 y no está basado en la secuencia de aminoácidos del polipéptido citado en la parte I, y, si está presente, no está basado en la secuencia de aminoácidos del polipéptido citado en la parte (i).

10 Concepto 19. La proteína de fusión o complejo de proteína de fusión de uno cualquiera de los Conceptos 14 a 18, en donde Py está fusionado al extremo C de la secuencia de aminoácidos de la parte I.

Alternativamente, Py se fusiona al extremo N de la secuencia de aminoácidos de la parte I.

15 Concepto 20. La proteína de fusión o complejo de proteína de fusión de uno cualquiera de los Conceptos 14 a 19, en donde Py es una enzima, seleccionada opcionalmente de una desaminasa (tal como una citosina o adenina desaminasa), un editor de bases, un *prime editor*, helicasa, una transcriptasa inversa, una metiltransferasa, metilasa, acetilasa, acetiltransferasa, un activador o desactivador de la transcripción, un activador o desactivador de la traducción o una nucleasa (tal como una nucleasa de ADN, una nucleasa de ARN, una nickasa o una nucleasa muerta, por ejemplo, dCas).

20 En cualquier opción o característica, Py es una nucleasa, como cualquiera de las nucleasas divulgadas en la presente memoria, en particular una nucleasa I-Tevl. Py puede ser una nucleasa de ADN. Py puede ser una nucleasa de ARN. Py puede ser una nickasa o una nucleasa muerta. En un ejemplo, Py es un editor de bases, tal como cualquiera de los editores de bases divulgados en la presente memoria, en particular una citidina desaminasa PmCDA1. En un ejemplo, Py es un *prime editor*.

Concepto 21. La proteína de fusión o complejo de proteína de fusión del Concepto 20, donde Py es una nucleasa que es una nucleasa I-Tevl.

25 En cualquier opción o característica, en la Parte I, la nucleasa I-Tevl se fusiona al extremo N de una secuencia de aminoácidos (por ejemplo, cualquier proteína CasS descrita en la presente memoria). En un ejemplo, en la Parte I, la nucleasa I-Tevl se fusiona al extremo N de una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idéntica a la secuencia de la SEQ ID No:1. La nucleasa I-Tevl puede ser como se describe en otra parte de la presente memoria (por ejemplo, como en el Concepto 14).

30 Concepto 22. La proteína de fusión o complejo de proteína de fusión del Concepto 21, en donde la nucleasa I-Tevl no comprende un dominio de unión al ADN completo.

I-Tevl comprende 245 aminoácidos y consiste en un dominio catalítico N-terminal y un dominio de unión al ADN C-terminal que están conectados por un conector largo y flexible. La estructura cristalina del dominio de unión al ADN de I-Tevl (que comprende los residuos 130 a 245) en complejo con la región de unión primaria de 20 pb de su ADN diana, revela la presencia de un dedo de zinc (que comprende los residuos 151 a 167) que hace contactos del núcleo con el ADN desde el surco menor, un segmento alargado que contiene una hélice  $\alpha$  de unión al surco menor (que comprende los residuos 183 a 194) y una hélice-giro-hélice (que comprende los residuos 204 a 245). Se mostró que el dominio catalítico N-terminal comprende los residuos 1-92. Los datos bioquímicos han mostrado que el dedo de zinc no contribuye a la afinidad de unión al ADN ni a la especificidad de la enzima, sino que tiene una función novedosa y actúa como un determinante de distancia que controla las posiciones relativas de los dominios catalíticos y de unión al ADN, véase, por ejemplo, Roey *et al.*, 2005, DOI:10.1007/0-387-27421-9\_7.

40 Kleinstiver *et al.*, G3 Genes/Genomes/Genetics, 4(6), 1155-1165, 2014, doi:https://doi.org/10.1534/g3.114.011445 mostraron que varios truncamientos del dominio catalítico N-terminal conservan la actividad cuando se fusionan en la construcción TALEN. La construcción más grande fue la de los residuos 1 a 206 de I-Tevl. La más pequeña que fue funcional cuando se fusionó a una TALEN fue la de los residuos 1 a 162.

45 Por lo tanto, en un ejemplo, la I-Tevl no comprende un dominio hélice-giro-hélice completo. La I-Tevl puede comprender la secuencia de aminoácidos de la SEQ ID No:45. La I-Tevl puede no comprender un dominio hélice-giro-hélice. La I-Tevl puede comprender los aminoácidos 1 a 195-203 de la SEQ ID No:45. La I-Tevl puede no comprender la hélice  $\alpha$  de unión al surco menor. La I-Tevl puede comprender los aminoácidos 1 a 167 de la SEQ ID No:45. La I-Tevl puede no comprender un dominio de dedo de zinc completo. La I-Tevl puede comprender los aminoácidos 1 a 162 de la SEQ ID No:45.

Concepto 23. La proteína de fusión o complejo de proteína de fusión del Concepto 21 o 22, en donde la nucleasa I-Tevl comprende un dominio catalítico N-terminal.

Por lo tanto, la I-Tevl puede comprender los aminoácidos 1 a 92 de la SEQ ID No:45. La I-Tevl puede comprender la

secuencia de aminoácidos de la SEQ ID No:45.

Concepto 24. La proteína de fusión o complejo de proteína de fusión de uno cualquiera de los Conceptos 21 a 23, en donde la nucleasa I-Tevl es una nucleasa I-Tevl de bacteriófago.

5 En cualquier opción o característica, la nucleasa I-Tevl puede ser una I-Tevl de bacteriófago T4. La nucleasa I-Tevl puede ser de un bacteriófago T4. La nucleasa I-Tevl puede estar codificada por el intrón del bacteriófago T4. La nucleasa I-Tevl puede comprender los aminoácidos 1 a 206 de una nucleasa I-Tevl de origen natural. La nucleasa I-Tevl puede comprender una secuencia de aminoácidos de la SEQ ID No:45. La nucleasa I-Tevl puede estar codificada por una secuencia de nucleótidos de la SEQ ID No:44.

10 Concepto 25. La proteína de fusión o complejo de proteína de fusión según la reivindicación 20, en donde Py comprende un editor de bases seleccionado de un editor de bases de adenina (ABE, por ejemplo, una adenosina desaminasa), un editor de bases de citosina (CBE, por ejemplo, una citidina desaminasa o una desaminasa de la familia del complejo de edición de ARNm de apolipoproteína B (APOBEC)), un editor de bases de citosina guanina (CGBE), un editor de bases de citosina adenina (CABE), un editor de bases de adenina citosina (ACBE), un editor de bases de adenina timina (ATBE), un editor de bases de timina adenina (TABE) y una proteína inhibidora de la uracil ADN glicosilasa (UGI).

15 En cualquier opción o característica de C relacionada con proteínas de fusión o complejos de proteínas de fusión, Py es un CBE. En un ejemplo, Py es una citidina desaminasa PmCDA1 (como se describe en otra parte de la presente memoria, por ejemplo, en el Concepto 14). En un ejemplo, Py es una citidina desaminasa PmCDA1 en combinación con una proteína UGI. En un ejemplo, Py es un CBE unido a través de un conector (como se describe en otra parte de la presente memoria) a una proteína UGI. En un ejemplo, Py es una citidina desaminasa PmCDA1 unida a través de un conector (como se describe en otra parte de la presente memoria) a una proteína UGI. En un ejemplo, Py es un CBE unido a través de un conector (como se describe en otra parte de la presente memoria) a una proteína UGI, y el CBE está unido a una proteína CasS descrita en la presente memoria a través de un conector (como se describe en otra parte de la presente memoria). En un ejemplo, Py es una citidina desaminasa PmCDA1 unida a través de un conector (como se describe en otra parte de la presente memoria) a una proteína UGI, y la PmCDA1 está unida a una proteína CasS descrita en la presente memoria a través de un conector (como se describe en otra parte de la presente memoria). En un ejemplo, Py es un CBE unido a través de un conector (como se describe en otra parte de la presente memoria) a una proteína UGI, y la proteína UGI está unida a una proteína CasS descrita en la presente memoria a través de un conector (como se describe en otra parte de la presente memoria). En un ejemplo, Py es una citidina desaminasa PmCDA1 unida a través de un conector (como se describe en otra parte de la presente memoria) a una proteína UGI, y la proteína UGI está unida a una proteína CasS descrita en la presente memoria a través de un conector (como se describe en otra parte de la presente memoria). En un ejemplo, la proteína CasS se selecciona de una proteína CasS1, una proteína CasS3 y una proteína CasS4.

30 Concepto 26. La proteína de fusión o complejo de proteína de fusión del Concepto 25, en donde el editor de bases es un editor de bases de lamprea marina.

En cualquier opción o característica, el editor de bases es una citidina desaminasa PmCDA1 que tiene la secuencia de aminoácidos codificada por la SEQ ID No:19. En un ejemplo, el editor de bases es de lamprea marina.

Concepto 27. La proteína de fusión o complejo de proteína de fusión del Concepto 25 o Concepto 26, en donde el editor de bases está fusionado al extremo C de la secuencia de aminoácidos de la parte I.

40 En cualquier opción o característica, el editor de bases (por ejemplo, citidina desaminasa PmCDA1) se fusiona a través de un conector peptídico a la secuencia de aminoácidos de la Parte I. El conector peptídico puede ser como se describe en el Concepto 14 o el Concepto 29. En cualquier opción o característica, el editor de bases (por ejemplo, citidina desaminasa PmCDA1) se fusiona a través de un conector peptídico a una proteína CasS. El conector puede tener una longitud de (aproximadamente) 10 a 20 aminoácidos. El conector puede tener una longitud de (aproximadamente) 10 a 22, de (aproximadamente) 11 a 21, de (aproximadamente) 12 a 20, de (aproximadamente) 13 a 19, de (aproximadamente) 14 a 18, de (aproximadamente) 15 a 17 aminoácidos. El conector puede tener una longitud de (aproximadamente) 16 aminoácidos. El conector puede ser un conector que tenga la secuencia de aminoácidos de la SEQ ID No:17. El conector puede ser cualquier conector descrito en la presente memoria.

50 Concepto 28. La proteína de fusión o complejo de proteína de fusión del Concepto 27, en donde una proteína UGI se fusiona al editor de bases.

En cualquier opción o característica, la proteína UGI tiene una secuencia de aminoácidos de la SEQ ID No:21. La proteína UGI puede ser como se describe en otra parte de la presente memoria (por ejemplo, en el Concepto 14). La proteína UGI puede fusionarse directamente al editor de bases. La proteína UGI puede fusionarse al editor de bases a través de un conector, por ejemplo, un conector peptídico como se describe en otra parte de la presente memoria.

55 Concepto 29. La proteína de fusión o complejo de proteína de fusión del Concepto 28, donde la proteína UGI se fusiona al editor de bases a través de un conector peptídico.

En cualquier opción o característica, la proteína UGI se fusiona al editor de bases a través de un conector peptídico con una longitud de (aproximadamente) 8 a 12 aminoácidos. El conector peptídico puede tener una longitud de (aproximadamente) 9 a 11 aminoácidos. El conector puede tener una longitud de (aproximadamente) 10 aminoácidos. El conector puede tener la secuencia de aminoácidos de la SEQ ID No:23. El conector puede fusionarse al extremo C del editor de bases, y la proteína UGI se fusiona al extremo C del conector. El conector puede fusionarse al extremo N del editor de bases, y la proteína UGI se fusiona al extremo N del conector. El conector puede ser como se describe en otra parte de la presente memoria (por ejemplo, en el Concepto 14 o 27).

Concepto 30. La proteína de fusión o complejo de proteína de fusión de uno cualquiera de los Conceptos 14 a 29, en donde la proteína de fusión o complejo de proteína de fusión está en combinación con un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana.

La secuencia de ARNcr, espaciador y/o protoespaciador puede tener cualquiera de las características descritas en otra parte de la presente memoria (por ejemplo, en el Concepto 10 o en el Primer Aspecto de la Descripción Detallada). En cualquier opción o característica, el protoespaciador no se encuentra en una bacteria que comprende secuencias de nucleótidos endógenas que codifican los polipéptidos de a) a e) como se recita en el Concepto 1A o 1B. En cualquier opción o característica, el protoespaciador es un protoespaciador de célula eucariota. En cualquier opción o característica, el protoespaciador es heterólogo a la fuente de Py. Por ejemplo, cuando Py es un polipéptido de origen natural, el protoespaciador es de una especie (o género) diferente al Py. En cualquier opción o característica, el protoespaciador es un protoespaciador de célula animal (opcionalmente humana), vegetal, de insecto o de hongo.

En un ejemplo, el complejo de proteína de fusión forma un complejo de ribonucleoproteína con un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana. En un ejemplo, la proteína de fusión forma un complejo de ribonucleoproteína con un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana.

Concepto 31. Uno o más vectores de ácido nucleico que comprenden una o más secuencias de nucleótidos que codifican una proteína de fusión o un complejo de proteína de fusión como se recita en uno cualquiera de los Conceptos 14 a 30.

El o los vectores pueden codificar además un ARNcr como se describe en otra parte de la presente memoria (por ejemplo, en el Primer Aspecto de la Descripción Detallada, en el Concepto 10 o en el Concepto 30). En un ejemplo, se proporciona un primer y un segundo vector que codifican un complejo de proteína de fusión como se describe en otra parte de la presente memoria. En un ejemplo, el primer vector codifica la proteína de fusión Px y el segundo vector codifica las proteínas adicionales del complejo de proteína de fusión.

El o los vectores pueden codificar un segundo, tercero, cuarto, etc., ARNcr que comprende un espaciador que es cognado a un segundo, tercero, cuarto, etc., protoespaciador en una secuencia diana. La expresión de múltiples ARNcr puede ser útil para la edición multiplex, por ejemplo, para dirigirse a múltiples genes en una célula u organismo.

Concepto 32. El o los vectores del Concepto 31, en donde la proteína de fusión está codificada por una única primera secuencia de nucleótidos y en donde los al menos dos polipéptidos adicionales están codificados por una segunda secuencia de nucleótidos.

En un ejemplo, las proteínas de fusión descritas en la presente memoria están codificadas por una única primera secuencia de nucleótidos comprendida por un primer vector, y los al menos dos polipéptidos CasS adicionales del complejo están codificados por una segunda secuencia de nucleótidos comprendida por un segundo vector. En un ejemplo, todos los polipéptidos CasS adicionales de los complejos de proteínas de fusión descritos en la presente memoria están codificados por una segunda secuencia de nucleótidos. En un ejemplo, todos los polipéptidos CasS adicionales de los complejos descritos en la presente memoria están codificados por una segunda secuencia de nucleótidos comprendida por un segundo vector.

Concepto 33. El o los vectores del Concepto 31, en donde el complejo de proteína de fusión está codificado por:

- a) una primera secuencia de nucleótidos que codifica Py fusionada a una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idéntica a la secuencia de la SEQ ID No:1, SEQ ID No:3 o SEQ ID No:4; y
- b) una segunda secuencia de nucleótidos que codifica los al menos dos polipéptidos adicionales que codifica secuencias de aminoácidos que son al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idénticas a las secuencias de las SEQ ID No:1, 2, 3, 4 y/o 5.

En cualquier opción o característica, los complejos de proteína de fusión descritos en la presente memoria comprenden polipéptidos que comprenden secuencias de aminoácidos al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idénticas a cada una de las secuencias de las SEQ ID No:1, 2, 3, 4 y 5. En cualquier opción o característica, los complejos de proteína de fusión descritos en la presente memoria comprenden polipéptidos que comprenden secuencias de aminoácidos al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idénticas a cada una de las secuencias de las SEQ ID No:2, 4 y 5. En cualquier opción o característica, los

- complejos de proteína de fusión descritos en la presente memoria comprenden polipéptidos que comprenden secuencias de aminoácidos al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idénticas a cada una de las secuencias de las SEQ ID No:2, 3, 4 y 5. En cualquier opción o característica, los complejos de proteína de fusión descritos en la presente memoria comprenden polipéptidos que comprenden secuencias de aminoácidos al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idénticas a cada una de las secuencias de las SEQ ID No:1, 2, 4 y 5.
- 5 Concepto 34A. El o los vectores del Concepto 32 o del Concepto 33, en donde el ARNcr está codificado por la primera secuencia de nucleótidos.
- 10 Concepto 34B. El o los vectores del Concepto 32 o del Concepto 33, en donde el ARNcr o está codificado por la segunda secuencia de nucleótidos.
- Cuando se codifican múltiples ARNcr para dirigirse a múltiples protoespaciadores, todos ellos pueden estar codificados por la primera secuencia de nucleótidos. Cuando se codifican múltiples ARNcr para dirigirse a múltiples protoespaciadores, todos ellos pueden estar codificados por la segunda secuencia de nucleótidos.
- 15 Concepto 35A. El o los vectores de uno cualquiera de los Conceptos 31 a 34, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es heterólogo a la al menos una de las secuencias de nucleótidos.
- Concepto 35B. El o los vectores de uno cualquiera de los Conceptos 31 a 34, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es un promotor de célula eucariota.
- El promotor de célula eucariota puede ser cualquiera de los descritos en la presente memoria.
- 20 Concepto 35C. El o los vectores de uno cualquiera de los Conceptos 31 a 34, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es un promotor animal.
- En un ejemplo, el promotor es un promotor de mamífero o humano. El promotor animal, de mamífero o humano puede ser cualquiera de los descritos en la presente memoria.
- 25 Concepto 35D. El o los vectores de uno cualquiera de los Conceptos 31 a 34, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es un promotor vegetal.
- El promotor vegetal puede ser cualquiera de los descritos en la presente memoria.
- Concepto 35E. El o los vectores de uno cualquiera de los Conceptos 31 a 34, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es un promotor fúngico.
- En un ejemplo, el promotor es un promotor de levadura. El promotor fúngico o de levadura puede ser cualquiera de los descritos en la presente memoria.
- 30 Concepto 35F. El o los vectores de uno cualquiera de los Conceptos 31 a 34, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es un promotor de insecto.
- El promotor de insecto puede ser cualquiera de los descritos en la presente memoria.
- 35 Concepto 35G. El o los vectores de uno cualquiera de los Conceptos 31 a 34, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es un promotor viral.
- En un ejemplo, el promotor es un promotor viral, de AAV o lentiviral. El promotor viral, de AAV o lentiviral puede ser cualquiera de los descritos en la presente memoria.
- Concepto 35H. El o los vectores de uno cualquiera de los Conceptos 31 a 34, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que es un promotor sintético.
- 40 El promotor sintético puede ser cualquiera de los descritos en la presente memoria.
- Concepto 36. El o los vectores de uno cualquiera de los Conceptos 31 a 35, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor constitutivo.
- Concepto 37. El o los vectores de uno cualquiera de los Conceptos 31 a 35, en donde al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor inducible.
- 45 Concepto 38A. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr, en donde el ARNcr comprende dos secuencias repetidas (por ejemplo, dos secuencias repetidas).
- El ARNcr y otras características de las secuencias repetidas pueden tener las otras características descritas en otras

partes de la presente memoria (por ejemplo, en el Primer Aspecto de la Descripción Detallada o en el Concepto 10).

Concepto 38B. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr, en donde el ARNcr comprende al menos una secuencia repetida que comprende la secuencia de nucleótidos de la SEQ ID No:6.

- 5 Concepto 38C. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr, en donde el ARNcr comprende dos secuencias repetidas que comprenden la secuencia de nucleótidos de la SEQ ID No:6.

- 10 Concepto 38D. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr, en donde el ARNcr comprende al menos una secuencia repetida que comprende la secuencia de nucleótidos de la SEQ ID No:70.

Concepto 38E. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr, en donde el ARNcr comprende dos secuencias repetidas que comprenden la secuencia de nucleótidos de la SEQ ID No:70.

- 15 Concepto 39. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, en donde el protoespaciador está inmediatamente adyacente a una secuencia de Resto Adyacente al Protoespaciador (PAM) en la secuencia diana seleccionada de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3', 5'-ACA-3', 5'-ACT-3', 5'-ATC-3', 5'-ATA-3', 5'-GAG-3', 5'-TAG-3', 5'-ACC-3', 5'-AGG-3', 5'-ATT-3', 5'-GAC-3' y 5'-GTG-3'.

- 20 En cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, el protoespaciador puede estar inmediatamente adyacente a una secuencia de Resto Adyacente al Protoespaciador (PAM) en la secuencia diana seleccionada de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3', 5'-ACA-3', 5'-ACT-3', 5'-ATC-3', 5'-ATA-3', 5'-GAG-3' y 5'-TAG-3'. En cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, el protoespaciador puede estar inmediatamente adyacente a una secuencia de Resto Adyacente al Protoespaciador (PAM) en la secuencia diana seleccionada de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3' y 5'-ACA-3'. En cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, el protoespaciador puede estar inmediatamente adyacente a una secuencia de Resto Adyacente al Protoespaciador (PAM) en la secuencia diana que es 5'-AAC-3'. En cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, el protoespaciador puede estar inmediatamente adyacente a una secuencia de Resto Adyacente al Protoespaciador (PAM) que tiene cualquiera de las características descritas en otra parte de la presente memoria.
- 25
- 30
- 35

- 40 Concepto 40. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, en donde la secuencia espaciadora tiene una longitud de (aproximadamente) 25 a 39 nucleótidos.

La secuencia espaciadora puede tener una longitud de (aproximadamente) 28 a 32 nucleótidos. La secuencia espaciadora puede tener una longitud de aproximadamente 32 nucleótidos. La secuencia espaciadora puede tener una longitud de 32 nucleótidos.

- 45 En cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, la secuencia espaciadora puede tener una longitud de 32 nucleótidos. En cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, la secuencia espaciadora puede tener cualquiera de las características descritas en otras partes de la presente memoria (por ejemplo, en el Primer Aspecto de la Descripción Detallada, en el Concepto 10 y en el Concepto 30).

- 50 Concepto 41. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, en donde los nucleótidos 1 a 28 del espaciador son idénticos al complemento de los nucleótidos 1 a 28 de la secuencia del protoespaciador que está inmediatamente en 5' de la secuencia PAM en la secuencia diana.

- 55 El protoespaciador puede tener cualquiera de las características descritas en otras partes de la presente memoria (por ejemplo, en el Primer Aspecto de la Descripción Detallada, en el Concepto 10 y en el Concepto 30).

Concepto 42. El o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, en donde el espaciador es idéntico al complemento del protoespaciador en la secuencia diana.

- 5 El espaciador puede ser idéntico al complemento del protoespaciador en la secuencia diana a lo largo de toda su longitud. El espaciador puede ser (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idéntico al complemento del protoespaciador en la secuencia diana a lo largo de toda su longitud. El espaciador puede ser idéntico al complemento del protoespaciador en la secuencia diana a lo largo de los primeros 1 a 28 nucleótidos que están inmediatamente en 5' de la secuencia PAM en la secuencia diana, y (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idéntico a lo largo del resto de los nucleótidos en el protoespaciador.

Concepto 43. El o los vectores, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprenden una proteína de fusión en donde Py es una nucleasa que es una nucleasa I-TevI, en donde la nucleasa I-TevI reconoce una secuencia de nucleótidos del sitio de escisión de I-TevI que está en 5' o 3' (por ejemplo, en 5') del protoespaciador en la secuencia diana.

- 15 La secuencia de nucleótidos del sitio de escisión de I-TevI puede ser como se describe en otra parte de la presente memoria. En un ejemplo, la secuencia de nucleótidos del sitio de escisión de I-TevI está en 5' del protoespaciador en la secuencia diana. La secuencia de nucleótidos del sitio de escisión de I-TevI puede tener una longitud entre (aproximadamente) 3 y 7 nucleótidos, por ejemplo, una longitud entre (aproximadamente) 4 y 6 nucleótidos. La secuencia de nucleótidos del sitio de escisión de I-TevI puede tener una longitud de aproximadamente 5 nucleótidos.
- 20 La secuencia de nucleótidos del sitio de escisión de I-TevI puede comprender el esto 5'-CNNNG-3'. La secuencia de nucleótidos del sitio de escisión de I-TevI puede seleccionarse de 5'-CCACG-3', 5'-CACCG-3', 5'-CATAG-3', 5'-CTAAG-3', 5'-CTGAG-3', 5'-CTTAG-3', 5'-CCTCG-3', 5'-CAGCG-3', 5'-CACTG-3', 5'-CGACG-3', 5'-CGATG-3', 5'-CTCCG-3', 5'-CTACG-3', 5'-CTGTG-3', 5'-CTGCG-3', 5'-CCGTG-3', 5'-CCTAG-3', 5'-CCATG-3', 5'-CAAAG-3', 5'-CAGAG-3', 5'-CATTG-3' y 5'-CATCG-3'. La secuencia de nucleótidos del sitio de escisión de I-TevI puede seleccionarse de 5'-CCACG-3', 5'-CACCG-3', 5'-CATAG-3', 5'-CTAAG-3', 5'-CTGAG-3', 5'-CTTAG-3', 5'-CCTCG-3' y 5'-CAGCG-3', o se selecciona de 5'-CCACG-3', 5'-CACCG-3', 5'-CATAG-3', 5'-CTAAG-3' y 5'-CTGAG-3'. La secuencia de nucleótidos del sitio de escisión de I-TevI puede seleccionarse de 5'-CCACG-3', 5'-CACCG-3' y 5'-CATAG-3'. La secuencia de nucleótidos del sitio de escisión de I-TevI puede seleccionarse de 5'-CCACG-3' y 5'-CACCG-3'. La secuencia de nucleótidos del sitio de escisión de I-TevI puede tener la secuencia de nucleótidos de la SEQ ID No:42.
- 30 Concepto 44. El o los vectores, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprenden una proteína de fusión en donde Py es una nucleasa que es una nucleasa I-TevI, en donde la nucleasa I-TevI reconoce una secuencia de nucleótidos espaciadora de I-TevI que está en 5' o 3' (por ejemplo, en 5') del protoespaciador en la secuencia diana,

- La secuencia de nucleótidos espaciadora de I-TevI puede ser como se describe en otra parte de la presente memoria.
- 35 En un ejemplo, la secuencia de nucleótidos espaciadora de I-TevI está en 5' del protoespaciador en la secuencia diana. La secuencia de nucleótidos espaciadora de I-TevI puede tener una longitud entre (aproximadamente) 10 y 35 nucleótidos, tal como una longitud de 15 a 35, o de 20 a 35. La secuencia de nucleótidos espaciadora de I-TevI puede tener una longitud entre (aproximadamente) 29 y 33 nucleótidos. La secuencia de nucleótidos espaciadora de I-TevI puede tener una longitud entre (aproximadamente) 30 y 32 nucleótidos. La secuencia de nucleótidos espaciadora de I-TevI puede tener una longitud de (aproximadamente) 31 nucleótidos. La secuencia de nucleótidos espaciadora de I-TevI puede ser al menos (aproximadamente) un 80 % (por ejemplo, (aproximadamente) un 90 %) idéntica a la secuencia de nucleótidos de la SEQ ID No:43. La secuencia de nucleótidos espaciadora de I-TevI puede ser un 100 % idéntica a la secuencia de nucleótidos de la SEQ ID No: 43.
- 40

- Concepto 45. El o los vectores, proteínas de fusión o complejos de proteína de fusión descritos en cualquier opción o característica que comprenden una proteína de fusión en donde Py es una nucleasa que es una nucleasa I-TevI, en donde la secuencia diana comprende, en orientación 5' a 3':
- 45

- a) una secuencia de nucleótidos del sitio de escisión de I-TevI;
- b) una secuencia de nucleótidos espaciadora de I-TevI;
- c) una secuencia PAM; y
- 50 d) una secuencia de protoespaciador.

- En un ejemplo, las secuencias a) a d) son cada una inmediatamente adyacentes entre sí. La secuencia de nucleótidos del sitio de escisión de I-TevI es como se describe en otra parte de la presente memoria (por ejemplo, como se describe en el Concepto 43). La secuencia de nucleótidos espaciadora de I-TevI es como se describe en otra parte de la presente memoria (por ejemplo, como se describe en el Concepto 44). La secuencia PAM es como se describe en otra parte de la presente memoria (por ejemplo, como se describe en el Concepto 39). La secuencia del protoespaciador es como se describe en otra parte de la presente memoria (por ejemplo, como en el Primer Aspecto de la Descripción Detallada, o en el Concepto 10, o en el Concepto 30).
- 55

Concepto 46A. Una célula que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria.

Concepto 46B. Una célula eucariota que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria.

- 5 Concepto 46C. Una célula animal, vegetal, de insecto o de hongo que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria.

Concepto 46D. Una célula procariota que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria.

- 10 Concepto 46E. Una célula procariota que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria, en donde la célula no es una célula bacteriana (por ejemplo, una célula de *E. coli*, *Pseudomonas* o *Klebsiella*) que comprende secuencias de nucleótidos endógenas que codifican los polipéptidos de a) a e) como se recita en el Concepto 1A.

- 15 En un ejemplo, la célula es una célula de vertebrado, de mamífero o humana.

Concepto 47A. Una composición que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria.

- 20 La composición puede ser como se describe en otra parte de la presente memoria. En un ejemplo, una composición descrita en la presente memoria puede ser una composición *in vitro*. En un ejemplo, una composición descrita en la presente memoria puede estar contenida en un recipiente médico.

También se proporciona una composición farmacéutica que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria y que comprende además un diluyente, excipiente o vehículo. Las composiciones farmacéuticas están destinadas a utilizarse en cualquiera de los métodos de tratamiento médico descritos en la presente memoria.

- 25 También se proporciona uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria para su uso en terapia.

- 30 También se proporciona uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica como se describe en otra parte de la presente memoria para su uso en el tratamiento de cualquier enfermedad descrita en la presente memoria mediante la realización de cualquiera de los métodos descritos en la presente memoria.

Las composiciones farmacéuticas pueden estar contenidas en un dispositivo médico (como una ampolla, una jeringa o un inhalador). Las composiciones farmacéuticas pueden estar formuladas en una tintura, una cápsula o una formulación de liberación lenta. Las composiciones farmacéuticas pueden ser un comprimido oral contenido en un envase blíster.

- 35 En un ejemplo, la composición farmacéutica que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión como se describe en la presente memoria se formula para administración oral o rectal. En un ejemplo, la composición farmacéutica se formula para administración oral. En un ejemplo, la composición farmacéutica se formula como una cápsula o un comprimido recubierto.

- 40 La formulación que comprende el uno o más vectores, proteínas de fusión o complejos de proteínas de fusión descritos en la presente memoria se puede liofilizar antes de la encapsulación. En un ejemplo, la composición farmacéutica que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión como se describe en la presente memoria es una formulación liofilizada. En un ejemplo, la composición farmacéutica que comprende uno o más vectores, proteínas de fusión o complejos de proteínas de fusión como se describe en la presente memoria es una formulación encapsulada para ser liberada en el intestino delgado de un sujeto.

- 45 Concepto 47B. Una composición (opcionalmente, una composición *in vitro* o en donde la composición está contenida en un recipiente médico) que comprende un ARNcr o ácido nucleico que codifica el ARNcr; y uno, más o todos de

- a) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- 50 b) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- c) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;

- d) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- e) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5

5 en donde

A: el ARNcr comprende un espaciador que es cognado a un primer protoespaciador, en donde el protoespaciador

- f) no se encuentra en *E. coli*;
- g) es un protoespaciador de célula eucariota; o
- h) es un protoespaciador de célula animal (opcionalmente, de mamífero o humana), vegetal o fúngica;

10 o

B: un protoespaciador de una *E. coli* que carece de secuencias de nucleótidos endógenas que codifican los polipéptidos citados en las partes a) a e) anteriores.

En un ejemplo, la composición comprende (a)-(e). En un ejemplo, la composición comprende (a)-(d). En un ejemplo, la composición comprende (a)-(c). En un ejemplo, la composición comprende (a)-(b). En un ejemplo, la composición comprende (d)-(e). En un ejemplo, la composición comprende (c)-(e). En un ejemplo, la composición comprende (b)-(e).

15

En un ejemplo, la composición comprende (a). En un ejemplo, la composición comprende (b). En un ejemplo, la composición comprende (c). En un ejemplo, la composición comprende (d). En un ejemplo, la composición comprende (e).

20

El ARNcr es operativo en una célula para guiar el o los polipéptidos, proteínas, proteínas de fusión y complejos descritos en la presente memoria a una secuencia diana contenida en un ácido nucleico en la célula. Se proporciona un complejo que comprende uno, más o todos los polipéptidos (a)-(e) y el ARNcr, en donde el ARNcr es capaz de guiar el complejo a un protoespaciador contenido en un ADN diana. Por ejemplo, el complejo comprende (a)-(e).

25

Concepto 47C. Una composición (opcionalmente, una composición *in vitro* o en donde la composición está contenida en un recipiente médico) que comprende un ARNcr o un ácido nucleico que codifica el ARNcr; y un ácido nucleico que codifica uno, más o todos de

- a) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- b) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- c) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;
- d) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- e) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5

30

35

en donde

A: el ARNcr comprende un espaciador que es cognado a un primer protoespaciador, en donde el protoespaciador

- f) no se encuentra en *E. coli*;
- g) es un protoespaciador de célula eucariota; o
- h) es un protoespaciador de célula animal (opcionalmente, de mamífero o humano), vegetal o fúngica;

40

o

B: un protoespaciador de una *E. coli* que carece de secuencias de nucleótidos endógenas que codifican los polipéptidos citados en las partes a) a e) anteriores.

45

En un ejemplo, la composición comprende un vector que codifica al menos 2, 3 o 4 de dichos polipéptidos. En un

ejemplo, (a)-(e) están codificados por el mismo vector. En un ejemplo, la composición comprende un vector que codifica todos los (a)-(e).

En un ejemplo, el ácido nucleico de la composición codifica (a)-(e). En un ejemplo, el ácido nucleico de la composición codifica (a)-(d). En un ejemplo, el ácido nucleico de la composición codifica (a)-(c). En un ejemplo, el ácido nucleico de la composición codifica (a)-(b). En un ejemplo, el ácido nucleico de la composición codifica (d)-(e). En un ejemplo, el ácido nucleico de la composición codifica (c)-(e). En un ejemplo, el ácido nucleico de la composición codifica (b)-(e).

El ARNcr es operativo en una célula para guiar el o los polipéptidos a una secuencia diana contenida en un ácido nucleico en la célula. Se proporciona un complejo que comprende uno, más o todos los polipéptidos (a)-(e) y el ARNcr, en donde el ARNcr es capaz de guiar el complejo a un protoespaciador contenido en un ADN diana. Por ejemplo, el complejo comprende (a)-(e).

También se proporcionan métodos. Cualquier método de la presente memoria puede ser un método *in vitro*. Opcionalmente, el método se lleva a cabo en una célula contenida en un sujeto, como un ser humano, un animal (por ejemplo, un mamífero, por ejemplo, un roedor, ratón o rata), planta, insecto u hongo (por ejemplo, levadura).

En la presente memoria, modificar (por ejemplo, modificar una secuencia diana de un ADN, polinucleótido o célula) puede ser cuando la modificación corta, edita, bloquea, marca o etiqueta la secuencia diana.

Concepto 48A. Un método para modificar una secuencia diana de ácido nucleico en una célula, comprendiendo el método

(I) poner en contacto la célula con uno o más vectores, proteínas, proteínas de fusión, complejos de proteínas de fusión o con una composición como la descrita en cualquier opción o característica que comprende un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana;

(II) permitir la introducción en la célula de

(i) el o los vectores, proteínas, proteínas de fusión o complejos de proteína de fusión, mediante los cuales el o los polipéptidos y el ARNcr se expresan en la célula, o

(ii) el ARNcr de la composición (o ácidos nucleicos que codifican el ARNcr mediante lo cual el ARNcr se expresa en la célula) y el ácido nucleico de la composición que codifica el o los polipéptidos o proteínas, mediante lo cual el o los polipéptidos o proteínas se expresan en la célula;

(III) en donde el ARNcr forma un complejo con el o los polipéptidos o proteínas para guiar el complejo a la secuencia diana.

Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas CasS o la introducción de la o las proteínas CasS, la modificación puede ser la regulación a la baja de un gen que comprende la secuencia diana. Cuando el método comprende la introducción de del o de los vectores que expresan la o las proteínas CasS o la introducción de la o las proteínas CasS, la modificación puede ser la regulación a la baja de un gen adyacente a (por ejemplo, a 2 kb de) la secuencia diana.

Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteína de fusión, la modificación puede ser cortar la secuencia diana (por ejemplo, cuando Py comprende una nucleasa o una nickasa).

Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteína de fusión, la modificación puede ser la introducción de una o más mutaciones en la secuencia diana (por ejemplo, cuando Py comprende una nucleasa, una nickasa, un *prime editor* o un editor de bases).

Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteína de fusión, la modificación puede ser la realización de una edición de bases en la secuencia diana (por ejemplo, cuando Py es un editor de bases).

Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteína de fusión, la modificación puede ser la realización de *prime editing* en la secuencia diana (por ejemplo, cuando Py es un *prime editor*).

Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser la desaminación de uno o más nucleótidos en la secuencia diana (por ejemplo, cuando Py

es una desaminasa, tal como una citosina o adenina desaminasa).

5 Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser la metilación de uno o más nucleótidos en la secuencia diana (por ejemplo, cuando Py es un metiltransferasa).

Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser la metilación de uno o más nucleótidos en la secuencia diana (por ejemplo, cuando Py es una metilasa).

10 Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser la acetilación de uno o más nucleótidos en la secuencia diana (por ejemplo, cuando Py es una acetilasa).

15 Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser someter el ADN de la secuencia diana a la actividad de acetiltransferasa (por ejemplo, cuando Py es una acetiltransferasa).

20 Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser la transcripción inversa de la secuencia diana, o la transcripción inversa de un ARN en la célula para producir un ADN que se inserta en la secuencia diana (por ejemplo, cuando Py es una transcriptasa inversa).

25 Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser la activación de la transcripción de un gen en la célula, tal como un gen que comprende o es adyacente a la secuencia diana (por ejemplo, cuando Py es un activador de la transcripción).

30 Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser la desactivación de la transcripción de un gen en la célula, tal como un gen que comprende o es adyacente a la secuencia diana (por ejemplo, cuando Py es un desactivador de la transcripción).

35 Cuando el método comprende la introducción del o de los vectores que expresan la o las proteínas de fusión o complejos de proteína de fusión, o la introducción de la o las proteínas de fusión o complejos de proteínas de fusión, la modificación puede ser la desactivación de la traducción de un gen en la célula, tal como un gen que comprende o es adyacente a la secuencia diana (por ejemplo, cuando Py es un desactivador de la traducción).

40 Concepto 48B. Un método para modificar una secuencia de ácido nucleico diana en una célula, comprendiendo el método

a) poner en contacto la célula con el o los vectores del Concepto 10 o uno cualquiera de los Conceptos 11 a 12 cuando dependan del Concepto 10, o una composición según el Concepto 47;

b) permitir la introducción en la célula de

(i) el o los vectores mediante lo cual se expresan el o los polipéptidos y el ARNcr en la célula, o

45 (ii) el ARNcr de la composición (o ácidos nucleicos que codifican el ARNcr mediante lo cual el ARNcr se expresa en la célula) y el ácido nucleico de la composición que codifica el o los polipéptidos, mediante lo cual el o los polipéptidos se expresan en la célula;

c) en donde el ARNcr forma un complejo con el o los polipéptidos para guiar el complejo a la secuencia diana.

50 Cabe destacar que encontramos que la orientación de los espaciadores no influye en la eficiencia del sistema CRISPR-Cas Tipo-S para regular a la baja la transcripción de un gen seleccionado, lo que no es el caso de otros sistemas CRISPR-Cas que deben dirigirse a la región codificante o no codificante de un gen para regular a la baja de manera eficiente su transcripción. Por lo tanto, el complejo, por ejemplo, modifica (por ejemplo, corta o edita) una secuencia diana codificante o una secuencia diana no codificante, opcionalmente en donde (i) se modifica una hebra codificante,

pero no una hebra no codificante de ADN o (ii) se modifica una hebra no codificante, pero no una hebra codificante de ADN. La modificación en la presente memoria puede modificar de manera útil (i) una hebra codificante, pero no una hebra no codificante de ADN o (ii) una hebra no codificante, pero no una hebra codificante de ADN.

Concepto 48C. El método del Concepto 48B, en donde el complejo se guía hasta la secuencia diana y modifica la secuencia diana, opcionalmente en donde el complejo comprende un componente seleccionado de

- a) una nucleasa que corta la secuencia diana;
- b) una desaminasa (como una citosina o adenina desaminasa) que lleva a cabo la desaminación de la secuencia diana;
- c) un editor bases que edita bases de la secuencia diana;
- d) un *prime editor* que realiza la edición principal de la secuencia diana;
- e) una transcriptasa inversa que transcribe de forma inversa la secuencia diana o que transcribe de forma inversa un ARN en la célula para producir un ADN que se inserta en la secuencia diana;
- f) una metiltransferasa que metila el ADN en la célula;
- g) una metilasa que metila el ADN en la célula;
- h) una acetilasa que acetila el ADN en la célula;
- i) una acetiltransferasa que somete el ADN de la célula a la actividad acetiltransferasa;
- j) un activador de la transcripción que activa la transcripción de un gen en la célula, tal como un gen que comprende o es adyacente a la secuencia diana; y
- k) un desactivador de la transcripción que desactiva la transcripción de un gen en la célula, tal como un gen que comprende o es adyacente a la secuencia diana;

opcionalmente, en donde el componente está codificado por el o los vectores o ácido nucleico de la composición.

En cualquier opción o característica relacionada con la modificación del ADN o ARN diana, el complejo comprende además un editor de bases (por ejemplo, como se describe en otra parte de la presente memoria).

En cualquier opción o característica relacionada con la modificación del ADN o ARN diana, el complejo comprende además un *prime editor* (por ejemplo, como se describe en otra parte de la presente memoria).

En cualquier opción o característica relacionada con la modificación del ADN o ARN diana, el complejo comprende además una nucleasa (por ejemplo, como se describe en otra parte de la presente memoria).

Concepto 49A. El método del Concepto 48C, en donde la célula es una célula eucariota, animal, vegetal, de insecto o fúngica.

Concepto 49B. El método del Concepto 48, en donde la célula es una célula procariota.

En un ejemplo, la célula es una célula procariota en donde la célula no es una célula procariota (por ejemplo, una célula de *E. coli*, *Pseudomonas* o *Klebsiella*) que comprende secuencias de nucleótidos endógenas que codifican los polipéptidos a) a e) como se cita en el Concepto 1.

En cualquier opción o característica de la presente memoria, un animal de la presente memoria es un mamífero o vertebrado no humano. Opcionalmente, una célula animal de la presente memoria es una célula de mamífero o vertebrado no humano. Opcionalmente, la célula es una célula vegetal, una célula bacteriana, una célula fúngica, una célula de mamífero, una célula de insecto o una célula de arqueas. Una célula de la presente memoria es *ex vivo*.

En cualquier opción o característica de la presente memoria, una célula de la presente memoria es una célula pluripotente, tal como una célula madre pluripotente. Opcionalmente, una célula de la presente memoria es una célula madre, tal como una célula madre embrionaria. Opcionalmente, una célula de la presente memoria es una célula totipotente no humana. En un ejemplo de estas opciones, la célula es una célula humana. Según la invención, la célula no es una célula humana contenida en un ser humano. Según la invención, la célula no es un embrión humano.

Ningún método de la presente memoria, según la invención, es ninguno de los siguientes:

- (a) un proceso para clonar un ser humano;
- (b) un proceso para modificar la identidad genética de la línea germinal de un ser humano;

- (c) una utilización de un embrión humano;
- (d) un proceso para modificar la identidad genética de un animal que probablemente le cause sufrimiento sin ningún beneficio médico sustancial para el ser humano o el animal.

5 Ningún método de la presente memoria es un método de tratamiento del cuerpo humano o animal. Ningún método de la presente memoria es un método quirúrgico.

Concepto 50. El método del Concepto 48 o 49, en donde la secuencia diana está contenida en un cromosoma o episoma en la célula.

10 En cualquier opción o característica de la presente memoria, la secuencia diana está contenida en un cromosoma. En cualquier opción o característica de la presente memoria, la secuencia diana está contenida en un plásmido. En cualquier opción o característica de la presente memoria, la secuencia diana no está contenida en un plásmido.

Concepto 51. El método del Concepto 50, en donde el método inhibe la replicación de cromosomas o episomas en la célula.

15 En un ejemplo, la inhibición es al menos (aproximadamente) del 20, 30, 40, 50, 60, 70, 80, 90 o 95 % en comparación con la replicación en una célula idéntica que no ha sido expuesta al vector o vectores o a la composición. En un ejemplo, la inhibición es del 100 %.

Concepto 52. El método de uno cualquiera de los Conceptos 48 a 51, en donde la secuencia diana está dentro de (aproximadamente) 2 kb aguas arriba o aguas abajo de un gen de interés.

20 En cualquier opción o característica relacionada con una secuencia diana, la secuencia diana está dentro de (aproximadamente) 2 kb aguas arriba (5') o aguas abajo (3') de un gen de interés. La secuencia diana puede estar aguas arriba (5') del gen de interés. La secuencia diana puede estar aguas abajo (3') del gen de interés. La secuencia diana puede estar dentro de (aproximadamente) 1,75 kb, dentro de (aproximadamente) 1,5 kb, dentro de (aproximadamente) 1,25 kb, dentro de (aproximadamente) 1 kb, dentro de (aproximadamente) 0,75 kb, dentro de (aproximadamente) 0,5 kb o dentro de (aproximadamente) 0,25 kb aguas arriba (5') o aguas abajo (3') de un gen de interés. La secuencia diana puede estar dentro de un gen de interés, por ejemplo, puede estar dentro del promotor, exón(es) y/o intrón(es) del gen de interés. En un ejemplo, la secuencia diana está dentro del promotor del gen de interés.

Concepto 53. El método del Concepto 52, en donde la secuencia diana del gen de interés está al menos 2kb aguas arriba y aguas abajo de cualquier promotor, exón e intrón de cualquier otro gen.

30 Como se muestra en el Ejemplo 4 de la presente memoria, la proteína CasS3 puede tener una amplia gama de influencia, mostrándose los efectos de unión en un rango de aproximadamente 2 kb a cada lado (es decir, 5' y 3') de la secuencia del protoespaciador en la secuencia diana. Por lo tanto, para aplicaciones en las que se utiliza el sistema CasS para CRISPRi (es decir, en cualquier método descrito en la presente memoria de regulación al alza o regulación a la baja de un gen de interés), puede ser beneficioso evitar efectos fuera de la diana si el protoespaciador en la secuencia diana está al menos 2 kb aguas arriba (5') y aguas abajo (3') de cualquier otro gen (es decir, promotor, exón o intrón).

Concepto 54. El método de uno cualquiera de los Conceptos 48 a 53, en donde la modificación es una sustitución, delección o inserción de uno o más nucleótidos.

40 En un ejemplo, la modificación es una sustitución de uno o más nucleótidos en la secuencia diana. En un ejemplo, la modificación es una o más sustituciones de C a T en los nucleótidos de la secuencia diana cuando Py es un CBE. En un ejemplo, el CBE es una citidina desaminasa PmCDA1, por ejemplo, como se describe en otra parte de la presente memoria.

45 Concepto 55. Un método para tratar o prevenir una enfermedad o afección que está mediada por células diana en un sujeto, comprendiendo el método llevar a cabo un método como se describe en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en uno cualquiera de los Conceptos 48 a 54) para modificar las células diana, en donde dicho contacto comprende administrar el o los vectores, proteínas (incluyendo proteínas de fusión o complejos de proteína de fusión) o composición al sujeto y en donde la modificación trata o previene la enfermedad o afección.

Las células diana pueden ser, por ejemplo, células bacterianas patógenas. Pueden ser células bacterianas resistentes a los antibióticos y la modificación vuelve a sensibilizar a la célula bacteriana al antibiótico.

50 Concepto 56. El método del Concepto 55, en donde el sujeto es un ser humano, un animal o una planta.

Concepto 57. El método del Concepto 56, en donde las células diana son células del sujeto que comprenden un defecto en el ácido nucleico y la modificación corrige el defecto.

Por ejemplo, se sabe que ciertas enfermedades están causadas por polimorfismos de un solo nucleótido (SNP) (por ejemplo, el síndrome de progeria y la displasia mandibuloacral están causados por un SNP sin sentido, SNP c.1580G>T, en el gen LMNA que codifica la proteína lamina A/C, y un SNP en el gen F5 causa trombofilia del Factor V Leiden). La corrección de estos SNP por sustituciones de dianas podría usarse para mejorar las enfermedades. En 2012, se demostró que 124 enfermedades genéticas estaban causadas por inserciones de retrotransposones, incluyendo fibrosis quística (Alu), hemofilia A (L1) y distonía-parkinsonismo ligado al cromosoma X (SVA), véase Hancks *et al.*, Curr. Opin. Genet. Dev., 22, 191-203, 2012, doi: 10.1016/j.gde.2012.02.006. Es posible utilizar ciertas ediciones dirigidas para corregir estos tipos de inserciones y mejorar la o las enfermedades asociadas.

Concepto 58A. El método del Concepto 56 o 57, en donde la modificación añade una nueva funcionalidad a las células.

- 10 En un ejemplo, la modificación regula al alza o regula a la baja la expresión de un gen en las células. En un ejemplo, la modificación regula a la baja la expresión de un gen en las células. En un ejemplo, la modificación inhibe la expresión de un gen en las células. En un ejemplo, la modificación bloquea la expresión de un gen en las células.

- 15 En otro ejemplo, la modificación es la inserción de un nuevo gen o una nueva vía de genes (por ejemplo, un operón), que se puede utilizar para producir sustancias que son beneficiosas para las células. Por ejemplo, los nuevos genes pueden producir metabolitos beneficiosos o productos terapéuticos. En otro ejemplo, la modificación es la inserción de un nuevo gen o una nueva vía de genes (por ejemplo, un operón), que se puede utilizar para eliminar sustancias que son perjudiciales para las células. Por ejemplo, los nuevos genes pueden metabolizar toxinas nocivas.

Concepto 58B. El método del Concepto 56 o 57, en donde la modificación añade una nueva secuencia de nucleótidos a los genomas de las células para la expresión de una proteína codificada por la nueva secuencia.

- 20 Por ejemplo, la proteína es una proteína heteróloga. Por ejemplo, la proteína es una proteína terapéutica, por ejemplo, un anticuerpo, un fragmento de anticuerpo (como un dominio variable único de anticuerpo), un sitio de unión de TCR (receptor de células T), un dominio variable de TCR, una hormona, una incretina, un factor de crecimiento, un agente anticanceroso, un neurotransmisor o una enzima.

- 25 Concepto 58C. El método del Concepto 56 o 57, en donde la modificación modula la expresión de una secuencia de nucleótidos contenida en un cromosoma o episoma de la célula.

En un ejemplo, la modificación modula la expresión de una secuencia de nucleótidos contenida en un plásmido. En un ejemplo, la modificación modula la expresión de una secuencia de nucleótidos que no está contenida en un plásmido. Por ejemplo, la expresión se regula al alza. Por ejemplo, la expresión se regula a la baja.

- 30 Concepto 59. El o los vectores, proteínas (incluyendo proteínas de fusión y complejos de proteína de fusión), células o composición como se describe en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en uno cualquiera de los Conceptos 1-58) para su uso en un método para tratar o prevenir una enfermedad o afección en un ser humano o animal, en donde el método es como se describe en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en uno cualquiera de los Conceptos 55 a 58).

Cualquier vector de la presente memoria (en cualquier opción o característica de la presente memoria) puede ser

- 35 a) un vector plasmídico (opcionalmente, un plásmido conjugativo)  
b) un vector de transposón (opcionalmente, un transposón conjugativo);  
c) un vector de virus (opcionalmente, un vector de fago, AAV o lentivirus);  
d) un fagémido (opcionalmente, un fagémido empaquetado); o  
e) una nanopartícula (opcionalmente, una nanopartícula lipídica).

- 40 Concepto 60. Un método para introducir una edición dirigida en un polinucleótido diana, comprendiendo el método poner en contacto el polinucleótido diana con una proteína de fusión o un complejo de proteína de fusión como se describe en cualquier opción o característica de la presente memoria, en donde la proteína o el complejo de proteína de fusión está en combinación con un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana (por ejemplo, como se describe en el Primer Aspecto de la Descripción Detallada, Concepto 10 o en el Concepto 30 o en uno cualquiera de los Conceptos 38 a 45 cuando dependan del Concepto 30)

en donde el primer protoespaciador en la secuencia diana está contenido en el polinucleótido, en donde el ARNcr se hibrida con el protoespaciador para guiar la proteína (por ejemplo, el complejo de ribonucleoproteína) mediante lo cual la proteína (por ejemplo, el complejo de ribonucleoproteína) edita el polinucleótido.

- 50 El método puede llevarse a cabo en una célula parental que comprende el polinucleótido. Se proporciona además una célula o un organismo de progeie que se deriva de la célula parental editada, en donde la célula o el organismo de progeie conserva la edición en su genoma. El método puede comprender una o más etapas de cultivo de la célula

editada para producir dicha célula de progenie o una pluralidad de dichas células de progenie.

Concepto 61A. El método del Concepto 60, en donde el método inserta, deleciona o sustituye una secuencia de bases o ácidos nucleicos en el polinucleótido.

5 Concepto 61B. El o los vectores de los Conceptos 10 o el Concepto 11 o 12 cuando dependen del Concepto 10, método de uno cualquiera de los Conceptos 48-54, o método del Concepto 60 o 61, en donde el ARNcr es cognado a un resto adyacente al protoespaciador (PAM) que tiene la secuencia 5'-AAG-3'.

Alternativamente, el PAM puede ser uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39).

10 Concepto 62A. Un contenedor que comprende una pluralidad de proteínas operables para su uso con un ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido, comprendiendo la o las proteínas cualquiera de la o las proteínas descritas en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en el Concepto 13),

15 en donde el complejo es operable con cualquiera de los restos adyacentes al protoespaciador (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), y el contenedor no es una célula, y las proteínas se mezclan con un tampón *in vitro*.

20 Concepto 62B. Un contenedor que comprende una pluralidad de proteínas operables para su uso con un ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido, comprendiendo las proteínas cualquiera de la o las proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en uno cualquiera de los Conceptos 14 a 29),

en donde el complejo es operable con cualquiera de los restos adyacentes al protoespaciador (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), y el contenedor no es una célula, y las proteínas se mezclan con un tampón *in vitro*.

25 Concepto 62C. Un contenedor que comprende una proteína o una pluralidad de proteínas operables para su uso con un ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido, comprendiendo la o las proteínas uno, más o todos los polipéptidos seleccionados de

- a) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- 30 b) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- c) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;
- d) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- 35 e) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5

en donde el complejo es operable con un resto adyacente al protoespaciador (PAM) que tiene la secuencia 5'-AAG-3', el contenedor no es una célula y las proteínas se mezclan con un tampón *in vitro*.

40 Alternativamente, el PAM puede ser uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39).

También se proporciona:

Una proteína o pluralidad de proteínas operables para su uso con un ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido, comprendiendo la o las proteínas uno, más o todos los polipéptidos seleccionados de

- 45 a) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- b) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- 50 c) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;

- d) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- e) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5

5 en donde el complejo es operable con un resto adyacente al protoespaciador (PAM) que tiene la secuencia 5'-AAG-3'.

Alternativamente, el PAM puede ser uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39).

10 Concepto 63. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico como se describe en cualquier opción o característica de la presente memoria (por ejemplo, según uno cualquiera de los Conceptos 1 a 12, o uno cualquiera de los Conceptos 31 a 45), en donde dicho o dichos polipéptidos o proteínas (incluyendo proteína(s) de fusión y complejos de proteínas de fusión) codificados por el o los vectores son operables para su uso con un ARNcr para formar un complejo de ribonucleoproteína para la toma como diana de protoespaciadores en un polinucleótido, en donde el complejo es operable con cualquiera de los restos adyacentes a protoespaciadores (PAM) descritos en  
15 cualquier opción o característica de la presente memoria (por ejemplo, que tienen una de las secuencias citadas en el Concepto 39).

En un ejemplo, el resto adyacente al protoespaciador (PAM) tiene la secuencia 5'-AAG-3'.

Concepto 64. Un método para tomar como diana un polinucleótido, comprendiendo el método

- 20 a) poner en contacto el polinucleótido con la o las proteínas, (incluyendo la o las proteínas de fusión y los complejos de proteína de fusión) como se describe en cualquier opción o característica de la presente memoria, en donde la o las proteínas están en combinación con un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana;
- b) permitir la formación de un complejo de ribonucleoproteína que comprende la o las proteínas y ARNcr, en donde el complejo es guiado a una secuencia diana contenida en el polinucleótido para modificar el  
25 polinucleótido o la replicación del mismo.

La diana puede estar en una célula. La diana puede ser *in vitro*.

Concepto 65. El método del Concepto 64, en donde el polinucleótido está contenido en un cromosoma o un episoma.

En un ejemplo, el polinucleótido está contenido en un cromosoma. En un ejemplo, el polinucleótido está contenido en un plásmido. En un ejemplo, el polinucleótido no está contenido en un plásmido.

30 Concepto 66A. El método del Concepto 65, en donde el método inhibe la replicación del polinucleótido.

Concepto 66B. El método del Concepto 65, en donde el método edita el polinucleótido.

La edición puede insertar una secuencia de bases o de ácidos nucleicos en el polinucleótido. La edición puede delecionar una secuencia de bases o de ácidos nucleicos en el polinucleótido. La edición puede sustituir una secuencia de bases o de ácidos nucleicos en el polinucleótido.

35 Concepto 67A. Un kit que comprende

- a) Una o más proteínas descritas en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en el Concepto 13); y
- b) Un ARNcr, (o uno o más ácidos nucleicos que codifican un ARNcr), en donde el ARNcr es cognado a cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por  
40 ejemplo, un PAM que tiene una de las secuencias citadas en el Concepto 39);

en donde dicho o dichos polipéptidos son operativos para su uso con el ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido.

Concepto 67B. Un kit que comprende

- 45 a) Una o más proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en uno cualquiera de los Conceptos 14 a 29), o uno o más ácidos nucleicos que codifican dicho(s) polipéptido(s); y
- b) un ARNcr, o uno o más ácidos nucleicos que codifican un ARNcr), en donde el ARNcr es cognado a cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, un PAM que tiene una de las secuencias citadas en el Concepto 39);

en donde dicho o dichos polipéptidos son operativos para su uso con el ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido.

Concepto 67C. Un kit que comprende

- 5 a) Uno o más vectores como se describe en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en uno cualquiera de los Conceptos 1 a 12, o uno cualquiera de los Conceptos 31 a 45), o uno o más ácidos nucleicos que codifican dicho(s) polipéptido(s); y
- b) Un ARNcr, o uno o más ácidos nucleicos que codifican un ARNcr, en donde el ARNcr es cognado a cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, un PAM que tiene una de las secuencias citadas en el Concepto 39);

10 en donde dicho o dichos polipéptidos son operativos para su uso con el ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido.

Concepto 67D. Un kit que comprende

- a) Uno o más polipéptidos como se cita en el Concepto 7, o uno o más ácidos nucleicos que codifican dicho(s) polipéptido(s); y
- 15 b) Un ARNcr, (o uno o más ácidos nucleicos que codifican un ARNcr, en donde el ARNcr es cognado a PAM que tiene la secuencia 5'-AAG-3' );

en donde dicho o dichos polipéptidos son operativos para su uso con el ARNcr o ARN guía para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido.

20 Se proporciona un complejo como se describe en la presente memoria. En un ejemplo, el complejo comprende además una o más proteínas Cas Cascada, por ejemplo, una, más o todas las proteínas CasA-E de Tipo I. El complejo puede comprender además una Cas3 y, opcionalmente, una o más proteínas Cascada de Tipo I cognadas. El complejo puede comprender además una Cas3, 9, 10, 12 o 13. Por lo tanto, la función de un sistema o componentes de Tipo-S se puede proporcionar junto con la función en una o más proteínas de un Tipo diferente de sistema CRISPR/Cas. El complejo puede comprender una helicasa fusionada a una nucleasa. La nucleasa puede ser, por ejemplo, una Cas3, 9, 10, 12 o 13. Por ejemplo, la nucleasa es una Cas3. Por ejemplo, la nucleasa es una Cas9.

En un ejemplo, el complejo comprende además una nucleasa Tefl, por ejemplo, un dominio de nucleasa I-Tefl. La I-Tefl puede ser como se describe en cualquier opción o característica de la presente memoria (por ejemplo, en el Concepto 14, o en los Conceptos 21 a 24, o en los Conceptos 43 a 45).

En un ejemplo, el complejo comprende además una proteína MthH.

30 En un ejemplo, el complejo es un complejo aislado. El complejo puede estar presente *in vitro*. El complejo puede estar contenido en un recipiente médico, como una bolsa IV, un vial médico o un dispositivo médico de inyección.

Concepto 68A. Un complejo de ribonucleoproteína que comprende

- a) Una o más proteínas como se describe en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en el Concepto 13); y
- 35 b) Un ARNcr que es cognado a cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39).

Concepto 68B. Un complejo de ribonucleoproteína que comprende

- 40 a) Una o más proteínas de fusión o complejos de proteínas de fusión descritos en cualquier opción o característica de la presente memoria (por ejemplo, como se describe en uno cualquiera de los Conceptos 14 a 29); y
- b) Un ARNcr que es cognado a cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39).

Concepto 68C. Un complejo de ribonucleoproteína que comprende

- 45 a) Uno o más polipéptidos como se cita en el Concepto 5, o uno o más ácidos nucleicos que codifican dicho(s) polipéptido(s); y
- b) Un ARNcr, o uno o más ácidos nucleicos que codifican un ARNcr, donde el ARNcr es cognado a PAM que tiene la secuencia 5'-AAG-3'.

Concepto 68D. El complejo del Concepto 68C, en donde el complejo está desprovisto de

- a) una nucleasa Cas;
- b) una nucleasa de ADN o una nucleasa de ARN; o
- c) una nucleasa.

5 Concepto 69. Un método de control transcripcional o replicación de un ADN diana, que comprende poner en contacto el ADN diana con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo está desprovisto de una nucleasa de ADN y el complejo se une al ADN diana, controlando así la transcripción o replicación del ADN diana.

10 Concepto 70A. Un método para controlar la replicación de un ARN diana, que comprende poner en contacto el ARN diana o un ADN que codifica el ARN con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo se une al ARN o ADN diana, controlando así la transcripción del ARN diana.

15 Concepto 70B. Un método para controlar la transcripción de un ARN diana, que comprende poner en contacto el ARN diana o un ADN que codifica el ARN con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo se une al ARN o ADN diana, controlando así la transcripción del ARN diana.

Concepto 71. El método del Concepto 69 o 70, en donde el ADN está contenido en un cromosoma o un episoma (opcionalmente, un plásmido).

En un ejemplo, el ADN está contenido en un cromosoma. En otro ejemplo, el ADN está contenido en un plásmido. En otro ejemplo, el ADN no está contenido en un plásmido.

20 Concepto 72. Un método para editar un ADN diana, que comprende poner en contacto el ADN diana con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo se une al ADN diana, editando así el ADN diana.

25 Concepto 73A. Un método para escindir un ADN bicatenario (ADNds), que comprende poner en contacto dicho ADNds con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo comprende una proteína de fusión en donde Py comprende una nucleasa, en donde el ADNds comprende una secuencia protoespaciadora flanqueada en su extremo 5' por uno cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), por lo que la nucleasa escinde el ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador.

30 Concepto 73B. Un método para escindir ARN, que comprende poner en contacto dicho ARN con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo comprende una proteína de fusión en donde Py comprende una nucleasa de ARN, en donde el ARN comprende una secuencia protoespaciadora flanqueada en su extremo 5' por uno cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), por lo que la nucleasa escinde el ARN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador.

35 Concepto 73C. Un método para escindir un ADN bicatenario (ADNds), que comprende poner en contacto dicho ADNds con el complejo del Concepto 68C o 68D, en donde el complejo comprende una nucleasa, en donde el ADNds comprende una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que la nucleasa escinde el ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador.

Alternativamente, el

- PAM puede ser AAN, ANG, NAG; o
- PAM puede ser AAG sin cambio de nucleótido o con un solo cambio de nucleótido.

45 La nucleasa puede ser cualquier nucleasa descrita en la presente memoria (como una nucleasa I-TevI como se describe en otra parte de la presente memoria). La nucleasa puede ser una nucleasa de ADN. La nucleasa puede ser una nucleasa de ARN. La nucleasa puede ser una TALEN. La nucleasa puede ser una proteína con dedos de zinc. La nucleasa puede ser una nucleasa Cas de cualquier sistema de origen natural. La nucleasa puede ser una nucleasa Cas sintética.

50 Concepto 74A. Un método para escindir un ADN monocatenario (ADNss), que comprende poner en contacto dicho ADNss con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo comprende una proteína de fusión en donde Py comprende una nucleasa, en donde el ADNss comprende una secuencia protoespaciadora flanqueada en su extremo 5' por cualquiera de los (PAM) descritos en

cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), por lo que la nickasa escinde una única cadena del ADNss en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador.

Concepto 74B. Un método para escindir una única cadena en un ADN bicatenario (ADNds), que comprende poner en contacto dicho ADNds con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo comprende una proteína de fusión en donde Py comprende una nickasa, en donde el ADNds comprende una secuencia protoespaciadora flanqueada en su extremo 5' por cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), por lo que la nickasa escinde una única cadena del ADNds en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador.

Concepto 74C. Un método para escindir un ADN monocatenario (ADNss), que comprende poner en contacto dicho ADN con el complejo del Concepto 68C o 68D, en donde el complejo comprende una nickasa, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que la nickasa escinde una única cadena del ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador.

La nickasa puede ser cualquier nucleasa descrita en la presente memoria. La nickasa puede ser una nickasa de ADN. La nickasa puede seleccionarse de Nt.BstNBI, Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nb.BbvCI, Nt.BbvCI y Nb.BsmI (disponibles en New England BioLabs). La nickasa es una nickasa Cas sintética (por ejemplo, nCas9 o Cas9D10A).

Concepto 75A. Un método para marcar o identificar una región de un ADN, que comprende poner en contacto dicho ADN con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el ADN comprende una secuencia protoespaciadora flanqueada en su extremo 5' por uno cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, y opcionalmente en donde el complejo comprende una proteína de fusión en donde Py comprende un marcador detectable.

Concepto 75B. Un método para marcar o identificar una región de un ADN, que comprende poner en contacto dicho ADN con el complejo del Concepto 68C o 68D, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, y opcionalmente en donde el complejo comprende un marcador detectable.

Concepto 76A. Un método para modificar la transcripción de una región de un ADN, que comprende poner en contacto dicho ADN con cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el ADN comprende una secuencia protoespaciadora flanqueada en su extremo 5' por uno cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el complejo regula al alza o a la baja la transcripción de la región de ADN o de un gen adyacente.

Concepto 76B. Un método para modificar la transcripción de una región de un ADN, que comprende poner en contacto dicho ADN con el complejo del Concepto 68C o 68D, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el complejo regula al alza o a la baja la transcripción de la región de ADN o de un gen adyacente.

En un ejemplo, el complejo regula a la baja la transcripción. La regulación a la baja puede ser de al menos aproximadamente el 50 % en comparación con la transcripción sin el complejo presente. En un ejemplo, la regulación a la baja es de al menos aproximadamente el 60 %, 70 %, 80 % o 90 %. En otro ejemplo, la regulación a la baja es de al menos aproximadamente el 95 %, 96 %, 97 %, 98 % o 99 %. En otro ejemplo, la regulación a la baja es del 100 %, es decir, la transcripción de la región de ADN se bloquea por completo.

En algunos ejemplos, la regulación a la baja puede ser deseable cuando un gen está sobreexpresado y es causante de un fenotipo negativo, pero la eliminación completa de la proteína sería perjudicial. La regulación a la baja de un gen diana puede ser útil para generar y estudiar fenotipos de inactivación donde la generación de una inactivación completa sería letal para la célula o el organismo.

En algunos ejemplos, el bloqueo de la transcripción de un gen esencial en una bacteria patógena se puede utilizar para matar a la bacteria patógena sin afectar a otras bacterias beneficiosas. Del mismo modo, las bacterias patógenas se pueden matar bloqueando la transcripción del origen de replicación. El bloqueo de genes no esenciales en una célula u organismo es útil para estudiar inactivaciones y el o los comportamientos fenotípicos.

Concepto 77A. Un método para modificar un ADNds diana de una célula sin introducir roturas en el ADNds, comprendiendo el método producir en la célula cualquiera de los complejos descritos en la presente memoria (por ejemplo, con el complejo del Concepto 68), en donde el complejo se dirige a un ADNds diana contenido en la célula, comprendiendo el ADNds diana una secuencia protoespaciadora flanqueada en su extremo 5' por uno cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), por lo que el complejo se une al ADNds en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el ADNds se modifica sin introducir roturas en el ADNds.

Concepto 77B. Un método para modificar un ADNds diana de una célula sin introducir roturas en el ADNds, comprendiendo el método producir en la célula el complejo del Concepto 68C o 68D, en donde el complejo se dirige a un ADN diana contenido en la célula, comprendiendo el ADN diana una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el ADN se modifica sin introducir roturas en el ADN.

Concepto 78. El método del Concepto 77, en donde el método edita el ADN, tal como donde la edición inserta, deleta o sustituye una secuencia de bases o ácidos nucleicos en el ADN.

La modificación puede ser la introducción de una o más sustituciones cuando el complejo es un complejo del Concepto 68B y Py es un editor de bases o un *prime editor* (como cualquiera de los editores de bases o *prime editors* descritos en otra parte de la presente memoria). La modificación puede ser la introducción de una o más inserciones y/o deleciones (INDELS) cuando el complejo es un complejo del Concepto 68B y Py es una nucleasa o nickasa y la célula es una célula que puede reparar la rotura en el ADNds o la rotura en el ADNss a través del mecanismo NHEJ.

Concepto 79A. Un método para inhibir el crecimiento o la proliferación de una célula sin introducir roturas letales en el ADNds, comprendiendo el método producir en la célula cualquiera de los complejos descritos en la presente memoria (por ejemplo, un complejo del Concepto 68), en donde el complejo se dirige a un ADN diana contenido en la célula, comprendiendo el ADN diana una secuencia protoespaciadora flanqueada en su extremo 5' por uno cualquiera de los (PAM) descritos en cualquier opción o característica de la presente memoria, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que se inhibe la replicación del ADN sin introducir roturas letales en el ADNds en el ADN, inhibiendo así el crecimiento o la proliferación de una célula.

Concepto 79B. Un método para inhibir el crecimiento o la proliferación de una célula sin introducir roturas letales en el ADNds, comprendiendo el método producir en la célula el complejo del Concepto 68C o 68D, en donde el complejo se dirige a un ADN diana contenido en la célula, comprendiendo el ADN diana una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que se inhibe la replicación del ADN sin introducir roturas en el ADN, inhibiendo así el crecimiento o la proliferación de una célula.

El crecimiento o la proliferación de una célula se puede inhibir mediante la introducción de una o más inserciones y/o deleciones (INDELS) cuando el complejo es un complejo del Concepto 68B y Py es una nickasa que introduce roturas monocatenarias en el ADN y la célula es una célula que es capaz de reparar la rotura en el ADNds o la rotura en el ADNss a través del mecanismo NHEJ. El crecimiento o la proliferación de una célula se puede inhibir mediante la regulación a la baja o el bloqueo de la transcripción de un gen esencial en la célula cuando el complejo es un complejo del Concepto 68A. Los expertos en la técnica conocen bien los genes esenciales. El crecimiento o la proliferación de una célula se puede inhibir mediante la regulación a la baja o el bloqueo de la transcripción de un origen cromosómico de replicación en la célula cuando el complejo es un complejo del Concepto 68A.

Concepto 80. Un método para tratar o prevenir una enfermedad o afección en un sujeto humano, animal, vegetal o fúngico, comprendiendo el método llevar a cabo el método de uno cualquiera de los Conceptos 69 a 79, en donde las células del sujeto comprenden dicho ADN o ARN y las células median la enfermedad o afección.

Concepto 81A. El método del Concepto 81, en donde el complejo modifica una secuencia diana codificante o una secuencia diana no codificante.

En un ejemplo, se modifica una cadena de ADN codificante, pero no una cadena de ADN no codificante. En otro ejemplo, se modifica una cadena de ADN no codificante, pero no una cadena de ADN codificante.

Concepto 81B. El método de uno cualquiera de los Conceptos 69 a 80, en donde el complejo modifica (por ejemplo, corta, edita, bloquea, marca o etiqueta) una secuencia diana codificante o una secuencia diana no codificante, opcionalmente en donde (i) se modifica una cadena codificante, pero no una cadena no codificante de ADN o (ii) se modifica una cadena no codificante, pero no una cadena codificante de ADN.

Concepto 82. El método del Concepto 81, en donde la modificación corta, edita, bloquea, marca o etiqueta la secuencia diana.

- Cuando el complejo es un complejo del Concepto 68A, la modificación puede ser de bloqueo. Cuando el complejo es un complejo del Concepto 68A, la modificación puede ser una regulación a la baja de la transcripción de una secuencia diana codificante. Cuando el complejo es un complejo del Concepto 68B en donde Py es una nucleasa o nickasa, la modificación puede ser un corte de una secuencia diana codificante. Cuando el complejo es un complejo del Concepto 68B en donde Py es un editor de bases o *prime editor*, la modificación puede ser una edición de una secuencia diana codificante. El editor de bases puede introducir una o más sustituciones en la secuencia diana codificante.
- Concepto 83A. Uno o más vectores de ácido nucleico o uno o más ácidos nucleicos que comprenden al menos una secuencia de nucleótidos seleccionada de las SEQ ID NO:7-11, en donde dicha secuencia de nucleótidos está conectada operativamente a un promotor heterólogo, promotor sintético, promotor eucariota o promotor no bacteriano.
- Concepto 83B. Uno o más vectores de ácido nucleico o uno o más ácidos nucleicos que comprenden al menos una secuencia de nucleótidos seleccionada de las SEQ ID NO:7-11, en donde dicha secuencia de nucleótidos está contenida en una célula que es una célula eucariota, una célula no bacteriana o una célula que no es una célula bacteriana (por ejemplo, una célula de *E. coli*, *Pseudomonas* o *Klebsiella*) que comprende una secuencia de nucleótidos endógena que comprende las SEQ ID NO:7-11.
- Concepto 83C. Uno o más vectores de ácido nucleico o uno o más ácidos nucleicos que comprenden al menos una secuencia de nucleótidos seleccionada de las SEQ ID NO:7-11, en donde dicha secuencia de nucleótidos está conectada operativamente a un promotor heterólogo, promotor sintético, promotor eucariota o promotor no bacteriano y dicha secuencia de nucleótidos está contenida en una célula que es una célula eucariota, una célula no bacteriana o una célula que no es una célula bacteriana (por ejemplo, una célula de *E. coli*, *Pseudomonas* o *Klebsiella*) que comprende una secuencia de nucleótidos endógena que comprende las SEQ ID NO:7-11.
- Concepto 84. Un vector de ácido nucleico o un ácido nucleico que comprende la SEQ ID NO:12 o una secuencia de ADN que es al menos (aproximadamente) un 80 % idéntica a la SEQ ID NO:12.
- Concepto 85. Uno o más ácidos nucleicos que codifican una pluralidad de proteínas Cas y que comprenden al menos una secuencia de nucleótidos para producir un ARNcr, en donde las proteínas Cas y el ARN son capaces de formar un complejo CRISPR/Cas de ribonucleoproteína en donde el ARN es capaz de guiar el complejo a un protoespaciador contenido en un ADN diana, en donde el extremo 5' del protoespaciador está flanqueado por uno cualquiera de los restos adyacentes al protoespaciador (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tienen una de las secuencias citadas en el Concepto 39), en donde
- el complejo carece de una nucleasa de ADN y es capaz de modificar el ADN sin introducir una rotura de doble cadena en el ADN;
  - las proteínas Cas no comprenden DinG, Cas3 y Cas10; y
  - el complejo no comprende todas las proteínas Cas de un complejo CRISPR/Cas Tipo I, II, III, IV, V o VI.
- Concepto 86. El o los ácidos nucleicos del Concepto 85, en donde el ácido nucleico está contenido en un vector o los ácidos nucleicos están contenidos en uno o más vectores.
- Concepto 87. El o los ácidos nucleicos del Concepto 85 o 86, en donde el ADN está contenido en un cromosoma.
- Concepto 88. Un complejo CRISPR/Cas de ribonucleoproteína que comprende una pluralidad de proteínas Cas y un ARNcr, en donde el ARN es capaz de guiar el complejo a un protoespaciador contenido en un ADN diana, en donde el extremo 5' del protoespaciador está flanqueado por uno cualquiera de los restos adyacentes al protoespaciador (PAM) descritos en cualquier opción o característica de la presente memoria (por ejemplo, que tiene una de las secuencias citadas en el Concepto 39), en donde
- el complejo carece de una nucleasa de ADN y es capaz de modificar el ADN sin introducir una rotura de doble cadena en el ADN;
  - el complejo está desprovisto de una Cas 3 y una Cas 10; y
  - el complejo no comprende todas las proteínas Cas de un complejo CRISPR/Cas Tipo I, II, III, IV, V o VI.
- Concepto 89. El o los ácidos nucleicos de uno cualquiera de los Conceptos 85-87 o el complejo del Concepto 88, en donde el ARNcr no es un ARNcr Tipo I, II, III, V o VI, respectivamente.
- Concepto 90. El o los ácidos nucleicos o complejo de uno cualquiera de los Conceptos 85 a 89, en donde el complejo es capaz de modificar (i) una cadena codificante, pero no una cadena no codificante de ADN o (ii) una cadena no codificante, pero no una cadena codificante de ADN.
- Concepto 91. El o los ácidos nucleicos o complejo de uno cualquiera de los Conceptos 85 a 90, en donde el complejo modifica (por ejemplo, corta, edita, bloquea, marca o etiqueta) una secuencia diana codificante o una secuencia diana no codificante del ADN.

Concepto 92. El o los ácidos nucleicos o complejo de cualquiera de los Conceptos 85 a 91, en donde la pluralidad de proteínas Cas comprende una Cas-S1, S2, S3, S4 y S5.

Por ejemplo, la pluralidad de proteínas Cas comprende Cas-S1.1, S2.1, S3.1, S4.1 y S5.1.

5 Concepto 93. El o los ácidos nucleicos o complejo de cualquiera de los Conceptos 85 a 92, en donde el complejo comprende además un dominio de proteína efectora para modificar ADN, opcionalmente en donde el dominio efector comprende actividad de nucleasa, actividad de nickasa, actividad de recombinasa, transcriptasa inversa, helicasa, actividad de desaminasa, actividad de metiltransferasa, actividad de metilasa, actividad de acetilasa, actividad de acetiltransferasa, actividad de activación transcripcional o actividad de represión transcripcional.

En un ejemplo, el complejo comprende un componente seleccionado de

- 10 a) una nucleasa que corta el sitio diana;
- b) una desaminasa (como una citosina o adenina desaminasa) que lleva a cabo la desaminación del sitio diana;
- c) un editor de bases que edita bases del sitio diana;
- d) un *prime editor* que produce *prime editing* del sitio diana;
- 15 e) una transcriptasa inversa que transcribe de forma inversa el sitio diana o que transcribe de forma inversa un ARN en la célula para producir un ADN que se inserta en el sitio diana;
- f) una metiltransferasa que metila el ADN en la célula;
- g) una metilasa que metila el ADN en la célula;
- h) una acetilasa que acetila el ADN en la célula;
- 20 i) una acetiltransferasa que somete el ADN de la célula a la actividad acetiltransferasa;
- j) un activador de la transcripción que activa la transcripción de un gen en la célula, tal como un gen que comprende o es adyacente al sitio diana; y
- k) un desactivador de la transcripción que desactiva la transcripción de un gen en la célula, tal como un gen que comprende o es adyacente al sitio diana.

25 Concepto 94. Una célula que comprende el o los ácidos nucleicos o complejo de uno cualquiera de los Conceptos 85 a 93, en donde el ADN está contenido en un cromosoma de la célula.

Concepto 95. Una célula que comprende el o los ácidos nucleicos o complejo de uno cualquiera de los Conceptos 85 a 93, en donde el ADN está contenido en un plásmido en la célula.

30 Concepto 96. La célula del Concepto 95, en donde el complejo es capaz de inhibir el crecimiento o proliferación de la célula.

Concepto 97. Un método para modificar un ADN, comprendiendo el método

- a) poner en contacto el ADN con el o los ácidos nucleicos definidos en uno cualquiera de los Conceptos 85 a 87 y 89 a 93;
- 35 b) permitir la formación de un complejo de ribonucleoproteína que comprende la(s) proteína(s) Cas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en el ADN para modificar el ADN.

Concepto 98. Un método para inhibir la replicación de un plásmido que comprende ADN, comprendiendo el método

- a) poner en contacto el ADN con el o los ácidos nucleicos definidos en uno cualquiera de los Conceptos 85 a 87 y 89 a 93;
- 40 b) permitir la formación de un complejo de ribonucleoproteína que comprende la(s) proteína(s) Cas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en el ADN para inhibir la replicación del plásmido.

Concepto 99. Un método para inhibir la transcripción de una secuencia de nucleótidos contenida en un ADN, comprendiendo el método

- 45 a) poner en contacto el ADN con el o los ácidos nucleicos definidos en uno cualquiera de los Conceptos 85 a 87 y 89 a 93;

- b) permitir la formación de un complejo de ribonucleoproteína que comprende la(s) proteína(s) Cas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en la secuencia de nucleótidos para inhibir la transcripción del mismo.

5 Concepto 100. Un método para inhibir el crecimiento o la proliferación de una célula (opcionalmente, una célula procariota, tal como una célula bacteriana) que comprende ADN, comprendiendo el método

- a) poner en contacto el ADN con el o los ácidos nucleicos definidos en uno cualquiera de los Conceptos 85 a 87 y 89 a 93;

10 b) permitir la formación de un complejo de ribonucleoproteína que comprende la(s) proteína(s) Cas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en el ADN para inhibir el crecimiento o la proliferación de la célula.

Concepto 101. Un método para tratar o prevenir una enfermedad o afección que está mediada por células diana en un sujeto humano o animal, comprendiendo el método llevar a cabo el método de uno cualquiera de los Conceptos 97 a 100 para modificar las células diana, en donde dicho contacto comprende administrar el o los ácidos nucleicos al sujeto y en donde la modificación trata o previene la enfermedad o afección.

15 Concepto 102. El método del Concepto 101, en donde las células diana son células del sujeto que comprenden un defecto en el ácido nucleico y la modificación corrige el defecto.

Concepto 103. El método del Concepto 101 o 102, en donde la modificación

20 a) añade una nueva funcionalidad a las células, opcionalmente la modificación regula al alza o regula a la baja la expresión de un gen en las células o añade una nueva secuencia de nucleótidos a los genomas de las células para la expresión de una proteína codificada por la secuencia; o

- b) modula la expresión de una secuencia de nucleótidos contenida en un cromosoma o episoma (opcionalmente, un plásmido) de la célula; o

- c) inhibe el crecimiento o proliferación de las células.

25 La célula puede ser una célula bacteriana de cualquier género o especie de la Tabla 3. El método puede modificar una célula de cualquier género o especie de la Tabla 3. El ADN o polinucleótido puede estar contenido en una célula de cualquier género o especie de la Tabla 3. El método puede matar o inhibir el crecimiento o la proliferación de una célula de cualquier género o especie de la Tabla 3. La célula puede ser una célula de un género o especie que sea diferente de uno de los géneros o especies, respectivamente, de la Tabla 3, por ejemplo, la célula no es una célula de *E. coli*, o la célula no es una célula de *Klebsiella*. La célula puede ser una arquea, como un metanógeno. El método 30 puede matar o inhibir el crecimiento o la proliferación de la arquea. El método puede editar el genoma de la célula bacteriana o arquea. Cualquier célula procariota (por ejemplo, célula bacteriana o arquea) o célula fúngica de la presente memoria puede estar contenida en un microbioma, como un microbioma de un ser humano, animal, planta o entorno (como el suelo o una vía fluvial). Por ejemplo, el microbioma está contenido en el tracto gastrointestinal (tracto GI) de un ser humano o animal. Por ejemplo, el microbioma está contenido en el sistema urinario de un ser humano o animal, por ejemplo, contenido en una vejiga, un riñón o una uretra. Por ejemplo, el microbioma está contenido en la sangre de un ser humano o animal. Por ejemplo, el microbioma está contenido en el ojo o nariz u oído de un ser humano o animal. Por ejemplo, el microbioma está contenido en el cabello de un ser humano o animal. Por ejemplo, el microbioma está contenido en la piel de un ser humano o animal. Por ejemplo, el microbioma está contenido en una hoja, raíz, tallo o semilla de una planta.

40

Una enfermedad o afección de la presente memoria puede seleccionarse de cualquiera de las siguientes:

- Una enfermedad o afección neurodegenerativa
- Una enfermedad o afección del cerebro;
- Una enfermedad o afección del SNC;
- Pérdida o deterioro de la memoria;
- Una enfermedad o afección cardíaca o cardiovascular, por ejemplo, ataque cardíaco, accidente cerebrovascular o fibrilación auricular
- Una enfermedad o afección del hígado;
- Una enfermedad o afección renal, por ejemplo, enfermedad renal crónica (CKD);
- Una enfermedad o afección del páncreas;
- Una enfermedad o afección pulmonar, por ejemplo, fibrosis quística o EPOC;
- Una enfermedad o afección gastrointestinal;
- Una enfermedad o afección de la garganta o de la cavidad oral;
- Una enfermedad o afección ocular;
- Una enfermedad o afección genital, por ejemplo, una enfermedad o afección vaginal, labial, del pene o escrotal;
- Una enfermedad o afección transmitida por vía sexual, por ejemplo, gonorrea, infección por VIH, sífilis o infección por Clamidia;
- Una enfermedad o afección del oído;
- Una enfermedad o afección de la piel;
- Neovascularización (por ejemplo, de un tumor o en el ojo);
- Enfriamiento común;
- Epilepsia;
- Fibrosis, por ejemplo, fibrosis hepática o pulmonar;
- Una enfermedad o afección fúngica, por ejemplo, candidiasis;
- Una enfermedad o afección metabólica, por ejemplo, obesidad,
- Una enfermedad o afección inflamatoria, por ejemplo, artritis reumatoide, psoriasis, eccema, asma, colitis ulcerosa, colitis, enfermedad de Crohn o IBD;
- Autismo;
- TDAH;
- Trastorno bipolar;
- ELA [Esclerosis lateral amiotrófica ];
- Osteoartritis;
- Un defecto o afección congénito o del desarrollo;
- Aborto espontáneo;
- Una afección de coagulación de la sangre;
- Bronquitis;
- AMD seca o húmeda;
- Enfermedad de Kawasaki;
- Enfermedad de Lyme;
- Una alergia, por ejemplo, una alergia a nuez, hierba, polen, polvo, ácaro, pelo o caspa de gato o perro;
- Malaria, fiebre tifoidea, tuberculosis o cólera;
- Depresión;
- Retraso mental;
- Microcefalia;
- Malnutrición;
- Conjuntivitis;
- Neumonía;
- Embolia pulmonar;

- anorexia, diabetes, diabetes de Tipo I o Tipo II.
- Úlcera(s), por ejemplo, ulceración gástrica o ulceración de la piel;
- Piel seca;
- Síndrome de Sjögren;
- Tormenta de citoquinas;
- Sordera, pérdida o deterioro auditivo;
- Metabolismo lento o rápido (es decir, más lento o más rápido que el promedio para el peso, el sexo y edad del sujeto);
- Trastorno de la concepción, por ejemplo, infertilidad o baja fertilidad;
- Ictericia;
- Erupción cutánea;
- Una enfermedad o afección del corazón;
- Una enfermedad o afección nasal;
- Una enfermedad o afección hematológica, por ejemplo, anemia, por ejemplo, anemia de enfermedad crónica o cáncer;
- Una infección viral;
- Una infección por bacterias patógenas;
- Un cáncer;
- Una enfermedad o afección autoinmune, por ejemplo, SLE;
- Hipertensión pulmonar;
- Un trastorno óseo;
- Sepsis o shock séptico;
- Sinusitis;
- Estrés (por ejemplo, estrés laboral);
- Talasemia, anemia, enfermedad de von Willebrand, o hemofilia;
- Herpes zóster o herpes labial;
- Menstruación; y
- Recuento bajo de espermatozoides.

#### Enfermedades o afecciones neurodegenerativas o del SNC para tratamiento o prevención mediante el método

En un ejemplo, la enfermedad o afección neurodegenerativa o del SNC se selecciona del grupo que consiste en enfermedad de Alzheimer, geriopsicosis, síndrome de Down, enfermedad de Parkinson, enfermedad de Creutzfeldt-Jakob, neuropatía diabética, síndrome de Parkinson, enfermedad de Huntington, enfermedad de Machado-Joseph, esclerosis lateral amiotrófica, neuropatía diabética y enfermedad de Creutzfeldt-Jakob. Por ejemplo, la enfermedad es enfermedad de Alzheimer. Por ejemplo, la enfermedad es síndrome de Parkinson.

En un ejemplo, cualquiera de los métodos descritos en la presente memoria se practica en un sujeto humano o animal para tratar una enfermedad o afección del SNC o neurodegenerativa, el método provoca la regulación a la baja de las células Treg en el sujeto, promoviendo así la entrada de macrófagos derivados de monocitos sistémicos y/o células Treg a través del plexo coroideo hacia el cerebro del sujeto, por lo que la enfermedad o afección (por ejemplo, la enfermedad de Alzheimer) se trata, se previene o se reduce su progresión. En un ejemplo, el método provoca un aumento de IFN-gamma en el sistema SNC (por ejemplo, en el cerebro y/o LCR) del sujeto. En un ejemplo, el método restaura la fibra nerviosa y/o reduce la progresión del daño de la fibra nerviosa. En un ejemplo, el método restaura la mielina nerviosa y/o reduce la progresión del daño de la mielina nerviosa. En un ejemplo, cualquiera de los métodos descritos en la presente memoria trata o previene una enfermedad o afección divulgada en W02015/136541 y/o el método se puede utilizar con cualquier método divulgado en W02015/136541 (en particular, en relación con su divulgación de dichos métodos, enfermedades, afecciones y agentes terapéuticos potenciales que se pueden administrar al sujeto para efectuar el tratamiento y/o la prevención de enfermedades y condiciones del SNC y neurodegenerativas, por ejemplo, agentes tales como inhibidores de puntos de control inmunes, por ejemplo, anti-PD-1, anti-PD-L1, anti-TIM3 u otros anticuerpos divulgados en el mismo).

### **Cánceres para tratamiento o prevención mediante el método**

Los cánceres que pueden tratarse incluyen tumores que no están vascularizados, o que no están sustancialmente vascularizados, así como tumores vascularizados. Los cánceres pueden comprender tumores no sólidos (tales como tumores hematológicos, por ejemplo, leucemias y linfomas) o pueden comprender tumores sólidos. Los tipos de cánceres que se van a tratar mediante los métodos, proteína(s) (incluyendo proteína(s) de fusión y complejos de proteínas de fusión), vector(es), complejos y composición(es) descritas en la presente memoria incluyen, pero no se limitan a, carcinoma, blastoma y sarcoma, y ciertas malignidades de leucemia o linfoides, tumores benignos y malignos y malignidades, por ejemplo, sarcomas, carcinomas y melanomas. También se incluyen tumores/cánceres en adultos y tumores/cánceres pediátricos.

- 5
- 10
- 15

Los tumores sólidos son masas anormales de tejido que, por lo general, no contienen quistes ni áreas líquidas. Los tumores sólidos pueden ser benignos o malignos. Los distintos tipos de tumores sólidos reciben su nombre según el tipo de células que los forman (como sarcomas, carcinomas y linfomas). Los ejemplos de tumores sólidos, como sarcomas y carcinomas, incluyen fibrosarcoma, mixosarcoma, liposarcoma, condrosarcoma, osteosarcoma y otros sarcomas, sinovioma, mesotelioma, tumor de Ewing, leiomiomasarcoma, rhabdomyosarcoma, carcinoma de colon, malignidad linfóide, cáncer de páncreas, cáncer de mama, cánceres de pulmón, cáncer de ovario, cáncer de próstata, carcinoma hepatocelular, carcinoma de células escamosas, carcinoma de células basales, adenocarcinoma, carcinoma de glándulas sudoríparas, carcinoma medular de tiroides, carcinoma papilar de tiroides, feocromocitomas, carcinoma de glándulas sebáceas, carcinoma papilar, adenocarcinomas papilares, carcinoma medular, carcinoma broncogénico, carcinoma de células renales, hepatoma, carcinoma del conducto biliar, coriocarcinoma, tumor de Wilms, cáncer de cuello uterino, tumor testicular, seminoma, carcinoma de vejiga, melanoma y tumores del SNC (como glioma (como glioma del tronco encefálico y gliomas mixtos), glioblastoma (también conocido como glioblastoma multiforme), astrocitoma, linfoma del SNC, germinoma, meduloblastoma, Schwannoma, craneofaringioma, ependimoma, pineaioma, hemangioblastoma, neurinoma acústico, oligodendroglioma, menangioma, neuroblastoma, retinoblastoma y metástasis cerebrales).

- 20
- 25
- 30

### **Enfermedades autoinmunes para tratamiento o prevención mediante el método**

## ES 3 014 020 T3

|    |                                                              |    |                                                                                           |
|----|--------------------------------------------------------------|----|-------------------------------------------------------------------------------------------|
|    | • Encefalomiелitis diseminada aguda (ADEM)                   | 35 | • Ostiomielitis multifocal crónica recurrente (CRMO)                                      |
|    | • Leucoencefalitis hemorrágica necrotizante aguda            |    | • Síndrome de Churg-Strauss                                                               |
|    | • Enfermedad de Addison                                      |    | • Penfigoide cicatricial/penfigoide benigno mucosal                                       |
|    | • Agammaglobulinemia                                         |    | • Enfermedad de Crohn enfermedad                                                          |
| 5  | • Alopecia areata                                            |    | • Síndrome de Cogan                                                                       |
|    | • Amilosis                                                   | 40 | • Enfermedad de aglutinina fría                                                           |
|    | • Espondilitis anquilosante                                  |    | • Bloqueo cardíaco congénito                                                              |
|    | • Nefritis anti-GBM/anti-TBM                                 |    | • Miocarditis por Coxsackie                                                               |
|    | • Síndrome antifosfolípido (APS)                             |    | • Enfermedad CREST                                                                        |
| 10 | • Angioedema autoinmune                                      |    | • Crioglobulinemia mixta esencial                                                         |
|    | • Anemia aplásica autoinmune                                 | 45 | • Neuropatías desmielinizantes                                                            |
|    | • Disautonomía autoinmune                                    |    | • Dermatitis herpetiforme                                                                 |
|    | • Hepatitis autoinmune                                       |    | • Dermatomiositis                                                                         |
|    | • Hiperlipidemia autoinmune                                  |    | • Enfermedad de Devic (neuromielitis óptica)                                              |
| 15 | • Inmunodeficiencia autoinmune                               |    | • Lupus discoide                                                                          |
|    | • Enfermedad autoinmune del oído interno (AIED)              | 50 | • Síndrome de Dressler                                                                    |
|    | • Miocarditis autoinmune                                     |    | • Endometriosis                                                                           |
|    | • Ooforitis autoinmune                                       |    | • Esofagitis eosinofílica                                                                 |
|    | • Pancreatitis autoinmune                                    |    | • Fascitis eosinofílica                                                                   |
| 20 | • Retinopatía autoinmune                                     |    | • Eritema nudoso                                                                          |
|    | • Púrpura trombocitopénica autoinmune (ATP)                  | 55 | • Encefalomiелitis alérgica experimental                                                  |
|    | • Enfermedad del tiroides autoinmune                         |    | • Síndrome de Evans                                                                       |
|    | • Urticaria autoinmune                                       |    | • Fibromialgia                                                                            |
|    | • Neuropatías axónicas y neuronales                          |    | • Alveolitis fibrosante                                                                   |
| 25 | • Enfermedad de Balo                                         |    | • Arteritis de células gigantes (arteritis temporal)                                      |
|    | • Enfermedad de Behcet                                       | 60 | • Miocarditis de célula gigante                                                           |
|    | • Penfigoide ampolloso                                       |    | • Glomerulonefritis                                                                       |
|    | • Miocardiopatía                                             |    | • Síndrome de Goodpasture                                                                 |
|    | • Enfermedad de Castleman                                    |    | • Granulomatosis con Poliangeítis (GPA) (anteriormente llamada Granulomatosis de Wegener) |
| 30 | • Enfermedad de celiaquía                                    | 65 | • Enfermedad de Graves                                                                    |
|    | • Enfermedad de Chagas                                       |    | • Síndrome de Guillain-Barre                                                              |
|    | • Síndrome de fatiga crónica                                 |    | • Encefalitis de Hashimoto                                                                |
|    | • Polineuropatía desmielinizante inflamatoria crónica (CIPD) |    | • Tiroiditis de Hashimoto                                                                 |

|    |                                                |    |                                                                                             |
|----|------------------------------------------------|----|---------------------------------------------------------------------------------------------|
|    | • Anemia hemolítica                            | 35 | • Neuritis óptica                                                                           |
|    | • Púrpura de Henoch- Schonlein                 |    | • Reumatismo palindrómico                                                                   |
|    | • Herpes gestacional                           |    | • PANDAS (Trastornos neuropsiquiátricos pediátricos autoinmunes asociados con Streptococcus |
|    | • Hipogammaglobulinemia                        |    | • Degeneración cerebelosa paraneoplásica                                                    |
| 5  | • Púrpura idiopática trombocitopénica (ITP)    | 40 | • Hemoglobinuria nocturna paroxismal (PNH)                                                  |
|    | • Nefropatía por IgA                           |    | • Síndrome de Parry Romberg                                                                 |
|    | • Enfermedad esclerosante relacionada con IgG4 |    | • Síndrome de Parsonnage-Turner                                                             |
|    | • Lipoproteínas inmunorreguladoras             |    | • Pars planitis (uveitis periférica)                                                        |
|    | • Miositis por cuerpos de inclusión            |    | • Pénfigo                                                                                   |
| 10 | • Cistitis intersticial                        | 45 | • Neuropatía periférica                                                                     |
|    | • Artritis Juvenil                             |    | • Encefalomielitis perivenosa                                                               |
|    | • Diabetes juvenil (diabetes Tipo 1)           |    | • Anemia perniciosa                                                                         |
|    | • Miositis juvenil                             |    | • Síndrome de POEMS                                                                         |
|    | • Síndrome de Kawasaki                         |    | • Poliarteritis nodosa                                                                      |
| 15 | • Síndrome de Lambert-Eaton                    | 50 | • Síndromes poliglandulares autoinmunes Tipo I, II y III                                    |
|    | • Vasculitis leucocitoclástica                 |    | • Polimialgia reumática                                                                     |
|    | • Liquen plano                                 |    | • Polimiositis                                                                              |
|    | • Liquen escleroso                             |    | • Síndrome post infarto de miocardio                                                        |
|    | • Conjuntivitis leñosa                         |    | • síndrome de postpericardiotomía                                                           |
| 20 | • Enfermedad por IgA lineal (LAD)              | 55 | • Dermatitis por progesterona                                                               |
|    | • Lupus (SLE)                                  |    | • Cirrosis biliar primaria                                                                  |
|    | • Enfermedad de Lyme, crónica                  |    | • Colangitis esclerosante primaria                                                          |
|    | • Enfermedad de Ménière                        |    | • Psoriasis                                                                                 |
|    | • Poliangitis microscópica                     |    | • Artritis psoriásica                                                                       |
| 25 | • Enfermedad mixta del tejido conectivo (MCTD) | 60 | • Fibrosis pulmonar idiopática                                                              |
|    | • Úlcera de Mooren                             |    | • Pioderma gangrenoso                                                                       |
|    | • Enfermedad de Mucha-Habermann                |    | • Aplasia pura de glóbulos rojos                                                            |
|    | • Esclerosis múltiple                          |    | • Fenómeno de Raynaud                                                                       |
|    | • Miastenia grave                              |    | • Artritis reactiva                                                                         |
| 30 | • Miositis                                     | 65 | • Distrofia simpática refleja                                                               |
|    | • Narcolepsia                                  |    | • Síndrome de Reiter                                                                        |
|    | • Neuromielitis óptica (de Devic)              |    | • Policondritis recidivante                                                                 |
|    | • Neutropenia                                  |    | • Síndrome de piernas inquietas                                                             |
|    | • Penfigoide cicatricial ocular                |    |                                                                                             |

- |                                                                                       |    |                                                                                     |
|---------------------------------------------------------------------------------------|----|-------------------------------------------------------------------------------------|
| • Fibrosis retroperitoneal                                                            | 15 | • Arteritis temporal/arteritis de células gigantes                                  |
| • Fiebre reumática                                                                    |    | • Púrpura trombocitopénica (TTP)                                                    |
| • Artritis reumatoide                                                                 |    | • Síndrome de Tolosa-Hunt                                                           |
| • Sarcoidosis                                                                         |    | • Mielitis transversal                                                              |
| 5 • Síndrome de Schmidt                                                               |    | • Diabetes Tipo I                                                                   |
| • Escleritis                                                                          | 20 | • Colitis ulcerosa                                                                  |
| • Escleroderma                                                                        |    | • Enfermedad del tejido conectivo no diferenciado (UCTD)                            |
| • Síndrome de Sjogren                                                                 |    | • Uveitis                                                                           |
| • Autoinmunidad del espermato y testicular                                            |    | • Vasculitis                                                                        |
| 10 • Síndrome de la persona rígida                                                    | 25 | • Dermatitis vesiculobullosa                                                        |
| • Endocarditis bacteriana subaguda (SBE)                                              |    | • Vitiligo                                                                          |
| • Síndrome de Susac                                                                   |    | • Granulomatosis de Wegener (ahora denominada Granulomatosis con Poliangitis (GPA)) |
| • Oftalmia simpática                                                                  |    |                                                                                     |
| • Arteritis de Takayasu                                                               |    |                                                                                     |
| 30 <b>Enfermedades inflamatorias para tratamiento o prevención mediante el método</b> |    |                                                                                     |
| • Enfermedad de Alzheimer                                                             |    |                                                                                     |
| • espondilitis anquilosante                                                           |    |                                                                                     |
| 35 • artritis (osteoartritis, artritis reumatoide (AR), artritis psoriásica)          |    |                                                                                     |
| • asma                                                                                |    |                                                                                     |
| • aterosclerosis                                                                      |    |                                                                                     |
| • enfermedad de Crohn                                                                 |    |                                                                                     |
| • colitis                                                                             |    |                                                                                     |
| 40 • colitis ulcerosa.                                                                |    |                                                                                     |
| • dermatitis                                                                          |    |                                                                                     |
| • diverticulitis                                                                      |    |                                                                                     |
| • fibromialgia                                                                        |    |                                                                                     |
| • hepatitis                                                                           |    |                                                                                     |
| 45 • síndrome del intestino irritable (IBS)                                           |    |                                                                                     |
| • lupus eritematoso sistémico (SLE)                                                   |    |                                                                                     |
| • nefritis                                                                            |    |                                                                                     |
| • Enfermedad de Parkinson                                                             |    |                                                                                     |

Se proporcionan las siguientes configuraciones y aspectos adicionales de la divulgación:

**En una primera configuración**

En un primer aspecto: -

- Una proteína Cas-S1 o un ácido nucleico que codifica una Cas-S1.
- 5 Una proteína Cas-S2 o un ácido nucleico que codifica una Cas-S2.
- Una proteína Cas-S3 o un ácido nucleico que codifica una Cas-S3.
- Una proteína Cas-S4 o un ácido nucleico que codifica una Cas-S4.
- Una proteína Cas-S5 o un ácido nucleico que codifica una Cas-S5.
- Uno o más ácidos nucleicos que codifican una proteína Cas-S1, S2, S3, S4 y S5.
- 10 Uno o más ácidos nucleicos que codifican una proteína Cas-S4 y S5.
- Uno o más ácidos nucleicos que codifican una proteína Cas-S3, S4 y S5.
- Uno o más ácidos nucleicos que codifican una proteína Cas-S2, S3, S4 y S5.
- Uno o más ácidos nucleicos que codifican una proteína Cas-S1 y S2.
- Uno o más ácidos nucleicos que codifican una proteína Cas-S1, S2 y S3.
- 15 Uno o más ácidos nucleicos que codifican una proteína Cas-S1, S2, S3 y S4.
- Uno o más ácidos nucleicos que comprenden las SEQ ID NO:7-11.
- Uno o más ácidos nucleicos que comprenden las SEQ ID NO:8-11.
- Uno o más ácidos nucleicos que comprenden las SEQ ID NO:9-11.
- Uno o más ácidos nucleicos que comprenden las SEQ ID NO:10 y 11.
- 20 Uno o más ácidos nucleicos que comprenden las SEQ ID NO:7-10.
- Uno o más ácidos nucleicos que comprenden las SEQ ID NO:7-9.
- Uno o más ácidos nucleicos que comprenden las SEQ ID NO:7 y 8.
- Un ácido nucleico que comprende la SEQ ID NO:7.
- Un ácido nucleico que comprende la SEQ ID NO:8.
- 25 Un ácido nucleico que comprende la SEQ ID NO:9.
- Un ácido nucleico que comprende la SEQ ID NO:10.
- Un ácido nucleico que comprende la SEQ ID NO:11.

30 Cada una de dichas secuencias puede estar conectada operativamente a un promotor heterólogo (es decir, un promotor que no está conectado operativamente de forma natural a la secuencia). Cada una de dichas secuencias puede estar conectada operativamente a un promotor que es un promotor eucariota o un promotor de virus (por ejemplo, fago).

En un segundo aspecto:

- Un complejo de proteína o ribonucleoproteína que comprende 1, 2, 3, 4 o la totalidad de una proteína Cas-S1, S2, S3, S4 y S5.
- 35 Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S1, S2, S3, S4 y S5.
- Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S4 y S5.
- Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S3, S4 y S5.
- Un complejo de proteína o ribonucleoproteína que comprende una proteína Cas-S2, S3, S4 y S5.

**En una segunda configuración**

En un primer aspecto:

Un (i) vector de ácido nucleico o una pluralidad de vectores de ácido nucleico o (ii) un ácido nucleico o una pluralidad de ácidos nucleicos, que comprende secuencias de nucleótidos expresables, en donde las secuencias comprenden

- a) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO:1;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO:2;
- c) una tercera secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO:3;
- d) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO:4; y
- e) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 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 % idéntico a la SEQ ID NO:5.

También se proporciona un i) vector de ácido nucleico o una pluralidad de vectores de ácido nucleico o (ii) un ácido nucleico o una pluralidad de ácidos nucleicos, que comprenden secuencias de nucleótidos expresables, en donde las secuencias comprenden

- a) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos un 94, 95, 96, 97, 98 o 99 % idéntico a la SEQ ID NO:1; y/o
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos un 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 o 99 % idéntico a la SEQ ID NO:2; y/o
- c) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos un 98 o 99 % idéntico a la SEQ ID NO:3; y/o
- d) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos un 99 % idéntico a la SEQ ID NO:4; y
- e) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos un 98 o 99 % idéntico a la SEQ ID NO:5.

En un segundo aspecto:

Un vector de ácido nucleico o un ácido nucleico que comprende una secuencia de nucleótidos expresable seleccionada de

- a) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- c) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;
- d) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- e) una secuencia de nucleótidos que codifica un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5.

En un tercer aspecto:

Un (i) vector de ácido nucleico o una pluralidad de vectores de ácido nucleico o (ii) un ácido nucleico o una pluralidad

de ácidos nucleicos según el Primer o Segundo Aspecto, en donde dicho o dichos polipéptidos codificados por el o los vectores son operativos para su uso con un ARNcr para formar un complejo de ribonucleoproteína para la toma como diana de protoespaciadores en un polinucleótido, en donde el complejo es operativo con un resto adyacente a protoespaciador (PAM) que tiene la secuencia 5'-AAG-3'.

- 5 El ARNcr puede ser como se describe en cualquier opción o característica de la presente memoria. El PAM puede ser, alternativamente, uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un cuarto aspecto:

Un kit que comprende

- 10 a) Uno o más polipéptidos descritos en la presente memoria, o uno o más ácidos nucleicos que codifican dicho(s) polipéptido(s); y
- b) Un ARNcr, o uno o más ácidos nucleicos que codifican un ARNcr), en donde el ARNcr es cognado a PAM que tiene la secuencia 5'-AAG-3';

en donde dicho o dichos polipéptidos son operativos para su uso con el ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido.

- 15 Los polipéptidos pueden comprender Cas-S1, S2, S3, S4 y S5. Los polipéptidos pueden ser cualquier polipéptido, proteína, proteína de fusión o complejo descrito en cualquier opción o característica de la presente memoria. El ARNcr puede ser como se describe en cualquier opción o característica de la presente memoria. El PAM puede ser, alternativamente, uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un quinto aspecto:

- 20 Un complejo de ribonucleoproteína que comprende

- a) Uno o más polipéptidos descritos en la presente memoria, o uno o más ácidos nucleicos que codifican dicho(s) polipéptido(s); y
- b) Un ARNcr, o uno o más ácidos nucleicos que codifican un ARNcr), donde el ARNcr es cognado a PAM que tiene la secuencia 5'-AAG-3'.

- 25 Los polipéptidos pueden ser cualquier polipéptido, proteína, proteína de fusión o complejo descrito en cualquier opción o característica de la presente memoria. El ARNcr puede ser como se describe en cualquier opción o característica de la presente memoria. El PAM puede ser, alternativamente, uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un sexto aspecto:

- 30 Uno o más vectores de ácido nucleico o uno o más ácidos nucleicos que comprenden al menos una secuencia de nucleótidos seleccionada de las SEQ ID NO:7-11, en donde dicha secuencia de nucleótidos está

- a) conectada operativamente a un promotor heterólogo, promotor sintético, promotor eucariota o promotor no bacteriano; y/o
- 35 b) contenida en una célula que es una célula eucariota, una célula no bacteriana o una célula que no es una célula bacteriana (por ejemplo, una célula de *E. coli*, *Pseudomonas* o *Klebsiella*) que comprende una secuencia de nucleótidos endógena que comprende las SEQ ID NO:7-11.

### En una tercera configuración

En un primer aspecto:

Una proteína de fusión que comprende un polipéptido (Px), donde Px

- 40 a) comprende una secuencia de aminoácidos que es al menos (aproximadamente) un 80 % idéntica a una secuencia seleccionada de las SEQ ID No:1-5; y
- b) se fusiona a un polipéptido heterólogo (Py).

Py puede ser como se describe en cualquier opción o característica de la presente memoria.

En un segundo aspecto:

- 45 Una proteína o pluralidad de proteínas operativas para su uso con un ARNcr para formar un complejo de ribonucleoproteína para la toma como diana del protoespaciador en un polinucleótido, comprendiendo la(s) proteína(s)

uno, más o todos los polipéptidos seleccionados de

- a) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- 5 b) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- c) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;
- d) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- 10 e) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5;

en donde el complejo es operativo con un resto adyacente al protoespaciador (PAM) que tiene la secuencia 5'-AAG-3'.

- 15 El ARNcr puede ser como se describe en cualquier opción o en otras de la presente memoria. El PAM puede ser, alternativamente, cualquiera de los PAM descritos en cualquier opción o en otras de la presente memoria.

#### En una cuarta configuración

Una célula, en donde la célula es

- a) una célula eucariota que comprende el o los vectores o proteína de la primera, segunda o tercera configuración;
- 20 b) una célula animal, vegetal, de insecto o de hongo que comprende el o los vectores o proteína de la primera, segunda o tercera configuración; o
- c) una célula procariota que comprende el o los vectores o proteína de la primera, segunda o tercera configuración, en donde la célula no es una célula bacteriana (por ejemplo, una célula *de E. coli*, *Pseudomonas* o *Klebsiella*) que comprende secuencias de nucleótidos endógenas que codifican los polipéptidos de las partes a) a e) como se recita en la segunda configuración.
- 25

#### En una quinta configuración

En un primer aspecto:

Una composición (opcionalmente una composición *in vitro* o en donde la composición está contenida en un recipiente médico) que comprende un ARNcr o ácido nucleico que codifica el ARNcr o ARN guía; y uno, más o todos de

- 30 a) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- b) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- 35 c) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;
- d) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- e) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5

40 en donde

A: el ARNcr comprende un espaciador que es cognado a un primer protoespaciador, en donde el protoespaciador

- f) no se encuentra en *E. coli*;
- g) es un protoespaciador de célula eucariota; o
- h) un protoespaciador de célula animal (opcionalmente, de mamífero o humano), vegetal o fúngico; o

- 45 B: un protoespaciador de una *E. coli* que carece de secuencias de nucleótidos endógenas que codifican los

polipéptidos de a) a e).

El ARNcr puede ser como se describe en cualquiera de las características de la presente memoria.

En un segundo aspecto:

5 Una composición (opcionalmente, una composición *in vitro* o en donde la composición está contenida en un recipiente médico) que comprende un ARNcr o un ácido nucleico que codifica el ARNcr; y un ácido nucleico que codifica uno, más o todos de

- a) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- 10 b) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- c) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;
- d) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- 15 e) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5

en donde

A: el ARNcr comprende un espaciador que es cognado a un primer protoespaciador, en donde el protoespaciador

- f) no se encuentra en *E. coli*;
- 20 g) es un protoespaciador de célula eucariota; o
- h) un protoespaciador de células animales (opcionalmente, de mamíferos o humanos), vegetales o fúngicas;

o

25 B: un protoespaciador de una *E. coli* que carece de secuencias de nucleótidos endógenas que codifican los polipéptidos de a) a e).

El ARNcr puede ser como se describe en cualquiera de las características de la presente memoria.

### En una sexta configuración

En un primer aspecto:

Un método para modificar un sitio diana de ácido nucleico en una célula, comprendiendo el método

- 30 a) poner en contacto la célula con el o los vectores, en donde el o los vectores comprenden una o más secuencias de nucleótidos para producir un ARNcr, en donde el ARNcr comprende un espaciador que es cognado a un primer protoespaciador, en donde el protoespaciador:  
no se encuentra en *E. coli*; un protoespaciador de célula eucariota; o un protoespaciador de célula animal (opcionalmente humana), vegetal, de insecto o de hongo;
- 35 b) permitir la introducción en la célula de
  - (i) el o los vectores mediante lo cual se expresan el o los polipéptidos y el ARNcr en la célula, o
  - (ii) el ARNc de la composición (o ácidos nucleicos que codifican el ARNc o el ARN guía mediante los cuales el ARNc se expresa en la célula) y el ácido nucleico de la composición que codifica el o los polipéptidos, mediante lo cual el o los polipéptidos se expresan en la célula;
- 40 c) en donde el ARNcr guía forma un complejo con el o los polipéptidos para guiar el complejo al sitio diana.

El o los vectores y el ARNcr pueden ser como se describe en cualquier opción o característica de la presente memoria. Los polipéptidos pueden ser cualquier polipéptido, proteína, proteína de fusión o complejo descrito en cualquier opción o característica de la presente memoria.

En un segundo aspecto:

Un método para tratar o prevenir una enfermedad o afección que está mediada por células diana en un sujeto, comprendiendo el método llevar a cabo cualquiera de los métodos descritos en la presente memoria para modificar las células diana, en donde dicho contacto comprende administrar el vector o vectores o la composición al sujeto y en donde la modificación trata o previene la enfermedad o afección.

El o los vectores y la o las composiciones pueden ser como se describe en cualquier opción o característica de la presente memoria.

En un tercer aspecto:

Un método para introducir una edición dirigida en un polinucleótido diana, comprendiendo el método poner en contacto el polinucleótido diana con un sistema CRISPR/Cas que comprende ARNcr y un complejo proteico que comprende uno, más o todos los polipéptidos seleccionados de

- a) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:1;
- b) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:2;
- c) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:3;
- d) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:4; y
- e) un polipéptido que comprende un aminoácido que es al menos (aproximadamente) un 80 % idéntico a la SEQ ID NO:5

en donde el ARNcr comprende un espaciador que es cognado a un primer protoespaciador contenido en el polinucleótido, en donde el ARNcr se hibrida con el protoespaciador para guiar el complejo mediante lo cual el complejo edita el polinucleótido.

En un cuarto aspecto:

Un método para tomar como diana un polinucleótido (opcionalmente en una célula o *in vitro*), comprendiendo el método

- a) poner en contacto el polinucleótido con la o las proteínas y el ARNcr del Segundo Aspecto de la Tercera Configuración;
- b) permitir la formación de un complejo de ribonucleoproteína que comprende la o las proteínas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en el polinucleótido para modificar el polinucleótido o la replicación del mismo.

En un quinto aspecto:

Un método de control transcripcional o replicación de un ADN diana, que comprende poner en contacto el ADN diana con el complejo, en donde el complejo está desprovisto de una nucleasa de ADN y el complejo se une al ADN diana, controlando así la transcripción o replicación del ADN diana.

En un sexto aspecto:

Un método para controlar la replicación de un ARN diana, que comprende poner en contacto el ARN diana o un ADN que codifica el ARN con el complejo, en donde el complejo se une al ARN o ADN diana, controlando así la transcripción del ARN diana.

En un séptimo aspecto:

Un método para editar un ADN diana, que comprende poner en contacto el ADN diana con el complejo, en donde el complejo se une al ADN diana, editando así el ADN diana.

En un octavo aspecto:

Un método para escindir un ADN bicatenario (ADNs), que comprende poner en contacto dicho ADNs con el complejo, en donde el complejo comprende una nucleasa, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, mediante lo cual la nucleasa escinde el ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador.

El PAM puede ser alternativamente uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un noveno aspecto:

- 5 Un método para escindir un ADN monocatenario (ADNss), que comprende poner en contacto dicho ADN con el complejo, en donde el complejo comprende una nickasa, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, mediante lo cual la nucleasa escinde una sola hebra del ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador.

- 10 El PAM puede ser alternativamente uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un décimo aspecto:

- 15 Un método para marcar o identificar una región de un ADN, que comprende poner en contacto dicho ADN con el complejo, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, y opcionalmente en donde el complejo comprende un marcador detectable.

El PAM puede ser alternativamente uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un undécimo aspecto:

- 20 Un método para modificar la transcripción de una región de un ADN, que comprende poner en contacto dicho ADN con el complejo, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el complejo regula al alza o a la baja la transcripción de la región de ADN o de un gen adyacente.
- 25 El PAM no es CGG. Opcionalmente,
- El PAM puede ser AAN, ANG, NAG; o
  - El PAM puede ser AAG sin cambio de nucleótido o con un solo cambio de nucleótido.

El PAM puede ser alternativamente uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

- 30 En un duodécimo aspecto:

- 35 Un método para modificar un ADNds diana de una célula sin introducir roturas en el ADNds, comprendiendo el método producir en la célula el complejo, en donde el complejo se dirige a un ADN diana contenido en la célula, comprendiendo el ADN diana una secuencia protoespaciadora flanqueada en su 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el ADN se modifica sin introducir roturas en el ADN.

El PAM puede ser alternativamente uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un decimotercer aspecto:

- 40 Un método para inhibir el crecimiento o la proliferación de una célula sin introducir roturas letales en el ADNds, comprendiendo el método producir en la célula el complejo, en donde el complejo se dirige a un ADN diana contenido en la célula, comprendiendo el ADN diana una secuencia protoespaciadora flanqueada en su 5' por un PAM que tiene la secuencia 5'-AAG-3' o un PAM que es idéntico excepto por un cambio de base, por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador,
- 45 por lo que se inhibe la replicación del ADN sin introducir roturas en el ADN, inhibiendo así el crecimiento o la proliferación de una célula.

El PAM puede ser alternativamente uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un decimocuarto aspecto:

Un método para tratar o prevenir una enfermedad o afección en un sujeto humano, animal, vegetal o fúngico, comprendiendo el método llevar a cabo el método de cualquiera de los Aspectos Quinto a Decimotercero, en donde las células del sujeto comprenden dicho ADN o ARN y las células median la enfermedad o afección.

**En una séptima configuración**

- 5 Un vector de ácido nucleico o un ácido nucleico que comprende la SEQ ID NO:12 o una secuencia de ADN que es al menos (aproximadamente) un 70 u 80 % idéntica a la SEQ ID NO:12.

**En una octava configuración**

En un primer aspecto:

- 10 Uno o más ácidos nucleicos que codifican una pluralidad de proteínas Cas y que comprenden al menos una secuencia de nucleótidos para producir un ARNcr, en donde las proteínas Cas y el ARN son capaces de formar un complejo CRISPR/Cas de ribonucleoproteína en donde el ARN es capaz de guiar el complejo a un protoespaciador contenido en un ADN diana, en donde el extremo 5' del protoespaciador está flanqueado por un resto adyacente al protoespaciador (PAM) que tiene la secuencia 5'-AAG-3', en donde

- 15 a) el complejo carece de una nucleasa de ADN y es capaz de modificar el ADN sin introducir una rotura bicatenaria en el ADN;
- b) las proteínas Cas no comprenden una DinG, una Cas3 y una Cas10; y
- c) el complejo no comprende todas las proteínas Cas de un complejo CRISPR/Cas Tipo I, II, III, V o VI.

El PAM puede ser alternativamente uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

- 20 En un segundo aspecto:

Un complejo CRISPR/Cas de ribonucleoproteína que comprende una pluralidad de proteínas Cas y un ARNcr, en donde el ARN es capaz de guiar el complejo a un protoespaciador contenido en un ADN diana, en donde el extremo 5' del protoespaciador está flanqueado por un resto adyacente al protoespaciador (PAM) que tiene la secuencia 5'-AAG-3', en donde

- 25 a) el complejo carece de una nucleasa de ADN y es capaz de modificar el ADN sin introducir una rotura bicatenaria en el ADN;
- b) el complejo está desprovisto de una DinG, una Cas3 y una Cas10; y
- c) el complejo no comprende todas las proteínas Cas de un complejo CRISPR/Cas Tipo I, II, III, V o VI.

- 30 El PAM puede ser alternativamente uno cualquiera de los PAM descritos en cualquier opción o característica de la presente memoria.

En un tercer aspecto:

Una célula que comprende el o los ácidos nucleicos o complejo descritos en la presente memoria, en donde el ADN está contenido en un cromosoma de la célula.

En un cuarto aspecto:

- 35 Una célula que comprende el o los ácidos nucleicos o complejo descritos en la presente memoria, en donde el ADN está contenido en un plásmido en la célula.

En un quinto aspecto:

Un método para modificar un ADN, comprendiendo el método

- a) poner en contacto el ADN con el o los ácidos nucleicos o vectores descritos en la presente memoria;
- 40 b) permitir la formación de un complejo de ribonucleoproteína que comprende la o las proteínas Cas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en el ADN para modificar el ADN.

El ARNcr y la o las proteínas Cas pueden ser como se describe en cualquier opción o característica de la presente memoria.

En un sexto aspecto:

- 45 Un método para inhibir la replicación de un plásmido que comprende ADN, comprendiendo el método

- a) poner en contacto el ADN con el o los ácidos nucleicos o vectores descritos en la presente memoria;
- b) permitir la formación de un complejo de ribonucleoproteína que comprende la o las proteínas Cas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en el ADN para inhibir la replicación del plásmido.

5 El ARNcr y la o las proteínas Cas pueden ser como se describe en cualquier opción o característica de la presente memoria.

En un séptimo aspecto:

Un método para inhibir la transcripción de una secuencia de nucleótidos contenida en un ADN, comprendiendo el método

- 10 a) poner en contacto el ADN con el o los ácidos nucleicos o vectores descritos en la presente memoria;
- b) permitir la formación de un complejo de ribonucleoproteína que comprende la o las proteínas Cas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en la secuencia de nucleótidos para inhibir su transcripción.

15 El ARNcr y la o las proteínas Cas pueden ser como se describe en cualquier opción o característica de la presente memoria.

En un octavo aspecto:

Un método para inhibir el crecimiento o la proliferación de una célula (opcionalmente una célula procariota, tal como una célula bacteriana) que comprende ADN, comprendiendo el método

- a) poner en contacto el ADN con el o los ácidos nucleicos o vectores descritos en la presente memoria;
- 20 b) permitir la formación de un complejo de ribonucleoproteína que comprende la o las proteínas Cas y ARNcr, en donde el complejo es guiado a un sitio diana contenido en el ADN para inhibir el crecimiento o la proliferación de la célula.

El ARNcr y la o las proteínas Cas pueden ser como se describe en cualquier opción o característica de la presente memoria.

25 En un noveno aspecto:

Un método para tratar o prevenir una enfermedad o afección que está mediada por células diana en un sujeto humano o animal, comprendiendo el método llevar a cabo cualquiera de los métodos descritos en la presente memoria para modificar las células diana, en donde dicho contacto comprende administrar el o los ácidos nucleicos al sujeto y en donde la modificación trata o previene la enfermedad o afección.

30 En un décimo aspecto:

Uno o más ácidos nucleicos del Primer Aspecto para su uso en un método para tratar o prevenir una enfermedad o condición que está mediada por células diana en un sujeto humano o animal, comprendiendo el método llevar a cabo cualquiera de los métodos descritos en la presente memoria para modificar las células diana, en donde dicho contacto comprende administrar el o los ácidos nucleicos al sujeto y en donde la modificación trata o previene la enfermedad o afección.

35 Los expertos en la técnica reconocerán, o podrán determinar mediante un mero estudio rutinario, numerosos equivalentes a los procedimientos específicos descritos en la presente memoria.

40 El uso de la palabra "un" o "una" cuando se utiliza junto con el término "que comprende" en las reivindicaciones y/o la memoria descriptiva puede significar "uno", pero también es coherente con el significado de "uno o más", "al menos uno" y "uno o más de uno". El uso del término "o" en las reivindicaciones se utiliza para significar "y/o" a menos que se indique explícitamente que se refiere únicamente a alternativas o que las alternativas son mutuamente excluyentes, aunque la divulgación respalda una definición que se refiere únicamente a alternativas e "y/o". A lo largo de esta solicitud, el término "aproximadamente" se utiliza para indicar que un valor incluye la variación inherente del error para el dispositivo, el método que se emplea para determinar el valor o la variación que existe entre los sujetos del estudio.

45 Tal como se utiliza en esta memoria descriptiva y en las reivindicaciones, las palabras "que comprende" (y cualquier forma de que comprende, como "comprenden" y "comprende"), "que tiene" (y cualquier forma de que tiene, como "tienen" y "tiene"), "que incluye" (y cualquier forma de que incluye, como "incluye" e "incluyen") o "que contiene" (y cualquier forma de que contiene, como "contiene" y "contienen") son inclusivas o de extremos abiertos y no excluyen elementos o etapas de métodos adicionales no citados.

50 El término "o combinaciones de los mismos" o similar, como se utiliza en la presente memoria, se refiere a todas las

permutaciones y combinaciones de los elementos enumerados que preceden al término. Por ejemplo, "A, B, C o combinaciones de los mismos" pretende incluir al menos uno de: A, B, C, AB, AC, BC o ABC, y si el orden es importante en un contexto particular, también BA, CA, CB, CBA, BCA, ACB, BAC o CAB. Siguiendo con este ejemplo, se incluyen expresamente las combinaciones que contienen repeticiones de uno o más elementos o términos, como BB, AAA, MB, BBC, AAABCCCC, CBBAAA, CABABB, etc. El experto en la técnica entenderá que, típicamente, no hay límite en el número de elementos o términos en cualquier combinación, a menos que sea evidente lo contrario por el contexto.

Cualquier parte de esta divulgación puede leerse en combinación con cualquier otra parte de la divulgación, a menos que sea evidente lo contrario por el contexto.

La presente invención se describe con más detalle en los siguientes Ejemplos no limitativos.

## Ejemplos

### Ejemplo 1

#### 1.1 Resumen ejecutivo

Este Ejemplo describe la identificación *in silico* y la caracterización funcional *in vivo* (en *E. coli*) del sistema CRISPR-Cas no caracterizado previamente, cuya actividad de direccionamiento no se basa en una endonucleasa Cas guiada por ARN predicha. Demostramos que el sistema i) se dirige con éxito a ADN plasmídico, con un mecanismo desconocido, lo que dificulta la resistencia a los antibióticos inducida por plásmidos, y ii) se une de manera eficiente al ADN cromosómico sin inducir la muerte celular. Denominamos a este sistema "sistema CRISPR/Cas Tipo-S".

##### 1.1.1 Objetivos

Caracterizar un sistema CRISPR que sea diferente de los sistemas caracterizados funcionalmente conocidos actualmente.

#### 1.2 Materiales y métodos

##### 1.2.1 Construcción de plásmidos y cepas

Los plásmidos y las cepas utilizadas en este estudio se enumeran en el Apéndice (véase la sección 1.5 a continuación). Los plásmidos se construyeron a partir de fragmentos de ADN generados por PCR mediante clonación InFusion HD™ (Takara). La inserción de las secuencias diana en el cromosoma de la cepa *C-1* de *E. coli* se realizó mediante ingeniería recombinante mediada por lambda-red. La cepa resultante es b5451 ( $\Delta$ lacZ).

##### 1.2.2 Ensayo de transformación para la inhibición de plásmidos específicos de secuencia

Las células se transformaron primero con el plásmido p1624 que se muestra en la Figura 1 y los plásmidos que se muestran en la Figura 3, que portan componentes de un nuevo sistema CRISPR/Cas (al que denominamos en la presente memoria, un sistema CRISPR/Cas Tipo-S). A continuación, para evaluar la funcionalidad de los plásmidos que portaban componentes del sistema Tipo-S, las células de *E. coli* se transformaron por separado con estos plásmidos y posteriormente se transformaron con una mezcla 1:1 de dos plásmidos, p1631 y p892. El plásmido diana p1631 portaba todas las secuencias diana naturales (5 protoespaciadores) de la matriz CRISPR del sistema y el gen *amiCP*, proporcionando un color morado a las colonias que portaban el plásmido. El plásmido p892 sirvió como un control no diana, tenía el mismo origen de replicación (p15A) y gen marcador de antibiótico (CmR) que el plásmido diana, pero no portaba las secuencias diana y el gen *amiCP*.

Las células que no portaban ningún componente del sistema Tipo-S también se transformaron con la mezcla de plásmidos p1631/p892 como control para evaluar la proporción de los dos plásmidos en la mezcla.

Las células transformadas se sembraron en placas de agar LB que contenían los antibióticos adecuados para seleccionar la presencia de los plásmidos. Las placas se incubaron a 37 °C durante la noche y se determinó la proporción de colonias moradas y blancas, que portaban el plásmido diana y no diana, respectivamente.

##### 1.2.3 Inhibición de plásmidos usando expresión controlada de los componentes del sistema Tipo-S

La cepa b5456 que porta el plásmido p1826, que es idéntico al p1793 (Figura 3) pero que contiene el sistema Tipo-S bajo el control de un promotor inducible por arabinosa (en lugar de pB<sub>ol</sub>A), se transformó por electroporación con una mezcla 1:1 de dos plásmidos, p1631 y p892. Se dejó que las células se recuperaran durante 1,5 horas y se sembraron en agar LB que contenía antibióticos apropiados, con y sin arabinosa al 0,2 %.

##### 1.2.4 Ensayo de inhibición del crecimiento

Se transformó una mezcla 1:1 de células competentes de *E. coli C-1* y b5451 con el plásmido p1826 que contenía los componentes requeridos del sistema CRISPR-Cas Tipo-S y se dejó que se recuperaran durante 1,5 horas antes de sembrarlas en agar LB que contenía 20 µg/ml de tetraciclina para seleccionar los transformantes que portaban los

plásmidos. Las placas también contenían X-gal (20 ng/ml) e IPTG (1 mM) para distinguir las colonias de *E. coli* C-1 (lacZ+, azul) y b5451 (lacZ-, blanco).

#### 1.2.5 Ensayos de aclaramiento de plásmidos

- 5 La cepa b5408 se transformó por separado con los plásmidos p1935, p1936, p1937, p1938, p1631 y p144 (véase la Tabla 2 y la Figura 7). Se sembraron las diluciones seriadas de las soluciones de transformación recuperadas en placas de agar LB suplementadas con 10 µg/mL de tetraciclina y 20 µg/mL de cloranfenicol. Las placas se incubaron durante la noche a 37 °C y se contaron las unidades formadoras de colonias resultantes para calcular el efecto del sistema Tipo-S en la eficiencia de transformación de la cepa b5408.

#### 1.2.6 Ensayos de unión cromosómica

- 10 La cepa b5872 se transformó con los plásmidos p1985, p1993, p1994, p1995 y p1996 (véase la Tabla 2). Las mezclas de transformación se sembraron en placas de agar LB suplementadas con 10 ng/mL de tetraciclina y se incubaron durante la noche a 37 °C. Se utilizaron tres colonias individuales seleccionadas al azar por transformación para la inoculación de cultivos LB suplementados con 10 ng/mL de tetraciclina. Se permitió que los cultivos alcanzaran la fase estacionaria tardía tras la incubación a 37 °C durante 20 horas con agitación rigurosa a 220 rpm. Se añadió 1 µL de  
15 cada cultivo a 99 µL de medio LB suplementado con 10 ng/mL de tetraciclina y 0 µg/mL, 20 µg/mL, 25 µg/mL, 30 µg/mL o 35 µg/mL de cloranfenicol. Los cultivos se transfirieron a los pocillos de una placa de microtitulación y se incubaron en un lector de placas a 37 °C durante 24 horas con agitación rigurosa (900 rpm). Las densidades ópticas de los cultivos se monitorizaron durante todo el período de incubación con lecturas a 600 nm cada 10 minutos.

### 1.3 Resultados

#### 20 1.3.1 Sistemas CRISPR/Cas Tipo-S

- Clonamos en un plásmido de baja copia ADN que codificaba una endonucleasa guiada por ARN (*rge*), dos marcos de lectura abiertos (ORF) aguas abajo del gen *rge*, una matriz CRISPR cognada y cuatro ORF inmediatamente aguas arriba (5') de la matriz CRISPR (Figura 1). Se predijo que los ORF 140 y 423 codificarían una ADN polimerasa propensa a errores, mientras que se predijo que los ORF 624 y 188 codificarían una helicasa y una ARN endonucleasa, respectivamente. El plásmido p1624 generado establece los genes insertados bajo el control transcripcional de un promotor constitutivo. Por otra parte, la matriz CRISPR portaba cinco espaciadores. Los protoespaciadores correspondientes estaban precedidos por un resto adyacente al protoespaciador (PAM) 5'-AAG-3' del sistema CRISPR-Cas Tipo-S. Construimos un plásmido diana (p1631), que portaba cuatro de los protoespaciadores predichos (correspondientes a los espaciadores no. 1, no. 2, no. 4 y no. 5) flanqueados en su extremo 5' por el PAM 5'-AAG-3' (Figura 1).  
30

- Posteriormente, ensayamos la funcionalidad *in vivo* del sistema en p1624. Los plásmidos p1624 y p1631 se emplearon para ensayos de inhibición de plásmidos específicos de secuencia, descritos en la sección 'Materiales y métodos' 1.2.2. En resumen, las células que portaban el plásmido p1624 se transformaron con una mezcla 1:1 del plásmido diana p1631 y un plásmido de control (no diana). p1631 expresa una proteína de color morado, lo que facilita el cribado.  
35 Tanto los plásmidos diana como no diana, se transformaron con eficiencias similares, lo que demuestra que la configuración actual del sistema no estaba induciendo el direccionamiento del plásmido. Por lo tanto, se planteó la hipótesis de que el sistema requiere subunidades adicionales para ser activo. Ampliamos nuestro estudio clonando seis ORF adicionales (ORF 83, 135, 136, 162, 145 y 286, véase la Figura 3) en p1624 aguas arriba (5') del ORF 188. Todos los ORF adicionales se transcribieron en la misma dirección que los ORF clonados previamente y la matriz CRISPR. Usamos el plásmido p1760 resultante en lugar de p1624 para repetir el ensayo de inhibición del plásmido. p1760 inhibió el crecimiento de colonias transformadas con el plásmido diana p1631, pero permitió el crecimiento de colonias transformadas con el plásmido no diana, ya que todas las colonias aparecieron blancas (Figura 2). El resultado indicó que p1760 portaba un sistema CRISPR-Cas activo.  
40

#### 1.3.2 Análisis de delección para la identificación del sistema funcional mínimo

- 45 Nuestro objetivo fue identificar el número mínimo de elementos requeridos para la inhibición/direccionamiento de plásmidos específicos de secuencia mediante el sistema CRISPR-Cas. Creamos una biblioteca de plásmidos realizando una serie de delecciones de genes en p1760 (Figura 3) y repetimos el ensayo de inhibición de plásmidos (Materiales y métodos, Sección 1.2.3).

- 50 Once de las doce nuevas construcciones que ensayamos mostraron inhibición del plásmido diana (Figura 4). No se observó inhibición del plásmido en ausencia de la matriz CRISPR, como lo demostraron los ensayos que utilizaron p1899 (fila inferior, panel derecho de la Figura 4). Se observó inhibición del plásmido con los ensayos que utilizaron p1793 (fila inferior, tercer panel de la Figura 4), que contienen un operón de solo cinco genes (que codifican Cas que llamamos Cas-S1, Cas-S2, Cas-S3, Cas-S4 y Cas-S5).

- 55 La delección por separado de cada uno de estos genes de p1793 dio como resultado la construcción de los plásmidos p2009 (ΔCas-S1), p2011 (ΔCas-S3), p2012 (ΔCas-S4) y p2013 (ΔCas-S5) (véase la Tabla 2). Los plásmidos se transformaron en células de *E. coli* NEB10b, generando las cepas b5700 (ΔCas-S1), b5702 (ΔCas-S3), b5703 (ΔCas-

S4) y b5704 ( $\Delta$ Cas-S5), respectivamente (véase la Tabla 1). Cada una de estas cepas, junto con las cepas de control de direccionamiento negativo y positivo, b4816 y b5408, se transformaron con el plásmido diana p1631 y se sembraron en placas de agar con la selección de antibióticos adecuada. La eficiencia de transformación de b5408 disminuyó aproximadamente 3 órdenes de magnitud, mientras que la eficiencia de transformación de todas las demás cepas no disminuyó (véase la Figura 5). Estos resultados demuestran que el sistema CRISPR-Cas Tipo-S requiere las cinco subunidades para la inhibición del plásmido. Cabe destacar que no se prevé ningún gen de endonucleasa de ADN guiada por ARN entre estos cinco genes, lo que indica que el mecanismo de acción del sistema CRISPR-Cas probablemente no se base en la escisión del ADN.

Para confirmar aún más la funcionalidad del sistema CRISPR-Cas Tipo-S, clonamos el fragmento de ADN que contenía el sistema mínimo y la matriz CRISPR aguas abajo de un promotor inducible por arabinosa. El plásmido resultante (p1826) se pudo mantener junto con el plásmido diana p1631 en la misma célula en ausencia de arabinosa (Figura 6, panel izquierdo). En presencia de arabinosa, el sistema CRISPR-Cas se expresó dando como resultado la inhibición de p1631 y, sorprendentemente, la inhibición del crecimiento bacteriano (Figura 6, panel derecho).

### 1.3.3 Los espaciadores en la matriz CRISPR Tipo-S exhiben eficiencias de direccionamiento similares

Evaluamos la eficiencia de direccionamiento de cada espaciador en la matriz CRISPR que utilizamos en estos experimentos en el sistema CRISPR-Cas Tipo-S. La matriz contenía cinco espaciadores y el plásmido diana p1631 contenía protoespaciadores para 4 de estos 5 espaciadores. Usando p1631 como nuestra base, construimos cuatro plásmidos diana adicionales, a saber, p1935, p1936, p1937 y p1938, cada uno conteniendo solo un protoespaciador (véase la Tabla 2 y la Figura 7, panel izquierdo). Transformamos por separado los 5 plásmidos diana y el plásmido no diana p144 (véase la Tabla 2) en la cepa b5408 y sembramos diluciones seriadas de las mezclas de transformación en placas de selección apropiadas (Figura 7, panel derecho). Al día siguiente, se contaron las unidades formadoras de colonias para el análisis; los 4 espaciadores ensayados demostraron eficiencias de direccionamiento similares, lo que redujo las eficiencias de transformación de los plásmidos diana correspondientes en 3 órdenes de magnitud, en relación con el plásmido no diana. Además, la eficiencia de direccionamiento combinada de los 4 espaciadores no fue significativamente mayor que la eficiencia de cada espaciador por separado.

### 1.3.4 El sistema CRISPR-Cas Tipo-S se une al ADN cromosómico

Determinamos que el sistema CRISPR-Cas Tipo-S no contiene una endonucleasa de ADN guiada por ARN prevista. Por lo tanto, decidimos examinar si las subunidades del sistema contienen un nuevo dominio de escisión de ADNds, que puede introducir de manera eficiente roturas letales en el ADNds en el cromosoma de *E. coli*. Construimos la cepa b5451 moviendo los 4 protoespaciadores ensayados previamente de p1631 al cromosoma de *E. coli C-1*, inactivando el gen *lacZ*. La cepa resultante se puede distinguir de la cepa original de *E. coli C-1* no diana en placas X-gal ya que esta última cepa metaboliza X-gal y la convierte en un tinte azul. Tanto el sistema de expresión CRISPR-Cas Tipo-S constitutivo como el inducible por arabinosa se ensayaron como se describe en la sección 'Materiales y métodos' 1.2.6. En resumen, una mezcla 1:1 de células competentes de *E. coli C-1* y b5451 se transformó con el plásmido p1826, que contenía el sistema CRISPR-Cas Tipo-S bajo el control de un promotor inducible por arabinosa. La proporción de colonias blancas a azules se mantuvo en 1:1 en condiciones adecuadas de selección con antibióticos, independientemente de las condiciones de inducción (Figura 8). Este resultado respaldó aún más la probable ausencia de un dominio de endonucleasa de ADNds en el sistema.

Posteriormente, nos preguntamos si el sistema CRISPR-Cas Tipo-S tiene una función de unión al ADN o su modo de acción es a nivel del ARN. Construimos la cepa bSNP5810 introduciendo el gen *crm* en una región intergénica de *E. coli C-1*. Posteriormente, construimos una cepa de control negativo, denominada p1985, y cuatro plásmidos de direccionamiento a *crm*, denominados p1993, p1994, p1995 y p1996 (véase la Tabla 2). Para este fin, sustituimos los espaciadores dentro de la matriz CRISPR de p1793 bien por un espaciador sin direccionamiento o por espaciadores individuales que se dirigen a cuatro posiciones diferentes en el promotor o la región codificante del gen *crm* (Figura 9). Transformamos la cepa bSNP5810 con cada uno de p1993, p1994, p1995 y p1996 y cultivamos las células en medios suplementados con diferentes concentraciones de cloranfenicol. En ausencia de cloranfenicol, las tasas de crecimiento de todas las cepas con los espaciadores de direccionamiento a *crm* fueron similares a la tasa de crecimiento de la cepa de control (Figura 10, primer gráfico). No obstante, las concentraciones de cloranfenicol crecientes redujeron las tasas de crecimiento de todas las cepas diana de manera proporcional (Figura 10, 2° – 4° gráficos). La tasa de crecimiento de la cepa de control no se vio afectada por la presencia de cloranfenicol en el medio. Estos resultados demuestran que el sistema CRISPR-Cas Tipo-S se une de manera constante al cromosoma y actúa como un inhibidor transcripcional sorprendentemente eficiente. Cabe destacar que la orientación de los espaciadores de direccionamiento a *crm* no influye en la eficiencia de unión del sistema CRISPR-Cas Tipo-S y la posterior regulación a la baja de la expresión de *crm*, lo que no es el caso de otros sistemas CRISPR-Cas conocidos previamente que deben dirigirse a la región codificante o no codificante (característica dependiente del sistema) de un gen para regular a la baja de manera eficiente su transcripción (véase Qi *et al.*, Cell, 152(5), 1173-1183, 2013, doi:<https://doi.org/10.1016/j.cell.2013.02.022>, Rath *et al.*, NAR, 43(1), 237-246, 2015, doi:<https://doi.org/10.1093/nar/gku1257> y Luo *et al.*, NAR, 43(1), 674-681, 2015, doi:<https://doi.org/10.1093/nar/gku971>).

## 1.4 Conclusiones

Se caracterizó un sistema CRISPR-Cas deficiente en nucleasas (al que llamamos CRISPR/Cas Tipo-S) y se determinaron sus elementos funcionales mínimos. Nuestros estudios revelaron que el CRISPR/Cas Tipo-S comprende cinco proteínas Cas (llamadas Cas-S1 a S5) y una matriz CRISPR cognada. Ninguna de las proteínas Cas mostró tener una actividad de nucleasa de ADN. Sorprendentemente, el sistema inhibe el crecimiento de células bacterianas que contienen al menos una secuencia diana adyacente a un PAM AAG. Además, el sistema podría programarse para unirse de manera eficiente a dianas cromosómicas y actuar como un inhibidor transcripcional, sin inducir roturas letales en el ADN. Con estos hallazgos, hemos identificado y caracterizado un sistema CRISPR/Cas funcional novedoso que se puede utilizar en su forma mínima para inhibir plásmidos, reducir el crecimiento de células bacterianas e inhibir la transcripción de un cromosoma. Podemos ver aplicaciones para Cas individuales del sistema, tales como mediante la creación de una fusión de un Cas Tipo-S con un dominio de nucleasa, con el fin de producir un corte dirigido del ADN. Otros compañeros de fusión incluyen editores de bases, *prime editors*, desaminasas y transcriptasas inversas.

## 1.5 Apéndice

Tabla 1: Cepas bacterianas utilizadas en este estudio

| Número de cepa | Características relevantes                                                                                              |
|----------------|-------------------------------------------------------------------------------------------------------------------------|
| b5451          | <i>E. coli</i> C-1 que porta secuencias diana de p1631 insertadas con cmR en lacZ, lac <sup>-</sup>                     |
| b3127          | NEB Turbo                                                                                                               |
| b52            | Lac+                                                                                                                    |
| b5456          | P1826                                                                                                                   |
| bSNP5810       | <i>E. coli</i> C-1 con el gen <i>crm</i> (marcador de resistencia al cloranfenicol) integrado en una región intergénica |
| b5700          | <i>E. coli</i> NEB10b que porta el plásmido p2009                                                                       |
| b5702          | <i>E. coli</i> NEB10b que porta el plásmido p2011                                                                       |
| b5703          | <i>E. coli</i> NEB10b que porta el plásmido p2012                                                                       |
| b5704          | <i>E. coli</i> NEB10b que porta el plásmido p2013                                                                       |
| b4816          | <i>E. coli</i> NEB10b que porta el plásmido p1624                                                                       |
| b5408          | <i>E. coli</i> NEB10b que porta el plásmido p1793                                                                       |

Tabla 2: Plásmidos bacterianos utilizados en este estudio

| No. de plásmido | Número de cepa huésped | Características relevantes                                                                                                                     |
|-----------------|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| p1624           | b4816                  | Para el contenido génico, véase la Figura 1                                                                                                    |
| p1631           | b4818                  | Plásmido diana, que contiene protoespaciadores diana de los espaciadores 1º, 2º, 4º y 5º de la matriz CRISPR Tipo-S natural (colonias moradas) |
| p892            | b2685                  | Plásmido no diana, contiene solo el origen de replicación p15a y el gen marcador de resistencia al cloranfenicol.                              |
| p144            |                        | Casi idéntico a p1631, pero no contiene protoespaciadores diana de la matriz CRISPR Tipo-S natural (colonias blancas)                          |
| p1783           | b5398                  | Para el contenido génico, véase la Figura 3                                                                                                    |
| p1784           | b5399                  | Para el contenido génico, véase la Figura 3                                                                                                    |
| p1785           | b5400                  | Para el contenido génico, véase la Figura 3                                                                                                    |
| p1786           | b5401                  | Para el contenido génico, véase la Figura 3                                                                                                    |
| p1787           | b5402                  | Para el contenido génico, véase la Figura 3                                                                                                    |
| p1788           | b5403                  | Para el contenido génico, véase la Figura 3                                                                                                    |
| p1789           | b5404                  | Para el contenido génico, véase la Figura 3                                                                                                    |
| p1790           | b5405                  | Para el contenido génico, véase la Figura 3                                                                                                    |

|       |       |                                                                                                                                                                                                                                                                                           |
|-------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| p1791 | b5406 | Para el contenido génico, véase la Figura 3                                                                                                                                                                                                                                               |
| p1792 | b5407 | Para el contenido génico, véase la Figura 3                                                                                                                                                                                                                                               |
| p1793 | b5408 | Sistema funcional mínimo para el direccionamiento de plásmidos + la matriz CRISPR Tipo-S natural (Figura 3)                                                                                                                                                                               |
| p1899 | b5538 | Para el contenido génico, véase la Figura 3                                                                                                                                                                                                                                               |
| p1826 | b5456 | Sistema de p1793 bajo promotor inducible por arabinosa                                                                                                                                                                                                                                    |
| p1935 | b5647 | Plásmido diana, casi idéntico a p1631, pero que contiene solo 1 protoespaciador diana del 1 <sup>er</sup> espaciador de la matriz CRISPR Tipo-S natural (colonias moradas)                                                                                                                |
| p1936 | b5648 | Plásmido diana, casi idéntico a p1631, pero que contiene solo 1 protoespaciador diana del 2 <sup>o</sup> espaciador de la matriz CRISPR Tipo-S natural (colonias moradas)                                                                                                                 |
| p1937 | b5649 | Plásmido diana, casi idéntico a p1631, pero que contiene solo 1 protoespaciador diana del 4 <sup>o</sup> espaciador de la matriz CRISPR Tipo-S natural (colonias moradas)                                                                                                                 |
| p1938 | b5650 | Plásmido diana, casi idéntico a p1631, pero que contiene solo 1 protoespaciador diana del 5 <sup>o</sup> espaciador de la matriz CRISPR Tipo-S natural (colonias moradas)                                                                                                                 |
| p1985 |       | Casi idéntico a p1793, pero contiene un solo espaciador no dirigido en la matriz CRISPR y ninguno de los espaciadores naturales de la matriz CRISPR Tipo-S                                                                                                                                |
| p1993 |       | Casi idéntico a p1793, pero contiene un solo espaciador (véase el no. 1 en la Figura 9) en la matriz CRISPR y ninguno de los espaciadores naturales de la matriz CRISPR Tipo-S. El espaciador se dirige a la cadena directa de la región promotora del gen <i>crm</i> en la cepa bSNP5810 |
| p1994 |       | Casi idéntico a p1793, pero contiene un solo espaciador (véase el no. 2 en la Figura 9) en la matriz CRISPR y ninguno de los espaciadores naturales de la matriz CRISPR Tipo-S. El espaciador se dirige a la cadena inversa del gen <i>crm</i> en la cepa bSNP5810                        |
| p1995 |       | Casi idéntico a p1793, pero contiene un solo espaciador (véase el no. 3 en la Figura 9) en la matriz CRISPR y ninguno de los espaciadores naturales de la matriz CRISPR Tipo-S. El espaciador se dirige a la cadena directa de la región promotora del gen <i>crm</i> en la cepa bSNP5810 |
| p1996 |       | Casi idéntico a p1793, pero contiene un solo espaciador (véase el no. 4 en la Figura 9) en la matriz CRISPR y ninguno de los espaciadores naturales de la matriz CRISPR Tipo-S. El espaciador se dirige a la cadena directa del gen <i>crm</i> en la cepa bSNP5810                        |
| p2009 | b5700 | Casi idéntico a p1793, pero CasS1 está delecionada                                                                                                                                                                                                                                        |
| p2011 | b5702 | Casi idéntico a p1793, pero CasS3 está delecionada                                                                                                                                                                                                                                        |
| p2012 | b5703 | Casi idéntico a p1793, pero CasS4 está delecionada                                                                                                                                                                                                                                        |
| p2013 | b5704 | Casi idéntico a p1793, pero CasS5 está delecionada                                                                                                                                                                                                                                        |

## Ejemplo 2: Identificación y caracterización de PAM para el sistema CRISPR/Cas Tipo-S

### 2.1 Antecedentes

SNIPR Biome ha identificado y caracterizado un sistema CRISPR/Cas Tipo-S como se describe en el Ejemplo 1.

- 5 Se realizó de manera interna un análisis *in silico* de bases de datos genómicas públicas, con el objetivo de identificar una secuencia PAM potencial para el sistema CRISPR/Cas Tipo-S del Ejemplo 1. Este análisis se basó en los espaciadores de la matriz CRISPR Tipo-S de tipo salvaje. Se descubrieron algunos protoespaciadores potenciales que comprendían un resto de secuencia 5'-AAG-3' en su extremo 5', pero ningún resto de secuencia común en su extremo 3'. Se planteó la hipótesis de que este resto 5'-AAG-3' en el extremo 5' de un protoespaciador podría servir
- 10 como un PAM válido para el sistema Tipo-S. Esta hipótesis se confirmó con ensayos de direccionamiento de plásmidos/muerte (como se describe en los Ejemplos 1.3.1-1.3.3). No obstante, las secuencias PAM de la mayoría de los sistemas CRISPR-Cas son promiscuas, por lo que se decidió investigar más experimentalmente la variedad de secuencias que el CRISPR/Cas Tipo-S reconoce como PAM válidos.

## 2.2 Diseño experimental

El proceso de identificación de PAM del sistema CRISPR/Cas Tipo-S se basó en los siguientes plásmidos:

1. **Biblioteca de plásmidos PAM** (p2259); cada miembro de esta biblioteca abarca la secuencia protoespaciadora presente en el plásmido diana p1935 (véase el Ejemplo 1.2.5) y una mezcla de PAM candidatos de 5 nt de longitud en el extremo 5' del protoespaciador.
2. **Plásmido Tipo-S de direccionamiento** (p1793); el plásmido expresa el sistema CRISPR/Cas Tipo-S (Cas-S1 a Cas-S5) y un espaciador que es complementario al protoespaciador de la biblioteca de plásmidos PAM.
3. **Plásmido Tipo-S sin direccionamiento (control)** (p1985); el plásmido expresa el sistema CRISPR/Cas Tipo-S (Cas-S1 a Cas-S5) y un espaciador que no es complementario al protoespaciador de la biblioteca de plásmidos PAM. Además, el espaciador no es complementario a ninguna parte del genoma de *E. coli*.

Las etapas del ensayo de identificación de PAM son las siguientes:

1. Construcción de la biblioteca de plásmidos PAM, denominada p2259 (véase la sección a continuación).
2. Construcción de la cepa de **direccionamiento *E. coli* NEBTurbo** (b5408), mediante la transformación de la cepa NEB Turbo b3127 (NEB) con el plásmido p1793 (direccionamiento Tipo-S).
3. Construcción de la **cepa sin direccionamiento *E. coli* NEBTurbo** (b6259) mediante la transformación de la cepa NEB Turbo b3127 (NEB) con el plásmido p1985 (Tipo-S control).
4. Transformación de b5408 (direccionamiento Tipo-S) y b6259 (Tipo-S control) con la biblioteca de plásmidos p2259 y cultivo en condiciones de selección de antibióticos adecuadas.
5. Aislamiento de la biblioteca de plásmidos p2259 de ambos cultivos y preparación de fragmentos NGS que contienen la región protoespaciadora PAM.
6. NGS y análisis de los resultados de la secuenciación.

### 2.2.1 Etapa 1: Construcción de la biblioteca PAM

El plásmido p2259 de la biblioteca PAM se diseñó para contener una región PAM variable de 5 nt de longitud en el extremo 5' del protoespaciador. Por lo tanto, la biblioteca debería estar compuesta por  $4^5$  (1.024) PAM diferentes.

La construcción del plásmido p2259 de la biblioteca PAM se basó en el plásmido diana p1935 descrito previamente (véase el Ejemplo 1.2.5). El plásmido p1935 se empleó como el molde de amplificación por PCR, con cebadores diseñados apropiadamente. Se utilizó la premezcla de PCR CloneAMP (Takara) para la reacción de PCR. El cebador directo se diseñó para introducir una secuencia degenerada de 5 nt de longitud en el extremo 5' del protoespaciador presente en el plásmido p1935. El producto de PCR purificado en gel se utilizó para la clonación In-Fusion, empleando 5 x premezcla de enzima In-Fusion HD (Takara). Se transformaron 5 µL de la reacción de clonación In-Fusion por electroporación en 50 µL de células electrocompetentes NEB10b (NEB). Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y agitando a 200 rpm. Las células recuperadas se transfirieron a 500 mL de medio líquido LB, suplementado con 20 µg/mL de cloranfenicol. Se retiraron 100 µL del cultivo de 500 mL y se sembraron en una placa de agar LB, suplementada con 20 µg/mL de cloranfenicol. El cultivo y la placa se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm). Al día siguiente, se contaron 536 colonias en la placa y posteriormente se calculó que se añadieron 2.680.000 células transformadas con éxito al cultivo de 500 mL. Para este cálculo, se utilizó la Ecuación 1.

Ecuación 1:

$$\text{No. de células transformadas en el cultivo} = \frac{\text{Volumen del cultivo}}{\text{Volumen del cultivo sembrado}} \times \text{No. de colonias en la placa de selección}$$

Este número de células transformadas en el cultivo corresponde a una cobertura de 2.617x de los miembros de la biblioteca PAM, según la Ecuación 2.

Ecuación 2:

$$\text{Cobertura de la biblioteca de plásmidos} = \frac{\text{Número de células transformadas en el cultivo}}{\text{Número de miembros de la biblioteca de plásmidos distintos}}$$

Después de preparar las soluciones madre de glicerol a partir del cultivo de 500 mL, la biblioteca de plásmidos p2259 construida se aisló del cultivo restante utilizando el kit de Qiagen plasmid midi.

### 2.2.2 Etapa 2: Transformación de la cepa b5408 (direccionamiento CRISPR-S) y b6259 (Tipo-S control) con la biblioteca de plásmidos p2259

Se espera que algunos miembros de la biblioteca de plásmidos p2259 contengan PAM que sean reconocidos por el sistema CRISPR/Cas Tipo-S. Estos plásmidos positivos para PAM serán dianas exitosas y su replicación será obstaculizada por el sistema Tipo-S. Por lo tanto, las células b5408 (direccionamiento Tipo-S) transformadas con los miembros positivos para PAM de la biblioteca de plásmidos p2259 no deberían sobrevivir en un medio suplementado con cloranfenicol, ya que la biblioteca de plásmidos p2259 porta el gen marcador de resistencia al cloranfenicol. Por otra parte, las células b5408 (direccionamiento Tipo-S) transformadas con miembros negativos para PAM de la biblioteca de plásmidos p2259 deberían sobrevivir en un medio suplementado con cloranfenicol. Por el contrario, ninguno de los plásmidos p2259 de la biblioteca será la diana del sistema CRISPR/Cas Tipo-S expresado en la cepa b6259 (Tipo-S control) y sobrevivirán en el medio suplementado con cloranfenicol.

La biblioteca de plásmidos p2259 se electroporó en células electrocompetentes b5408 (direccionamiento Tipo-S) y b6259 (Tipo-S control), respectivamente. Las células electrocompetentes de ambas cepas se prepararon utilizando el protocolo descrito en (<https://barricklab.org/wiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Se realizaron cinco reacciones de transformación por cepa, utilizando 40 ng de la biblioteca de plásmidos p2259 por reacción. Se dejó que las células se recuperaran en 500 µL de medio SOC durante 1 hora a 37 °C y agitando a 200 rpm. Las recuperaciones de cada cepa se combinaron y cada una de las dos recuperaciones finales (una por cepa) se utilizó para la inoculación de 500 mL de cultivos de medio líquido LB, suplementado con 20 µg/mL de cloranfenicol y 10 µg/mL de tetraciclina. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa a 200 rpm. Al día siguiente, se prepararon soluciones madre de glicerol de cada cultivo y se aislaron las bibliotecas de plásmidos p2259 de cada cepa utilizando el kit de Qiagen plasmid midi.

El aislamiento de plásmidos de cultivos de b5408 (direccionamiento Tipo-S) transformados con p2259, seguido de NGS, mostrará cuáles de las secuencias PAM no son reconocidas por el sistema CRISPR/Cas Tipo-S y, posteriormente, cuáles son las secuencias PAM que requiere el sistema. El aislamiento de plásmidos de cultivos b6259 (Tipo-S control) transformados con p2259, seguido de NGS, indicará la abundancia de cada miembro del plásmido en la biblioteca p2259. Esta información se utilizará para la normalización de los resultados de NGS de la cepa b5408 (direccionamiento Tipo-S).

### 2.2.3 Etapa 3: Identificación de PAM mediante NGS

El p2259 aislado de b6259 (Tipo-S control) debe contener todos los miembros de la biblioteca PAM p2259, proporcionando información cuantitativa sobre la abundancia de cada plásmido PAM en la biblioteca.

Los plásmidos p2259 aislados de b5408 (direccionamiento Tipo-S) y b6259 (Tipo-S control) se sometieron a NGS. La 1ª etapa de amplificación por PCR de NGS en ambas preparaciones de p2259 se realizó utilizando cebadores diseñados adecuadamente en 50 µL de reacciones KAPA HiFi Hotstart ReadyMix (Roche) (25 ciclos, 30 s de desnaturalización, 20 s de hibridación, 15 s de extensión). Los productos de reacción se purificaron con reactivo AMPure XP (Beckman Coulter). Posteriormente, se añadieron índices de Illumina y secuencias de unión de celdas de flujo a los productos purificados con una 2ª etapa de amplificación por PCR de NGS. El 1º producto de PCR de NGS de la cepa b5408 (direccionamiento Tipo-S) se sometió a la 2ª etapa de amplificación por PCR de NGS utilizando cebadores diseñados adecuadamente. El 1º producto de PCR de NGS de la cepa b6259 (CRISPR-S control) se sometió a una 2ª etapa de amplificación por PCR de NGS utilizando cebadores diseñados adecuadamente. Ambas amplificaciones fueron reacciones de 50 µL de KAPA HiFi Hotstart ReadyMix (Roche) (10 ciclos, 20 ng de ADN molde, 30 s de desnaturalización, 20 s de hibridación, 15 s de extensión).

Ambos productos de reacción se purificaron con el reactivo AMPure XP (Beckman Coulter). Los productos purificados se secuenciaron utilizando Illumina MiSeq® con 150 ciclos de extremo único.

## 2.3 Resultados

Los datos de NGS se utilizaron para comparar la abundancia relativa de cada miembro de la biblioteca p2259 en las cepas b5408 (direccionamiento Tipo-S) y b6259 (control). El análisis de los datos reveló una clara preferencia del sistema CRISPR/Cas Tipo-S por las siguientes secuencias PAM (Figuras 11 y 12):

5'-AHN-3'

5'- KAG-3'

5'- AGG-3'

5'-CAG-3'

5'-GTG-3'

**Ejemplo 3: Tolerancia al error de emparejamiento con el protoespaciador en el sistema CRISPR/Cas Tipo S****3.1 Antecedentes**

Los sistemas CRISPR/Cas presentan diversos requisitos de complementariedad entre espaciador-protoespaciador para la unión y direccionamiento a ADN/ARN. Por ejemplo, el sistema CRISPR/Cas Tipo I-E de *Escherichia coli* tolera errores de emparejamiento espaciador-protoespaciador fuera de la región de la "semilla" sin comprometer su función como un sistema inmune (véase Semenova *et al.*, PNAS, 108(25), 10098-10103, 2011, doi:https://doi.org/10.1073/pnas.110414410), y Cas9 de *Streptococcus pyogenes* requiere una complementariedad espaciador-protoespaciador de 15 nt para la escisión del ADN *in vitro* (véase Jinek *et al.*, Science, 337(6096), 816-821, 2012, doi:10.1126/science.12258), pero tolera múltiples errores de emparejamiento espaciador-protoespaciador (hasta 15 nucleótidos) para la unión del ADN en bacterias (véase Ciu *et al.*, Nature Comm., 9(1912), 2018, doi:https://doi.org/10.1038/s41467-018-04209-5).

El sistema CRISPR/Cas Tipo-S no contiene un dominio de nucleasa de ADNds predicho. Se ha mostrado previamente que la actividad de aclaramiento de plásmidos del sistema CRISPR/Cas Tipo-S requiere la presencia de la subunidad de helicasa Cas-S3 (véase el Ejemplo 1.2.5). Se espera que los errores de emparejamiento espaciador-protoespaciador afecten negativamente la estabilidad de unión a la diana de ADN del complejo CRISPR-Cas del sistema CRISPR/Cas Tipo-S pre-helicasa (el complejo CRISPR-Cas Tipo-S sin la subunidad de helicasa, que comprende las siguientes subunidades: Cas-S1, S2, S4 y S5). Se espera que la reducción de la eficiencia de unión al ADN del sistema CRISPR/Cas Tipo-S dé como resultado un aclaramiento reducido del plásmido. Por lo tanto, los ensayos descritos en este Ejemplo fueron diseñados para emplear la eficiencia de curación/transformación de plásmidos como el resultado mensurable del efecto que tienen los errores de emparejamiento espaciador-protoespaciador sobre la eficiencia de direccionamiento CRISPR-Cas del sistema CRISPR/Cas Tipo-S.

**3.2 Diseño experimental**

El plásmido p1935 se diseñó previamente (véase el Ejemplo 1.2.5) y se construyó para contener un protoespaciador de 32 pb y una secuencia PAM (5'-AAG-3') que es la diana del sistema CRISPR/Cas Tipo-S. En este Ejemplo, se diseñaron y construyeron diferentes variantes del plásmido p1935, que contenían varias combinaciones de mutaciones en la región del protoespaciador. Estas mutaciones disrumpen parcialmente la complementariedad del protoespaciador con el espaciador dirigido. Los gráficos de eficiencia de transformación adjuntos (Figura 13) representan estas mutaciones en detalle.

Las diversas mutaciones del protoespaciador se introdujeron en la secuencia del protoespaciador del plásmido p1935 mediante mutagénesis dirigida al sitio Q5® (NEB) utilizando cebadores diseñados apropiadamente. Los productos de PCR resultantes se circularizaron utilizando reacciones de mezcla enzimática KLD (NEB). Las mezclas de reacción KLD se transformaron en células electrocompetentes NEB-10β (NEB). Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y agitando a 200 rpm. Las células recuperadas se sembraron en placas de agar LB suplementadas con 20 µg/mL de cloranfenicol para la selección del plásmido. Las placas se incubaron durante la noche a 37 °C. Se inocularon colonias individuales de cada placa en cultivos de medio líquido LB de 10 mL suplementados con 20 µg/mL de cloranfenicol para la selección del plásmido. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm). Se tomaron muestras de cada cultivo para preparar las soluciones madre de glicerol, que se almacenaron a -80 °C. El contenido restante del cultivo se utilizó para el aislamiento de plásmidos utilizando el minikit de plásmidos de Qiagen.

b5408 es una cepa NEB Turbo de *E. coli* que contiene el plásmido p1793 (que contiene el sistema CRISPR/Cas Tipo-S de tipo salvaje, Cas-S1 a Cas-S5, incluida la matriz CRISPR Tipo-S). Se electroporaron 50 ng de cada miembro de la biblioteca de plásmidos p1935 en células electrocompetentes b5408 (Tipo-S wt). Las células electrocompetentes de ambas cepas se prepararon utilizando el protocolo descrito en (https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells). Se dejó que las células se recuperaran en 500 µL de medio SOC durante 1 hora a 37 °C y agitando a 200 rpm. Las recuperaciones de cada cepa se utilizaron para la preparación de diluciones seriadas, que posteriormente se sembraron en placas de agar LB suplementadas con los antibióticos apropiados, 20 µg/mL de cloranfenicol y 10 µg/mL de tetraciclina, para la selección de plásmidos. Las transformaciones se realizaron en triplicados biológicos. Se contaron las colonias formadas a partir de cada transformación, y la reducción en la eficiencia de transformación de la cepa b5408 (Tipo-S wt), en comparación con su eficiencia de transformación cuando se transformó con un plásmido de control negativo (no diana), indicó cómo los errores de emparejamiento espaciador-protoespaciador correspondientes afectaban la eficiencia de direccionamiento del sistema CRISPR/Cas Tipo-S (véase la Figura 13).

Controles :

- El plásmido original p1935 es el plásmido de control positivo, que siempre generó casi ningún transformante (solo mutantes de escape).
- El plásmido p2361 es el plásmido de control negativo. Es idéntico a p1935, excepto por la secuencia PAM 5'-AAG-3' aguas arriba (5') del protoespaciador, que se intercambia con la secuencia 5'-CGG-3' que no se prevé

que sea un PAM de CRISPR/Cas Tipo-S (véase el Ejemplo 2). Por lo tanto, la eficiencia de transformación de p2361 (control negativo) debería ser la más alta de todos los miembros de la biblioteca mencionados en este Ejemplo.

### 3.3 Resultados

5 La Figura 13 muestra el análisis de la eficiencia de transformación para cada uno de los errores de emparejamiento. Un examen cuidadoso de los gráficos rindió las siguientes conclusiones, con respecto a cómo los errores de emparejamiento espaciador-protoespaciador afectan la eficiencia de aclaramiento del plásmido del sistema CRISPR/Cas Tipo-S:

1. Se toleran bien cinco errores de emparejamiento distales de PAM consecutivos.
- 10 2. 8 errores de emparejamiento distales de PAM consecutivos eliminan por completo el direccionamiento.
3. Los errores de emparejamiento individuales en las posiciones 1ª o 2ª proximales de PAM reducen el direccionamiento en 3 órdenes de magnitud.
4. Los errores de emparejamiento individuales en las posiciones 3ª a 10ª proximales de PAM reducen el direccionamiento entre 0 y 1 órdenes de magnitud.
- 15 5. Los errores de emparejamiento dobles en las posiciones 1ª a 3ª proximales de PAM reducen el direccionamiento entre 3 y 4 órdenes de magnitud.
6. Los errores de emparejamiento dobles en las posiciones 4ª a 10ª proximales de PAM reducen el direccionamiento entre 0 y 1 órdenes de magnitud.
- 20 7. Los errores de emparejamiento triples en las posiciones 1ª a 5ª proximales de PAM reducen el direccionamiento entre 3 y 4 órdenes de magnitud.
8. Los errores de emparejamiento triples en las posiciones 4ª a 9ª proximales de PAM reducen el direccionamiento entre 0 y 1 órdenes de magnitud.
9. Los errores de emparejamiento triples en las posiciones 8ª a 10ª proximales de PAM reducen el direccionamiento en 4 órdenes de magnitud.
- 25 10. Los errores de emparejamiento cuádruples en las posiciones 1ª a 8ª proximales de PAM reducen el direccionamiento entre 3 y 4 órdenes de magnitud.
11. Los errores de emparejamiento cuádruples en las posiciones 6ª a 9ª proximales de PAM reducen el direccionamiento en 1 orden de magnitud.
- 30 12. Los errores de emparejamiento cuádruples en las posiciones 7ª a 10ª proximales de PAM reducen el direccionamiento en 5 órdenes de magnitud.

### Ejemplo 4: El sistema CRISPR/Cas Tipo-S para aplicaciones CRISPRi

#### 4.1 Objetivo

Este estudio evaluó i) la eficiencia del sistema CRISPR/Cas Tipo-S para unirse a dianas cromosómicas y plasmídicas de *E. coli*, actuando como una herramienta CRISPRi, ii) qué subunidades son esenciales para la unión al ADN.

#### 35 4.2 Introducción

Previamente, demostramos que el sistema CRISPR/Cas Tipo-S dirige la inhibición dependiente de ARNcr de la replicación de plásmidos (véase el Ejemplo 1.3.2 y 1.3.3) mientras que carece de un dominio de nucleasa de ADN. La actividad de unión al ADN guiada por ARNcr del sistema CRISPR/Cas Tipo-S se puede explotar para el desarrollo de una herramienta de interferencia CRISPR (CRISPRi) en *E. coli*. Esperamos que también se pueda utilizar en otros tipos de células. En este Ejemplo, se evaluó el potencial de CRISPRi del sistema CRISPR/Cas Tipo-S a través de la regulación a la baja dirigida de la transcripción y la posterior traducción de un gen informador de proteína fluorescente verde (*gfp*). El gen *gfp* diana estaba integrado cromosómicamente o era portado por un plásmido. Además, este Ejemplo reveló qué subunidades del sistema CRISPR/Cas Tipo-S son esenciales para la actividad de unión al ADN del sistema CRISPR/Cas Tipo-S.

#### 45 4.3 Métodos y resultados

Se ensayaron cinco variaciones del sistema CRISPR/Cas Tipo-S para su potencial de CRISPRi;

- i) el sistema CRISPR/Cas Tipo-S de tipo salvaje (wt)

- ii) el sistema CRISPR/Cas Tipo-S mutante con delección de Cas-S1 ( $\Delta$ Cas-S1)
- iii) el sistema CRISPR/Cas Tipo-S mutante con delección de Cas-S3 ( $\Delta$ Cas-S3)
- iv) el sistema CRISPR/Cas Tipo-S mutante con delección de Cas-S4 ( $\Delta$ Cas-S4)
- vi) el sistema CRISPR/Cas Tipo-S mutante con delección doble de Cas-S1/Cas-S3 ( $\Delta$ Cas-S1  $\Delta$ Cas-S3).

5 Todas las variantes se colocaron bajo el control transcripcional del promotor pBoA (SEQ ID No:14) y se clonaron junto con un módulo de expresión de ARNcr no dirigido, generando los plásmidos p1985 (wt), p2062 ( $\Delta$ Cas-S1), p2064 ( $\Delta$ Cas-S3), p2065 ( $\Delta$ Cas-S4) y p2160 ( $\Delta$ Cas-S1  $\Delta$ Cas-S3), respectivamente (Figura 14). El extremo 5' y 3' del espaciador del módulo de ARNcr contenía sitios de reconocimiento para la enzima de restricción BsmBI-v2 (NEB), lo que facilitó la clonación de los espaciadores dirigidos (Figura 14). Todos los plásmidos contenían el gen de la proteína de replicación *repA101*, el origen de replicación de bajo número de copias SC101 y el *gen* marcador de resistencia a la tetraciclina *tetR* (Figura 14).

#### 4.4 CRISPRi cromosómico

15 Las cepas MG1655 y b5815 (cepa *gfp* positiva) se emplearon para los ensayos CRISPRi del sistema CRISPR/Cas Tipo-S cromosómico. b5815 es un derivado de *E. coli* MG1655, construido para contener el gen *gfp* bajo el control del promotor p70a (SEQ ID No:15) y posicionado entre los genes *ybcB* e *ybcC* (Figura 15). Dieciséis protoespaciadores dentro o cerca del gen *gfp* de la cepa b5815 (*gfp* positiva) se seleccionaron y tomaron como diana para los ensayos CRISPRi cromosómicos (Figura 15, Tabla 4, Seq ID No:25 a 41). Ocho de los protoespaciadores estaban localizados en la cadena codificante de *gfp* (a saber, Protoespaciador U2kF, U1kF, S1F, S2F, S3F, S4F, S5F, S6F) y ocho en la cadena no codificante (a saber, Protoespaciador U2kR, U1kR, S1R, S2R, S3R, S4R, S5R, S6R).

##### 4.4.1 Métodos

25 La construcción de los plásmidos que expresaban las diferentes variantes del sistema CRISPR/Cas Tipo-S y que tenían como diana los protoespaciadores antes mencionados se realizó digiriendo los plásmidos p1985 (wt), p2062 ( $\Delta$ Cas-S1), p2064 ( $\Delta$ Cas-S3), p2065 ( $\Delta$ Cas-S4) y p2160 ( $\Delta$ Cas-S1  $\Delta$ Cas-S3) con la enzima de restricción BsmBI-v2 (NEB) y luego ligando oligos de ADNss complementarios prehibridados que codificaban los espaciadores correspondientes en los plásmidos digeridos utilizando la ADN ligasa T4 (NEB). Se transformaron 5  $\mu$ L de cada reacción de ligación con electroporación en 50  $\mu$ L de células electrocompetentes NEB10 $\beta$  (NEB). Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500  $\mu$ L) durante 1 hora a 37 °C y agitando a 200 rpm. Se extrajeron 100  $\mu$ L de cada recuperación y se sembraron en placas de agar LB suplementadas con 20  $\mu$ g/mL de cloranfenicol. Las placas se incubaron durante la noche a 37 °C.

30 Al día siguiente, se seleccionaron al azar colonias individuales y se utilizaron para la inoculación de cultivos de 10 mL de medio líquido LB suplementado con 10  $\mu$ g/mL de tetraciclina. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm).

35 Al día siguiente, se prepararon soluciones madre de glicerol para cada cultivo y los cultivos restantes se usaron para el aislamiento de plásmidos con el mini kit de plásmidos de Qiagen. Cada plásmido se verificó por secuenciación con NGS. Cada uno de los plásmidos se transformó en células electrocompetentes b5815 (cepa positiva para *gfp*). Las células electrocompetentes se prepararon utilizando el protocolo descrito en (<https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Las reacciones de transformación se realizaron utilizando 40 ng de cada plásmido por reacción. Se dejó que las células se recuperaran en 500  $\mu$ L de medio SOC durante 1 hora a 37 °C y agitando a 200 rpm. Se sembraron 100  $\mu$ L de cada recuperación en placas de agar LB suplementadas con 10  $\mu$ g/mL de tetraciclina. Las placas se incubaron durante la noche a 37 °C.

40 Al día siguiente, se seleccionaron al azar tres colonias individuales de cada transformación y se utilizaron para inocular 5 mL de cultivos de medio mínimo M9 suplementado con glicerol (0,4 % p/v) y 10  $\mu$ g/mL de tetraciclina. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm).

45 Al día siguiente, se transfirieron 2  $\mu$ L de cada cultivo a 198  $\mu$ L de medio M9 suplementado con glicerol (0,4 % p/v) y 10  $\mu$ g/mL de tetraciclina. Los cultivos de 200  $\mu$ L resultantes se transfirieron a distintos pocillos en placas de fondo óptico de 96 pocillos Thermo Scientific™ Nunc microWell con base de polímero. Las placas se sellaron con membranas de sellado Breathe-Easy® (Sigma-Aldrich) y se cargaron en lectores de microplacas Synergy H1 (Biotek) para incubación a 37 °C con agitación continua. Las densidades ópticas (absorbancia a 600 nm) y las emisiones de fluorescencia verde (excitación a 485 nm, emisión a 516 nm) de los cultivos se midieron en intervalos de 10 minutos durante un período de 24 horas. Los datos de emisiones de fluorescencia de cada cultivo y cada punto de tiempo se normalizaron con los datos de DO600 nm correspondientes. Los datos normalizados de los cultivos que pertenecen a triplicados biológicos se promediaron y se calcularon las desviaciones estándar correspondientes. Los datos finales se representaron en los gráficos presentados en la Figura 16.

#### 4.4.2 Resultados

El sistema CRISPR/Cas Tipo-S reguló a la baja la expresión de GFP desde los niveles de b5815 (cepa positiva para *gfp*) hasta la cepa de control sin fluorescencia (b230), cuando se toman como diana los protoespaciadores al comienzo del orf de *gfp* y en la región promotora p70a (Protoespaciadores S1F, S1R, S2F, S2R, S3F, S3R) (Figura 16A) o protoespaciadores al final del orf de *gfp* y las regiones de 200 pb aguas arriba y aguas abajo de *gfp* (Protoespaciadores S4F, S4R, S5F, S6F, S6R) (Figura 16B), lo que demuestra un rendimiento de CRISPRi extremadamente eficiente y ajustado. La única excepción fue la toma como diana del protoespaciador S5R, que reguló a la baja la expresión de GFP en aproximadamente un 50 %. La expresión de GFP también se reguló a la baja cuando el sistema CRISPR/Cas Tipo-S se dirigió a los protoespaciadores ubicados 1 kb aguas arriba del orf de *gfp* (Protoespaciadores U1kF, U1kR), mientras que la expresión de GFP no se vio afectada en absoluto para los protoespaciadores 2 kb aguas arriba del orf de *gfp* (U2kF, U2kR) (Figura 16C).

El sistema CRISPR/Cas Tipo-S  $\Delta$ Cas-S3 deficiente en helicasa reguló a la baja de forma extensa o completamente la expresión de GFP de b5815 (cepa positiva para *gfp*) cuando se dirigió a la región del promotor p70a o a la cadena directa en el orf de *gfp* (Protoespaciadores S1F, S1R, S2F, S2R, S3F), pero no afectó la expresión de GFP cuando se dirigió a la cadena inversa en el orf de *gfp* (Protoespaciador S3R) (Figura 17A). Aquí, se demostró que Cas-S3 no puede desenrollar el ADN y, por consiguiente, bloquear la transcripción de genes que residen a una distancia de 2 kb del protoespaciador diana.

Para confirmar que la función CRISPRi del sistema CRISPR/Cas Tipo-S depende de Cas-S3 cuando se toman como diana protoespaciadores en el extremo final de, o fuera de, el orf de un gen, se repitieron los ensayos de direccionamiento en los protoespaciadores S4F, S4R, S5F, S5R, S6F y S6R empleando el sistema CRISPR/Cas Tipo-S  $\Delta$ Cas-S3 (Figura 17B). De hecho, la expresión de GFP de b5815 (cepa positiva para *gfp*) no se vio afectada para las dianas seleccionadas en ausencia de Cas-S3. Estos resultados indican que Cas-S3 no es esencial para la formación del complejo RNP del sistema CRISPR/Cas Tipo-S y su unión al ADN diana.

El sistema CRISPR/Cas Tipo-S  $\Delta$ Cas-S4 no afectó la expresión de GFP de b5815 (cepa positiva para *gfp*), lo que indica que la subunidad es esencial para la formación del complejo RNP del sistema CRISPR/Cas Tipo-S y/o su unión al ADN diana (Figura 18A).

El sistema CRISPR/Cas Tipo-S  $\Delta$ Cas-S1 y el sistema CRISPR/Cas Tipo-S  $\Delta$ Cas-S1,  $\Delta$ Cas-S3 redujo la expresión de GFP de b5815 (cepa positiva para *gfp*) de manera similar y solo cuando se dirigieron (completa o parcialmente) a la cadena inversa del promotor p70a (Figura 18B, 18C). Estos resultados indican que Cas-S1 no es esencial para la formación del complejo RNP del sistema CRISPR/Cas Tipo-S y su unión al ADN diana, aunque es importante para su estabilidad y/o la eficiencia de su actividad de unión al ADN. Además, estos resultados indican que Cas-S1 es vital para la carga de Cas-S3 al complejo.

#### 4.5 CRISPRi plasmídico

Se empleó la cepa b6386 para los ensayos CRISPRi del sistema CRISPR/Cas Tipo-S plasmídico. b6386 es una cepa de *E. coli* MG1655 transformada con el plásmido p2370 que contiene el gen *gfp* (bajo el control del promotor p70a, SEQ ID No:15), el origen de replicación p15a y el gen marcador de resistencia al cloranfenicol *cmR*. Solo cuatro de las cinco variantes de Tipo-S ensayadas para CRISPRi cromosómico en el Ejemplo 4.4 también se ensayaron para CRISPRi plasmídico. El sistema CRISPR/Cas Tipo-S de tipo salvaje que comprende Cas-S1-Cas-S5 no se ensayó debido a su actividad demostrada para inhibir la replicación del plásmido (véase el Ejemplo 1.3.2 y 1.3.3). Se seleccionaron seis protoespaciadores dentro de la región promotora p70a y el gen *gfp* en b6386 y se tomaron como diana para los ensayos CRISPRi en plásmido, al igual que para los ensayos CRISPRi cromosómicos; tres de los protoespaciadores estaban localizados en la cadena codificante de *gfp* (a saber, Protoespaciador S1F, S2F, S3F) y tres en la cadena no codificante (a saber, Protoespaciador S1R, S2R, S3R) (véase la Figura 15).

##### 4.5.1 Métodos

La construcción de los diversos plásmidos que expresan Tipo-S que se dirigen a estos protoespaciadores se realizó como se describió previamente en el Ejemplo 4.4.1.

La transformación de los plásmidos se realizó como se describe en el Ejemplo 4.4.1, con los dos cambios siguientes:

1. Se sembraron 100  $\mu$ L de cada recuperación en placas de agar LB suplementadas con 20  $\mu$ g/mL de cloranfenicol, en lugar de 10  $\mu$ g/mL de tetraciclina.
2. Los cultivos de inoculación se suplementaron con 20  $\mu$ g/mL de cloranfenicol, en lugar de 10  $\mu$ g/mL de tetraciclina.
3. Los datos finales se representaron en los gráficos presentados en la Figura 19.

##### 4.5.2 Resultados

El Tipo-S  $\Delta$ Cas-S3 reguló a la baja, de manera completa o eficiente (pero no completa), la expresión de GFP de b6386

(cepa positiva para *gfp*) solo cuando se dirigió a los protoespaciadores dentro de la región del promotor p70a (S1F y S1R), independientemente de la orientación de la cadena diana (Figura 19A). Ninguna de las otras variantes ensayadas reguló a la baja la expresión de GFP (Figura 19B, 19C, 19D). Esto contrasta completamente con los resultados obtenidos mediante los experimentos CRISPRi cromosómicos.

## 5 Ejemplo 5: Sistema CRISPR/Cas Tipo-S en aplicaciones de edición de bases

### 5.1 Introducción

Las herramientas de edición de bases CRISPR/Cas desarrolladas y reportadas actualmente se basan en los módulos efectores de subunidad única de diferentes sistemas CRISPR/Cas de Clase 2. Los sistemas CRISPR/Cas de Clase 1 tienen módulos efectores con múltiples subunidades, lo que complica el desarrollo de herramientas de edición de bases. En consecuencia, las capacidades de edición de bases de los editores de bases de Clase 1, en lo que respecta, por ejemplo, a la especificidad, eficiencia y ventana de edición, han permanecido sin explorar.

En este Ejemplo, se estudió ampliamente el potencial del sistema CRISPR/Cas Tipo-S para transformarse en una plataforma de edición de base flexible. Se diseñó y construyó una colección de editores de bases fusionando, con un conector XTEN de 16 aminoácidos de longitud (SEQ ID No:16 y 17, véase Schellenberger *et al.*, Nat. Biotechnol., 27(12), 1186-1190, 2009, doi:10.1038/nbt.1588), la citidina desaminasa PmCDA1 de lamprea marina (SEQ ID No:18 y 19, véase Nishida *et al.*, Science, 353(6305), aaf8729, 2016, doi:10.1126/science.aaf8729) al extremo N o C de las subunidades Cas-S1, Cas-S3 o Cas-S4 del sistema CRISPR/Cas Tipo-S. La proteína inhibidora de la uracil ADN glicosilasa (UGI) (SEQ ID No:20 y 21, véase Mol *et al.*, Cell, 82(5), 701-8, 1995, doi:10.1016/0092-8674(95)90467-0) se fusionó al extremo C de cada quimera con un conector de 10 aminoácidos de longitud (SEQ ID No:22 y 23). Posteriormente, se ensayó la capacidad de estos editores de bases para generar modificaciones cromosómicas dirigidas de C a T.

### 5.2 Diseño experimental

#### 5.2.1 Descripción general de la cepa de cribado y los plásmidos:

Se insertó un gen *gfp* en el cromosoma de *E. coli* MG1655 bajo el control transcripcional del promotor p70a (SEQ ID No:15). La cepa b5815 resultante (positiva para *gfp*) se empleó como cepa de ensayo para todos los ensayos de edición de bases. Los seis protoespaciadores diana para los ensayos de edición de bases se seleccionaron para que tuvieran un PAM 5'-AAG-3' en su extremo 5' y se ubicaran dentro de una región de 249 pb de largo, compuesta por los últimos 52 pb del promotor p70a y los primeros 197 pb de la secuencia de *gfp*. Tres de los protoespaciadores estaban en la cadena molde (con respecto a la transcripción) y tres estaban en la cadena no molde (Figura 15).

Se diseñaron y construyeron dos categorías de plásmidos para los ensayos de edición de bases:

La 1ª categoría contenía los plásmidos "editores de bases" responsables de la expresión del gen *pm\_cda1* que se fusionó al extremo N o C del gen de Cas-S1, Cas-S3 o Cas-S4 (véase la Figura 20, lado izquierdo). La expresión de las quimeras se estableció bajo el control transcripcional del promotor inducible por arabinosa (*pBAD*, SEQ ID No:24, véase Guzman *et al.*, J. Bacteriol., 177(14), 4121-4130, 1995, doi: 10.1128/jb.177.14.4121-4130.1995). El núcleo de los plásmidos contenía el gen marcador de resistencia al cloranfenicol (*crm<sup>R</sup>*, para la selección con el antibiótico cloranfenicol, y el origen de replicación *p15a*.

Todas las reacciones de PCR se realizaron con la Mezcla Maestra Q5 Hi Fi 2X (NEB), con cebadores y moldes de PCR diseñados apropiadamente para construir los plásmidos. Los fragmentos de PCR se emplearon para las reacciones de la Mezcla Maestra de Ensamblaje de ADN NEBuilder® HiFi. Se transformaron 5 µL de cada reacción de ensamblaje con electroporación en 50 µL de células electrocompetentes NEB10β (NEB). Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y agitando a 200 rpm. Se retiraron 100 µL de cada recuperación y se sembraron en placas de agar LB suplementadas con 20 µg/mL de cloranfenicol. Las placas se incubaron durante la noche a 37 °C.

Al día siguiente, se seleccionaron al azar colonias individuales y se utilizaron para la inoculación de cultivos de 10 mL de medio líquido LB suplementado con 20 µg/mL de cloranfenicol. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm).

Al día siguiente, se prepararon las soluciones madre de glicerol para cada cultivo y los cultivos restantes se utilizaron para el aislamiento de plásmidos con el mini kit de plásmidos de Qiagen. La secuencia de cada plásmido se verificó con NGS.

La 2ª categoría contenía los plásmidos "guía" responsables de la expresión de i) los módulos de expresión de ARNcr (uno para cada uno de los controles no dirigidos, S1F, S1R, S2F, S2R, S3F y S3R) y ii) diferentes versiones del operón Cas Tipo-S con la delección de los genes Cas-S1, Cas-S3 o Cas-S4 (véase la Figura 20, plásmidos de la derecha). La expresión de los ARNcr se estableció bajo el control de la secuencia líder (promotora) original tal como se encontró en el locus CRISPR del sistema CRISPR/Cas Tipo-S. La expresión de los operones se estableció bajo el control transcripcional del promotor constitutivo del gen *bolA* (SEQ ID No:14) presente en el genoma de *E. coli* (*p<sub>bolA</sub>*). El

núcleo de los plásmidos contiene el gen marcador de resistencia a la tetraciclina ( $tet^R$ ), para la selección con el antibiótico tetraciclina, y el origen de replicación pSC101.

La metodología de construcción para los plásmidos "guía" fue la misma que para la 1ª categoría, excepto que (i) se utilizaron 10 µg/mL de tetraciclina para la siembra en agar LB, en lugar de 20 µg/mL de cloranfenicol y (ii) se utilizaron 10 µg/mL de tetraciclina para la inoculación del medio líquido LB, en lugar de 20 µg/mL de cloranfenicol.

Posteriormente, los plásmidos p2062 (control no dirigido,  $\Delta$ Cas-S1), p2064 (control no dirigido,  $\Delta$ Cas-S3), p2065 (control no dirigido,  $\Delta$ Cas-S4) y p2160 (control no dirigido,  $\Delta$ Cas-S1,  $\Delta$ Cas-S3) se sometieron a digestión de restricción con BsmBI-v2 (NEB) y los productos de digestión se purificaron en gel con el kit de Recuperación de ADN en Gel Zymoclean (Zymoresearch). Los oligos diseñados adecuadamente se hibridaron y se ligaron en los plásmidos digeridos con ADN ligasa T4 (NEB). Se transformaron 5 µL de cada reacción de ligación con electroporación en 50 µL de células electrocompetentes NEB10β (NEB). Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y agitando a 200 rpm. Se retiraron 100 µL de cada recuperación y se sembraron en placas de agar LB suplementadas con 10 µg/mL de tetraciclina. Las placas se incubaron durante la noche a 37 °C.

Al día siguiente, se seleccionaron al azar colonias individuales y se utilizaron para la inoculación de cultivos de 10 mL de medio líquido LB suplementado con 10 µg/mL de tetraciclina. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm).

Al día siguiente, se prepararon las soluciones madre de glicerol para cada cultivo y los cultivos restantes se utilizaron para el aislamiento de plásmidos con el mini kit de plásmidos de Qiagen. Se verificó la secuencia de cada plásmido con NGS.

### 5.2.2 Ensayos de edición de bases

Cada uno de los 6 plásmidos "editores de bases" que tenían la citidina desaminasa PmCDA1 fusionada al extremo N o C de cada una de las subunidades Tipo-S wt Cas-S1, Cas-S3 y Cas-S4 (véase la Figura 20, lado izquierdo) se transformó en células electrocompetentes b5815. Las células electrocompetentes se prepararon utilizando el protocolo descrito en <https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Las reacciones de transformación se realizaron utilizando 40 ng de cada plásmido por reacción. Se dejó que las células se recuperaran en 500 µL de medio SOC durante 1 hora a 37 °C y agitando a 200 rpm. Se sembraron 100 µL de cada recuperación en placas de agar LB suplementadas con 20 µg/mL de cloranfenicol. Las placas se incubaron durante la noche a 37 °C.

Al día siguiente, se seleccionaron al azar colonias individuales y se utilizaron para la inoculación de cultivos de 10 mL de medio líquido LB suplementado con 20 µg/mL de cloranfenicol. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm).

Al día siguiente, se prepararon células electrocompetentes para cada una de las seis cepas b5815 producidas con los plásmidos editores de bases, según el protocolo descrito en (<https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Posteriormente, cada una de las seis cepas se electro-transformó con el grupo correspondiente de plásmidos guía, como se presenta en el lado derecho de la Figura 20. Estos plásmidos eran responsables de la expresión de i) subunidades Tipo-S que no son expresadas por los plásmidos editores de bases de apareamiento y ii) ARNcr de direccionamiento y no direccionamiento. Las reacciones de transformación se realizaron utilizando 40 ng de cada plásmido por reacción. Se dejó que las células se recuperaran en 500 µL de medio SOC durante 1 hora a 37 °C y agitando a 200 rpm. Se sembraron 100 µL de cada recuperación en placas de agar LB suplementadas con 20 µg/mL de cloranfenicol, 10 µg/mL de tetraciclina y arabinosa al 0,2 % v/p. Las placas se incubaron durante la noche a 37 °C.

Al día siguiente, se seleccionaron al azar colonias individuales y se las sometió a PCR de colonias con cebadores específicos del genoma diseñados apropiadamente. Los productos de PCR se secuenciaron mediante secuenciación de Sanger y los resultados se analizaron para detectar modificaciones de C a T o modificaciones de G a A cuando se tomaba como diana y se modificaba la cadena inversa.

## 5.3 Resultados

Los resultados representativos de los ensayos de edición de bases se presentan en la Figura 21.

A continuación, se proporciona un análisis detallado de los datos de secuenciación de Sanger de todos los ensayos de edición de bases:

### 5.3.1 PmCDA1 ligado al extremo N de

- Cas-S1 (Tipo-S WT): sin edición de bases
- Cas-S3 (Tipo-S WT): mutación única de C a T en solo 1 de los protoespaciadores diana

- Cas-S4 (Tipo-S WT): mutación única de C a T en solo 1 de los protoespaciadores diana (igual que para el CDA N term de Cas-S3)

### 5.3.2 PmCDA1 ligado al extremo C de

- **Cas-S1:**
  - 5
    - Protoespaciador S1F: Sin edición de bases
    - Protoespaciador S1R: Sin edición de bases
    - Protoespaciador S2F: Edición de bases de múltiples posiciones (ventana de edición de 105 nt) con diferentes niveles de eficiencia para cada posición. No solo conversiones de C a T, sino también algunas conversiones de G a A
  - 10
    - Protoespaciador S2R: 4 posiciones con edición de bases (ventana de edición de 29 nt) con diferentes niveles de eficiencia para cada posición. No solo conversión de G a A sino también 1 de C a T
    - Protoespaciador S3F: 8 posiciones con edición de bases (ventana de edición de 35 nt) con diferentes niveles de eficiencia para cada posición. No solo conversión de C a T sino también 1 de G a A
    - 15
      - Protoespaciador S3R: pocas posiciones con edición de bases (fuera de la región del protoespaciador) con baja eficiencia. El rayado mejoró la eficiencia y extendió la ventana de edición (96 nt). Se detectó una mezcla de conversiones de C a T y de G a A.
- **Cas-S1 (en ausencia de Cas-S3: Tipo-S<sub>Δ</sub>Cas-S3):**
  - Protoespaciador S1F: Sin edición de bases
  - Protoespaciador S1R: Modificación única de G a A dentro del protoespaciador
  - 20
    - Protoespaciador S2F: Se modificaron múltiples posiciones de C a T, con diferentes niveles de eficiencia, dentro y 7 nt aguas abajo de la región del protoespaciador. Se detectaron 2 conversiones de C a T adicionales dentro de una ventana de edición de 102 nt aguas arriba del extremo 5' del protoespaciador.
  - 25
    - Protoespaciador S2R: 4 posiciones con edición de bases (ventana de edición de 48 nt) con diferentes niveles de eficiencia para cada posición. Solo 1 modificación débil de G a A en la posición proximal de 3 PAM y 3 modificaciones eficientes de C a T aguas abajo del protoespaciador
    - Protoespaciador S3F: múltiples modificaciones muy débiles
    - Protoespaciador S3R: se detectaron algunas modificaciones muy débiles (mezcla de conversiones de C a T y de G a A), así como 1 modificación de C a T muy eficiente 40 nt aguas abajo del extremo 3' del protoespaciador.
  - 30
- **Cas-S3:**
  - Protoespaciador S1F: Sin edición de bases
  - Protoespaciador S1R: La posición dominante única con edición de bases en la mitad del protoespaciador y algunas más fueron editadas de manera ineficiente. El rayado extendió la ventana de edición de bases (50 nt), mientras que se detectaron conversiones no solo de G a A sino también algunas conversiones de C a T
  - 35
    - Protoespaciador S2F: Sin edición de bases
    - Protoespaciador S2R: Múltiples posiciones con edición de bases de manera ineficiente en una ventana de edición amplia (hasta 208 nt). No solo se detectaron conversiones de G a A, sino principalmente de C a T. El rayado mejoró la eficiencia, ya que dio como resultado 3 conversiones limpias (tanto de G a A como de C a T) dentro de una ventana de 178 nt.
    - 40
      - Protoespaciador S3F: Múltiples posiciones con edición de bases (ventana de edición de 388 nt). Las conversiones dentro de la región del protoespaciador fueron las más limpias. Solo se detectaron 2 posiciones con conversiones de G a A.
    - 45
      - Protoespaciador S3R: Inicialmente no se detectó edición de bases, pero el rayado generó múltiples conversiones de G a A y de C a T dentro de una ventana de edición de 388 nt.

- **Cas-S4:**

- Todos los protoespaciadores diana se editaron en múltiples posiciones dentro de la región del protoespaciador. La excepción fue S2R, que también se editó aguas abajo de la región del protoespaciador, y S3R, que tenía una única mutación de G a A en la 2ª posición distal de PAM del protoespaciador.
- Los protoespaciadores S1F y S2R tenían mezclas de modificaciones de C a T y de G a A, mientras que el resto de los protoespaciadores tenían modificaciones de C a T o de G a A, dependiendo de la dirección del protoespaciador.

#### 5.4 Conclusiones

Este ejemplo generó la primera plataforma CRISPR/Cas de Clase I reportada para una edición de bases eficiente. Los siete editores de bases desarrollados emplearon la citidina desaminasa PmCDA1. Seis de los editores de bases emplearon el sistema CRISPR/Cas Tipo-S de tipo salvaje y uno de los editores de bases empleó el sistema CRISPR/Cas Tipo-S  $\Delta$ Cas-S3.

La fusión de PmCDA1 al extremo N de Cas-S1 no generó ninguna modificación detectable.

La fusión de PmCDA1 al extremo N de Cas-S4 y Cas-S3 generó solo una única modificación detectable y mixta (genotipo de tipo salvaje/mutante) para uno de los espaciadores ensayados. Estos sistemas podrían utilizarse cuando se requieren mutaciones puntuales muy específicas, pero el bajo número de espaciadores que se pueden utilizar restringe la aplicabilidad de estos sistemas.

La fusión de PmCDA1 al extremo C de Cas-S1, Cas-S4 y Cas-S3 generó tres sistemas de edición de bases con resultados de edición de bases muy diferentes (ventanas de edición, eficiencias de edición y posiciones editadas) incluso en los mismos protoespaciadores. En general, los editores de bases basados en Cas-S1 y Cas-S3 tenían ventanas de edición muy amplias y pueden ser muy valiosos en la generación de bibliotecas de aminoácidos para una sola proteína o la introducción de codones de parada. En ausencia de Cas-S3, la ventana de edición del editor de bases basado en Cas-S1 se redujo y se observó un desplazamiento en las posiciones editadas. El editor de bases basado en Cas-S4 tenía una ventana de edición de bases más estrecha y fue el único editor de bases que introdujo modificaciones en todos los protoespaciadores diana.

#### Ejemplo 6: Funcionalización del sistema CRISPR/Cas Tipo-S para la escisión de ADNds

##### 6.1 Objetivo

El Tipo-S es un sistema CRISPR-Cas deficiente en nucleasas que no posee dominios de nucleasa predichos bioinformáticamente ni determinados experimentalmente (véanse los Ejemplos 1 y 4 en la presente memoria) en ninguna de sus subunidades. El objetivo de este estudio fue expandir el repertorio de reactivos de edición del genoma mediante la creación de nucleasas de ADN guiadas por ARNcr sintéticas al fusionar el dominio de nucleasa de la endonucleasa de localización I-Tevl (SEQ ID No: 44 y 45) con los extremos N de las subunidades CasS1 y CasS4 del sistema Tipo-S deficiente en helicasas ( $\Delta$ CasS3, véase la Figura 22). Las quimeras desarrolladas se ensayaron en dianas de ADN cromosómicas y portadas por plásmidos y se estableció un conjunto de reglas para la escisión eficiente del ADNds.

##### 6.2 Introducción

I-Tevl es una endonucleasa dirigida que se compone de un dominio de nucleasa N-terminal (1-92 aminoácidos (aa)), un dominio espaciador que contiene dedos de zinc (92-169 aa) y un dominio de unión al ADN (169-245 aa) (véase Edgell *et al.*, PNAS, 98(14), 7898-7903, 2001, doi: <https://doi.org/10.1073/pnas.141222498> y Kleinstiver *et al.*, PNAS, 109(21), 8061-8066, 2012, doi: <https://doi.org/10.1073/pnas.111798410>). I-Tevl se une a su ADN diana como un monómero e introduce un corte en el ADNds escalonado específico de secuencia en la escisión de I-Tevl a través del resto GIY-YIG presente en su dominio de nucleasa N-terminal. La secuencia del sitio de escisión de I-Tevl de origen natural es 5'-CAACG-3' (cadena directa)/5'-CGTTG-3' (cadena inversa). Sin embargo, se ha reportado que I-Tevl reconoce un sitio de escisión de I-Tevl 5'-CNNNG-3' (cadena directa)/5'-CNNNG-3' (cadena inversa) más flexible (véase Edgell *et al.*, 2001 y Kleinstiver *et al.* 2012, ambos *supra*). Se planteó la hipótesis de que al fusionar los dominios N-terminal y de dedo de zinc/conector de I-Tevl al extremo N de CasS1 o CasS4, sería posible crear nucleasas de ADN guiadas por ARNcr TevCasSx (x=1 o 4) sintéticas.

En el Ejemplo 4 se demostró -mediante experimentos CRISPRi- que el sistema Tipo-S  $\Delta$ CasS3 deficiente en helicasa retiene una fuerte actividad de unión al ADN. Además, en el mismo ejemplo, se demostró que el sistema Tipo-S  $\Delta$ CasS1 que expresa helicasa muestra la misma eficiencia de unión a CRISPRi/ADN que el sistema Tipo-S  $\Delta$ CasS3 deficiente en helicasa, mientras que el sistema Tipo-S  $\Delta$ CasS1 $\Delta$ CasS3 muestra una actividad de unión a CRISPRi/ADN reducida. Estos resultados respaldan un modelo en donde CasS1 está involucrada en el reclutamiento de CasS3 al complejo Tipo-S. Por lo tanto, se decidió emplear el sistema Tipo-S  $\Delta$ CasS3 deficiente en helicasa para la construcción de las nucleasas sintéticas TevCasSx, con el objetivo de evitar el impedimento de la actividad nucleasa

de las quimeras TevCasSx por el reclutamiento o la actividad helicasa de CasS3.

### 6.3 Diseño experimental

#### 6.3.1 Ensayos de escisión del ADN del plásmido por TevCasS

Se diseñaron, construyeron y emplearon tres categorías de plásmidos para los ensayos de escisión de ADN plasmídico por TevCasS; i) los plásmidos diana de TevCasS, ii) los plásmidos que expresan TevCasSx y iii) los plásmidos guía Tipo-S.

Se diseñaron 23 plásmidos diana de TevCasS (denominados p2497 - p2510, p2575 - p2578 y p2655 - p2659, véase la Figura 22, plásmidos de la derecha) para contener varios sitios diana de TevCasS (SEQ ID No: 46 a 68, respectivamente, véase la Tabla 4) clonados por separado en una estructura de plásmido de bajo número de copias, compuesta por el replicón pBBR1 de bajo número de copias y el gen marcador de resistencia a la kanamicina. El 1<sup>er</sup> plásmido diana de TevCasS, denominado p2497 (Tabla 4), se diseñó para abarcar un sitio diana de 71 nt de longitud con la siguiente estructura -en dirección 5' a 3'-: i) el sitio de escisión de I-TevI de 5 nt de longitud 5'-CAACG-3' (SEQ ID No:42) seguido de ii) la secuencia espaciadora de ADN natural de I-TevI de 31 nt de longitud 5'-CTCAGTAGATGTTTTCTTGGGTCTACCGTTA-3' (SEQ ID No:46), que interactúa con el dominio de dedo de zinc de I-TevI, seguido de iii) la secuencia PAM Tipo-S de 3 nt de longitud 5'-AAG-3' (SEQ ID No:13) seguida de iv) la secuencia protoespaciadora de 32 nt de longitud 5'-TCGTGTTGTCCACGGTTACCCGCTGGCTGGAA-3' (SEQ ID No:69), que es complementaria a la secuencia del 1<sup>er</sup> espaciador en la matriz CRISPR natural del sistema CRISPR-Cas Tipo-S. Los sitios diana de los plásmidos diana de TevCasS adicionales (Tabla 4, SEQ ID No: 46 a 68) tenían el mismo sitio de escisión de I-TevI, porciones PAM y protoespaciador que la diana en el plásmido p2497, pero sus porciones de espaciador de ADN I-TevI eran diferentes; se eliminaron o añadieron varios números de nucleótidos al extremo 3' de la secuencia espaciadora de ADN I-TevI de p2497, lo que resultó en la construcción de plásmidos diana con secuencias espaciadoras de ADN de I-TevI con una longitud de 5nt (SEQ ID No:59), 7nt (SEQ ID No:58), 9nt (SEQ ID No:57), 11nt (SEQ ID No:56), 13nt (SEQ ID No:55), 15nt (SEQ ID No:54), 17nt (SEQ ID No:53), 19nt (SEQ ID No:52), 21nt (SEQ ID No:51), 23nt (SEQ ID No:50), 25nt (SEQ ID No:49), 27nt (SEQ ID No:48), 28nt (SEQ ID No:64), 29nt (SEQ ID No:47), 30nt (SEQ ID No:65), 31nt (SEQ ID No: 46), 32nt (SEQ ID No: 66), 33nt (SEQ ID No: 60), 34nt (SEQ ID No: 67), 35nt (SEQ ID No: 61), 36nt (SEQ ID No: 68), 37nt (SEQ ID No: 62) o 39nt (SEQ ID No: 63).

Para la construcción de los plásmidos diana de I-TevI, todas las reacciones de PCR se realizaron con Mezcla Maestra Q5 Hi Fi 2x (NEB) y cebadores y moldes de PCR diseñados apropiadamente. Los fragmentos de PCR se emplearon para las reacciones de Mezcla Enzimática KLD (NEB). Se transformaron 5 µL de cada reacción con electroporación en 50 µL de células electrocompetentes NEB10β (NEB). Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y agitando a 200 rpm. Se retiraron 100 µL de cada recuperación y se sembraron en placas de agar LB suplementadas con 50 µg/mL de kanamicina. Las placas se incubaron durante la noche a 37 °C.

Al día siguiente, se seleccionaron al azar colonias individuales y se utilizaron para la inoculación de cultivos de 10 mL de medio líquido LB suplementado con 50 µg/mL de kanamicina. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm).

Al día siguiente, se prepararon las soluciones madre de glicerol para cada cultivo y los cultivos restantes se utilizaron para el aislamiento de plásmidos con el mini kit de plásmidos de Qiagen. La secuencia de cada plásmido se verificó con NGS.

Los plásmidos que expresaban TevCasS1 (p2461) y TevCasS4 (p2462) (véase la Figura 22, plásmidos del medio) se diseñaron para tener los 618 pb iniciales del gen *I-tevI* (SEQ ID No:44), que expresan los 206 aminoácidos N-terminales de la nucleasa I-TevI (SEQ ID No:45), ligados al extremo N de los genes *casS1* y *casS4* con una secuencia que expresa el conector XTEN (SEQ ID No:16 y 17). La expresión de las quimeras se estableció bajo el control transcripcional del promotor inducible por arabinosa (pBAD, SEQ ID No:24) (véase Guzman *et al.*, J. Bacteriol., 177(14), 4121-4130, 1995, doi:10.1128/jb.177.14.4121-4130.1995). La estructura principal de los plásmidos construidos contenía el gen marcador de resistencia al cloranfenicol (*crm<sup>R</sup>*), para la selección con el antibiótico cloranfenicol, y el origen de replicación p15a.

La metodología de construcción para los plásmidos de expresión fue la misma que para los plásmidos diana, excepto que (i) se utilizó la Mezcla Maestra de Ensamblaje de ADN HiFi NEBuilder® en lugar de la Mezcla Enzimática KLD (NEB), (ii) se utilizaron 20 µg/mL de cloranfenicol para la siembra en agar LB, en lugar de 50 µg/mL de kanamicina, y (iii) se utilizaron 20 µg/mL de cloranfenicol para la inoculación del medio líquido LB, en lugar de 50 µg/mL de kanamicina.

Se diseñaron cuatro plásmidos guía Tipo-S (véase la Figura 22, plásmidos de la izquierda); dos controles no dirigidos (negativos) y dos plásmidos dirigidos. Los plásmidos guía fueron responsables de la expresión de i) los módulos de expresión de ARNcr (ya sea un ARNcr de control no dirigido o un ARNcr que se dirige al protoespaciador dentro de la secuencia diana de TevCasS) y ii) la versión ΔCasS1 ΔCasS3 o la versión ΔCasS1 ΔCasS4 del operón Cas Tipo-S (Tabla 2). La expresión de los ARNcr se estableció bajo el control de la secuencia líder (promotora) original tal como se encontró en el locus CRISPR del sistema CRISPR/Cas Tipo-S de tipo salvaje. La expresión de los operones se

estableció bajo el control transcripcional del promotor constitutivo del gen *bolA* (SEQ ID No:14) presente en el genoma de *E. coli* (*pbolA*). Los plásmidos guía comparten la misma estructura principal, que contiene el gen marcador de resistencia a la tetraciclina (*tetR*), para la selección con el antibiótico tetraciclina, y el origen de replicación pSC101 de bajo número de copias.

- 5 La metodología de construcción para los plásmidos guía fue la misma que para los plásmidos de expresión, excepto que (i) se utilizaron 10 µg/mL de tetraciclina para la siembra en agar LB, en lugar de 20 µg/mL de cloranfenicol y (ii) se utilizaron 10 µg/mL de tetraciclina para la inoculación del medio líquido LB, en lugar de 20 µg/mL de cloranfenicol.

Para los ensayos de escisión del ADN plasmídico por TevCasS, las siguientes combinaciones de plásmidos se transformaron con electroporación en la cepa de *E. coli* MG1655 (véase también la Figura 22, lado izquierdo):

- 10 1. p2461 (que expresa TevCasS1) y p2160 (guía de control no dirigido)
2. p2461 (que expresa TevCasS1) y p2511 (guía dirigida de TevCasS)
3. p2462 (que expresa TevCasS4) y p2460 (guía de control no dirigido)
4. p2462 (que expresa TevCasS4) y p2512 (guía dirigida de TevCasS)

15 Las células electrocompetentes MG1655 se prepararon utilizando el protocolo descrito en (<https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Las reacciones de transformación se realizaron utilizando 40 ng de cada plásmido por reacción. Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y agitando a 200 rpm. Se retiraron 100 µL de cada recuperación y se sembraron en placas de agar LB suplementadas con 20 µg/mL de cloranfenicol y 10 µg/mL de tetraciclina. Las placas se incubaron durante la noche a 37 °C. Al día siguiente, se seleccionaron al azar colonias individuales y se utilizaron para la inoculación de cultivos de 10 mL de medio líquido LB suplementado con 20 µg/mL de cloranfenicol y 10 µg/mL de tetraciclina. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm). Al día siguiente, se prepararon soluciones madre de glicerol para cada cultivo y los cultivos se utilizaron para la preparación de plásmidos de células electrocompetentes según el protocolo descrito en (<https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Cada una de las cuatro cepas se transformó con cada uno de los plásmidos diana (Figura 22 lado derecho) y el plásmido de control p2027, que es idéntico a los plásmidos diana, pero no contiene ninguna secuencia diana de TevCasS. Las recuperaciones de cada cepa se utilizaron para la preparación de diluciones seriadas, que posteriormente se sembraron en placas de agar LB suplementadas con 20 µg/mL de cloranfenicol, 10 µg/mL de tetraciclina y 50 µg/mL de kanamicina, para la selección de plásmidos. Las transformaciones se realizaron en triplicados biológicos. La introducción exitosa de una rotura en ADNs en un plásmido diana daría como resultado la eliminación del plásmido y, posteriormente, la inhibición del crecimiento en presencia de kanamicina.

Se contaron las colonias formadas a partir de cada transformación. La reducción en las eficiencias de transformación de las dos cepas que expresan el ARNcr de direccionamiento, en comparación con las dos cepas que expresan el ARNcr no de direccionamiento, se calculó en 2 etapas: 1) Normalización de la eficiencia de transformación de cada cepa para cada plásmido diana por división por la eficiencia de transformación correspondiente del plásmido de control p2027, y 2) División de las eficiencias de transformación normalizadas de las cepas con el ARNcr de direccionamiento por las eficiencias de transformación normalizadas correspondientes de las cepas con el ARNcr no de direccionamiento.

### 6.3.2 Ensayos de escisión de ADN cromosómico por TevCasS

40 Las secuencias diana de TevCasS con espaciadores I-TevI con una longitud de 29 nt (SEQ ID No:47), 30 nt (SEQ ID No:65), 31 nt (SEQ ID No:46), 32 nt (SEQ ID No:66) y 33 nt (SEQ ID No:60) se insertaron en el cromosoma de la cepa de *E. coli* MG1655, generando cinco cepas diana cromosómicas de TevCasS, cada una conteniendo una diana diferente. Todas las cepas construidas se transformaron con el plásmido p2461 (que expresa TevCasS1). Las células electrocompetentes se prepararon utilizando el protocolo descrito en (<https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Las reacciones de transformación se realizaron utilizando 40 ng de cada plásmido por reacción. Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y agitando a 200 rpm. Se retiraron 100 µL de cada recuperación y se sembraron en placas de agar LB suplementadas con 20 µg/mL de cloranfenicol. Las placas se incubaron durante la noche a 37 °C.

50 Al día siguiente, se seleccionaron al azar colonias individuales y se utilizaron para la inoculación de cultivos de 10 mL de medio líquido LB suplementado con 20 µg/mL de cloranfenicol. Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm).

Al día siguiente, se prepararon soluciones madre de glicerol para cada cultivo y los cultivos se utilizaron para la preparación de plásmidos de células electrocompetentes según el protocolo descrito en (<https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Cada una de las cepas resultantes se transformó con los plásmidos p2160 (guía de control no dirigida) y p2511 (guía dirigida de TevCasS) (véase la Figura

22). Las recuperaciones de cada cepa se utilizaron para la preparación de diluciones seriadas, que posteriormente se sembraron en placas de agar LB suplementadas con 20 µg/mL de cloranfenicol y 10 µg/mL de tetraciclina. Las transformaciones se realizaron en triplicados biológicos. La introducción exitosa de roturas en el ADNds en las dianas cromosómicas de TevCasS serían letales y darían como resultado una reducción en el número de colonias sobrevivientes después de la siembra.

Se contaron las colonias formadas a partir de cada transformación. Las eficiencias de transformación de las cepas transformadas con el plásmido guía de ARNcr de direccionamiento se dividieron por las eficiencias de transformación de las cepas transformadas con el plásmido guía de ARNcr de direccionamiento. Estas divisiones revelaron la reducción en las eficiencias de transformación de las cepas y, posteriormente, la eficiencia de direccionamiento del sistema TevCasS1 sobre dianas cromosómicas.

### 6.3.3 Identificación del sitio de escisión de TevCasS

I-TevI reconoce un sitio de escisión flexible 5'-CNNNG-3' (cadena directa)/5'-CNNNG-3' (cadena inversa). Se decidió identificar los sitios de escisión de I-TevI que TevCasS1 reconoce y escinde, mediante un ensayo de aclaramiento de plásmidos similar al descrito en el Ejemplo 2 para la identificación de PAM Tipo-S.

Se diseñó y construyó una biblioteca de plásmidos diana de TevCasS; cada miembro de esta biblioteca es idéntico al plásmido diana p2497, excepto por la secuencia del sitio de escisión de I-TevI, que, en lugar de 5'-CAACG-3' es 5'-CNNNG-3', donde N es cualquier nucleótido. Por lo tanto, la biblioteca debería estar compuesta por 4<sup>3</sup> (64) sitios de escisión de I-TevI diferentes.

Para la construcción de la biblioteca, p2497 se amplificó por PCR (premezcla de PCR CloneAMP) con cebadores diseñados apropiadamente. El producto de PCR purificado en gel se utilizó para la reacción InFusion (Takarabio) y 5 µL de la mezcla de reacción se transformaron en 50 µL de células electrocompetentes NEB10β (NEB). Las células transformadas se recuperaron en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y 200 rpm y posteriormente se transfirieron a 500 mL de medio LB suplementado con 50 µg/mL de kanamicina. 100 µL del cultivo de 500 mL se sembraron en placas de agar LB suplementadas con 50 µg/mL de kanamicina. El cultivo y la placa se incubaron durante la noche a 37 °C, el primero con agitación rigurosa (200 rpm).

Al día siguiente, se contaron las colonias en la placa (180 colonias) y se calculó que se añadieron 900.000 células transformadas al cultivo de 500 mL, utilizando la Ecuación 1.

Con este número de células transformadas en el cultivo, la cobertura de los miembros de la biblioteca se calculó en 14.062x según la Ecuación 2.

Se prepararon soluciones madre de glicerol a partir del cultivo de 500 mL y se realizaron preparaciones midi para aislar la biblioteca de plásmidos diana de TevCasS1. La biblioteca de plásmidos se transformó posteriormente en las cepas de *E. coli* MG1655 que albergan las siguientes combinaciones de plásmidos:

Cepa no. 1. p2461 (que expresa TevCasS1) y p2160 (guía de control no dirigido)

Cepa no. 2. p2461 (que expresa TevCasS1) y p2511 (guía dirigida de TevCasS)

Se espera que algunos o todos los miembros de la biblioteca de plásmidos diana contengan sitios de escisión que sean reconocidos por TevCasS1. Estos plásmidos diana deberían ser atacados de manera eficiente tras la transformación en la cepa no. 2, dificultando el crecimiento de las células transformadas en condiciones adecuadas de selección con antibióticos. Ninguno de los miembros de la biblioteca diana debería ser atacado tras la transformación de la biblioteca en la cepa no. 1.

El aislamiento de plásmidos de cultivos de la cepa no. 2 transformada con la biblioteca de plásmidos diana, seguido de NGS, muestra qué secuencias del sitio de escisión de TevCasS no son reconocidas por TevCasS y, posteriormente, cuáles son las secuencias que el sistema reconoce y escinde.

El aislamiento de plásmidos a partir de cultivos de la cepa no. 1 transformados con la biblioteca de plásmidos diana, seguido de NGS, indicará la abundancia de cada miembro plasmídico en la biblioteca de plásmidos diana. Esta información se utilizará para la normalización de los resultados de NGS de la cepa no. 2.

Experimentalmente, se prepararon células electrocompetentes para ambas cepas utilizando el protocolo descrito en (<https://barricklab.org/twiki/bin/view/Lab/ProtocolsElectrocompetentCells>). Se realizaron cinco reacciones de transformación por cepa, utilizando 40 ng de la biblioteca de plásmidos por reacción. Posteriormente, se dejó que las células se recuperaran en medio SOC (volumen final de 500 µL) durante 1 hora a 37 °C y agitando a 200 rpm. Las células recuperadas de cada cepa se combinaron y se utilizaron para la inoculación de cultivos de 500 mL de LB suplementados con 50 µg/mL de kanamicina (un cultivo por cepa). Los cultivos se incubaron durante la noche a 37 °C con agitación rigurosa (200 rpm). Al día siguiente, los cultivos se utilizaron para la preparación de soluciones madre de glicerol y se sometieron a aislamiento de plásmidos. Los plásmidos aislados se emplearon para la preparación de muestras de NGS, siguiendo el proceso descrito en el Ejemplo 2, y las muestras preparadas se sometieron a NGS.

## 6.4 Resultados

### 6.4.1 Ensayos de direccionamiento de TevCasS en dianas plasmídicas

Las cepas que expresan la quimera TevCasS1 y el ARNcr de direccionamiento o el ARNcr no de direccionamiento se transformaron con los plásmidos diana de TevCasS y se compararon sus eficiencias de transformación (Figura 23).

5 La eficiencia de transformación de la cepa diana cayó:

- i) 4 órdenes de magnitud cuando se transforma con plásmidos diana con espaciadores I-Tevl de 30 nt y 31 nt de longitud,
- ii) 3 órdenes de magnitud cuando se transforma con el plásmido diana con espaciador I-Tevl de 32 nt de longitud,
- 10 iii) 2 órdenes de magnitud cuando se transforma con el plásmido diana con espaciador I-Tevl de 33 nt de longitud, y
- iv) 1 orden de magnitud cuando se transforma con el plásmido diana con espaciador I-Tevl de 29 nt de longitud (Figura 23).

15 Ninguno de los plásmidos diana restantes fue diana de manera eficiente (al menos 1 orden de magnitud de caída en la eficiencia de transformación) por la quimera TevCasS1. Ninguno de los plásmidos diana fue diana de manera eficiente (al menos 1 orden de magnitud) por la quimera TevCasS4. Por lo tanto, la quimera TevCasS1 es altamente eficiente en la selección de dianas de TevCasS portadas por plásmidos.

### 6.4.2 Ensayos de direccionamiento de TevCasS en dianas cromosómicas

20 Cinco cepas con integración cromosómica de dianas de TevCasS, que también portaban el plásmido que expresa TevCasS1, se transformaron con el plásmido guía de direccionamiento o con el plásmido guía sin direccionamiento y se compararon sus eficiencias de transformación (Figura 24). Las eficiencias de transformación de las cepas con dianas de TevCasS que tienen:

- i) espaciadores I-Tevl de 30 nt, 31 nt y 32 nt de longitud se redujeron en 4 órdenes de magnitud cuando se transformaron con el plásmido guía de direccionamiento.
- 25 ii) espaciadores I-Tevl de 29 nt y 33 nt de longitud se redujeron en 1 orden de magnitud cuando se transformaron con el plásmido guía de direccionamiento.

Por lo tanto, la quimera TevCasS1 es muy eficiente para tomar como diana, dianas cromosómicas de TevCasS.

### 6.4.3 Identificación del sitio de escisión de TevCasS

30 Los datos generados por NGS se utilizaron para comparar la abundancia relativa de cada miembro de la biblioteca de plásmidos diana en las cepas no. 1 y no. 2. Este análisis reveló las preferencias de TevCasS por secuencias 5'-CNNNG-3' específicas (Figura 25):

5'-CASHG-3'

5'-CATNG-3'

5'-CAAMG-3'

35 5'-CCAYG-3'

5'-CCTHG-3'

5'-CCGTG-3'

5'-CTDHG-3'

5'-CTCCG-3'

40 5'-CGGCG-3'

5'-CGAYG-3'

TABLA 3: BACTERIAS EJEMPLARES

|                                   |                                |                               |                                |
|-----------------------------------|--------------------------------|-------------------------------|--------------------------------|
| <b>Abiotrophia</b>                | <i>Acetobacter ghanensis</i>   | <i>Acetobacterium</i>         | <i>Acholeplasma</i>            |
| <i>Abiotrophia defectiva</i>      | <i>Acetobacter</i>             | <i>dehalogenans</i>           | <i>multilocale</i>             |
| <b>Acaricomes</b>                 | <i>indonesiensis</i>           | <i>Acetobacterium</i>         | <i>Acholeplasma oculi</i>      |
| <i>Acaricomes phytosejuli</i>     | <i>Acetobacter lovanensis</i>  | <i>finnetarium</i>            | <i>Acholeplasma palmae</i>     |
| <b>Acetitomaculum</b>             | <i>Acetobacter malorum</i>     | <i>Acetobacterium</i>         | <i>Acholeplasma parvum</i>     |
| <i>Acetitomaculum ruminis</i>     | <i>Acetobacter</i>             | <i>malicum</i>                | <i>Acholeplasma pleciae</i>    |
| <b>Acetivibrio</b>                | <i>nitrogenifigens</i>         | <i>Acetobacterium</i>         | <i>Acholeplasma vituli</i>     |
| <i>Acetivibrio cellulolyticus</i> | <i>Acetobacter oeni</i>        | <i>paludosum</i>              | <b>Achromobacter</b>           |
| <i>Acetivibrio</i>                | <i>Acetobacter orientalis</i>  | <i>Acetobacterium tundrae</i> | <i>Achromobacter</i>           |
| <i>ethanolignens</i>              | <i>Acetobacter orleanensis</i> | <i>Acetobacterium</i>         | <i>denitrificans</i>           |
| <i>Acetivibrio multivorans</i>    | <i>Acetobacter</i>             | <i>wielingae</i>              | <i>Achromobacter insolitus</i> |
| <b>Acetoanaerobium</b>            | <i>pasteurianus</i>            | <i>Acetobacterium woodii</i>  | <i>Achromobacter</i>           |
| <i>Acetoanaerobium</i>            | <i>Acetobacter pomorum</i>     | <b>Acetofilamentum</b>        | <i>piechaudii</i>              |
| <i>noterae</i>                    | <i>Acetobacter</i>             | <i>Acetofilamentum</i>        | <i>Achromobacter</i>           |
| <b>Acetobacter</b>                | <i>senegalensis</i>            | <i>rigidum</i>                | <i>ruhlandii</i>               |
| <i>Acetobacter aceti</i>          | <i>Acetobacter xylinus</i>     | <b>Acetohalobium</b>          | <i>Achromobacter spanius</i>   |
| <i>Acetobacter cerevisiae</i>     | <b>Acetobacterium</b>          | <i>Acetohalobium</i>          | <b>Acidaminobacter</b>         |
| <i>Acetobacter</i>                | <i>Acetobacterium bakii</i>    | <i>arabaticum</i>             | <i>Acidaminobacter</i>         |
| <i>cibinongensis</i>              | <i>Acetobacterium</i>          | <b>Acetomicrobium</b>         | <i>hydrogenoformans</i>        |
| <i>Acetobacter estunensis</i>     | <i>carbinolicum</i>            | <i>Acetomicrobium faecale</i> | <b>Acidaminococcus</b>         |
| <i>Acetobacter fabarum</i>        |                                |                               | <i>Acholeplasma morum</i>      |

|                                                                                                                                                                                                                                    |                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Acidaminococcus fermentans</i>                                                                                                                                                                                                  | <b>Acidisphaera</b><br><i>Acidisphaera rubrifaciens</i>                                                                                      | <i>Acidothermus cellulolyticus</i>                                                                                                                                                                                                                                                                                                 | <i>Acinetobacter haemolyticus</i>                                                                                                                                                                                                                                                                                                                                                                                                          | <i>Acrocarpospora pleiomorpha</i>                                                                                                                                                                       |
| <i>Acidaminococcus intestinalis</i>                                                                                                                                                                                                | <b>Acidithiobacillus</b><br><i>Acidithiobacillus albertensis</i><br><i>Acidithiobacillus caldus</i><br><i>Acidithiobacillus ferrooxidans</i> | <b>Acidovorax</b><br><i>Acidovorax anthracis</i><br><i>Acidovorax caeni</i><br><i>Acidovorax cailleae</i><br><i>Acidovorax citrulli</i><br><i>Acidovorax defluvii</i><br><i>Acidovorax delafieldii</i><br><i>Acidovorax facilis</i><br><i>Acidovorax korjaci</i><br><i>Acidovorax temperans</i><br><i>Acidovorax valerianellae</i> | <i>Acinetobacter johnsonii</i><br><i>Acinetobacter junii</i><br><i>Acinetobacter lwoffii</i><br><i>Acinetobacter parvus</i><br><i>Acinetobacter</i><br><i>radioresistens</i><br><i>Acinetobacter schindleri</i><br><i>Acinetobacter soli</i><br><i>Acinetobacter tandoii</i><br><i>Acinetobacter</i><br><i>tjernbergiae</i><br><i>Acinetobacter townieri</i><br><i>Acinetobacter ursingii</i><br><i>Acinetobacter</i><br><i>venetianus</i> | <b>Actibacter</b><br><i>Actibacter sediminis</i><br><b>Actinoalloteichus</b><br><i>Actinoalloteichus cyanogriseus</i><br><i>Actinoalloteichus hymeniacidinis</i><br><i>Actinoalloteichus spitiensis</i> |
| <b>Acidimicrobium</b><br><i>Acidimicrobium ferrooxidans</i>                                                                                                                                                                        | <b>Acidobacterium</b><br><i>Acidobacterium</i><br><i>capsulatum</i>                                                                          | <b>Acinetobacter</b><br><i>Acinetobacter baumannii</i><br><i>Acinetobacter baylyi</i><br><i>Acinetobacter bouvetii</i><br><i>Acinetobacter calcoaceticus</i><br><i>Acinetobacter gernerii</i>                                                                                                                                      | <b>Actinobacillus</b><br><i>Actinobacillus capsulatus</i><br><i>Actinobacillus delphiniicola</i><br><i>Actinobacillus hominis</i><br><i>Actinobacillus intollicus</i><br><i>Actinobacillus lignieresii</i><br><i>Actinobacillus minor</i><br><i>Actinobacillus muris</i>                                                                                                                                                                   |                                                                                                                                                                                                         |
| <b>Acidiphilium</b><br><i>Acidiphilium acidophilum</i><br><i>Acidiphilium angustum</i><br><i>Acidiphilium cryptum</i><br><i>Acidiphilium multivorum</i><br><i>Acidiphilium</i><br><i>organovorum</i><br><i>Acidiphilium rubrum</i> | <b>Acidocella</b><br><i>Acidocella aminolytica</i><br><i>Acidocella facilis</i><br><b>Acidomonas</b><br><i>Acidomonas methanolytica</i>      | <b>Acrocarpospora</b><br><i>Acrocarpospora corrugata</i><br><i>Acrocarpospora macrocephala</i>                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                         |
| <b>Acidisoma</b><br><i>Acidisoma sibiricum</i><br><i>Acidisoma tundrae</i>                                                                                                                                                         | <b>Acidothermus</b>                                                                                                                          |                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                         |

|                                        |                                      |                                      |                                    |                                  |
|----------------------------------------|--------------------------------------|--------------------------------------|------------------------------------|----------------------------------|
| <i>Actinobacillus pleuropneumoniae</i> | <i>Actinocorallia aurantiaca</i>     | <i>Actinomadura echinospore</i>      | <i>Actinomadura verrucosospore</i> | <i>Actinomyces mayeri</i>        |
| <i>Actinobacillus porcini</i>          | <i>Actinocorallia aurea</i>          | <i>Actinomadura fibrosa</i>          | <i>Actinomadura vinea</i>          | <i>Actinomyces naeslundii</i>    |
| <i>Actinobacillus rossii</i>           | <i>Actinocorallia cavernae</i>       | <i>Actinomadura formosensis</i>      | <i>Actinomadura viridilutea</i>    | <i>Actinomyces neui</i>          |
| <i>Actinobacillus scotiae</i>          | <i>Actinocorallia glomerata</i>      | <i>Actinomadura hibisca</i>          | <i>Actinomadura viridis</i>        | <i>Actinomyces odontolyticus</i> |
| <i>Actinobacillus seminis</i>          | <i>Actinocorallia herbida</i>        | <i>Actinomadura kijaniata</i>        | <i>Actinomadura yumaensis</i>      | <i>Actinomyces oris</i>          |
| <i>Actinobacillus succinogenes</i>     | <i>Actinocorallia libanotica</i>     | <i>Actinomadura atina</i>            | <b>Actinomyces</b>                 | <i>Actinomyces radigae</i>       |
| <i>Actinobacillus suis</i>             | <i>Actinocorallia longicatena</i>    | <i>Actinomadura livida</i>           | <i>Actinomyces bovis</i>           | <i>Actinomyces slackii</i>       |
| <i>Actinobacillus ureae</i>            | <b>Actinomadura</b>                  | <i>Actinomadura luteofluorescens</i> | <i>Actinomyces denticolens</i>     | <i>Actinomyces turicensis</i>    |
| <b>Actinobaculum</b>                   | <i>Actinomadura alba</i>             | <i>Actinomadura macra</i>            | <i>Actinomyces europaeus</i>       | <i>Actinomyces viscosus</i>      |
| <i>Actinobaculum massiliense</i>       | <i>Actinomadura atramentaria</i>     | <i>Actinomadura madurae</i>          | <i>Actinomyces georgiae</i>        | <b>Actinoplanes</b>              |
| <i>Actinobaculum schaalii</i>          | <i>Actinomadura bangladeshiensis</i> | <i>Actinomadura oligospora</i>       | <i>Actinomyces gerencseriae</i>    | <i>Actinoplanes aurenticolor</i> |
| <i>Actinomyces urinae</i>              | <i>Actinomadura catellatispora</i>   | <i>Actinomadura pelletieri</i>       | <i>Actinomyces hordeovulvulae</i>  | <i>Actinoplanes brasiliensis</i> |
| <b>Actinocatenispora</b>               | <i>Actinomadura chibensis</i>        | <i>Actinomadura rubrobrunea</i>      | <i>Actinomyces howellii</i>        | <i>Actinoplanes consetensis</i>  |
| <i>Actinocatenispora rupis</i>         | <i>Actinomadura chokoriensis</i>     | <i>Actinomadura rugatobispora</i>    | <i>Actinomyces hyovaginalis</i>    | <i>Actinoplanes decanensis</i>   |
| <i>Actinocatenispora thailandica</i>   | <i>Actinomadura civea</i>            | <i>Actinomadura umbrina</i>          | <i>Actinomyces israelii</i>        | <i>Actinoplanes dentriticus</i>  |
| <i>Actinocatenispora sera</i>          | <i>Actinomadura coerulea</i>         |                                      | <i>Actinomyces johnsonii</i>       | <i>Actinoplanes digitatus</i>    |
| <b>Actinocorallia</b>                  |                                      |                                      |                                    |                                  |

|                                     |                                    |                                    |                                   |                                          |
|-------------------------------------|------------------------------------|------------------------------------|-----------------------------------|------------------------------------------|
| <i>Actinoplanes durhamensis</i>     | <i>Actinopolyspora halophila</i>   | <i>Aeromonas bestiarum</i>         | <b>Agromyces</b>                  | <i>Alcanivorax borkumensis</i>           |
| <i>Actinoplanes ferrugineus</i>     | <i>Actinopolyspora mortalis</i>    | <i>Aeromonas caviae</i>            | <i>Agromyces fucosus</i>          | <i>Alcanivorax jadensis</i>              |
| <i>Actinoplanes globosporus</i>     | <b>Actinosynnema</b>               | <i>Aeromonas enterepologenes</i>   | <i>Agromyces luteolus</i>         | <b>Algicola</b>                          |
| <i>Actinoplanes humidus</i>         | <i>Actinosynnema mirum</i>         | <i>Aeromonas eucnerophila</i>      | <i>Agromyces mediolanus</i>       | <i>Algicola bacteriolytica</i>           |
| <i>Actinoplanes italicus</i>        | <b>Actinotalea</b>                 | <i>Aeromonas ichthiosmia</i>       | <i>Agromyces ramosus</i>          | <b>Alicyclobacillus</b>                  |
| <i>Actinoplanes liguricus</i>       | <i>Actinotalea fermentans</i>      | <i>Aeromonas jandaei</i>           | <i>Agromyces rhizosphaerae</i>    | <i>Alicyclobacillus disulfidooxidans</i> |
| <i>Actinoplanes lobatus</i>         | <b>Aerococcus</b>                  | <i>Aeromonas media</i>             | <b>Akkermansia</b>                | <i>Alicyclobacillus sendaiensis</i>      |
| <i>Actinoplanes missouriensis</i>   | <i>Aerococcus sanguinicola</i>     | <i>Aeromonas popoffii</i>          | <i>Akkermansia muciniphila</i>    | <i>Alicyclobacillus vulcanalis</i>       |
| <i>Actinoplanes palleronii</i>      | <i>Aerococcus urinae</i>           | <i>Aeromonas sobria</i>            | <b>Albidiferax</b>                | <b>Alishewanella</b>                     |
| <i>Actinoplanes philippinensis</i>  | <i>Aerococcus urinaequi</i>        | <i>Aeromonas veronii</i>           | <i>Albidiferax ferritreducens</i> | <i>Alishewanella fetalis</i>             |
| <i>Actinoplanes rectilineatus</i>   | <b>Aeromicrobium</b>               | <b>Agrobacterium</b>               | <b>Albidovulum</b>                | <b>Alkalibacillus</b>                    |
| <i>Actinoplanes regularis</i>       | <i>Aeromicrobium erythreum</i>     | <i>Agrobacterium gelatinovorum</i> | <i>Albidovulum inexpectatum</i>   | <i>Alkalibacillus haloalkaliphilus</i>   |
| <i>Actinoplanes tchonomyceticus</i> | <b>Aeromonas</b>                   | <b>Agrococcus</b>                  | <b>Alcaligenes</b>                | <b>Alkalilimnicola</b>                   |
| <i>Actinoplanes utahensis</i>       | <i>Aeromonas allosaccharophila</i> | <i>Agrococcus citreus</i>          | <i>Alcaligenes denitrificans</i>  | <i>Alkalilimnicola ehrlichii</i>         |
| <b>Actinopolyspora</b>              |                                    |                                    | <i>Alcaligenes faecalis</i>       | <b>Alkaliphilus</b>                      |
|                                     |                                    |                                    | <b>Alcanivorax</b>                | <i>Alkaliphilus oremlandii</i>           |

|                                |                              |                             |                               |                               |
|--------------------------------|------------------------------|-----------------------------|-------------------------------|-------------------------------|
| <i>Alkaliphilus</i>            | <i>Aminobacter</i>           | <i>Amicolatopsis</i>        | <b>Anaerobaculum</b>          | <i>Anaeromyxobacter</i>       |
| <i>transvaalensis</i>          | <i>aminovorans</i>           | <i>albidiflavus</i>         | <i>Anaerobaculum mobile</i>   | <i>dehalogenans</i>           |
| <b>Allochromatium</b>          | <i>Aminobacter</i>           | <i>Amicolatopsis azurea</i> | <i>Anaerobiospirillum</i>     | <b>Anaerorhabdus</b>          |
| <i>Allochromatium</i>          | <i>nigidaensis</i>           | <i>Amicolatopsis</i>        | <i>Anaerobiospirillum</i>     | <i>Anaerorhabdus furcosa</i>  |
| <i>vinosum</i>                 | <b>Aminobacterium</b>        | <i>coloradensis</i>         | <i>succiniciproducens</i>     | <b>Anaerosinus</b>            |
| <b>Alloicoccus</b>             | <i>Aminobacterium mobile</i> | <i>Amicolatopsis lunda</i>  | <i>Anaerobiospirillum</i>     | <i>Anaerosinus glycerini</i>  |
| <i>Alloicoccus otitis</i>      | <b>Aminomonas</b>            | <i>Amicolatopsis</i>        | <i>thomasi</i>                | <b>Anaerovirgula</b>          |
| <b>Allokutzneria</b>           | <i>Aminomonas</i>            | <i>mediterranei</i>         | <b>Anaerococcus</b>           | <i>Anaerovirgula</i>          |
| <i>Allokutzneria alбата</i>    | <i>paucivorans</i>           | <i>Amicolatopsis</i>        | <i>Anaerococcus</i>           | <i>multivorans</i>            |
| <b>Altererythrobacter</b>      | <b>Ammoniphilus</b>          | <i>rifamycinica</i>         | <i>hydrogenalis</i>           | <b>Ancalomicrobium</b>        |
| <i>Altererythrobacter</i>      | <i>Ammoniphilus</i>          | <i>Amicolatopsis rubida</i> | <i>Anaerococcus</i>           | <i>Ancalomicrobium</i>        |
| <i>ishigakiensis</i>           | <i>oxalaticus</i>            | <i>Amicolatopsis</i>        | <i>lactolyticus</i>           | <i>adatum</i>                 |
| <b>Altermonas</b>              | <i>Ammoniphilus</i>          | <i>sulphurea</i>            | <i>Anaerococcus prevotii</i>  | <b>Ancylobacter</b>           |
| <i>Altermonas haloplanktis</i> | <i>oxalivorans</i>           | <i>Amicolatopsis</i>        | <i>Anaerococcus tetradius</i> | <i>Ancylobacter aquaticus</i> |
| <i>Altermonas macleodii</i>    | <b>Amphibacillus</b>         | <i>tolypomycinica</i>       | <i>Anaerococcus vaginalis</i> | <b>Aneurinibacillus</b>       |
| <b>Alysiella</b>               | <i>Ampyibacillus xylanus</i> | <b>Anabaena</b>             | <b>Anaerofustis</b>           | <i>Aneurinibacillus</i>       |
| <i>Alysiella crassa</i>        | <b>Amphritea</b>             | <i>Anabaena cylindrica</i>  | <i>Anaerofustis</i>           | <i>aneurinolyticus</i>        |
| <i>Alysiella filiformis</i>    | <i>Amphritea balenae</i>     | <i>Anabaena flos-aquae</i>  | <i>stercorihominis</i>        | <i>Aneurinibacillus</i>       |
| <b>Aminobacter</b>             | <i>Amphritea japonica</i>    | <i>Anabaena variabilis</i>  | <b>Anaeromusa</b>             | <i>migulanus</i>              |
| <i>Aminobacter</i>             | <b>Amicolatopsis</b>         | <b>Anaerococcus</b>         | <i>Anaeromusa</i>             | <i>Aneurinibacillus</i>       |
| <i>agardensis</i>              | <i>Amicolatopsis alba</i>    | <i>Anaerococcus</i>         | <i>acidaminophila</i>         | <i>thermoaerophilus</i>       |
|                                |                              | <i>burkinensis</i>          | <b>Anaeromyxobacter</b>       | <b>Angiococcus</b>            |

|                                  |                                |                                |                                |
|----------------------------------|--------------------------------|--------------------------------|--------------------------------|
| <i>Angiobococcus disciformis</i> | <i>Arcanobacterium</i>         | <i>Arthrobacter</i>            | <i>Asticcacaulis</i>           |
| <b>Angulomicrobium</b>           | <i>pyogenes</i>                | <i>crystallopiletes</i>        | <i>excentricus</i>             |
| <i>Angulomicrobium</i>           | <b>Archangium</b>              | <i>Arthrobacter cummingsii</i> | <b>Atopobacter</b>             |
| <i>tetraedrale</i>               | <i>Archangium gephyra</i>      | <i>Arthrobacter</i>            | <i>Atopobacter phocae</i>      |
| <b>Anoxybacillus</b>             | <b>Arcobacter</b>              | <i>globiformis</i>             | <b>Atopobium</b>               |
| <i>Anoxybacillus</i>             | <i>Arcobacter butzleri</i>     | <i>Arthrobacter</i>            | <i>Atopobium fossor</i>        |
| <i>pushtinchensis</i>            | <i>Arcobacter</i>              | <i>histidinovorans</i>         | <i>Atopobium minutum</i>       |
| <b>Aquabacterium</b>             | <i>cryaerophilus</i>           | <i>Arthrobacter ilicis</i>     | <i>Atopobium parvulum</i>      |
| <i>Aquabacterium</i>             | <i>Arcobacter halophilus</i>   | <i>Arthrobacter luteus</i>     | <i>Atopobium rimae</i>         |
| <i>commune</i>                   | <i>Arcobacter nitrofigilis</i> | <i>Arthrobacter</i>            | <i>Atopobium vaginae</i>       |
| <i>Aquabacterium parvum</i>      | <i>Arcobacter skirrowii</i>    | <i>methylobrophus</i>          | <b>Aureobacterium</b>          |
| <b>Aquaspirillum</b>             | <b>Arhodomonas</b>             | <i>Arthrobacter mysorens</i>   | <i>Aureobacterium barkeri</i>  |
| <i>Aquaspirillum</i>             | <i>Arhodomonas aquaedlei</i>   | <i>Arthrobacter nicotianae</i> | <b>Aurobacterium</b>           |
| <i>polymorphum</i>               | <b>Arsenophonus</b>            | <i>Arthrobacter</i>            | <i>Aurobacterium</i>           |
| <i>Aquaspirillum</i>             | <i>Arsenophonus nasoniae</i>   | <i>nicotinovorans</i>          | <i>liquefaciens</i>            |
| <i>putridiconchylium</i>         | <b>Arthrobacter</b>            | <i>Arthrobacter oxydans</i>    | <b>Avibacterium</b>            |
| <i>Aquaspirillum serpens</i>     | <i>Arthrobacter agilis</i>     | <i>Arthrobacter pascens</i>    | <i>Avibacterium avium</i>      |
| <b>Aquimarina</b>                | <i>Arthrobacter albus</i>      | <i>Arthrobacter</i>            | <i>Avibacterium gallinarum</i> |
| <i>Aquimarina latercula</i>      | <i>Arthrobacter aurescens</i>  | <i>phenanthrenivorans</i>      | <i>Avibacterium</i>            |
| <b>Arcanobacterium</b>           | <i>Arthrobacter</i>            | <i>Arthrobacter</i>            | <i>paragallinarum</i>          |
| <i>Arcanobacterium</i>           | <i>chlorophenolicus</i>        | <i>polychromogenes</i>         | <i>Avibacterium volantium</i>  |
| <i>haemolyticum</i>              | <i>Arthrobacter citreus</i>    | <i>biprosthecium</i>           | <b>Azoarcus</b>                |

|                                 |                                 |                                 |                               |                                 |
|---------------------------------|---------------------------------|---------------------------------|-------------------------------|---------------------------------|
| <i>Azoarcus indigenus</i>       | <b>Azotobacter</b>              | <i>Bacteroides stercoris</i>    | <i>Bartonella elizabethae</i> | <i>Beilina caldissulae</i>      |
| <i>Azoarcus toluyticus</i>      | <i>Azotobacter beijerinckii</i> | <i>Bacteroides suis</i>         | <i>Bartonella grahamii</i>    | <b>Belnapia</b>                 |
| <i>Azoarcus toluyticus</i>      | <i>Azotobacter</i>              | <i>Bacteroides testus</i>       | <i>Bartonella henselae</i>    | <i>Belnapia moabensis</i>       |
| <b>Azohydromonas</b>            | <i>Chroococcum</i>              | <i>Bacteroides</i>              | <i>Bartonella rochalimae</i>  | <b>Bergeriella</b>              |
| <i>Azohydromonas australica</i> | <i>Azotobacter nigrificans</i>  | <i>thetataenion</i>             | <i>Bartonella vinsonii</i>    | <i>Bergeriella dentrificans</i> |
| <i>Azohydromonas lata</i>       | <i>Azotobacter salinis</i>      | <i>Bacteroides uniformis</i>    | <b>Bavaricoccus</b>           | <b>Beutenbergia</b>             |
| <b>Azomonas</b>                 | <i>Azotobacter vinelandii</i>   | <i>Bacteroides ureolyticus</i>  | <i>Bavaricoccus seileri</i>   | <i>Beutenbergia cavernae</i>    |
| <i>Azomonas agilis</i>          | <b>Bacteriovorax</b>            | <i>Bacteroides vulgaris</i>     | <b>Bdellovibrio</b>           | <b>Bibersteinia</b>             |
| <i>Azomonas insignis</i>        | <i>Bacteriovorax stolpii</i>    | <b>Balnearium</b>               | <i>Bdellovibrio</i>           | <i>Bibersteinia trehalosi</i>   |
| <i>Azomonas</i>                 | <b>Bacteroides</b>              | <i>Balnearium</i>               | <i>bacteriovorus</i>          | <b>Bifidobacterium</b>          |
| <i>macrococcus</i>              | <i>Bacteroides caecae</i>       | <i>lithotrophicum</i>           | <i>Bdellovibrio exovorus</i>  | <i>Bifidobacterium</i>          |
| <b>Azorhizobium</b>             | <i>Bacteroides coagulans</i>    | <b>Balneatrix</b>               | <b>Beggiatoa</b>              | <i>adolescentis</i>             |
| <i>Azorhizobium</i>             | <i>Bacteroides eggertii</i>     | <i>Balneatrix alpica</i>        | <i>Beggiatoa alba</i>         | <i>Bifidobacterium</i>          |
| <i>caulinodans</i>              | <i>Bacteroides fragilis</i>     | <b>Balneola</b>                 | <b>Beijerinckia</b>           | <i>angulatum</i>                |
| <b>Azorhizophilus</b>           | <i>Bacteroides</i>              | <i>Balneola vulgaris</i>        | <i>Beijerinckia densii</i>    | <i>Bifidobacterium</i>          |
| <i>Azorhizophilus paspali</i>   | <i>galacturonicus</i>           | <b>Barnesiella</b>              | <i>Beijerinckia ruminans</i>  | <i>animalis</i>                 |
| <b>Azospirillum</b>             | <i>Bacteroides heliogenes</i>   | <i>Barnesiella viscericola</i>  | <i>Beijerinckia indica</i>    | <i>Bifidobacterium</i>          |
| <i>Azospirillum brasilense</i>  | <i>Bacteroides ovatus</i>       | <b>Bartonella</b>               | <i>Beijerinckia mobilis</i>   | <i>asteroides</i>               |
| <i>Azospirillum</i>             | <i>Bacteroides</i>              | <i>Bartonella alsatica</i>      | <b>Belliella</b>              | <i>Bifidobacterium bifidum</i>  |
| <i>halopraeferens</i>           | <i>pectinophilus</i>            | <i>Bartonella bacilliformis</i> | <i>Belliella batika</i>       | <i>Bifidobacterium boum</i>     |
| <i>Azospirillum irakense</i>    | <i>Bacteroides pyogenes</i>     | <i>Bartonella clarridgeae</i>   | <b>Beilina</b>                | <i>Bifidobacterium breve</i>    |
|                                 | <i>Bacteroides salyersiae</i>   | <i>Bartonella doshiae</i>       |                               |                                 |

|                                          |                                          |                                  |                                     |                                          |
|------------------------------------------|------------------------------------------|----------------------------------|-------------------------------------|------------------------------------------|
| <i>Bifidobacterium catenulatum</i>       | <i>Bifidobacterium pseudolongum</i>      | <i>Elastococcus aggregatus</i>   | <i>Bordetella holmesii</i>          | <i>Brachybacterium paraconglomeratum</i> |
| <i>Bifidobacterium choerinum</i>         | <i>Bifidobacterium pullorum</i>          | <i>Elastococcus savosidentis</i> | <i>Bordetella pertussis</i>         | <i>Brachybacterium rhamnosum</i>         |
| <i>Bifidobacterium coryneforme</i>       | <i>Bifidobacterium ruminantium</i>       | <b>Blastochloris</b>             | <i>Bordetella trematum</i>          | <i>Brachybacterium tyrofermentans</i>    |
| <i>Bifidobacterium cuniculi</i>          | <i>Bifidobacterium saeculare</i>         | <i>Blastochloris viridis</i>     | <b>Borrelia</b>                     | <b>Brachyspira</b>                       |
| <i>Bifidobacterium dentium</i>           | <i>Bifidobacterium thermophilum</i>      | <b>Blastomonas</b>               | <i>Borrelia afzelii</i>             | <i>Brachyspira alvinipulli</i>           |
| <i>Bifidobacterium gallium</i>           | <i>Bifidobacterium bifidum</i>           | <i>Blastomonas ratoria</i>       | <i>Borrelia burgdorferi</i>         | <i>Brachyspira hyodysenteriae</i>        |
| <i>Bifidobacterium gallinarum</i>        | <i>Bifidobacterium bifidum</i>           | <b>Blastopirellula</b>           | <i>Borrelia cardinensis</i>         | <i>Brachyspira innoxens</i>              |
| <i>Bifidobacterium indicum</i>           | <i>Bifidobacterium longum</i>            | <i>Blastopirellula marina</i>    | <i>Borrelia coriaceae</i>           | <i>Brachyspira murchisonii</i>           |
| <i>Bifidobacterium longum</i>            | <i>Bifidobacterium longum</i>            | <b>Blautia</b>                   | <i>Borrelia garinii</i>             | <i>Brachyspira pilosicoli</i>            |
| <i>Bifidobacterium magnum</i>            | <i>Bifidobacterium merycicum</i>         | <i>Blautia coecoides</i>         | <i>Borrelia japonica</i>            | <b>Bradyrhizobium</b>                    |
| <i>Bifidobacterium minimum</i>           | <i>Bifidobacterium minimum</i>           | <i>Blautia hansenii</i>          | <b>Bosea</b>                        | <i>Bradyrhizobium canariense</i>         |
| <i>Bifidobacterium pseudocatenulatum</i> | <i>Bifidobacterium pseudocatenulatum</i> | <i>Blautia producta</i>          | <i>Bosea minatitanensis</i>         | <i>Bradyrhizobium elkanii</i>            |
|                                          |                                          | <i>Blautia wexlerae</i>          | <i>Bosea thiooxidans</i>            | <i>Bradyrhizobium japonicum</i>          |
|                                          |                                          | <b>Bogoriella</b>                | <b>Brachybacterium</b>              | <i>Bradyrhizobium liaoningense</i>       |
|                                          |                                          | <i>Bogoriella caseolytica</i>    | <i>Brachybacterium alimentarium</i> | <b>Brenneria</b>                         |
|                                          |                                          | <b>Bordetella</b>                | <i>Brachybacterium faecium</i>      |                                          |
|                                          |                                          | <i>Bordetella avium</i>          |                                     |                                          |
|                                          |                                          | <i>Bordetella bronchiseptica</i> |                                     |                                          |
|                                          |                                          | <i>Bordetella hinzii</i>         |                                     |                                          |

|                                 |                                 |                               |                                   |                                   |
|---------------------------------|---------------------------------|-------------------------------|-----------------------------------|-----------------------------------|
| <i>Brenneria alni</i>           | <i>Brevibacterium</i>           | <i>Brevundimonas alta</i>     | <b>Bryobacter</b>                 | <i>Burkholderia</i>               |
| <i>Brenneria nigrifluens</i>    | <i>aurantiacum</i>              | <i>Brevundimonas</i>          | <i>Bryobacter aggregatus</i>      | <i>phenazinum</i>                 |
| <i>Brenneria quercina</i>       | <i>Brevibacterium celere</i>    | <i>aurantiaca</i>             | <b>Burkholderia</b>               | <i>Burkholderia plantarii</i>     |
| <i>Brenneria quercina</i>       | <i>Brevibacterium</i>           | <i>Brevundimonas</i>          | <i>Burkholderia ambifaria</i>     | <i>Burkholderia pyrocinia</i>     |
| <i>Brenneria salicis</i>        | <i>epidermidis</i>              | <i>diminuta</i>               | <i>Burkholderia</i>               | <i>Burkholderia</i>               |
| <b>Brevibacillus</b>            | <i>Brevibacterium</i>           | <i>Brevundimonas</i>          | <i>andropogonis</i>               | <i>silvaticus</i>                 |
| <i>Brevibacillus agri</i>       | <i>frigoritolerans</i>          | <i>intermedia</i>             | <i>Burkholderia anthina</i>       | <i>Burkholderia stabilis</i>      |
| <i>Brevibacillus</i>            | <i>Brevibacterium</i>           | <i>Brevundimonas</i>          | <i>Burkholderia caledonica</i>    | <i>Burkholderia</i>               |
| <i>burstelensis</i>             | <i>halotolerans</i>             | <i>subvibrioides</i>          | <i>Burkholderia caryophylli</i>   | <i>thailandensis</i>              |
| <i>Brevibacillus brevis</i>     | <i>Brevibacterium iodinum</i>   | <i>Brevundimonas</i>          | <i>Burkholderia</i>               | <i>Burkholderia tropica</i>       |
| <i>Brevibacillus</i>            | <i>Brevibacterium linens</i>    | <i>vancouverii</i>            | <i>cenazepeana</i>                | <i>Burkholderia unananae</i>      |
| <i>centrosporus</i>             | <i>Brevibacterium lyticum</i>   | <i>Brevundimonas</i>          | <i>Burkholderia cepacia</i>       | <i>Burkholderia</i>               |
| <i>Brevibacillus</i>            | <i>Brevibacterium</i>           | <i>variabilis</i>             | <i>Burkholderia</i>               | <i>vietnamiensis</i>              |
| <i>choshinensis</i>             | <i>mbrellinii</i>               | <i>Brevundimonas</i>          | <i>cocovenenans</i>               | <b>Buttiauxella</b>               |
| <i>Brevibacillus invocatus</i>  | <i>Brevibacterium obitidis</i>  | <i>vesicularis</i>            | <i>Burkholderia dolosa</i>        | <i>Buttiauxella agrestis</i>      |
| <i>Brevibacillus</i>            | <i>Brevibacterium oxydans</i>   | <b>Brochothrix</b>            | <i>Burkholderia fungorum</i>      | <i>Buttiauxella brennerae</i>     |
| <i>laterosporus</i>             | <i>Brevibacterium</i>           | <i>Brochothrix campestris</i> | <i>Burkholderia gladii</i>        | <i>Buttiauxella ferrugulifera</i> |
| <i>Brevibacillus parabravis</i> | <i>paucivorans</i>              | <i>Brochothrix</i>            | <i>Burkholderia glumae</i>        | <i>Buttiauxella gaviniae</i>      |
| <i>Brevibacillus reuszeri</i>   | <i>Brevibacterium stationis</i> | <i>thermosphaera</i>          | <i>Burkholderia graminis</i>      | <i>Buttiauxella izardii</i>       |
| <b>Brevibacterium</b>           | <b>Brevinema</b>                | <b>Brucella</b>               | <i>Burkholderia kururienensis</i> | <i>Buttiauxella noackiae</i>      |
| <i>Brevibacterium abidum</i>    | <i>Brevinema andersonii</i>     | <i>Brucella canis</i>         | <i>Burkholderia</i>               | <i>Buttiauxella</i>               |
| <i>Brevibacterium album</i>     | <b>Brevundimonas</b>            | <i>Brucella neotomae</i>      | <i>multivorans</i>                | <i>warmholdiae</i>                |

|                                     |                                |                                                           |                             |                            |  |
|-------------------------------------|--------------------------------|-----------------------------------------------------------|-----------------------------|----------------------------|--|
| <b>Butyrivibrio</b>                 |                                |                                                           |                             |                            |  |
| <i>Butyrivibrio fibrissolvens</i>   | <i>B. alkalidiazotrophicus</i> | <i>B. fengquensis</i>                                     | <i>B. amyolyticus</i>       | <i>B. beringensis</i>      |  |
| <i>Butyrivibrio hungatei</i>        | <i>B. alkalinitrilicus</i>     | <i>B. firmus</i>                                          | <i>B. andreesenii</i>       | <i>B. berkeleyi</i>        |  |
| <i>Butyrivibrio proteoclasticus</i> | <i>B. alkalisediminis</i>      | <i>B. flexus</i>                                          | <i>B. aneurinilyticus</i>   | <i>B. beveridgei</i>       |  |
|                                     | <i>B. alkalitelluris</i>       | <i>B. foraminis</i>                                       | <i>B. anthracis</i>         | <i>B. bogoriensis</i>      |  |
| <b>Bacillus</b>                     | <i>B. altitudinis</i>          | <i>B. fordii</i>                                          | <i>B. aquimaris</i>         | <i>B. boroniphilus</i>     |  |
| <i>B. acidicer</i>                  | <i>B. alveayuensis</i>         | <i>B. formosus</i>                                        | <i>B. arenosi</i>           | <i>B. borstelensis</i>     |  |
| <i>B. acidicola</i>                 | <i>B. alvei</i>                | <i>B. fortis</i>                                          | <i>B. arseniciselenatis</i> | <i>B. brevis Migula</i>    |  |
| <i>B. acidiproducens</i>            | <i>B. amyloliquefaciens</i>    | <i>B. fumarii</i>                                         | <i>B. arsenicus</i>         | <i>B. butanolivorans</i>   |  |
| <i>B. acidocaldarius</i>            | <i>B. a. subsp.</i>            | <i>B. funiculus</i>                                       | <i>B. aurantiacus</i>       | <i>B. canaveraius</i>      |  |
| <i>B. acidoterrestris</i>           | <i>Amyloliquefaciens</i>       | <i>B. fusiformis</i>                                      | <i>B. anvi</i>              | <i>B. carboniphilus</i>    |  |
| <i>B. aeolus</i>                    | <i>B. a. subsp. Plantarum</i>  | <i>B. galactophilus</i>                                   | <i>B. anyabhatai</i>        | <i>B. cecembensis</i>      |  |
| <i>B. aerius</i>                    | <i>B. dipsosauri</i>           | <i>B. galactosidilyticus</i>                              | <i>B. asahii</i>            | <i>B. cellulosilyticus</i> |  |
| <i>B. aerophilus</i>                | <i>B. drenensis</i>            | <i>B. gallicensis</i>                                     | <i>B. atrophaeus</i>        | <i>B. centrosporus</i>     |  |
| <i>B. agaradhaerens</i>             | <i>B. edaphicus</i>            | <i>B. gelatini</i>                                        | <i>B. axarquiensis</i>      | <i>B. cereus</i>           |  |
| <i>B. agri</i>                      | <i>B. ehimensis</i>            | <i>B. gibsonii</i>                                        | <i>B. azotofixans</i>       | <i>B. chagannorensis</i>   |  |
| <i>B. aldingensis</i>               | <i>B. eiseniae</i>             | <i>B. ginsengi</i>                                        | <i>B. azotoformans</i>      | <i>B. chitinolyticus</i>   |  |
| <i>B. alibai</i>                    | <i>B. endensis</i>             | <i>B. ginsengihumi</i>                                    | <i>B. badius</i>            | <i>B. chondroitinus</i>    |  |
| <i>B. alcalophilus</i>              | <i>B. endophyticus</i>         | <i>B. ginsengisoli</i>                                    | <i>B. barbaricus</i>        | <i>B. choshinensis</i>     |  |
| <i>B. algicola</i>                  | <i>B. endoradicis</i>          | <i>B. globisporus (p. ej. B. g. subsp. Globisporus; o</i> | <i>B. bataviensis</i>       | <i>B. chungangensis</i>    |  |
| <i>B. alginolyticus</i>             | <i>B. farraginis</i>           | <i>B. g. subsp. Marinus)</i>                              | <i>B. beijingensis</i>      | <i>B. cibi</i>             |  |
|                                     | <i>B. fastidiosus</i>          |                                                           | <i>B. benzoovorans</i>      | <i>B. circulans</i>        |  |
|                                     |                                | <i>B. aminovorans</i>                                     |                             |                            |  |

|                             |                                |                            |                                   |                               |
|-----------------------------|--------------------------------|----------------------------|-----------------------------------|-------------------------------|
| <i>B. clarkii</i>           | <i>B. halodurans</i>           | <i>B. jeotgali</i>         | <i>B. siamensis</i>               | <i>B. thermoaerophilus</i>    |
| <i>B. clausii</i>           | <i>B. halophilus</i>           | <i>B. kaustophilus</i>     | <i>B. silvestris</i>              | <i>B. thermoamylovorans</i>   |
| <i>B. coagulans</i>         | <i>B. halosaccharovorans</i>   | <i>B. kobensis</i>         | <i>B. simplex</i>                 | <i>B. thermocatenulatus</i>   |
| <i>B. coahuilensis</i>      | <i>B. hemicellulosilyticus</i> | <i>B. kochii</i>           | <i>B. siralis</i>                 | <i>B. thermocloacae</i>       |
| <i>B. cohnii</i>            | <i>B. hemientrofi</i>          | <i>B. kokeshiiformis</i>   | <i>B. smithii</i>                 | <i>B. thermocopriae</i>       |
| <i>B. compositi</i>         | <i>B. herbersteinensis</i>     | <i>B. koreensis</i>        | <i>B. soli</i>                    | <i>B. thermodenitrificans</i> |
| <i>B. curdianolyticus</i>   | <i>B. horikoshii</i>           | <i>B. korlensis</i>        | <i>B. solinangrovii</i>           | <i>B. thermoglucosidasius</i> |
| <i>B. cycloheptanicus</i>   | <i>B. horneckiae</i>           | <i>B. kribbensis</i>       | <i>B. solisalsi</i>               | <i>B. thermolactis</i>        |
| <i>B. cytotoxicus</i>       | <i>B. horti</i>                | <i>B. krulwichiae</i>      | <i>B. songkiensis</i>             | <i>B. thermoleovorans</i>     |
| <i>B. daliensis</i>         | <i>B. huizhouensis</i>         | <i>B. laevolacticus</i>    | <i>B. sonorensis</i>              | <i>B. thermophilus</i>        |
| <i>B. decisifrons</i>       | <i>B. humi</i>                 | <i>B. larvae</i>           | <i>B. sphaericus</i>              | <i>B. thermoruber</i>         |
| <i>B. decolorationis</i>    | <i>B. hwajirpoensis</i>        | <i>B. laterosporus</i>     | <i>B. sporothermodurans</i>       | <i>B. thermosphaericus</i>    |
| <i>B. deserti</i>           | <i>B. idriensis</i>            | <i>B. salexigens</i>       | <i>B. stearothermophilus</i>      | <i>B. thiaminolyticus</i>     |
| <i>B. glucanolyticus</i>    | <i>B. indicus</i>              | <i>B. saliphilus</i>       | <i>B. stratosphericus</i>         | <i>B. thioparans</i>          |
| <i>B. gordonae</i>          | <i>B. infantis</i>             | <i>B. schlegelii</i>       | <i>B. subterraneus</i>            | <i>B. thuringiensis</i>       |
| <i>B. gothelii</i>          | <i>B. infernus</i>             | <i>B. sediminis</i>        | <i>B. subtilis</i> (p. ei., B. s. | <i>B. tianshenii</i>          |
| <i>B. graminis</i>          | <i>B. insolitus</i>            | <i>B. selenatarsenatis</i> | subsp. <i>Inaquosorum</i> ; o     | <i>B. trypanoxylicola</i>     |
| <i>B. halmapatus</i>        | <i>B. invictae</i>             | <i>B. selenitireducens</i> | <i>B. s. subsp. Spizizeni</i> ; o | <i>B. tusciae</i>             |
| <i>B. haloalkaliphilus</i>  | <i>B. iranensis</i>            | <i>B. seohaeanensis</i>    | <i>B. s. subsp. Subtilis</i> )    | <i>B. validus</i>             |
| <i>B. halochares</i>        | <i>B. isabellae</i>            | <i>B. shacheensis</i>      | <i>B. taeanensis</i>              | <i>B. vallismortis</i>        |
| <i>B. halodenitrificans</i> | <i>B. isronensis</i>           | <i>B. shackletonii</i>     | <i>B. tequilensis</i>             | <i>B. vedderi</i>             |
|                             |                                |                            | <i>B. thermantarcticus</i>        |                               |

|                              |                               |                            |                            |                                 |
|------------------------------|-------------------------------|----------------------------|----------------------------|---------------------------------|
| <i>B. velezensis</i>         | <i>B. psychrodurans</i>       | <i>B. licheniformis</i>    | <i>B. mytilus</i>          | <i>B. oshimensis</i>            |
| <i>B. vietnamensis</i>       | <i>B. psychrophilus</i>       | <i>B. ligniphilus</i>      | <i>B. majavensis</i>       | <i>B. patuli</i>                |
| <i>B. urei</i>               | <i>B.</i>                     | <i>B. litoralis</i>        | <i>B. mucilaginosus</i>    | <i>B. pakistanensis</i>         |
| <i>B. vitani</i>             | <i>psychrosaccharolyticus</i> | <i>B. locisalis</i>        | <i>B. muralis</i>          | <i>B. pallidus</i>              |
| <i>B. wakoensis</i>          | <i>B. psychrotolerans</i>     | <i>B. luciferensis</i>     | <i>B. murimartini</i>      | <i>B. pallidus</i>              |
| <i>B. weihenstephanensis</i> | <i>B. pulvificans</i>         | <i>B. luteolus</i>         | <i>B. mycoides</i>         | <i>B. panacisoli</i>            |
| <i>B. xiamenensis</i>        | <i>B. pumilus</i>             | <i>B. luteus</i>           | <i>B. naganovensis</i>     | <i>B. panacterrae</i>           |
| <i>B. xiaoniensis</i>        | <i>B. purgationiresistens</i> | <i>B. macauensis</i>       | <i>B. nanhaiensis</i>      | <i>B. pantothenticus</i>        |
| <i>B. shanghaiensis</i>      | <i>B. pycnus</i>              | <i>B. nacerans</i>         | <i>B. nanhaiisediminis</i> | <i>B. parabravis</i>            |
| <i>B. peoniae</i>            | <i>B. qingdaonensis</i>       | <i>B. macquariensis</i>    | <i>B. nealsonii</i>        | <i>B. paraflexus</i>            |
| <i>B. persopolensis</i>      | <i>B. qingshengii</i>         | <i>B. macyae</i>           | <i>B. neidei</i>           | <i>B. pasteuri</i>              |
| <i>B. persicus</i>           | <i>B. reuszeri</i>            | <i>B. malacitensis</i>     | <i>B. nelzhoensis</i>      | <i>B. patagoniensis</i>         |
| <i>B. pervagus</i>           | <i>B. rhizosphaerae</i>       | <i>B. mannaniolyticus</i>  | <i>B. niabensis</i>        | <b>Caenimonas</b>               |
| <i>B. placentidis</i>        | <i>B. rigui</i>               | <i>B. marisflavi</i>       | <i>B. niacini</i>          | <i>Caenimonas koreensis</i>     |
| <i>B. pocheonensis</i>       | <i>B. ruri</i>                | <i>B. marismortui</i>      | <i>B. novalis</i>          | <b>Caldalkalibacillus</b>       |
| <i>B. polygoni</i>           | <i>B. salensis</i>            | <i>B. marmarensis</i>      | <i>B. oceanisediminis</i>  | <i>Caldalkalibacillus</i>       |
| <i>B. polymyxa</i>           | <i>B. salarius</i>            | <i>B. massiliensis</i>     | <i>B. odysseyi</i>         | <i>uzonensis</i>                |
| <i>B. papillae</i>           | <i>B. lautus</i>              | <i>B. megaterium</i>       | <i>B. okthensis</i>        | <b>Caldanaerobacter</b>         |
| <i>B. pseudocalophilius</i>  | <i>B. lehensis</i>            | <i>B. mesonae</i>          | <i>B. okuhidensis</i>      | <i>Caldanaerobacter</i>         |
| <i>B. pseudofirmus</i>       | <i>B. lentimorbus</i>         | <i>B. methanolicus</i>     | <i>B. oleronius</i>        | <i>subterraneus</i>             |
| <i>B. pseudomycoides</i>     | <i>B. lentus</i>              | <i>B. methylotrophicus</i> | <i>B. oryzaecorticis</i>   | <b>Caldanaerobius</b>           |
|                              |                               |                            |                            | <i>Caldanaerobius fijiensis</i> |

|                               |                             |                               |                               |                             |
|-------------------------------|-----------------------------|-------------------------------|-------------------------------|-----------------------------|
| <i>Caldanaerobius</i>         | <i>Campylobacter</i>        | <i>Capnocytophaga</i>         | <i>Carnobacterium mobile</i>  | <i>Catenuloplanes</i>       |
| <i>polysaccharolyticus</i>    | <i>hyointestinalis</i>      | <i>haemolytica</i>            | <i>Carnobacterium</i>         | <i>nepalensis</i>           |
| <i>Caldanaerobius zeae</i>    | <i>Campylobacter jejuni</i> | <i>Capnocytophaga</i>         | <i>viridans</i>               | <i>Catenuloplanes niger</i> |
| <b>Caldanaerovirga</b>        | <i>Campylobacter lari</i>   | <i>ocitracea</i>              | <b>Caryophanon</b>            | <b>Chryseobacterium</b>     |
| <i>Caldanaerovirga</i>        | <i>Campylobacter</i>        | <i>Capnocytophaga</i>         | <i>Caryophanon latum</i>      | <i>Chryseobacterium</i>     |
| <i>acetigenens</i>            | <i>mucosalis</i>            | <i>sputigena</i>              | <i>Caryophanon tenue</i>      | <i>balustinum</i>           |
| <b>Caldicellulosiruptor</b>   | <i>Campylobacter rectus</i> | <b>Cardiobacterium</b>        | <b>Catellatospora</b>         | <b>Citrobacter</b>          |
| <i>Caldicellulosiruptor</i>   | <i>Campylobacter showae</i> | <i>Cardiobacterium</i>        | <i>Catellatospora citrea</i>  | <i>C. amalonaticus</i>      |
| <i>bescii</i>                 | <i>Campylobacter</i>        | <i>hominis</i>                | <i>Catellatospora</i>         | <i>C. braakii</i>           |
| <i>Caldicellulosiruptor</i>   | <i>sputorum</i>             | <b>Carnimonas</b>             | <i>methionotrophica</i>       | <i>C. diversus</i>          |
| <i>kristjanssonii</i>         | <i>Campylobacter</i>        | <i>Carnimonas nigrificans</i> | <b>Catenococcus</b>           | <i>C. farneri</i>           |
| <i>Caldicellulosiruptor</i>   | <i>upsaliensis</i>          | <b>Carnobacterium</b>         | <i>Catenococcus thiocylli</i> | <i>C. freundii</i>          |
| <i>owensensis</i>             | <b>Capnocytophaga</b>       | <i>Carnobacterium</i>         | <b>Catenuloplanes</b>         | <i>C. gillenii</i>          |
| <b>Campylobacter</b>          | <i>Capnocytophaga</i>       | <i>alterfunditum</i>          | <i>Catenuloplanes</i>         | <i>C. koseri</i>            |
| <i>Campylobacter coli</i>     | <i>canimorsus</i>           | <i>Carnobacterium</i>         | <i>acrovinosus</i>            | <i>C. murliniae</i>         |
| <i>Campylobacter concisus</i> | <i>Capnocytophaga</i>       | <i>divergens</i>              | <i>Catenuloplanes</i>         | <i>C. pasteurei</i>         |
| <i>Campylobacter curvus</i>   | <i>cynodegmi</i>            | <i>Carnobacterium</i>         | <i>castaneus</i>              | <i>C. rodentium</i>         |
| <i>Campylobacter fetus</i>    | <i>Capnocytophaga</i>       | <i>funditum</i>               | <i>Catenuloplanes crispus</i> | <i>C. sedlakii</i>          |
| <i>Campylobacter gracilis</i> | <i>gingivalis</i>           | <i>Carnobacterium</i>         | <i>Catenuloplanes indicus</i> | <i>C. werkmanii</i>         |
| <i>Campylobacter</i>          | <i>Capnocytophaga</i>       | <i>gallinarum</i>             | <i>Catenuloplanes</i>         | <i>C. youngae</i>           |
| <i>helveticus</i>             | <i>granulosa</i>            | <i>Carnobacterium</i>         | <i>japonicus</i>              | <b>Coccochloris</b>         |
| <i>Campylobacter hominis</i>  |                             | <i>maltonaticum</i>           |                               | <i>Coccochloris elabens</i> |

|                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Corynebacterium</b> | <i>Clostridium aldenense</i> ,<br><i>Clostridium aldrichii</i> ,<br><i>Clostridium algidicani</i> ,<br><i>Clostridium</i><br><i>algidiviriditum</i> ,<br><i>Clostridium algaefaciens</i> ,<br><i>Clostridium</i><br><i>algoriphilum</i> ,<br><i>Clostridium</i><br><i>alkalicellulosi</i> ,<br><i>Clostridium</i><br><i>aminophilum</i> ,<br><i>Clostridium</i><br><i>aminovalepticum</i> ,<br><i>Clostridium</i><br><i>amygdalinum</i> ,<br><i>Clostridium</i><br><i>amyolyticum</i> ,<br><i>Clostridium</i><br><i>arbusci</i> ,<br><i>Clostridium arcticum</i> ,<br><i>Clostridium</i><br><i>argentiniense</i> ,<br><i>Clostridium</i><br><i>asparagiforme</i> , | <i>Clostridium</i><br><i>aurantiobutyricum</i> ,<br><i>Clostridium</i><br><i>autoethanogenum</i> ,<br><i>Clostridium</i><br><i>baculi</i> ,<br><i>Clostridium</i><br><i>barkeri</i> ,<br><i>Clostridium</i><br><i>bardettii</i> ,<br><i>Clostridium</i><br><i>beijerinckii</i> ,<br><i>Clostridium</i><br><i>bifermentans</i> ,<br><i>Clostridium</i><br><i>boltaee</i> ,<br><i>Clostridium</i><br><i>borminense</i> ,<br><i>Clostridium</i><br><i>botulinum</i> ,<br><i>Clostridium</i><br><i>bownanii</i> ,<br><i>Clostridium</i><br><i>bryantii</i> ,<br><i>Clostridium</i><br><i>butyricum</i> ,<br><i>Clostridium</i><br><i>cadaveris</i> ,<br><i>Clostridium</i><br><i>caenicola</i> ,<br><i>Clostridium</i><br><i>caminithermale</i> ,<br><i>Clostridium</i><br><i>carboxidivorans</i> ,<br><i>Clostridium</i><br><i>carnis</i> , | <i>Clostridium cavendishii</i> ,<br><i>Clostridium celatum</i> ,<br><i>Clostridium</i><br><i>celerecrescens</i> ,<br><i>Clostridium</i><br><i>celluloparum</i> ,<br><i>Clostridium</i><br><i>cellulofragmentans</i> ,<br><i>Clostridium</i><br><i>cellulolyticum</i> ,<br><i>Clostridium</i><br><i>cellulosi</i> ,<br><i>Clostridium</i><br><i>cellulovorans</i> ,<br><i>Clostridium</i><br><i>charatabidum</i> ,<br><i>Clostridium</i><br><i>chauvoei</i> ,<br><i>Clostridium</i><br><i>chromiireducens</i> ,<br><i>Clostridium</i><br><i>citroniae</i> ,<br><i>Clostridium clariflavum</i> ,<br><i>Clostridium</i><br><i>clausenense</i> ,<br><i>Clostridium</i><br><i>fallax</i> ,<br><i>Clostridium</i><br><i>feilsineum</i> ,<br><i>Clostridium</i><br><i>fervidum</i> , | <i>cochlearium</i> ,<br><i>Clostridium</i><br><i>colicantis</i> ,<br><i>Clostridium</i><br><i>colinum</i> ,<br><i>Clostridium</i><br><i>collagenovorans</i> ,<br><i>Clostridium</i><br><i>cylindrosporum</i> ,<br><i>Clostridium</i><br><i>difficile</i> ,<br><i>Clostridium</i><br><i>diolis</i> ,<br><i>Clostridium</i><br><i>disporicum</i> ,<br><i>Clostridium</i><br><i>drakei</i> ,<br><i>Clostridium</i><br><i>durum</i> ,<br><i>Clostridium</i><br><i>estertheticum</i> ,<br><i>Clostridium</i><br><i>estertheticum</i> ,<br><i>Clostridium</i><br><i>estertheticum</i> ,<br><i>Clostridium</i><br><i>estertheticum</i> ,<br><i>Clostridium</i><br><i>laramiense</i> ,<br><i>Clostridium</i><br><i>fallax</i> ,<br><i>Clostridium</i><br><i>feilsineum</i> ,<br><i>Clostridium</i><br><i>fervidum</i> , |
| <b>Curtobacterium</b>  | <i>Curtobacterium albidum</i><br><i>Curtobacterium citreus</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <b>Clostridium</b>     | <i>Clostridium</i><br><i>absorbum</i> ,<br><i>Clostridium</i><br><i>aceticum</i> ,<br><i>Clostridium</i><br><i>acetireducens</i> ,<br><i>Clostridium</i><br><i>acetobutylicum</i> ,<br><i>Clostridium</i><br><i>acidisoli</i> ,<br><i>Clostridium</i><br><i>aciditolerans</i> ,<br><i>Clostridium</i><br><i>acidurici</i> ,<br><i>Clostridium</i><br><i>aerotolerans</i> ,<br><i>Clostridium</i><br><i>aestuarii</i> ,<br><i>Clostridium</i><br><i>akagii</i> ,                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

|                                        |                                  |                  |                               |                 |                                         |                               |
|----------------------------------------|----------------------------------|------------------|-------------------------------|-----------------|-----------------------------------------|-------------------------------|
| <i>finetianum</i> , <i>Clostridium</i> | <i>Clostridium</i>               | <i>haskellii</i> | <i>Clostridium</i>            | <i>leptum</i>   | <i>Clostridium</i>                      | <i>populeti</i>               |
| <i>formicacepticum</i> ,               | <i>Clostridium</i>               | <i>hungatei</i>  | <i>Clostridium</i>            | <i>linosum</i>  | <i>orbiscindens</i>                     | <i>Clostridium</i>            |
| <i>Clostridium frigidicarnis</i> ,     | <i>Clostridium</i>               |                  | <i>Clostridium</i>            | <i>litorale</i> | <i>Clostridium oreiticum</i>            | <i>propionicum</i>            |
| <i>Clostridium frigoris</i> ,          | <i>hydrogeniformans</i> ,        |                  | <i>Clostridium</i>            |                 | <i>Clostridium oxalicum</i>             | <i>Clostridium</i>            |
| <i>Clostridium</i>                     | <i>Clostridium</i>               |                  | <i>lituseburense</i>          |                 | <i>Clostridium</i>                      | <i>proteoclasticum</i>        |
| <i>ganghwense</i> ,                    | <i>hydroxybenzolicum</i> ,       |                  | <i>Clostridium jungdahlii</i> |                 | <i>papyrosolvans</i>                    | <i>Clostridium</i>            |
| <i>Clostridium gasigenes</i> ,         | <i>Clostridium hylemoniae</i> ,  |                  | <i>Clostridium tortetii</i>   |                 | <i>Clostridium paradoxum</i>            | <i>proteolyticum</i>          |
| <i>Clostridium ghorii</i> ,            | <i>Clostridium jejuense</i> ,    |                  | <i>Clostridium lundense</i>   |                 | <i>Clostridium</i>                      | <i>Clostridium</i>            |
| <i>Clostridium glycolicum</i> ,        | <i>Clostridium indolis</i> ,     |                  | <i>Clostridium magnum</i>     |                 | <i>paraperfringens</i> (Alias:          | <i>psychrophilum</i>          |
| <i>Clostridium</i>                     | <i>Clostridium innocuum</i> ,    |                  | <i>Clostridium</i>            |                 | <i>C. welchii</i> ), <i>Clostridium</i> | <i>Clostridium puniceum</i>   |
| <i>glycylrhiziniyticum</i> ,           | <i>Clostridium intestinale</i> , |                  | <i>malenominatum</i>          |                 | <i>paraputrificum</i>                   | <i>Clostridium</i>            |
| <i>Clostridium grantii</i> ,           | <i>Clostridium irregulare</i> ,  |                  | <i>Clostridium mangelarii</i> |                 | <i>Clostridium pasculi</i>              | <i>pumiliyticum</i>           |
| <i>Clostridium</i>                     | <i>Clostridium isafidis</i> ,    |                  | <i>Clostridium mayombei</i>   |                 | <i>Clostridium</i>                      | <i>Clostridium</i>            |
| <i>haemolyticum</i> ,                  | <i>Clostridium jesusi</i> ,      |                  | <i>Clostridium</i>            |                 | <i>pasteurianum</i>                     | <i>putrefaciens</i>           |
| <i>Clostridium halophilum</i> ,        | <i>Clostridium kluyveri</i> ,    |                  | <i>methoxybenzovorans</i>     |                 | <i>Clostridium</i>                      | <i>Clostridium putrificum</i> |
| <i>Clostridium hastiforme</i> ,        | <i>Clostridium</i>               |                  | <i>Clostridium</i>            |                 | <i>peptidivorans</i>                    | <i>Clostridium</i>            |
| <i>Clostridium hathewayi</i> ,         | <i>lactatifermentans</i> ,       |                  | <i>methylospinosum</i>        |                 | <i>Clostridium perenne</i>              | <i>quercicolum</i>            |
| <i>Clostridium</i>                     | <i>Clostridium</i>               |                  | <i>Clostridium</i>            |                 | <i>Clostridium perfringens</i>          | <i>Clostridium quini</i>      |
| <i>herbivorans</i> ,                   | <i>lacusfryxellense</i> ,        |                  | <i>neopropionicum</i>         |                 | <i>Clostridium plenigii</i>             | <i>Clostridium ramosum</i>    |
| <i>Clostridium hiranonis</i> ,         | <i>Clostridium laramiense</i> ,  |                  | <i>Clostridium</i>            | <i>nexile</i>   | <i>Clostridium</i>                      | <i>Clostridium rectum</i>     |
| <i>Clostridium</i>                     | <i>Clostridium levalense</i> ,   |                  | <i>Clostridium</i>            |                 | <i>phytofermentans</i>                  | <i>Clostridium roseum</i>     |
| <i>histolyticum</i> ,                  | <i>Clostridium lentocellum</i> , |                  | <i>nitrophilolicum</i>        |                 | <i>Clostridium piliforme</i>            | <i>Clostridium</i>            |
| <i>Clostridium</i>                     | <i>Clostridium</i>               |                  | <i>Clostridium</i>            | <i>navyi</i>    | <i>Clostridium</i>                      | <i>saccharobutylicum</i>      |
| <i>homopropionicum</i> ,               | <i>lentiputrescens</i> ,         |                  | <i>Clostridium oceanicum</i>  |                 | <i>polysaccharolyticum</i>              | <i>Clostridium</i>            |

|                                 |                      |                        |                            |                             |                                |
|---------------------------------|----------------------|------------------------|----------------------------|-----------------------------|--------------------------------|
| <i>saccharogumia,</i>           | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>Clostridium</i>         | <i>Clostridium</i>          | <i>Dactylosporangium</i>       |
| <i>Clostridium</i>              | <i>leptospartum,</i> | <i>Clostridium</i>     | <i>tetanomorphum,</i>      | <i>thermosuccinogenes,</i>  | <i>matsuzakense</i>            |
| <i>saccharolyticum,</i>         | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermaceficum,</i>      | <i>thermosulfurigenes,</i>  | <i>Dactylosporangium</i>       |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>Clostridium</i>         | <i>thiosulfatireducens,</i> | <i>roseum</i>                  |
| <i>saccharoperbutylacetoni-</i> | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermautotrophicum,</i> | <i>Clostridium</i>          | <i>Dactylosporangium</i>       |
| <i>icum</i>                     | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>Clostridium</i>         | <i>thiosulfatireducens,</i> | <i>thailandense</i>            |
| <i>sardinense,</i>              | <i>Clostridium</i>   | <i>thermolacticum,</i> | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Dactylosporangium</i>       |
| <i>santagatiforme,</i>          | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>tyrobutylicum,</i>       | <i>Dactylosporangium</i>       |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>vinaceum</i>                |
| <i>scatogenes,</i>              | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <b>Deinococcus</b>             |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Deinococcus aerius</i>      |
| <i>schmackerense,</i>           | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Deinococcus apachensis</i>  |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Deinococcus aquaticus</i>   |
| <i>scindens,</i>                | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Deinococcus aquatilis</i>   |
| <i>septicum,</i>                | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Deinococcus caeni</i>       |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Deinococcus</i>             |
| <i>sordellii,</i>               | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>radiodurans</i>             |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Deinococcus radiophilus</i> |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <b>Delftia</b>                 |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Delftia acidovorans</i>     |
| <i>sporosphæroides,</i>         | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <b>Desulfotribrio</b>          |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>Desulfotribrio</i>          |
| <i>stercorarium,</i>            | <i>Clostridium</i>   | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          | <i>desulfuricans</i>           |
| <i>Clostridium</i>              | <i>stercorarium</i>  | <i>stercorarium</i>    | <i>thermoaerophilum,</i>   | <i>Clostridium</i>          |                                |

|                                |                              |                                 |                                 |                              |
|--------------------------------|------------------------------|---------------------------------|---------------------------------|------------------------------|
| <b>Diplococcus</b>             | <i>E. ludwigii</i>           | <b>Fangia</b>                   | <i>Flavobacterium</i>           | <i>Geobacillus</i>           |
| <i>Diplococcus</i>             | <i>E. mori</i>               | <i>Fangia hongkongensis</i>     | <i>degetachei</i>               | <i>thermoglucoosidans</i>    |
| <i>pneumoniae</i>              | <i>E. nimpresuralis</i>      | <b>Fastidiosipila</b>           | <i>Flavobacterium</i>           | <i>Geobacillus</i>           |
| <b>Echinicola</b>              | <i>E. oryzae</i>             | <i>Fastidiosipila sanguinis</i> | <i>denitrificans</i>            | <i>stearothermophilus</i>    |
| <i>Echinicola pacifica</i>     | <i>E. pulveris</i>           | <b>Fusobacterium</b>            | <i>Flavobacterium filum</i>     | <b>Geobacter</b>             |
| <i>Echinicola vietnamensis</i> | <i>E. pyrinus</i>            | <i>Fusobacterium</i>            | <i>Flavobacterium flavense</i>  | <i>Geobacter benidjensis</i> |
| <b>Enterobacter</b>            | <i>E. radicincitans</i>      | <i>nucleatum</i>                | <i>Flavobacterium</i>           | <i>Geobacter brennensis</i>  |
| <i>E. aerogenes</i>            | <i>E. taylorae</i>           | <b>Flavobacterium</b>           | <i>frigidarium</i>              | <i>Geobacter chapeliei</i>   |
| <i>E. amnigenus</i>            | <i>E. tunicensis</i>         | <i>Flavobacterium</i>           | <i>Flavobacterium nitutali</i>  | <i>Geobacter grubiatae</i>   |
| <i>E. agglomerans</i>          | <i>E. sakazakii</i>          | <i>antarcticum</i>              | <i>Flavobacterium</i>           | <i>Geobacter</i>             |
| <i>E. arachidis</i>            | <i>Enterobacter soli</i>     | <i>Flavobacterium aquatile</i>  | <i>okeanokites</i>              | <i>hydrogenophilus</i>       |
| <i>E. asburiae</i>             | <b>Enterococcus</b>          | <i>Flavobacterium</i>           | <b>Gaetbulibacter</b>           | <i>Geobacter lovleyi</i>     |
| <i>E. cancerogenus</i>         | <i>Enterococcus durans</i>   | <i>aquidulense</i>              | <i>Gaetbulibacter</i>           | <i>Geobacter</i>             |
| <i>E. cloacae</i>              | <i>Enterococcus faecalis</i> | <i>Flavobacterium</i>           | <i>saemankumensis</i>           | <i>metallireducens</i>       |
| <i>E. cowanii</i>              | <i>Enterococcus faecium</i>  | <i>balustinum</i>               | <b>Gallibacterium</b>           | <i>Geobacter pelophilus</i>  |
| <i>E. dissolvens</i>           | <b>Erwinia</b>               | <i>Flavobacterium</i>           | <i>Gallibacterium anatis</i>    | <i>Geobacter pickeringii</i> |
| <i>E. gergoviae</i>            | <i>Erwinia hapontici</i>     | <i>croceum</i>                  | <b>Gallicola</b>                | <i>Geobacter</i>             |
| <i>E. helveticus</i>           | <b>Escherichia</b>           | <i>Flavobacterium cucumis</i>   | <i>Gallicola barnesae</i>       | <i>sulfurreducens</i>        |
| <i>E. hommaechei</i>           | <i>Escherichia coli</i>      | <i>Flavobacterium</i>           | <b>Gardella</b>                 | <b>Geodermatophilus</b>      |
| <i>E. intermedius</i>          | <b>Faecalibacterium</b>      | <i>daejeonense</i>              | <i>Gardella nitratireducens</i> | <i>Geodermatophilus</i>      |
| <i>Enterobacter kobei</i>      | <i>Faecalibacterium</i>      | <i>Flavobacterium defluvi</i>   | <b>Geobacillus</b>              | <i>obscurus</i>              |
|                                | <i>praesuisii</i>            |                                 |                                 | <b>Gluconacetobacter</b>     |

|                                      |                                      |                                 |                                          |                              |
|--------------------------------------|--------------------------------------|---------------------------------|------------------------------------------|------------------------------|
| <i>Glucacetobacter xylinus</i>       | <i>Haemophilus pitmaniae</i>         | <i>Ignatzschineria larvae</i>   | <i>Jannaschia rubra</i>                  | <b>Kiloniella</b>            |
| <b>Gordonia</b>                      | <b>Hafnia</b>                        | <b>Ignavigranum</b>             | <b>Janthinobacterium</b>                 | <i>Kiloniella lanternae</i>  |
| <i>Gordonia rubripertincta</i>       | <i>Hafnia alvei</i>                  | <i>Ignavigranum ruoffiae</i>    | <i>Janthinobacterium agaricidamnosum</i> | <b>Klebsiella</b>            |
| <b>Haemophilus</b>                   | <b>Hahella</b>                       | <b>Humatobacter</b>             | <i>Janthinobacterium lividum</i>         | <i>K. granulomatis</i>       |
| <i>Haemophilus aegyptius</i>         | <i>Hahella ganghwensis</i>           | <i>Humatobacter fluminis</i>    |                                          | <i>K. oxytoca</i>            |
| <i>Haemophilus aphrophilus</i>       | <b>Halaalkalibacillus</b>            | <b>Ilyobacter</b>               |                                          | <i>K. pneumoniae</i>         |
| <i>Haemophilus felis</i>             | <i>Halaalkalibacillus halophilus</i> | <i>Ilyobacter delafieldii</i>   | <b>Jejuna</b>                            | <i>K. terrigena</i>          |
| <i>Haemophilus gallinarum</i>        | <b>Helicobacter</b>                  | <i>Ilyobacter insuetus</i>      | <i>Jejuna pallidulutea</i>               | <i>K. variicola</i>          |
| <i>Haemophilus haemolyticus</i>      | <i>Helicobacter pylori</i>           | <i>Ilyobacter polytropus</i>    | <b>Jeotgallibacillus</b>                 | <b>Kluyvera</b>              |
| <i>Haemophilus influenzae</i>        | <b>Ideonella</b>                     | <i>Ilyobacter tartarus</i>      | <i>Jeotgallibacillus alimentarius</i>    | <i>Kluyvera ascorbata</i>    |
| <i>Haemophilus paracuniculus</i>     | <i>Ideonella azotifigens</i>         | <b>Janibacter</b>               | <b>Jeotgallibacillus</b>                 | <b>Kocuria</b>               |
| <i>Haemophilus parahaemolyticus</i>  | <b>Idiomarina</b>                    | <i>Janibacter anophelis</i>     | <i>Jeotgallibacillus halotolerans</i>    | <i>Kocuria rosea</i>         |
| <i>Haemophilus parainfluenzae</i>    | <i>Idiomarina abyssalis</i>          | <i>Janibacter limosus</i>       |                                          | <i>Kocuria varians</i>       |
| <i>Haemophilus paraphaemolyticus</i> | <i>Idiomarina baltica</i>            | <i>Janibacter melonis</i>       | <b>Kaistia</b>                           | <b>Kurthia</b>               |
| <i>Haemophilus paraprofundus</i>     | <i>Idiomarina fontislapidosi</i>     | <i>Janibacter terrae</i>        | <i>Kaistia adipata</i>                   | <i>Kurthia zapfii</i>        |
| <i>Haemophilus paratuberculosis</i>  | <i>Idiomarina loihiensis</i>         | <b>Jannaschia</b>               | <i>Kaistia soli</i>                      | <b>Labedella</b>             |
| <i>Haemophilus paratuberculosis</i>  | <i>Idiomarina rambicola</i>          | <i>Jannaschia cystogens</i>     | <b>Kangiella</b>                         | <i>Labedella gwakgiensis</i> |
| <i>Haemophilus paratuberculosis</i>  | <i>Idiomarina seosinensis</i>        | <i>Jannaschia heiglandensis</i> | <i>Kangiella aquimarina</i>              | <b>Labrenzia</b>             |
| <i>Haemophilus paratuberculosis</i>  | <i>Idiomarina zobellii</i>           | <i>Jannaschia pohangensis</i>   | <i>Kangiella koreensis</i>               | <i>Labrenzia aggregata</i>   |
| <i>Haemophilus paratuberculosis</i>  | <b>Ignatzschineria</b>               |                                 | <b>Kerstersia</b>                        | <i>Labrenzia alba</i>        |
|                                      |                                      |                                 | <i>Kerstersia gyllorum</i>               | <i>Labrenzia alexandrii</i>  |

|                             |                                |                                 |                                |                             |
|-----------------------------|--------------------------------|---------------------------------|--------------------------------|-----------------------------|
| <i>Labrenzia marina</i>     | <i>Listeria ivanovi</i>        | <i>Methylophilus</i>            | <i>Nannocystis pusilla</i>     | <i>Nocardia coralina</i>    |
| <b>Labrys</b>               | <i>L. marthii</i>              | <i>methylobrophus</i>           | <b>Natranaerobius</b>          | <i>Nocardia</i>             |
| <i>Labrys</i>               | <i>L. monocytogenes</i>        | <b>Microbacterium</b>           | <i>Natranaerobius</i>          | <i>otitidiscaviarum</i>     |
| <i>methylaniniphilus</i>    | <i>L. newyorkensis</i>         | <i>Microbacterium</i>           | <i>thermophilus</i>            | <b>Lactobacillus</b>        |
| <i>Labrys miyagiensis</i>   | <i>L. riparia</i>              | <i>ammoniaophilum</i>           | <i>Natranaerobius trueperi</i> | <i>L. acetotolerans</i>     |
| <i>Labrys monachus</i>      | <i>L. rocourtae</i>            | <i>Microbacterium</i>           | <b>Naxibacter</b>              | <i>L. acidifarinae</i>      |
| <i>Labrys okinawensis</i>   | <i>L. seeligeri</i>            | <i>arborescens</i>              | <i>Naxibacter</i>              | <i>L. acidiphilis</i>       |
| <i>Labrys portucalensis</i> | <i>L. weihenstephanensis</i>   | <i>Microbacterium</i>           | <i>alkalitolerans</i>          | <i>L. acidophilus</i>       |
| <b>Laceyella</b>            | <i>L. weishimeri</i>           | <i>liquefaciens</i>             | <b>Neisseria</b>               | <i>Lactobacillus agilis</i> |
| <i>Laceyella putida</i>     | <b>Listonella</b>              | <i>Microbacterium</i>           | <i>Neisseria cinerea</i>       | <i>L. algidus</i>           |
| <b>Lechevalieria</b>        | <i>Listonella anguillarum</i>  | <i>oxydans</i>                  | <i>Neisseria denitrificans</i> | <i>L. alimentarius</i>      |
| <i>Lechevalieria</i>        | <b>Macrocooccus</b>            | <b>Microcooccus</b>             | <i>Neisseria gonorrhoeae</i>   | <i>L. amylolyticus</i>      |
| <i>aerodolorigenes</i>      | <i>Macrocooccus bovicus</i>    | <i>Microcooccus luteus</i>      | <i>Neisseria lactamica</i>     | <i>L. amylophilus</i>       |
| <b>Listeria</b>             | <b>Marinobacter</b>            | <i>Microcooccus lysiae</i>      | <i>Neisseria mucosa</i>        | <i>L. amyloprophicus</i>    |
| <i>L. aquatica</i>          | <i>Marinobacter algicola</i>   | <b>Moraxella</b>                | <i>Neisseria sicca</i>         | <i>L. amylovorus</i>        |
| <i>L. booriae</i>           | <i>Marinobacter</i>            | <i>Moraxella bovis</i>          | <i>Neisseria subflava</i>      | <i>L. animalis</i>          |
| <i>L. cornellensis</i>      | <i>bryozorum</i>               | <i>Moraxella</i>                | <b>Neptunomonas</b>            | <i>L. antri</i>             |
| <i>L. fleischmannii</i>     | <i>Marinobacter flavimaris</i> | <i>nonliquefaciens</i>          | <i>Neptunomonas japonica</i>   | <i>L. apodemii</i>          |
| <i>L. floridensis</i>       | <b>Meiothermus</b>             | <i>Moraxella osloensis</i>      | <b>Nesterenkonia</b>           | <i>L. aviarum</i>           |
| <i>L. grandensis</i>        | <i>Meiothermus ruber</i>       | <b>Nakamurella</b>              | <i>Nesterenkonia holobla</i>   | <i>L. bifementans</i>       |
| <i>L. gravi</i>             | <b>Methylophilus</b>           | <i>Nakamurella multipartita</i> | <b>Nocardia</b>                | <i>L. brevis</i>            |
| <i>L. innocua</i>           |                                | <b>Nannocystis</b>              | <i>Nocardia argentinensis</i>  |                             |

|                              |                              |                           |                          |                            |
|------------------------------|------------------------------|---------------------------|--------------------------|----------------------------|
| <i>L. buchneri</i>           | <i>L. delbrueckii</i> subsp. | <i>L. oris</i>            | <i>L. pontis</i>         | <i>L. sanfranciscensis</i> |
| <i>L. camelliae</i>          | <i>Lactis</i>                | <i>L. panis</i>           | <i>L. protectus</i>      | <i>L. satsumensis</i>      |
| <i>L. casei</i>              | <i>L. dextrinicus</i>        | <i>L. pantheris</i>       | <i>L. psittaci</i>       | <i>L. secaliphilus</i>     |
| <i>L. kitasatonis</i>        | <i>L. diolivorans</i>        | <i>L. parabrueis</i>      | <i>L. renini</i>         | <i>L. sharpeae</i>         |
| <i>L. kunkelii</i>           | <i>L. equi</i>               | <i>L. parabuchneri</i>    | <i>L. reuteri</i>        | <i>L. siliginis</i>        |
| <i>L. leichmannii</i>        | <i>L. equigenensis</i>       | <i>L. paracasei</i>       | <i>L. rhamnosus</i>      | <i>L. spicheri</i>         |
| <i>L. lindneri</i>           | <i>L. farraginis</i>         | <i>L. paracollinoides</i> | <i>L. rinae</i>          | <i>L. suebicus</i>         |
| <i>L. malefermentans</i>     | <i>L. farcininis</i>         | <i>L. parafarraginis</i>  | <i>L. rogosa</i>         | <i>L. thailandensis</i>    |
| <i>L. cateniformis</i>       | <i>L. fermentum</i>          | <i>L. homohiochii</i>     | <i>L. rossiae</i>        | <i>L. uftunensis</i>       |
| <i>L. ceti</i>               | <i>L. formicis</i>           | <i>L. iners</i>           | <i>L. ruminis</i>        | <i>L. vaccinostercus</i>   |
| <i>L. coleohominis</i>       | <i>L. fructivorans</i>       | <i>L. ingluviel</i>       | <i>L. saermonii</i>      | <i>L. vaginalis</i>        |
| <i>L. collinoides</i>        | <i>L. frumenti</i>           | <i>L. intestinalis</i>    | <i>L. jensenii</i>       | <i>L. versmoldensis</i>    |
| <i>L. composti</i>           | <i>L. mali</i>               | <i>L. fuchuenensis</i>    | <i>L. johnsonii</i>      | <i>L. vini</i>             |
| <i>L. concavus</i>           | <i>L. manihotivorans</i>     | <i>L. gallinarum</i>      | <i>L. kalixensis</i>     | <i>L. vitulinus</i>        |
| <i>L. coryniformis</i>       | <i>L. mundensis</i>          | <i>L. gasserii</i>        | <i>L. kefiranoformis</i> | <i>L. zeae</i>             |
| <i>L. crispatus</i>          | <i>L. mucosae</i>            | <i>L. parakefirii</i>     | <i>L. kefir</i>          | <i>L. zymae</i>            |
| <i>L. crustorum</i>          | <i>L. murinus</i>            | <i>L. paralimentarius</i> | <i>L. avianus</i>        | <i>L. gastricus</i>        |
| <i>L. curvatus</i>           | <i>L. nagelii</i>            | <i>L. paraplantarum</i>   | <i>L. helveticus</i>     | <i>L. ghanensis</i>        |
| <i>L. delbrueckii</i> subsp. | <i>L. namurensis</i>         | <i>L. pentosus</i>        | <i>L. hilgardii</i>      | <i>L. graminis</i>         |
| <i>Bulgicus</i>              | <i>L. nantensis</i>          | <i>L. perolens</i>        | <i>L. sakei</i>          | <i>L. hammesii</i>         |
| <i>L. delbrueckii</i> subsp. | <i>L. oligofermentans</i>    | <i>L. plantarum</i>       | <i>L. salivarius</i>     | <i>L. hamsteri</i>         |
| <i>Delbrueckii</i>           |                              |                           |                          |                            |

|                                |                                  |                                  |                                 |                                |
|--------------------------------|----------------------------------|----------------------------------|---------------------------------|--------------------------------|
| <i>L. harbinensis</i>          | <i>Legionella dumoffii</i>       | <i>Legionella lytica</i>         | <i>Legionella spiritensis</i>   | <b>Oceaniserpentilla</b>       |
| <i>L. hayakiiensis</i>         | <i>Legionella erythra</i>        | <i>Legionella maceachernii</i>   | <i>Legionella steelei</i>       | <i>Oceaniserpentilla</i>       |
| <b>Legionella</b>              | <i>Legionella fairfieldensis</i> | <i>Legionella massiliensis</i>   | <i>Legionella steigerwaltii</i> | <i>hanoiensis</i>              |
| <i>Legionella adelaidensis</i> | <i>Legionella fallonii</i>       | <i>Legionella micdadei</i>       | <i>Legionella taurinensis</i>   | <b>Oceanisphaera</b>           |
| <i>Legionella arisa</i>        | <i>Legionella feeleyi</i>        | <i>Legionella moravica</i>       | <i>Legionella tucsonensis</i>   | <i>Oceanisphaera</i>           |
| <i>Legionella bellardensis</i> | <i>Legionella geestiana</i>      | <i>Legionella moravica</i>       | <i>Legionella tunisiensis</i>   | <i>donghaensis</i>             |
| <i>Legionella</i>              | <i>Legionella</i>                | <i>Legionella</i>                | <i>Legionella watsworthii</i>   | <i>Oceanisphaera litoralis</i> |
| <i>birninghamensis</i>         | <i>genomospecies</i>             | <i>nagasakiensis</i>             | <i>Legionella waltersii</i>     | <b>Oceanithermus</b>           |
| <i>Legionella bozemanae</i>    | <i>Legionella gormanii</i>       | <i>Legionella nautarum</i>       | <i>Legionella worsteiensis</i>  | <i>Oceanithermus</i>           |
| <i>Legionella brunensis</i>    | <i>Legionella gratiana</i>       | <i>Legionella norlandica</i>     | <i>Legionella yabuuchiiae</i>   | <i>desulfurans</i>             |
| <i>Legionella busanensis</i>   | <i>Legionella gresliensis</i>    | <i>Legionella oakridgensis</i>   | <b>Oceanibulbus</b>             | <i>Oceanithermus</i>           |
| <i>Legionella cardiaca</i>     | <i>Legionella hackeliae</i>      | <i>Legionella parisiensis</i>    | <i>Oceanibulbus indolifex</i>   | <i>profundus</i>               |
| <i>Legionella cherrii</i>      | <i>Legionella impletisoli</i>    | <i>Legionella</i>                | <b>Oceanicaulis</b>             | <b>Oceanobacillus</b>          |
| <i>Legionella</i>              | <i>Legionella israelensis</i>    | <i>pittsburghensis</i>           | <i>Oceanicaulis alexandrii</i>  | <i>Oceanobacillus caeni</i>    |
| <i>cincinnatiensis</i>         | <i>Legionella</i>                | <i>Legionella pneumophila</i>    | <b>Oceanicola</b>               | <b>Oceanospirillum</b>         |
| <i>Legionella clemsonensis</i> | <i>jamestowniensis</i>           | <i>Legionella quateiensis</i>    | <i>Oceanicola batsensis</i>     | <i>Oceanospirillum litrum</i>  |
| <i>Legionella donaldsonii</i>  | <i>Candidatus Legionella</i>     | <i>Legionella quinivanii</i>     | <i>Oceanicola granulosus</i>    | <b>paenibacillus</b>           |
|                                | <i>jeonii</i>                    | <i>Legionella rowbothamii</i>    | <i>Oceanicola nanhaiensis</i>   | <i>Paenibacillus</i>           |
|                                | <i>Legionella jordanis</i>       | <i>Legionella rubifluens</i>     | <b>Oceanimonas</b>              | <i>thiaminolyticus</i>         |
| <i>Legionella francourtii</i>  | <i>Legionella lansingensis</i>   | <i>Legionella saintthelensis</i> | <i>Oceanimonas</i>              | <b>Pantoea</b>                 |
| <i>Legionella dresdenensis</i> | <i>Legionella londiniensis</i>   | <i>Legionella santacrucis</i>    | <i>baumannii</i>                | <i>Pantoea agglomerans</i>     |
| <i>Legionella grozanskii</i>   | <i>Legionella longbeachae</i>    | <i>Legionella shakopeae</i>      |                                 | <b>Paracoccus</b>              |

|                                      |                                    |                                    |                                      |
|--------------------------------------|------------------------------------|------------------------------------|--------------------------------------|
| <i>Paracoccus alcaliphilus</i>       | <i>Pectobacterium dentalis</i>     | <b>Providencia</b>                 | <i>Psychrobacter phenylpyruvicus</i> |
| <b>Paucimonas</b>                    | <i>wasabiae</i>                    | <i>Providencia stuartii</i>        |                                      |
| <i>Paucimonas lemoignei</i>          | <b>Planococcus</b>                 | <b>Pseudomonas</b>                 | <b>Quadriflustra</b>                 |
| <b>Pectobacterium</b>                | <i>Planococcus citreus</i>         | <i>Pseudomonas aeruginosa</i>      | <i>Quadriflustra granulorum</i>      |
| <i>Pectobacterium aridiarum</i>      | <b>Planomicrobium</b>              | <i>Pseudomonas alcaligenes</i>     | <b>Quatrinicoccus</b>                |
| <i>Pectobacterium atrosepticum</i>   | <i>Planomicrobium okeanokoites</i> | <i>Pseudomonas anguilliseptica</i> | <i>Quatrinicoccus australiensis</i>  |
| <i>Pectobacterium betavasculorum</i> | <b>Plesiomonas</b>                 | <i>Pseudomonas pseudomonas</i>     | <b>Quinella</b>                      |
| <i>Pectobacterium cacticida</i>      | <i>Plesiomonas shigelloides</i>    | <i>Pseudomonas fluorescens</i>     | <i>Quinella ovalis</i>               |
| <i>Pectobacterium carnegleana</i>    | <b>Prevotella</b>                  | <i>Pseudomonas halopterae</i>      | <b>Raistonia</b>                     |
| <i>Pectobacterium carotovorum</i>    | <i>Prevotella albensis</i>         | <i>Pseudomonas maltophilia</i>     | <i>Raistonia eutropha</i>            |
| <i>Pectobacterium chrysanthemi</i>   | <i>Prevotella amnii</i>            | <i>Pseudomonas maltophilia</i>     | <i>Raistonia insidiosa</i>           |
| <i>Pectobacterium cypripedii</i>     | <i>Prevotella bergensis</i>        | <i>Pseudomonas maltophilia</i>     | <i>Raistonia mannitolitica</i>       |
| <i>Pectobacterium rhapontici</i>     | <i>Prevotella bivia</i>            | <i>Pseudomonas maltophilia</i>     | <i>Raistonia pickettii</i>           |
|                                      | <i>Prevotella brevis</i>           | <i>Pseudomonas maltophilia</i>     | <i>Raistonia</i>                     |
|                                      | <i>Prevotella bryantii</i>         | <i>Pseudomonas maltophilia</i>     | <i>pseudosolanacearum</i>            |
|                                      | <i>Prevotella buccae</i>           | <i>Pseudomonas putida</i>          | <i>Raistonia syzygii</i>             |
|                                      | <i>Prevotella buccalis</i>         | <i>Pseudomonas tutzeri</i>         | <i>Raistonia solanacearum</i>        |
|                                      | <i>Prevotella copri</i>            | <i>Pseudomonas syringae</i>        | <b>Ramlibacter</b>                   |
|                                      |                                    | <b>Psychrobacter</b>               | <i>Ramlibacter hutchinsonii</i>      |
|                                      |                                    | <i>Psychrobacter faecalis</i>      |                                      |

|                                |                          |                                |                                   |                              |
|--------------------------------|--------------------------|--------------------------------|-----------------------------------|------------------------------|
| <i>Ramlibacter</i>             | <b>Saccharomonospora</b> | <i>Saccharopolyspora</i>       | <i>Salegentibacter</i>            | <i>Sanguibacter keddiei</i>  |
| <i>tataouensis</i>             | <i>Saccharomonospora</i> | <i>spinosa</i>                 | <i>selegens</i>                   | <i>Sanguibacter suereii</i>  |
| <b>Raoultella</b>              | <i>azurea</i>            | <i>Saccharopolyspora</i>       | <b>Salimicrobium</b>              | <b>Saprosira</b>             |
| <i>Raoultella</i>              | <i>Saccharomonospora</i> | <i>taberi</i>                  | <i>Salimicrobium album</i>        | <i>Saprosira grandis</i>     |
| <i>ornithinolytica</i>         | <i>cyanea</i>            | <b>Saccharothrix</b>           | <b>Salinibacter</b>               | <b>Sarcina</b>               |
| <i>Raoultella planticola</i>   | <i>Saccharomonospora</i> | <i>Saccharothrix</i>           | <i>Salinibacter ruber</i>         | <i>Sarcina maxima</i>        |
| <i>Raoultella terrigena</i>    | <i>viridis</i>           | <i>australiensis</i>           | <b>Salinicoccus</b>               | <i>Sarcina ventriculi</i>    |
| <b>Rathayibacter</b>           | <b>Saccharophagus</b>    | <i>Saccharothrix</i>           | <i>Salinicoccus alkaliophilus</i> | <b>Sebaldella</b>            |
| <i>Rathayibacter canis</i>     | <i>Saccharophagus</i>    | <i>coeruleofusca</i>           | <i>Salinicoccus hispanicus</i>    | <i>Sebaldella termitidis</i> |
| <i>Rathayibacter festucae</i>  | <i>degradans</i>         | <i>Saccharothrix</i>           | <i>Salinicoccus roseus</i>        | <b>Serratia</b>              |
| <i>Rathayibacter iranicus</i>  | <b>Saccharopolyspora</b> | <i>espanensis</i>              | <b>Salinispora</b>                | <i>Serratia fonticola</i>    |
| <i>Rathayibacter rathayi</i>   | <i>Saccharopolyspora</i> | <i>Saccharothrix</i>           | <i>Salinispora arenicola</i>      | <i>Serratia marcescens</i>   |
| <i>Rathayibacter toxicus</i>   | <i>erythraea</i>         | <i>longispora</i>              | <i>Salinispora tropica</i>        | <b>Sphaerotilus</b>          |
| <i>Rathayibacter tritici</i>   | <i>Saccharopolyspora</i> | <i>Saccharothrix mutabilis</i> | <b>Salinivibrio</b>               | <i>Sphaerotilus natans</i>   |
| <b>Rhodobacter</b>             | <i>gregorii</i>          | <i>Saccharothrix syringae</i>  | <i>Salinivibrio costicola</i>     | <b>Sphingobacterium</b>      |
| <i>Rhodobacter</i>             | <i>Saccharopolyspora</i> | <i>Saccharothrix</i>           | <b>Salmonella</b>                 | <i>Sphingobacterium</i>      |
| <i>sphaeroides</i>             | <i>hirsuta</i>           | <i>tangerinus</i>              | <i>Salmonella bongori</i>         | <i>multivorum</i>            |
| <b>Ruegeria</b>                | <i>Saccharopolyspora</i> | <i>Saccharothrix</i>           | <i>Salmonella enterica</i>        | <b>Stenotrophomonas</b>      |
| <i>Ruegeria gelatinovorans</i> | <i>hordei</i>            | <i>texasensis</i>              | <i>Salmonella subterranea</i>     | <i>Stenotrophomonas</i>      |
| <b>Saccharococcus</b>          | <i>Saccharopolyspora</i> | <b>Sagittula</b>               | <i>Salmonella typhi</i>           | <i>malophilia</i>            |
| <i>Saccharococcus</i>          | <i>rectivirgula</i>      | <i>Sagittula stieliae</i>      | <b>Sanguibacter</b>               | <b>Streptomyces</b>          |
| <i>thermophilus</i>            |                          | <b>Salegentibacter</b>         |                                   |                              |

|                               |                                |                                 |                        |                             |
|-------------------------------|--------------------------------|---------------------------------|------------------------|-----------------------------|
| <i>Streptomyces</i>           | <i>Streptomyces</i>            | <i>Tepidibacter</i>             | <i>S. epidermidis</i>  | <i>S. pettenkoferi</i>      |
| <i>achromogenes</i>           | <i>thermodiastaticus</i>       | <i>formicigenes</i>             | <i>S. equorum</i>      | <i>S. piscifermentans</i>   |
| <i>Streptomyces cesaiibis</i> | <i>Streptomyces</i>            | <i>Tepidibacter thalassicus</i> | <i>S. felis</i>        | <i>S. pseudointermedius</i> |
| <i>Streptomyces</i>           | <i>tuberculosis</i>            | <b>Thermus</b>                  | <i>S. fleurettii</i>   | <i>S. pseudolugdunensis</i> |
| <i>cescaeplisus</i>           | <b>Tatlockia</b>               | <i>Thermus aquaticus</i>        | <i>S. gallinarum</i>   | <i>S. pulvereri</i>         |
| <i>Streptomyces</i>           | <i>Tatlockia maceachernii</i>  | <i>Thermus filiformis</i>       | <i>S. haemolyticus</i> | <i>S. rostri</i>            |
| <i>cesdiastaticus</i>         | <i>Tatlockia micdadei</i>      | <i>Thermus thermophilus</i>     | <i>S. hominis</i>      | <i>S. saccharolyticus</i>   |
| <i>Streptomyces</i>           | <b>Tenacibaculum</b>           | <b>Staphylococcus</b>           | <i>S. hyicus</i>       | <i>S. saprophyticus</i>     |
| <i>cesexfoliatus</i>          | <i>Tenacibaculum</i>           | <i>S. arlettae</i>              | <i>S. intermedius</i>  | <i>S. schleiferi</i>        |
| <i>Streptomyces</i>           | <i>amycolyticum</i>            | <i>S. agnetis</i>               | <i>S. kloosii</i>      | <i>S. sciuri</i>            |
| <i>fimbriatus</i>             | <i>Tenacibaculum discolor</i>  | <i>S. aureus</i>                | <i>S. leei</i>         | <i>S. simiae</i>            |
| <i>Streptomyces fradiae</i>   | <i>Tenacibaculum</i>           | <i>S. auricularis</i>           | <i>S. lentus</i>       | <i>S. simulans</i>          |
| <i>Streptomyces</i>           | <i>gallaicum</i>               | <i>S. capitis</i>               | <i>S. lugdunensis</i>  | <i>S. stepanovicii</i>      |
| <i>fulvissimus</i>            | <i>Tenacibaculum lutinaris</i> | <i>S. caprae</i>                | <i>S. lubrae</i>       | <i>S. succinus</i>          |
| <i>Streptomyces</i>           | <i>Tenacibaculum</i>           | <i>S. carnosus</i>              | <i>S. lyticans</i>     | <i>S. vitulinus</i>         |
| <i>griseoruber</i>            | <i>mesophilum</i>              | <i>S. caseolyticus</i>          | <i>S. massiliensis</i> | <i>S. warneri</i>           |
| <i>Streptomyces griseus</i>   | <i>Tenacibaculum</i>           | <i>S. chromogenes</i>           | <i>S. microbi</i>      | <i>S. xylosus</i>           |
| <i>Streptomyces</i>           | <i>skagerakense</i>            | <i>S. cohnii</i>                | <i>S. muscae</i>       | <b>Streptococcus</b>        |
| <i>lavendulae</i>             | <b>Tepidanaerobacter</b>       | <i>S. condimenti</i>            | <i>S. nepalensis</i>   | <i>Streptococcus</i>        |
| <i>Streptomyces</i>           | <i>Tepidanaerobacter</i>       | <i>S. delphini</i>              | <i>S. pasteurii</i>    | <i>agalactiae</i>           |
| <i>phaeochromogenes</i>       | <i>syntrophicus</i>            | <i>S. devriesei</i>             | <i>S. petrasii</i>     | <i>Streptococcus</i>        |
|                               | <b>Tepidibacter</b>            |                                 |                        | <i>anginosus</i>            |

|                                   |                                       |                                      |                                     |                                   |
|-----------------------------------|---------------------------------------|--------------------------------------|-------------------------------------|-----------------------------------|
| <i>Streptococcus bovis</i>        | <i>Streptococcus</i>                  | <i>Uliginosibacterium</i>            | <i>Vagococcus fluvialis</i>         | <i>Venenivibrio</i>               |
| <i>Streptococcus canis</i>        | <i>parasanguinis</i>                  | <i>gangwonense</i>                   | <i>Vagococcus lutrae</i>            | <i>stagnispumantis</i>            |
| <i>Streptococcus constellatus</i> | <i>Streptococcus peroris</i>          | <b>Ulvibacter</b>                    | <i>Vagococcus</i>                   | <b>Verminephrobacter</b>          |
| <i>Streptococcus downei</i>       | <i>Streptococcus pneumoniae</i>       | <i>Ulvibacter litoralis</i>          | <i>salmicinarum</i>                 | <i>Verminephrobacter eiseniae</i> |
| <i>Streptococcus dysgalactiae</i> | <i>Streptococcus pseudopneumoniae</i> | <b>Umezawaea</b>                     | <b>Varlovorax</b>                   | <b>Verrucomicrobium</b>           |
| <i>Streptococcus equines</i>      | <i>Streptococcus pyogenes</i>         | <i>Umezawaea angema</i>              | <i>Varlovorax</i>                   | <i>Verrucomicrobium spinosum</i>  |
| <i>Streptococcus faecalis</i>     | <i>Streptococcus rattus</i>           | <b>Undibacterium</b>                 | <i>Varlovorax dokdonensis</i>       | <b>Vibrio</b>                     |
| <i>Streptococcus ferus</i>        | <i>Streptococcus salivarius</i>       | <b>Ureaplasma</b>                    | <i>Varlovorax paradoxus</i>         | <i>Vibrio aerogenes</i>           |
| <i>Streptococcus infantarius</i>  | <i>Streptococcus thermophilus</i>     | <i>Ureaplasma urealyticum</i>        | <i>Varlovorax soli</i>              | <i>Vibrio aestuarius</i>          |
| <i>Streptococcus iniae</i>        | <i>Streptococcus sanguinis</i>        | <b>Ureibacillus</b>                  | <b>Veillonella</b>                  | <i>Vibrio albensis</i>            |
| <i>Streptococcus intermedius</i>  | <i>Streptococcus sobrinus</i>         | <i>Ureibacillus composti</i>         | <i>Veillonella atypica</i>          | <i>Vibrio alginolyticus</i>       |
| <i>Streptococcus lactarius</i>    | <i>Streptococcus uberis</i>           | <i>Ureibacillus suwonensis</i>       | <i>Veillonella caviae</i>           | <i>Vibrio campbellii</i>          |
| <i>Streptococcus milleri</i>      | <i>Streptococcus vestibularis</i>     | <i>Ureibacillus terrenus</i>         | <i>Veillonella criceti</i>          | <i>Vibrio cholerae</i>            |
| <i>Streptococcus mutans</i>       | <i>Streptococcus viridans</i>         | <i>Ureibacillus thermophilus</i>     | <i>Veillonella dispar</i>           | <i>Vibrio cincinnatiensis</i>     |
| <i>Streptococcus oralis</i>       | <i>Streptococcus zooepidemicus</i>    | <i>Ureibacillus thermosphaericus</i> | <i>Veillonella montpellierensis</i> | <i>Vibrio coralliilyticus</i>     |
| <i>Streptococcus orisratti</i>    | <b>Uliginosibacterium</b>             | <b>Vagococcus</b>                    | <i>Veillonella parvula</i>          | <i>Vibrio cyclitrophicus</i>      |
|                                   |                                       | <i>Vagococcus carnegii</i>           | <i>Veillonella ratti</i>            | <i>Vibrio diazotrophicus</i>      |
|                                   |                                       | <i>Vagococcus elongatus</i>          | <b>Venenivibrio</b>                 | <i>Vibrio fluvialis</i>           |
|                                   |                                       | <i>Vagococcus fessus</i>             |                                     | <i>Vibrio furnissii</i>           |
|                                   |                                       |                                      |                                     | <i>Vibrio gazogenes</i>           |

|                                       |                                       |                                        |                                      |                                       |
|---------------------------------------|---------------------------------------|----------------------------------------|--------------------------------------|---------------------------------------|
| <i>Vibrio</i> <b>halotitolicus</b>    | <i>Virgibacillus</i>                  | <i>Winogradskyella</i>                 | <i>Xanthomonas</i>                   | <i>Xanthomonas</i> <b>axipull</b>     |
| <i>Vibrio</i> <b>harveyi</b>          | <i>halodenitrificans</i>              | <i>thalassocola</i>                    | <i>altivirens</i>                    | <i>Xanthomonas</i> <b>theicola</b>    |
| <i>Vibrio</i> <b>ichthyocyteri</b>    | <i>Virgibacillus</i>                  | <b>Wolbachia</b>                       | <i>Xanthomonas</i> <b>alfalfae</b>   | <i>Xanthomonas</i>                    |
| <i>Vibrio</i> <b>mediterranei</b>     | <i>panthoentis</i>                    | <i>Wolbachia</i> <b>persica</b>        | <i>Xanthomonas</i> <b>arboricola</b> | <i>translucens</i>                    |
| <i>Vibrio</i> <b>metschnikovii</b>    | <b>Weissella</b>                      | <b>Wolinella</b>                       | <i>Xanthomonas</i>                   | <i>Xanthomonas</i>                    |
| <i>Vibrio</i> <b>mytili</b>           | <i>Weissella</i> <b>cibaria</b>       | <i>Wolinella</i> <b>succinogenes</b>   | <i>axinopodis</i>                    | <i>vesicatoria</i>                    |
| <i>Vibrio</i> <b>natriegens</b>       | <i>Weissella</i> <b>confusa</b>       | <b>Zobellia</b>                        | <i>Xanthomonas</i>                   | <b>Xylella</b>                        |
| <i>Vibrio</i> <b>navarrensis</b>      | <i>Weissella</i> <b>halotolerans</b>  | <i>Zobellia</i> <b>galactanivorans</b> | <i>campestris</i>                    | <i>Xylella</i> <b>fastidiosa</b>      |
| <i>Vibrio</i> <b>neris</b>            | <i>Weissella</i> <b>hellenica</b>     | <i>Zobellia</i> <b>uliginosa</b>       | <i>Xanthomonas</i> <b>citri</b>      | <b>Xylophilus</b>                     |
| <i>Vibrio</i> <b>nigripulchritudo</b> | <i>Weissella</i> <b>kandieri</b>      | <b>Zoogloea</b>                        | <i>Xanthomonas</i> <b>codiae</b>     | <i>Xylophilus</i> <b>ampelinus</b>    |
| <i>Vibrio</i> <b>ordalii</b>          | <i>Weissella</i> <b>koreensis</b>     | <i>Zoogloea</i> <b>ramigera</b>        | <i>Xanthomonas</i>                   | <b>Xenophilus</b>                     |
| <i>Vibrio</i> <b>orientalis</b>       | <i>Weissella</i> <b>minor</b>         | <i>Zoogloea</i> <b>resiniphila</b>     | <i>curvibatae</i>                    | <i>Xenophilus</i> <b>azovorum</b>     |
| <i>Vibrio</i> <b>parahaemolyticus</b> | <i>Weissella</i>                      | <b>Xanthobacter</b>                    | <i>Xanthomonas</i>                   | <i>Xenorhabdus</i>                    |
| <i>Vibrio</i> <b>pectenicauda</b>     | <i>paramesenteroides</i>              | <i>Xanthobacter</i> <b>agilis</b>      | <i>euvasicaria</i>                   | <i>Xenorhabdus</i> <b>bestdingii</b>  |
| <i>Vibrio</i> <b>penaeicida</b>       | <i>Weissella</i> <b>soii</b>          | <i>Xanthobacter</i>                    | <i>Xanthomonas</i> <b>fragariae</b>  | <i>Xenorhabdus</i> <b>bovienii</b>    |
| <i>Vibrio</i> <b>proteolyticus</b>    | <i>Weissella</i> <b>thailandensis</b> | <i>aminoxidans</i>                     | <i>Xanthomonas</i> <b>fuscans</b>    | <i>Xenorhabdus</i>                    |
| <i>Vibrio</i> <b>shilonii</b>         | <i>Weissella</i> <b>viridescens</b>   | <i>Xanthobacter</i>                    | <i>Xanthomonas</i> <b>gardneri</b>   | <i>cabanillasii</i>                   |
| <i>Vibrio</i> <b>splendendus</b>      | <b>Williamsia</b>                     | <i>autotrophicus</i>                   | <i>Xanthomonas</i> <b>hortorum</b>   | <i>Xenorhabdus</i> <b>doucetiae</b>   |
| <i>Vibrio</i> <b>tubiaschii</b>       | <i>Williamsia</i> <b>marianensis</b>  | <i>Xanthobacter</i> <b>flavus</b>      | <i>Xanthomonas</i> <b>hyacinthii</b> | <i>Xenorhabdus</i> <b>griffithiae</b> |
| <i>Vibrio</i> <b>vulnificus</b>       | <i>Williamsia</i> <b>maris</b>        | <i>Xanthobacter</i> <b>tagetidis</b>   | <i>Xanthomonas</i> <b>perforans</b>  | <i>Xenorhabdus</i> <b>hormickii</b>   |
| <b>Virgibacillus</b>                  | <i>Williamsia</i> <b>sermedensis</b>  | <i>Xanthobacter</i> <b>viscosus</b>    | <i>Xanthomonas</i> <b>phaseoli</b>   | <i>Xenorhabdus</i>                    |
|                                       | <b>Winogradskyella</b>                | <b>Xanthomonas</b>                     | <i>Xanthomonas</i> <b>pisi</b>       | <i>koppenhoeferi</i>                  |

|                                |                                 |                               |
|--------------------------------|---------------------------------|-------------------------------|
| <i>Xenorhabdus</i>             | <i>Yersinia philomiragia</i>    | <b>Zymophilus</b>             |
| <i>nematophila</i>             | <i>Yersinia pestis</i>          | <i>Zymophilus paucivorans</i> |
| <i>Xenorhabdus poinarii</i>    | <i>Yersinia</i>                 | <i>Zymophilus</i>             |
| <b>Xylanibacter</b>            | <i>pseudotuberculosis</i>       | <i>raffinivorans</i>          |
| <i>Xylanibacter oryzae</i>     | <i>Yersinia rohdei</i>          | <b>Zobellia</b>               |
| <b>Yangia</b>                  | <i>Yersinia ruckeri</i>         | <i>Zobellia denitrificans</i> |
| <i>Yangia pacifica</i>         | <b>Yokenella</b>                | <i>Zobellia taiwanensis</i>   |
| <b>Yaniella</b>                | <i>Yokenella regensburgeri</i>  | <b>Zeaxanthinibacter</b>      |
| <i>Yaniella flava</i>          | <b>Yonghaparkia</b>             | <i>Zeaxanthinibacter</i>      |
| <i>Yaniella halotolerans</i>   | <i>Yonghaparkia alkaliphila</i> | <i>enostimensis</i>           |
| <b>Yeosuana</b>                | <b>Zavarzinia</b>               | <b>Zhihengliuella</b>         |
| <i>Yeosuana</i>                | <i>Zavarzinia compransoris</i>  | <i>Zhihengliuella</i>         |
| <i>aromativorans</i>           | <b>Zooshikella</b>              | <i>halotolerans</i>           |
| <b>Yersinia</b>                | <i>Zooshikella</i>              | <b>Xylanibacterium</b>        |
| <i>Yersinia aldovae</i>        | <i>ganghivensis</i>             | <i>Xylanibacterium litni</i>  |
| <i>Yersinia bercovieri</i>     | <b>Zunongwangia</b>             |                               |
| <i>Yersinia enterocolitica</i> | <i>Zunongwangia</i>             |                               |
| <i>Yersinia entomophaga</i>    | <i>profunda</i>                 |                               |
| <i>Yersinia frederiksenii</i>  | <b>Zymobacter</b>               |                               |
| <i>Yersinia intermedia</i>     | <i>Zymobacter palmae</i>        |                               |
| <i>Yersinia kristensenii</i>   | <b>Zymomonas</b>                |                               |
| <i>Yersinia mollaretii</i>     | <i>Zymomonas mobilis</i>        |                               |

**TABLA 4: SECUENCIAS**

Las secuencias de nucleótidos se escriben en dirección 5' a 3'; las secuencias de aminoácidos se escriben en dirección N a C-terminal.

| <b>SEQ ID NO:</b> | <b>Descripción</b>                   | <b>Secuencia</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-------------------|--------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1                 | Secuencia de aminoácidos de Cas-S1.1 | LNHPVESVY SALT SILL PVMGEVPVQVRNCSCGRAPSEFDGVGFELVNAVRRVYVHCRPCQTFVSAPELMGVENRKPTTG QKFGMWSGVGAVINVEDNSSVLLAPQGVVINKLPEHFDHVEVITATSGQHLEYLNFTELKPLTYIQNFGVKTYELVRSLSL SADAYTCADQLL TRQNEVLYMLDLTKAKELHQEIKNYSKEIDIFIRIVTLLAYSRTTPEAASNEFKNNLPLLLLLPTDPHQRL SILHLKKV                                                                                                                                                                                                                                                                                                                                                                         |
| 2                 | Secuencia de aminoácidos de Cas-S2.1 | MIYYDNAFRLRIRHTSVYQIHQHLDFFMQERKGEKLPYSFKIPFGTGGDSLLLVRTATALELPGKRELLISEGHEIKFTTSMAL FHKGTKSEGRRRQFAPSEEA SYQLALTKLAKAGFKPGQIVVSGPKFVHVTKGNAGRGFTLPVFTVQGTATISNQQAEEVGIVY GVGPKRVFGCGFMHLAQ                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 3                 | Secuencia de aminoácidos de Cas-S3.1 | MNTSAVFESAGLSLRKVKQDYIEAAGALTDQHKVALISAETGVGKTLGYLVPALLILKNPEAKFVJATNSHALMHQIFRSRDRP LLEQIAEQCGIKVTFSRIMGKANYVSLKVRGLLMDFTDLDTVKVLEKLANWSKPLVEFEEY GELPAQITPEMVTYSIWDD IQDIDDIRLINALSANETVTTTHAMVMVDCMCHRLGDKENMYLIIDEADIFVDMLEVMWKQRRFNRLRELTSAFNEHIPRNGVHV IDQLMNDVTSIAGDLHFCSTPAVAALFDNSFNALSKVGREKNEAARKAFDCIYSWEMGLSGGQKGVGVSNKRRREPALLAV NPFIGMNVGRYCTQWRSALLTSATLSITSTPETGMELCKALGLTSDTISRKIFSPDVYGSMLKTIAGADFPKVFNIDPKEQIF SGQWLKAVVEQLSCIQGPALVLTASHYETRMIANQLGEV/SQPVYIQKAGQALSEIKQYQEPGILISAGASVGVSPRGENGEGQ IFQDLITRIPFLPPDRMKAESLYGYLKERGYSRTFEAVNRNIVLDNLRKVRKAKQSIGRIRSENDTVRIIILDRPREPTDLSS KHRSLHHIPVRFRRAYRSCEILSPAYCEEIQ |
| 4                 | Secuencia de aminoácidos de Cas-S4.1 | VLNFKPYRVIMSSLTPVVISGIAPSLDGILYEALSQAIPSNEPGW/LARLKEILLFNDELGVFHASSLRFGITPEQGTGATTTSVRCD YLSPEKLTAMFSPRIHRGVFTRVLLTGGPTKRRMTTRPAYSAPLYTFDFVGSSEAVEILLNHAHVGVGYDFFSAAANGEFNVT ILPLDIDTSSISNEGMAIRPVVPVNSGLNGIKGVSP LIPPYFVGEKLNIVHPAPVRTQLSSLRG                                                                                                                                                                                                                                                                                                                                                                                              |
| 5                 | Secuencia de aminoácidos de Cas-S5.1 | MRTLNFNGKISTLEPLTVTVKNAVSTSGHRLPRNGGFNAAPYFPGTSIRGTLRHAHKVTVDRVGLNADGKSSFDLAEHFMLA QGVQINGEAEITFAPGEINAGAEELRSKNPLISLFGRWGLSGKVGIGNAIPDGDWQWGMFGGARSIMFQRDESIMFELETQV DRLERLLEEQAESAVDISQIKTEQDALKKAMKSADKDTKAEIQKVRLEDEKIQARKDQKQESRESIRRPIDPYEAFITGAELSH RMSIKNATDEEAGLFI SALIRFAAEPRFGGHANHCGLVEAHWTVTTWKPGELVPVTLGEIVTPNGVEITGDELFA MVKAFNE NQSFDFIAR                                                                                                                                                                                                                                                                                   |
| 6                 | Secuencia repetida                   | GTATTCGCCCCCGTGTGCGGGGGTTATCGG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |

|   |                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|---|---------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 7 | Secuencia de ADN<br>de Cas-S1.1 | TTGAATCATCTCTAGAGTCAGTATATAGCGCTCTGACGAGTATTTTATTGCGCTATATGGGTGAGCCTGTGCCTGTGCAA<br>CGAAATTGCTCATGCTGGCGAGAGCGCCTCAGAGTTGACGGGGTAGGCTTCGAGCTGGTCAACGCTTACAGGGAAC<br>GTGTTGTGCAATTGCGTCTTGCCAGACATTTTGTGTTCTGCTCAGAACTGATGGGCGTGGAGAACCCCAAAACCAA<br>CAACCGCCAGAAATTTGGCATGTGGTCTGGGCTGTTAATAAGCTTGAAGACAACTTCTCGGTCTTTTAGCC<br>CCTCAGGGAGTGTCAATAAATACCTAGAGCATTTTTCGATCAGTAGAAGTATCAGGCCACCGGTCAGCACCTTG<br>GAATACCTGTTTAAACCCAACTTAAATTCGCGTGAATTTATATCAAACTTCGGCGTGAACCTACGAGCTTGTTCGC<br>TCACTGGGGTCACTGAGTGGCAGCAATTTATACCTGTGCGATCAACTACCGCGGAGAACGAGTCCCTTTAC<br>ATGCTGGAATTAACAAAGGCTAAGAACTCCATCAGGAAATAAAAACTACTCCAAAAGAGATAGACATCTTTATCAGA<br>ACAGTAACCTGCTGGCCTATTTCGGGATAACTCTGAAGCTGCTCGAATGAGTTTAAAGAAAACAACTCATCTCCGTTA<br>CTGCTGCTTCTGCCGACTGATC                                                                                                                                                                                                                                                                                                                                            |
| 8 | Secuencia de ADN<br>de Cas-S2.1 | ATGATTTACTAGATAATGCGTTTCGGTTGCGTATCGTCAACAAGTGTACCAGATCCACAGCACCTGGATTTTTC<br>ATGAGGAACGTAAGGGAGAGAGCTGCCGTACTCTTTAAATTTTCCGGTACCGGGATGATTCGTACTCCTGGT<br>CCGACAGCCACAGCTCTGGAGTTGCCAGGGGAGAAAAGAGAGCTTATCTCTGAGGGCCATGAATCAAGTTCA<br>TTACGTAAATGGCCATCTTTATAAAGGAACAAGTCAGAGGCGTGGCCGACCGAGTTTGTCTCCATCCGAAGAGGCT<br>AGTTATCAGTTGGCCCTACCAAAATAGCGAAGGCTGGTTTAAACCGGTCAATCGTCTCAGTGGCCCGAAAATTCGTC<br>CATATCGTCAAGGAAATGCTGGACGCGGTTTACGCTACCGGTATTCAGGTACAGGTACCGCCATAATCAGCAATCAA<br>CAGGAAGCTGAAGTTGGAATCGTTTATGGCGTAGGGCCCTAAACGCGTATTTGGTTGCGGTTTTCATGCACTCTGGCGGGCA<br>GTAG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| 9 | Secuencia de ADN<br>de Cas-S3.1 | ATGAATACATCAGCGGTATTTGAGTCTGCCGATTTGCTTACGTAAGGTCCACAGGATTAATCGAGGCTGCAGCTGG<br>TGGCTGACTCAGGATCATAAAGTTGCACTGATTAGTGTGAACAGGTGTAGGAAAACGCTGGGCTATCTGGTACCGG<br>CGCTCTGATTTCTGTGAACACCCAGAGGCAAAAATTTGTGATAGCCACGAATTTCTCATGCCCTGATGCACCAGATTTCA<br>GAAGCGATCGTCTCTCTTGAACAGATAGCTGAACAAATGCGGCATAAAGTTACTTTTCCCGGCTTATGGGGAAGCAA<br>ATTACGTATCCCTGGAGAAAGTACGTGGAATGTTACTCATGGACGAAATTTACCGATCTGGATACGGTTAAGGTCTTGAAA<br>AACTCGCTAACTGGTCAAACTCTGGTTGAGTTTGAAGAAGATATGGTGAACCTGCCAGCGCAGATCACGCCAGAGATG<br>GTCACCTACTCCATCTGGGATGATATACAGGACATTTGATGATTTGCGCTGAATGCTCTCAGTGTAACTTTTATAGTTACG<br>AOCCTATGCGATGGTAATGGTCGACTGTAATGTAATCACCGCATCTCGGTGACAAGAAAATATGTACCTCAATTTATTGAT<br>GAGGCAGACATTTTGTAGATATGCTTGAGGTCTGGAACACAGCGCGCTTCAACCTCAGAGAGTTGACCAGTGGCTCAA<br>TGAACATATCCACGTAATGGCGTGCATGTCATCGACCAGTTAATGAACGACGTGACTTCGATAGCGGGGATCTGCAATTT<br>TTGTAGTACGCTGCAGCGGTTGCCCTGTTTGACAAATAGCTTTAACGCCCTTATCAAGGTAGGGCGGGGAAATAAAAAATG<br>AAGCGGCCCGGAAAGCTTCTTTGACTGTATCTACAGCTGGGAAATGCTCGGATTTGTCGGGGGGCAAAAGGGGTGAGG<br>CGTTTCCAAATAACGTCTGTAACCCGCTGTAAACCGGTTTATCGGGATGAATGTGGGCAGATACTGCACCTCA |

|    |                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|----|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 10 | Secuencia de ADN<br>de Cas-S4.1 | <p>ATGGCGTAGTGGCTTACTGACGTGGGTACGCTCTCAATTACCAGTACCCAGAGACCGGAATGGAGTGGTTATGTAAG<br/>CTCTGGGGTTAACCAGCGATACGATTTCAATCAGGAAATATCTCCAGAGCTCTATGGCTCAATGAAGCTGACCATCG<br/>CCGGTCTGATTTCCAAAGGTGTTCAATGACCCGAAAGAACAGATATTTTCAGGTCAATGGTTAAAGGCAGTAGTGAAC<br/>AATTGTCTGTATTCAGGGGCCAGCTCTGGTATTAAGTCCAGCCCTGTCGAAATTTAATTAACAGTATCAGGAGATACCAAG<br/>GAGGTCTCCCAACCAAGTTATATTCAGAAAGCAGGTACAGCCCTGTCGAAATTTAATTAACAGTATCAGGAGATACCAAG<br/>GATTCATTTCTGCCGTGCTTCTGTTGGTGTAGCCACAGGCGAGAAATGGGGAACAGATTTTCCAGGATTTAATCA<br/>TTACCCGGATACCTTTCTACCTCCGACAGAAATGAAGCTGAAGCCCTTATGGGTATCTGAAGGAACGAGGCTATAGT<br/>CGTACCTTTGAGGCAGTTAACCGTAACATCTACCTGGATAATTTGCGGAAGTTATTCGTAAGCAAGCAGTCTATAGGG<br/>CGGGGATTCGTAGTGAAATGATACCGTACGGATAATATCTCTGATCCGCTTTCGGAACCTACTGACCTTTCTGCTCT<br/>AAACATCGGTGCTTGAACACATTAATCCCGTTTCGATTTCGTCGTGCATATCGTTCTGTGAATTTTATCTCCGGCGTAT<br/>TGTGAAGAGGATATCCAGTG</p>                                                                |
| 11 | Secuencia de ADN<br>de Cas-S5.1 | <p>GTGTTAAATTTAAACCGTATCGGGTATCATGTCTCTGACACCGGTGGTCAATCAGTGGGATAGCCCCATCGCTGGAC<br/>GGCATTCGTATGAAGCACTGTCTCAGGGGATACCCAGTACGAGCCAGGAGTCGTATTTGGCTCGTCTTAAGGAAATACT<br/>TCTATTAACGATGAAGTGGTGTCTTCCATGCTTCTATCTACGTTTGGCATAAACCCGGAACAGGGGATTTGGGGCGA<br/>CAACAAGGTACGTTGGGATTTCTCAGTCCCGAAATCTGACACTGCAATGTTCTCTCTGATCCACAGGGGGTGT<br/>TTACGGTGTCTCTGACCGGGGGCCCCACGAAGAGAGGATGACAAACCGTCCAGCTTATCTGCAACCTATTTTAAAG<br/>TTTGATTTTGTGGCTCTTCTGAGGCTGTAGAAATACTGCTTAACCATGCCCAGTTGGGCTTGTATGACTTTTCTCT<br/>GCAGCTAAATGGAGAGTTTAAACATGTGACAAATCTTCTCTGATATCGACACGAGTATATCCACAGGGGTATGGCATTTG<br/>CGTCCGTACCTGTTAATTCGGGTCTGAATGGGTATCAAGGAGTATCACCCTTATTCCTCCCTATTTTCGTTGGTGAAAG<br/>CTAAACATTTGCCACCCAGCACCACTGCTACTCAATTTCTCTTTATTTGCGAGGCTAA</p>                                                                                                                                                                                                      |
| 11 | Secuencia de ADN<br>de Cas-S5.1 | <p>ATGAGAACTTTAACTTCAACGGTAAATTTTCGACTCTGGAACCACTGACTGTAAACAGTAAATAATGCGGTCTCAACTTCC<br/>GGCATCGCTTGCCTCGTAACGGTGGATTCATATGACGCCCGTACTTTCAGGTACAGCAATTCGGGGAACCTCTTCGTCA<br/>TGCTGGCATAAAGGTCAATTTGTCGATCGTGTGGGCTTAAATGACAGCGGAAATCATCTTCGACCTGGGAGAACATTTCA<br/>TGCTGGCCAGGGGTGGATATCAATGGTGAAGCCGAAACCTTCGCGCAGGTGAATCAATGCTGGTGTGTAACCTACGT<br/>AGCAAAACCCGTTAATCAGCTTGTTTGGTCTTGGGATTAAGCGGCAAGTTGGTATAGGGAATGCTATCCCGGATGG<br/>TGATAACCAAGTGGGGATGTTTGGTGGTGGTGGGCTCAATTAATGTTTCAGCGTGACGAAAGTCTGATGGAGTTCTTG<br/>AGACAGACCAGTTGATCGTCTTGAACCGCTGCTCGAGGAACAGGCTGAGGCAAGTGGGATATCTCTCAGATCAAAACA<br/>GAGCAAGATGCACTCAAAAGGGGATGAAGTCGGCTGATTAAGACACCAAGCTGAGCTTCAATCAAGTACGGGAGCT<br/>CGATGAAATAATTCAGGCCCGCAAGATCAGAAGCAGGAATCTCGGAGTCAATTCGCCGCCGATTTGACCCGATATGAAG<br/>CGTTTATCACCGGGCGAAGTCAAGCATCGCATGAGTATTAATAATGCGACTGATGAGGAAGCAGGGCTTTTCATTTCT<br/>GCATTAATCCGCTTTGCAGCCGAAACCAAGTGTGGGGTCTATGCGAATCATAACTGCGGTCTGGTGGAGGCTCACTGGAC</p> |

|                                                                                      |  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------------------------------------------------------------------------------------|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                      |  | <p>AGTTACGACCTGGAAGCGGGTGAACCTGGTACCAGTTACACTTGGAGAAAATCGTCAACACCGAATGGTGTGAGATTACCGGGACGAGTTGTTTGTATGGTAAAGGCATTCAATGAAAATCAATCTTTTGTATTTCACTGGCCCCGTAA</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <p>12</p> <p>Secuencia de ADN<br/>que codifica Cas-S1.1,<br/>2.1, 3.1, 4.1 y 5.1</p> |  | <p>TTGAATCATCTGTAGAGTCAGTATATAGCGCTCTGACGAGTATTTTATTTGCCCTATATGGGTGAGCCTGTGCCTGTGCAA<br/>CGAAATTGCTCATGCTGGGACAGAGCGCCCTCAGAGTTCGAGGGGTAGGCTTCGAGCTGGTCAACGCCCTACAGGGAAC<br/>GTGTGTGCAATTGGCGTCTTGGCAGACATTTTGTGTTCTGCTCCAGAACTGATGGCGTGGAGAACCCCAAATAACCAA<br/>CAACGGCCAGAAAATTTGGCATGTGGTCTGGTGTGGGCTGTTATTAACGTTGAAGACAAATCTTCGGTCTTTTAGCC<br/>CCTCAGGGAGTGGTCAATAAATACCTGAGCATTTTTCGATCACTAGAGTCAACGCGCACAGCGGTACGCACTG<br/>GAATACCTGTTTAAACACCGAACTTAAATTCGCGCTGATTTATTTCAAAACTTCGGGTGAAACCTACGAGCTTGTTCGC<br/>TCACTGGGGTCACTGAGTGGCGACGCAATTTATACCTGTGCGATCAACTACTGACCCGGCAGAACGAAGTCTTTAC<br/>ATGCTGGAATTAACAAGGCTAAGAATCTCATCAGGAATAAATAACTACTCCAAAAAGAGATAGACATCTTTATCAGA<br/>ACAGTAACCTGCTGGCTATTCGCGGATAACTCTGAAGCTGCCTCGAATGAGTTTAAAGAAAACAACTCATTCCTGTTA<br/>CTGCTGCTTCTGCCGACTGATCCCATCAACGCCCTGAGCATTTTGCACTCTTGAAAGAGTATAACCATGATTTACTACG<br/>ATAATGCGTTTCGGTTCGATTCGCTACACAAAGTGTACCAGATCCACGACACCTGGATTTTTCATGCGAGGAACGTA<br/>AGGAGAGAAGCTGCCGTACTCTTTTAAATTTTTCGCGTACCGGGATGATTCGTTACTCTGCTGCCGACAGCCACA<br/>GCTCTGGAGTTGCCAGGGAGAAAAGAGAGCTTATTTCTCTGAGGGCCATGAATCAAGTTCAATACGTCAATGGC<br/>CATCTTTATAAAGGAACAAAGTCAGAAGGCCGTGGCCGACGCCAGTTTGCTCATCCGAAGAGGCTAGTTATCAGTTGG<br/>CCCTCACCAATTAGCGAAGGCTGTTTAAACCCGGTCAATCGTCTGAGTGGCCGAAATTCGTCCATATCGTCAAAAG<br/>GAAATGCTGGACCGGTTTACGCTACCGTATTCACGGTACAGGTTACCGCCATAATCAGCAATCAACAGGAAGCTGAA<br/>GTTGGAAATGTTTATGGCGTAGGCCCTAAACGGTATTTGGTTTGGGTTTCATGATCTGGCGGGCAGTAGTCATGAAT<br/>ACATCAGGGTATTTGAGTCTGCCGATTTGCTTACGTAAAGTCCACAGGATATATCGAGGCTGCAGCTGGTGGCT<br/>GACTCAGGATCAATAAGTTGCACTGATTAGTCTGAACAGGTTAGGAAACCGCTGGCTATCTGGTACCGGGCTTC<br/>TGATTTCTGCTGAACACCCAGAGGCAAAATTTGTGATAGCCACGAATTCATGCCCTGATGCACCAGATATTCAGAAAGCG<br/>ATCGTCTCTCTTGAACAGATAGCTGAACAATGCGGCAATAAAGTTACTTTTCCGGCTTATGGGAAAGCAAAATTACG<br/>TATCCCTGGAGAAAGTACGTGGACTGTTACTCATGGACGAATTTACCGATCTGGATACGGTTAAGGTCTTGAAAACTCG<br/>CTAACTGGTCAAAACCTCTGGTTGAGTTTGAAGAAGAGTATGGTGAATGCCAGGAGATCAACGCCAGAGATGGTCAACC<br/>TACTCCATCTGGGATGATATACAGGACATTTGATGATATTCGCTGAATGCTCTCAGTGTCACTTATAGTTACGACCCAT<br/>GCGATGTTATGGTCTGACTGTATGTGTAATCACCGCATCTCGGTGACAAAGAAATATGTACCTCATTTATGATGAGGCA<br/>GACATTTTGTAGATATGCTTGAGGTCTGGAAACAGCGGCGCTTCAACCTCAGAGAGTTGACCAGTGGTCAATGAACA<br/>TATCCACGTAATGGCGTGCATGTATCGACCACTTAATGAACGACGTGACTTCGATAGCGGGGATCTGCATTTTGTGA<br/>GTACGCTGCAGCGGTTGCCCTGTTGTGACAATAGCTTTAACGCCCTTATCAAGGTTAGGGGGGAAATAAAAAATGAAGCG<br/>GCCCGAAAGCTTCTTTGACTGTATCTACAGCTGGGAATGCTCGGATTTGTCGGGGGGCAAAAGGGTGTAGCGGTTTC</p> |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |  |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|
| CAATAAAGCTGCTGAAACAGCCCTGATCGCTGTAACCCGTTATACGGGATGAAATGTTGGGACAGATACGCACTCAATGGC<br>GTAGTGGTTACTGACGTGGCTACGCTCTCAATACAGTACCCAGAGACCGGAATGGAGTGGTTAATGTAAGCTCTG<br>GGTTAACCAGCGATACGATTTCAATCAGGAAATATTTCCCCAGACGCTATAGGCTCAATGAAGCTGACCAATCGCGGT<br>GCTGATTTCCAAAGGTGTTCAATGACCCGAAAGAACAGATATTTACGGTCAATGGTTAAAGGACAGTTGAACAATTG<br>TCCTGATTCAGGGGCGAGCTCTGGTATTAATCTCCAGCCATTAATGAACAGGATGATGCCAATCAGCTGGGTGAGGT<br>CTCGAACCACTTTATATTCAGAAAGCAGGTACGGCTCTCCGAAATTAATTAACAGTATCAGGATACCCAGGATTTCT<br>TATTTCTGCCGGTCTCTGTTGGTGTAGCCACAGGAGGAGATGGGAAACAGATTTCCAGGATTTAATCAATTACCC<br>GGATACCTTTCTTACCTCCGGACAGAAATGAAGGCTGAAGCCCTTAATGGGTATCTGAAGAACAGGCTATAGTCTGTAAC<br>TTTGAGGAGTTAACGTAACCTACCTGGATTAATTTGGGAAAGTTATTCGTAAAGAACAGCTATAGGCGGGG<br>GATTCGTAGTGAAATGATACCGTCAGGATTAATTAATCTCGATCCGGTTCCGGAACTACTGACCTTTGCTCTAAACA<br>TCGGTGGCTGAACACATTAATCCCGTTGCAATTCGTCTGTCATATCGTTCTGCTGATGAAATTTATCTCCGGCGTATTGTGA<br>AGAGGATATCCAGTGTAAATTTTAAACCGTATCGGGTCACTGCTCTGACACCGGTGGTCAATCAGTGGGATAGCC<br>CCATCGCTGAGCGGATTTCTGTATGAAGCACTGCTCAGGCGATACCCAGTAACGAGCCAGGATGCTATTTGGCTGCTCT<br>TAAGGAAATACCTTTATTAACGATGAACCTGGGTGTTTCCATGCTTCACTCTACGTTTGGGATAACGCCGGAACAGGG<br>GATTTGGGGGACAAACAGGTACGTTGGGATTAATCTCAGTCCGAAACAGTGAACCTGCAATGTTCTCTCTCTGCTATCCA<br>CAGGGGGGTGTTACGGGTGCTCTCTGACCGGGGCGCCACCAAGAGAACGATGACACCGCTCCAGCTTATTTCTGAC<br>CCTATTTAAGCTTTGATTTTGTGGCTCTCTGAGGCTGTAGAAATACCTGTTAACCATGCCCATGTTGGGCTGGTTATG<br>ACTTTTCTGACGCTAATGGAGAGTTTAACAATGTGACATTTCTCTCTCTGATATGACACAGATATATCAACGAGG<br>GTATGGCAATGGTCCGGTACCTGTTAATTCGGGTCTGAATGGTATCAAGGAGTATCACCCTTATTCCTCCCTATTTCG<br>TTGGTGAAGAAGCTAAACATTTGTCACCCAGCAACGATTCGTACTCAATTTGATTTCTTTATTTGGAGGCTAATTTTCATG<br>AGAACTTTAACTTCAACGGTAAATTTCCGACTCTGGAAACCTGACTGTAAACAGTAAACAAATGCGGCTCAACTTCGGC<br>CATCGCTTGCTCGTAACGGTGGATTTCAATGACCGCCGTTACTTTCCAGGTACCAAGCTTCCGGGAACTCTTCGTCTATGCT<br>GGCATAAGGTCATTTGTCGATCGTGTGGGCTTAATGACAGACGAAATCATCTTCGACCTGGCAGAACATTTTCATGCT<br>GGCCAGGGGGTGGATATCAATGGTGAAGCCGAAACCTTCGGCCAGGTGAATCAATGCTGGTGGTGAACCTACGTAGCA<br>AAACCCGTTAATCAGCTTGTGTTGGTCTGTTGGGATTAAGCGGCAAGTTGGTATAGGGAATGCTATCCCGGATGGTGAT<br>AACCAGTGGGGATGTTTGGTGGTGGTGGCTCAATTAATGTTACGCTGACGAAAGTCTGATGGAGTTCTTGAGAC<br>AGACCAGGTGATCGCTTTGAACGCTGCTCGAGGAAACAGGCTGAGGCAAGTGTGGATATCTCTCAGATCAAAACAGAGC<br>AAGATGCACTCAAAAGGGGATGAAGTGGCTGATAAAGACACCAAGCTGAGCTTCAATCAAAAGTACGGGAGCTCGAT<br>GAAAAATTCAGGCTCCAAAGATCAGAGGAGGATCTCGCGAGTCAATTCGCGCCGATGACCCGATGAAGCGGTT<br>TATCACCAGGCGAGAACTCAGCCATCGCATGAGTATTAATGCGACTGATGAGGAAGCAGGCTTTTCAATTTCTGCAAT<br>AATCCGCTTTGACGCGCAACACGTTTGGCGGTTCATGCGAATCATAACTGCGGTCTGGTGGAGCTCACTGGACAGTTA<br>CGACCTGGAAAGCCGGGTGAACCTGGTACGAGTTACACTTGGAGAAATCGTCAACACCGAATGGGTGTTGAGATTACCGGG |  |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|

|    |                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|----|------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                | GACGAGTTGTTTGCTATGGTAAAGGCAATCAATGAAAAATCAATCTTTGATTTCACTGCCCGCTAACTCCAAAAAGCTGGTG<br>GATTTTAGTGGGCTATTTAATATTTTATAATCAACCGGTATTTTAGAGTATTCCTCCCGTGTGCGGGGTTATCGGTG<br>GGCGTGTGTTGAATAAATAGATATATGCTGCGGGCTACCAAAACGTCGGTCTTTCCCAACGCAAAATCAGTAGGTAA<br>GGTTAAAGTCGGGAGCTACTACAAATAGTGACAGCATCTTTGACTAAAGCAGGCGGCCCAAAACAGCATGATCCCTATT<br>ATTCAACAAAGCCTATCGGTGGTCTCTCAACCGTCAACCGTCAACCGTCAACCGTCAACCGTCAACCGTCAACCGTCAACCGT<br>TCGTGTGTCACCGTTACCGCTGGTGGAGTATTCCTCCCGCATCGGGGTTATCGGTACCAATGCGGGAATAATC<br>TTCAATTTGTAATGATTTCCCTCCCGCATCGGGGTTATCGGAACATCAGTGGAAATCCACTGCGGCGTATTCCTCCCGCA<br>TCCGGGGTATTCGGCAAAACGGCAACCTTCATAAAACGCTCTTTGATTCCTCCCGCTGTCGGGGGTTATCGGCCGA<br>GATTGAGTAAAGCAAGTAACGGCGGTGGTATTCCTCCCGCATCGGGGTTATCGGCAGGTTATCTGGCAAAACGCTCG<br>ATG |
| 13 | PAM                                            | AAG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 14 | promotor pBoIA                                 | aacctaaatattgttgaagctgcaatgaaacgghaaaagcggcgtagattttaaaaggGatggatgacatdcaagcgttgcg                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| 15 | promotor P70a                                  | CCTGATGCGTGAACGTGACGGACGTAAACCACCGCGACATGTGTGCTGTTCGGCTGGGCATGCTGAGCTAACACCGTGC<br>GTGTTGACAATTTACCTCTGCGCGTGATTAATGGTTGCAGCTAGCAATAATTTTGTTTAACTTTAAGAAGGAGATATACC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 16 | conector XTEN<br>(nucleótidos)                 | agcggcagagagactccgggacccctcagagtcgcccacacccgaaagt                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 17 | conector XTEN<br>(aminoácidos)                 | SGSETPGTSEATPES                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 18 | citidina desaminasa<br>PmCDA1<br>(nucleótidos) | ATGACCCGACGCTGAGTACGTGAGAAATCCATGAGAAGTTGGACATCTACACGTTTAAGAAACAGTTTTTCAACAACAAAAA<br>TCCGTGTCGCATAGATGCTACGTTCTCTTTGAATTAACACGACGGGTGAACGTAGAGCGTGTTTTGGGGCTATGCTGT<br>GAATAAACCCACAGAGCGGGACAGAACGTGGCAATCACGCCGAAATCTTTAGCATTAAGAAAAGTCGAAGAAATACCTGCGCG<br>ACAAACCCCGGACAAATTCACGATAAATGGTACTCATCTCGAGTCTTTGTCAGATTGCGCTGAAAAGATCTTAGAATGGT<br>ATAACCCAGGAGCTGCGGGGGAACGGCCACACTTTGAAAATCTGGGCTTGCAAACTCTATTACGAGAAAAATGCGAGGAAT<br>CAAAATGGGCTGTGGAAATCTCAGAGATAACGGGGTTGGGTGAATGTAATGGTAAGTGAACACTACCAATGTTGCAGGAA<br>AATATTCATCCAATCGTCGCACAAATCAATTAATGAGAAATAGATGGCTTGAGAAGACTTTGAAGCGAGCTGAAAAACGACG<br>GACCGAGTTGTCATTATGATTCAGGTAAAAATACTCCACACCACCTAAGAGTCTGCTGTT                                                                                          |
| 19 | citidina desaminasa<br>PmCDA1<br>(aminoácidos) | MTDAEYVRIHEKLDIVTFKKQFFNNKKSVSHRVYVLFELKRRRGERACFWGVAVNKPQSGTERGIHAEIFSIRKVEEYLRDNP<br>QFTINWYSSWSPCADCAEKILEWYNQELRNGNHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNWMVSEHYQCCRKIFIQ<br>SSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

|    |                                                                      |                                                                                                                                                                                                                                                                                |
|----|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 20 | Proteína inhibidora de la uracil ADN glicosilasa (UGI) (nucleótidos) | ATGACCAACCTTTCCGACATCATAGAGAAGGAAACAGGCAACAGTTGGTTCATCCAAGAGTCGATCTCATGCTTCCTGAA<br>GAAGTTGAGGAGGTCAATTGGGAATAAGCCGGAAGTGACATTTCTCGTACACACTGCGTATGATGAGAGCACCGATGAGAA<br>CGTGATGCTGCTCACGTCAGATGCCCCAGAGTACAAACCCCTGGGCTCTGGTGATTCAGGACTCTTAATGGAGAGAACAAAGA<br>TCAAGATGCTATAA |
| 21 | Proteína inhibidora de la uracil ADN glicosilasa (UGI) (aminoácidos) | MTNLSDIIEKETGKQLVIQESILMPVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENIKIKML                                                                                                                                                                                             |
| 22 | Conector (nucleótidos)                                               | TCGTGGTGGATCTGGAGGTTCTGGTGGATCT                                                                                                                                                                                                                                                |
| 23 | Conector (aminoácidos)                                               | SGGSGGSGGS                                                                                                                                                                                                                                                                     |
| 24 | Promotor pBAD (nucleótidos)                                          | gacccaaggtcattgacaaaaaacgcgtacacaaaagtgctataataacagggcaggaagtcacattgattatgacagg-gtgcacacttgctatgcccatagcaatt<br>tttatcatagagatagcgggattccttacccttgacgctttttatagcaactctctactgtttctccataccccgttttttagggaaattc-gagctctaaggggttataaaaa                                             |
| 25 | Secuencia de protoespaciador no diana                                | GGAGACGTCATGTCAAGCTCATAGGCGTCTCG                                                                                                                                                                                                                                               |
| 26 | Secuencia del protoespaciador U1kF                                   | ACATACGCTTCTTGTGTGAGTACGTGAGTTTCAC                                                                                                                                                                                                                                             |
| 27 | Secuencia del protoespaciador U2kF                                   | CAGATTGCTGGTCAGGAGCATATTGATCCGCT                                                                                                                                                                                                                                               |
| 28 | Secuencia del protoespaciador S1F                                    | AAGGAGATATACCATGGAGCTTTTTCACCTGGCG                                                                                                                                                                                                                                             |
| 29 | Secuencia del protoespaciador S2F                                    | TTCAGCGTGTCCGGCGAGGGCGAGGCGCATGC                                                                                                                                                                                                                                               |
| 30 | Secuencia del protoespaciador S3F                                    | CTGCCCGTGGCCTTGGCCACCCCTCGTGACCAC                                                                                                                                                                                                                                              |
| 31 | Secuencia del protoespaciador S4F                                    | GAGAAAGCCGATCATGGTCTGCTGCTGGAGTT                                                                                                                                                                                                                                               |
| 32 | Secuencia del protoespaciador S5F                                    | TACCAGATGGCATTTGGCGCATCTGGCAGAGTG                                                                                                                                                                                                                                              |

|    |                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|----|----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 33 | Secuencia del<br>protoespaciador S6F         | AAAGTCAGGTTGCGCAGGCAGGGTAATCGGGG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 34 | Secuencia del<br>protoespaciador U1kR        | TCACGTACTCAGCAAGAGCGTATGTGTTTGC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 35 | Secuencia del<br>protoespaciador U2kR        | ACCTCTGTCAGCTCGCTGACATTCGCCGCCAG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 36 | Secuencia del<br>protoespaciador S1R         | TTAAACAAAATTATTGCTAGCTGCAACCAITTA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 37 | Secuencia del<br>protoespaciador S2R         | CTCCATGGTATATCTCCTTCTTAAAGTTAAAC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 38 | Secuencia del<br>protoespaciador S3R         | CACTGCACGCCGTAGGTCAGGGTGGTCACGAG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 39 | Secuencia del<br>protoespaciador S4R         | TCCAGCAGGACCATGTGATCGGCGTCTCTCGTT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 40 | Secuencia del<br>protoespaciador S5R         | CACTCTGCCAGATGGCGCAATGCCATCTGGTA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 41 | Secuencia del<br>protoespaciador S6R         | CTGACTTTGTTGTGCGGCCCGCGGCGCACTCAA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 42 | Secuencia del sitio de<br>escisión de I-TexI | CAACG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 43 | Secuencia de espaciador<br>de ADN de I-TexI  | CTCAGTAGATGTTTTCTTTGGGTCTACCGTTA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| 44 | Secuencia de I-TexI<br>(nucleótidos)         | ATGAAAAGCGGAATTTATCAGATTAAAAATACITTTAAACAATAAAGTATATGTAGGAAGTGCTAAAGA TTTTGAAAAGAGA<br>TGGAAAGAGGCATTTTAAAGATTTAAGAAAAGGATGCCATTTCTTCTATAAACTTCAGAGGTCITTTTAACAACATGGTAAT<br>GTGTTTGAATGTTCTATTTTGGAAAGAAATCCATATGAGAAAGATTGATTTATTGAACGAGAAAAATTTTGGATTAAAGAG<br>CTTAATCTAAAATTAAATGGATACAATATTGCTGATGCAACGTTTGGTGATACATGTTCTACGCATCCATTAAAAAGAAGAA<br>ATTATTAAAGAAACGTTCTGAAACTGTTAAAGCTAAGATGCTTAACTTGGACCTGATGGTCGGAAGCTCTTTACAGTAAA<br>CCCGGAAGTAAAAACGGCGGTGGGATCCAGAAACCCATAAGTTTTGTAAAGTCGGGTGTTGCA TACAAACTTCTGCTTAT |

|    |                                                                                                   |                                                                                                                                                                                                                        |
|----|---------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                                                   | ACTTGTAGTAAATGCAGAAATCGTTTCAGGTGAAAAATAATTCATTCTTTAATCATAAGCATTTCAGACATAACTAAATCTAAAA<br>TATCAGAAAAAGATGAAAGGTAAAAAGCCTAGTAATATTTAAAAAGATTTC                                                                           |
| 45 | Secuencia de I-TevI<br>(aminoácidos)                                                              | MKSGTYQIKNTLNNKYVVGSAKDFEKRWKRHFKDLEKGCHSSIKLQPSFNKHGNVFECSSILEEIPYEKDLIIERENFWIKELNSK<br>INGVNIADATFGDTCSTHPLKEEIIKKRSETVKAKMLKLGPDGRKALYSKPGSKNGRWNPETHKFKCGVRIQTSAITCSKCR<br>NRSGENNSFFNHHKSDITKSKISEKMKGKKPSNIKKIS |
| 46 | Espaciador de ADN de<br>I-TvI de 31 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2497)  | CTCAGTAGATGTTTTCTTGGGTCTACCGTTA                                                                                                                                                                                        |
| 47 | Espaciador de ADN de<br>I-TvI de 29 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2498)  | CTCAGTAGATGTTTTCTTGGGTCTACCGT                                                                                                                                                                                          |
| 48 | Espaciador de ADN de<br>I-TvI de 27 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2499)  | CTCAGTAGATGTTTTCTTGGGTCTACC                                                                                                                                                                                            |
| 49 | Espaciador de ADN de<br>I-TvI de 25 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2500)  | CTCAGTAGATGTTTTCTTGGGTCTA                                                                                                                                                                                              |
| 50 | Espaciador de ADN de<br>I-TvI de 23 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2501)  | CTCAGTAGATGTTTTCTTGGGTC                                                                                                                                                                                                |
| 51 | Espaciador de ADN de<br>I-TvI de 21 nt de longitud<br>en el plásmido diana de<br>TevCasS (p24502) | CTCAGTAGATGTTTTCTTGGG                                                                                                                                                                                                  |

|    |                                                                                         |                     |
|----|-----------------------------------------------------------------------------------------|---------------------|
| 52 | Espaciador de ADN de<br>19 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2503) | CTCAGTAGATGTTTTCTTG |
| 53 | Espaciador de ADN de<br>17 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2504) | CTCAGTAGATGTTTTCT   |
| 54 | Espaciador de ADN de<br>15 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2505) | CTCAGTAGATGTTTT     |
| 55 | Espaciador de ADN de<br>13 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2506) | CTCAGTAGATGTT       |
| 56 | Espaciador de ADN de<br>11 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2507) | CTCAGTAGATG         |
| 57 | Espaciador de ADN de<br>9 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2508)  | CTCAGTAGA           |
| 58 | Espaciador de ADN de<br>7 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2509)  | CTCAGTA             |
| 59 | Espaciador de ADN de<br>5 nt de longitud                                                | CTCAG               |

|    |                                                                                                   |                                        |
|----|---------------------------------------------------------------------------------------------------|----------------------------------------|
|    | en el plásmido diana de<br>TevCasS (p2510)                                                        |                                        |
| 60 | Espaciador de ADN de<br>1-Tvel de 33 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2575) | CTCAGTAGATGTTTCTTGGGTCTACCGTTACG       |
| 61 | Espaciador de ADN de<br>1-Tvel de 35 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2576) | CTCAGTAGATGTTTCTTGGGTCTACCGTTACGGA     |
| 62 | Espaciador de ADN de<br>1-Tvel de 37 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2577) | CTCAGTAGATGTTTCTTGGGTCTACCGTTACGGATT   |
| 63 | Espaciador de ADN de<br>1-Tvel de 39 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2578) | CTCAGTAGATGTTTCTTGGGTCTACCGTTACGGATTAG |
| 64 | Espaciador de ADN de<br>1-Tvel de 28 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2655) | CTCAGTAGATGTTTCTTGGGTCTACCG            |
| 65 | Espaciador de ADN de<br>1-Tvel de 30 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2656) | CTCAGTAGATGTTTCTTGGGTCTACCGTT          |
| 66 | Espaciador de ADN de<br>1-Tvel de 32 nt de longitud<br>en el plásmido diana de<br>TevCasS (p2657) | CTCAGTAGATGTTTCTTGGGTCTACCGTTAC        |

|    |                                                                                                                              |                                       |
|----|------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|
| 67 | <p>Espaciador de ADN de 1-Tel de 34nt de longitud en el plásmido diana de TevCasS (p2658)</p>                                | CTCAGTAGATGTTTTCTTTGGGTCTACCGTTACGG   |
| 68 | <p>Espaciador de ADN de 1-Tel de 36nt de longitud en el plásmido diana de TevCasS (p2659)</p>                                | CTCAGTAGATGTTTTCTTTGGGTCTACCGTTACGGAT |
| 69 | <p>1º espaciador en la matriz CRISPR<br/>Tipo-S de tipo salvaje (32nt), usado también como proespaciador en el Ejemplo 6</p> | TCGTGTTGTCCACGGTTACCCGCTGGCTGGAA      |
| 70 | <p>Secuencia repetida</p>                                                                                                    | GTATTCGCCCGCATGCGGGGGTTATCGG          |

## REIVINDICACIONES

1. Un vector de ácido nucleico o una pluralidad de vectores de ácido nucleico que comprenden secuencias de nucleótidos expresables, en donde las secuencias comprenden

- 5 a) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:1;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:2;
- c) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:3;
- 10 d) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:4; y
- e) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:5;

15 en donde el o los vectores comprenden una o más secuencias de nucleótidos para producir un ARNcr, en donde el ARNcr comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana.

2. El o los vectores de la reivindicación 1, en donde:

(a) al menos una de las secuencias de nucleótidos está conectada operativamente a un promotor que

- i. es heterólogo a la o las al menos una secuencia de nucleótidos;
- ii. no es un promotor de *E. coli*, *Pseudomonas* o *Klebsiella*;
- 20 iii. es un promotor de célula eucariota;
- iv. es un promotor animal (opcionalmente, un promotor de mamífero o humano);
- v. es un promotor vegetal;
- vi. es un promotor fúngico (opcionalmente, un promotor de levadura);
- vii. es un promotor de insecto;
- 25 viii. es un promotor de virus (opcionalmente, un promotor de virus, AAV o lentivirus); o
- ix. es un promotor sintético; y/o

(b) al menos una de las secuencias de nucleótidos está conectada operativamente a (i) un promotor constitutivo; o (ii) un promotor inducible.

3. Un vector de ácido nucleico que comprende una secuencia de nucleótidos expresable que comprende

- 30 a) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:2;
- b) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:4; y
- 35 c) una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:5

en donde el vector comprende además una o más secuencias de nucleótidos para producir un ARNcr, en donde el ARNcr comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, y opcionalmente en donde:

- 40 (i) el vector comprende además una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:1; y/o
- (ii) el vector comprende además una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:3.

4. El o los vectores de cualquier reivindicación anterior, en donde

(a) el protoespaciador

- (i) no se encuentra en *E coli*, *Pseudomonas* y/o *Klebsiella*;
- (ii) no se encuentra en una bacteria que comprende secuencias de nucleótidos endógenas que codifican los polipéptidos de a) a e) como se cita en la reivindicación 1;
- (iii) es un protoespaciador de célula eucariota; o
- (iv) es un protoespaciador de célula animal (opcionalmente, humana), vegetal, de insecto o fúngica;

(b) el o los vectores comprenden

- (i) una secuencia de localización nuclear (NLS);
- (ii) una secuencia de empaquetamiento de fagos (opcionalmente, un sitio *pac* o *cos*);
- (iii) un origen de replicación plasmídico;
- (iv) un origen de transferencia plasmídico;
- (v) una estructura de plásmido bacteriano;
- (vi) una estructura de plásmido eucariota;
- (vii) un gen de proteína estructural de un virus humano (opcionalmente, AAV o lentivirus);
- (viii) una secuencia *rep* y/o *cap* de un virus (opcionalmente, AAV o lentivirus);
- (ix) un marcador de selección o secuencia que codifica un marcador seleccionable;
- (x) un promotor eucariota; o
- (xi) una secuencia de nucleótidos de un gen humano, animal, vegetal o fúngico; y/o

(c) cada vector es

- i) un vector plasmídico (opcionalmente, un plásmido conjugativo)
- ii) un vector de transposón (opcionalmente, un transposón conjugativo);
- iii) un vector de virus (opcionalmente, un vector de fago, AAV o lentivirus);
- iv) un fagémido (opcionalmente, un fagémido empaquetado); o
- v) una nanopartícula (opcionalmente, una nanopartícula lipídica).

5. Un complejo aislado de proteína y ARNcr que comprende:

- a) un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:2; y
- b) un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:4; y
- c) un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:5; y
- d) opcionalmente, un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:1; y
- e) opcionalmente, un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la SEQ ID NO:3;

y un ARNcr que comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana.

6. El o los vectores de una cualquiera de las reivindicaciones 1 a 4, o el complejo de la reivindicación 5, en donde:

- (a) el ARNcr comprende dos secuencias repetidas, y/o en donde el ARNcr comprende una secuencia repetida que comprende la secuencia de nucleótidos de la SEQ ID No:6 o SEQ ID No:70; y/o

- (b) el protoespaciador está inmediatamente adyacente a una secuencia de resto adyacente al protoespaciador (PAM) en la secuencia diana seleccionada de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3', 5'-ACA-3', 5'-ACT-3', 5'-ATC-3', 5'-ATA-3', 5'-GAG-3', 5'-TAG-3', 5'-ACC-3', 5'-AGG-3', 5'-ATT-3', 5'-GAC-3' y 5'-GTG-3', por ejemplo, seleccionada de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3', 5'-ACA-3', 5'-ACT-3', 5'-ATC-3', 5'-ATA-3', 5'-GAG-3' y 5'-TAG-3', tal seleccionada de 5'-AAC-3', 5'-ATG-3', 5'-AAA-3', 5'-AAG-3', 5'-ACG-3', 5'-AAT-3' y 5'-ACA-3', por ejemplo, 5'-AAC-3' y 5'-ATG-3', y en particular el PAM es 5'-AAC-3'; y/o
- (c) la secuencia espaciadora tiene una longitud de 25 a 39 nucleótidos (por ejemplo, una longitud de de 28 a 32 nucleótidos), o tiene una longitud de aproximadamente 32 nucleótidos (por ejemplo, tiene una longitud de 32 nucleótidos); opcionalmente, en donde los pares de bases 1 a 28 del espaciador son idénticos al complemento de los nucleótidos 1 a 28 de la secuencia protoespaciadora que está inmediatamente en 5' de la secuencia PAM en la secuencia diana; y/o
- (d) el espaciador es idéntico al complemento del protoespaciador en la secuencia diana.
7. Una célula, en donde la célula es
- A. una célula eucariota que comprende el o los vectores o complejo de cualquier reivindicación anterior;
  - B. una célula animal, vegetal, de insecto o fúngica que comprende el o los vectores o complejo de cualquier reivindicación anterior; o
  - C. una célula procariota que comprende el vector o los vectores o complejo de cualquier reivindicación anterior, en donde la célula no es una célula de *Klebsiella* (opcionalmente, en donde la célula no es una célula bacteriana, por ejemplo, una célula de *E coli* o *Pseudomonas*) que comprende secuencias de nucleótidos endógenas que codifican los polipéptidos de a) a e) como se cita en la reivindicación 1.
8. Una composición (opcionalmente, una composición *in vitro* o en donde la composición está contenida en un recipiente médico) que comprende el o los vectores citados en una cualquiera de las reivindicaciones 1 a 4 o 6, o el complejo citado en la reivindicación 5 o 6, o la célula citada en la reivindicación 7.
9. Vector(es), complejo, célula(s) o composición de una cualquiera de las reivindicaciones 1 a 8 para su uso en un método para tratar o prevenir una enfermedad o afección en un ser humano, animal o planta, en donde el método comprende
- I) poner en contacto la célula con el o los vectores, complejo, células o composición de una cualquiera de las reivindicaciones 1-8;
  - II) permitir la introducción en la célula del o de los vectores o complejo, mediante lo cual los polipéptidos y el ARNcr se expresan en la célula;
- en donde el ARNcr forma un complejo con el o los polipéptidos para guiar el complejo a la secuencia diana; en donde el complejo modifica las células diana, en donde dicho contacto comprende administrar el o los vectores, complejo, células o composición al sujeto y en donde la modificación trata o previene la enfermedad o afección;
- opcionalmente, en donde:
- el sujeto es un ser humano o un animal, y/o
- la modificación: (i) añade una nueva funcionalidad a las células, opcionalmente la modificación regula al alza o regula a la baja (en particular, regula a la baja, por ejemplo, bloquea) la expresión de un gen en las células; o (ii) modula la expresión de una secuencia de nucleótidos contenida en un cromosoma o episoma (opcionalmente, un plásmido) de la célula.
10. Un método *in vitro* para tomar como diana un polinucleótido (opcionalmente, en una célula), comprendiendo el método
- a) poner en contacto el polinucleótido con el complejo de la reivindicación 5 o la reivindicación 6;
  - b) en donde el complejo es guiado a una secuencia diana contenida en el polinucleótido para modificar el polinucleótido o la replicación del mismo; opcionalmente
- en donde el polinucleótido está contenido en un cromosoma o un episoma (opcionalmente, un plásmido); además, opcionalmente, en donde el método inhibe la replicación del polinucleótido.
11. Un kit que comprende
- a) los polipéptidos como se citan en la reivindicación 5 o la reivindicación 6; o el o los vectores como se citan en una cualquiera de las reivindicaciones 1 a 4; y

- b) un crRNA, o uno o más ácidos nucleicos que codifican un ARNcr, en donde el ARNcr comprende un espaciador que es cognado a un primer protoespaciador en una secuencia diana, y en donde el ARNcr es cognado a un PAM que tiene una de las secuencias citadas en la reivindicación 6(b);

en donde dichos polipéptidos son operativos para su uso con el ARNcr para formar un complejo de ribonucleoproteína para tomar como diana el protoespaciador en un polinucleótido.

- 5 12. Un método *in vitro* de control transcripcional o replicación de un ADN diana, que comprende poner en contacto el ADN diana con el complejo de la reivindicación 5 o la reivindicación 6, en donde el complejo está desprovisto de una nucleasa de ADN y el complejo se une al ADN diana, controlando de este modo la transcripción o replicación del ADN diana; opcionalmente, en donde el ADN está contenido en un cromosoma o un episoma (opcionalmente, un plásmido).
- 10 13. Un método *in vitro* para controlar la transcripción de un ARN diana, que comprende poner en contacto el ARN diana o un ADN que codifica el ARN con el complejo de la reivindicación 5 o la reivindicación 6, en donde el complejo se une al ARN o ADN diana, controlando de este modo la transcripción del ARN diana; opcionalmente, en donde el ADN está contenido en un cromosoma o un episoma (opcionalmente, un plásmido).
- 15 14. Un método *in vitro* para modificar la transcripción de una región de un ADN, que comprende poner en contacto dicho ADN con el complejo de la reivindicación 5 o la reivindicación 6, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene una de las secuencias citadas en la reivindicación 6(b), por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el complejo regula al alza o a la baja la transcripción de la región de ADN o de un gen adyacente, en particular regula a la baja la transcripción.
- 20 15. Un método *in vitro* para modificar un ADNds diana de una célula sin introducir roturas en el ADNds, comprendiendo el procedimiento producir en la célula el complejo de la reivindicación 5 o la reivindicación 6, en donde el complejo se dirige a un ADNds diana contenido en la célula, comprendiendo el ADNds diana una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene una de las secuencias citadas en la reivindicación 6(b), por lo que el complejo se une al ADNds en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el ADNds se modifica sin introducir roturas en el ADNds.
- 25 16. Un método *in vitro* para inhibir el crecimiento o la proliferación de una célula sin introducir roturas letales en el ADNds, comprendiendo el procedimiento producir en la célula el complejo de la reivindicación 5 o la reivindicación 6, en donde el complejo se dirige a un ADN diana contenido en la célula, comprendiendo el ADN diana una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene una de las secuencias citadas en la reivindicación 6(b), por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que se inhibe la replicación del ADN sin introducir roturas letales en el ADNds en el ADN, inhibiendo de este modo el crecimiento o la proliferación de una célula.
- 30 17. Un complejo de la reivindicación 5 o la reivindicación 6 para su uso en un método para tratar o prevenir una enfermedad o afección en un sujeto humano, animal, vegetal o fúngico, comprendiendo el método:
- 35 (a) un método de control transcripcional o replicación de un ADN diana, que comprende poner en contacto el ADN diana con el complejo, en donde el complejo está desprovisto de una nucleasa de ADN y el complejo se une al ADN diana, controlando de este modo la transcripción o replicación del ADN diana; opcionalmente, en donde el ADN está contenido en un cromosoma o un episoma (opcionalmente, un plásmido);
- 40 (b) un método para controlar la transcripción de un ARN diana, que comprende poner en contacto el ARN diana o un ADN que codifica el ARN con el complejo, en donde el complejo se une al ARN o ADN diana, controlando de este modo la transcripción del ARN diana; opcionalmente, en donde el ADN está contenido en un cromosoma o un episoma (opcionalmente, un plásmido);
- 45 (c) un método para modificar la transcripción de una región de un ADN, que comprende poner en contacto dicho ADN con el complejo, en donde el ADN comprende una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene una de las secuencias citadas en la reivindicación 6(b), por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el complejo regula al alza o a la baja la transcripción de la región de ADN o de un gen adyacente, en particular regula a la baja la transcripción;
- 50 (d) un método para modificar un ADNds diana de una célula sin introducir roturas en el ADNds, comprendiendo el método producir en la célula el complejo, en donde el complejo se dirige a un ADNds diana contenido en la célula, comprendiendo el ADNds diana una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene una de las secuencias citadas en la reivindicación 6(b), por lo que el complejo se une al ADNds en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que el ADNds se modifica sin introducir roturas en el ADNds; o
- 55 (e) un método para inhibir el crecimiento o la proliferación de una célula sin introducir roturas letales en el ADNds, comprendiendo el método producir en la célula el complejo, en donde el complejo se dirige a un ADN diana contenido

5 en la célula, comprendiendo el ADN diana una secuencia protoespaciadora flanqueada en su extremo 5' por un PAM que tiene una de las secuencias citadas en la reivindicación 6(b), por lo que el complejo se une al ADN en la región definida por la unión complementaria de una secuencia espaciadora del ARNcr al protoespaciador, por lo que se inhibe la replicación del ADN sin introducir roturas letales en el ADN, inhibiendo de este modo el crecimiento o la proliferación de una célula;

en donde las células del sujeto comprenden dicho ADN o ARN y las células median la enfermedad o afección.

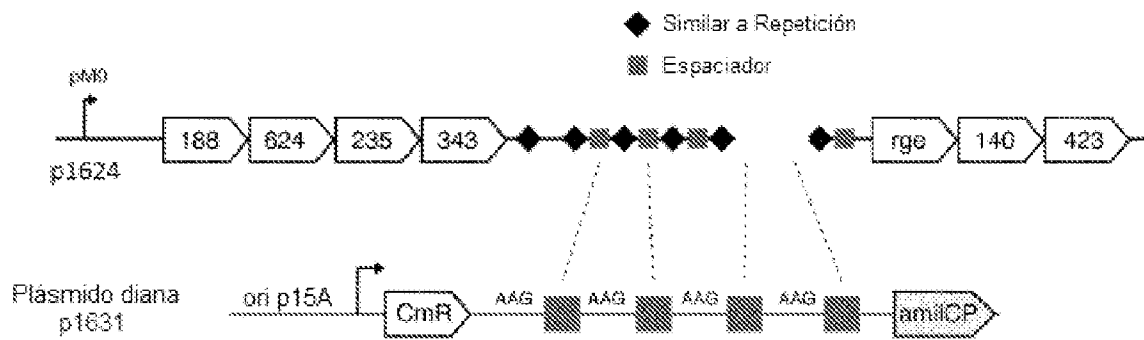


Figura 1

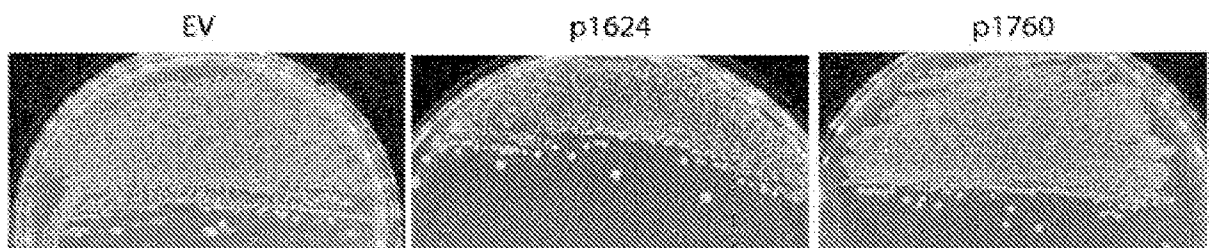
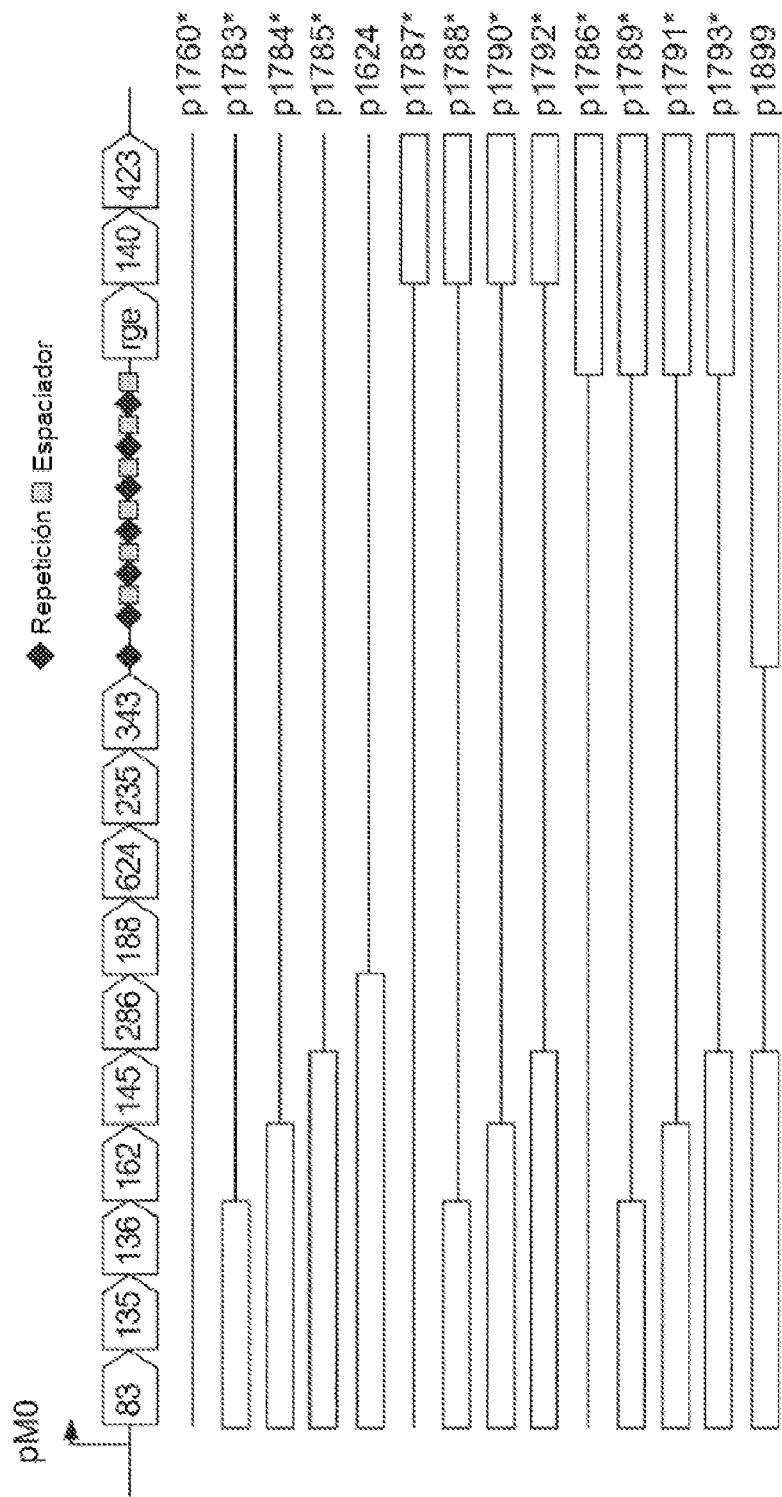
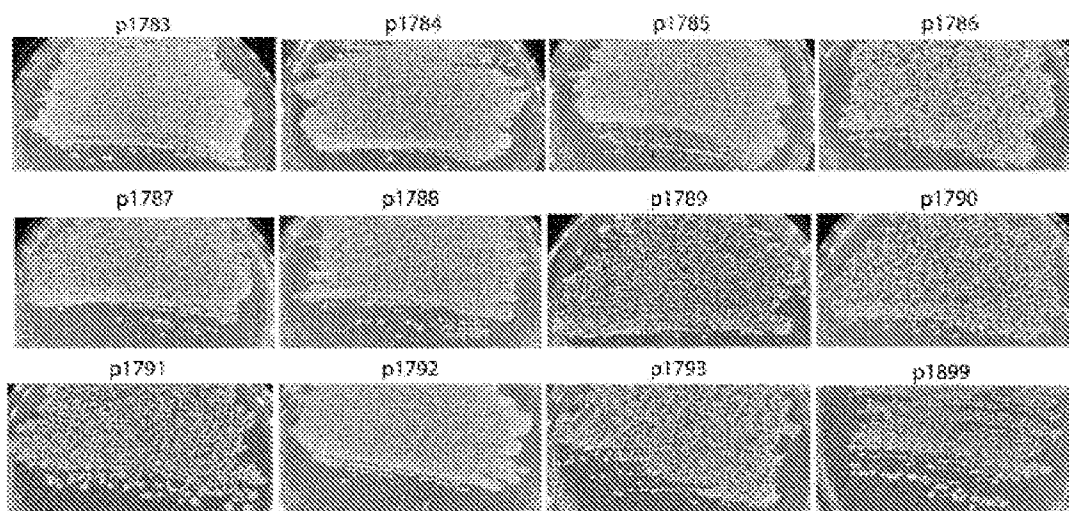


Figura 2

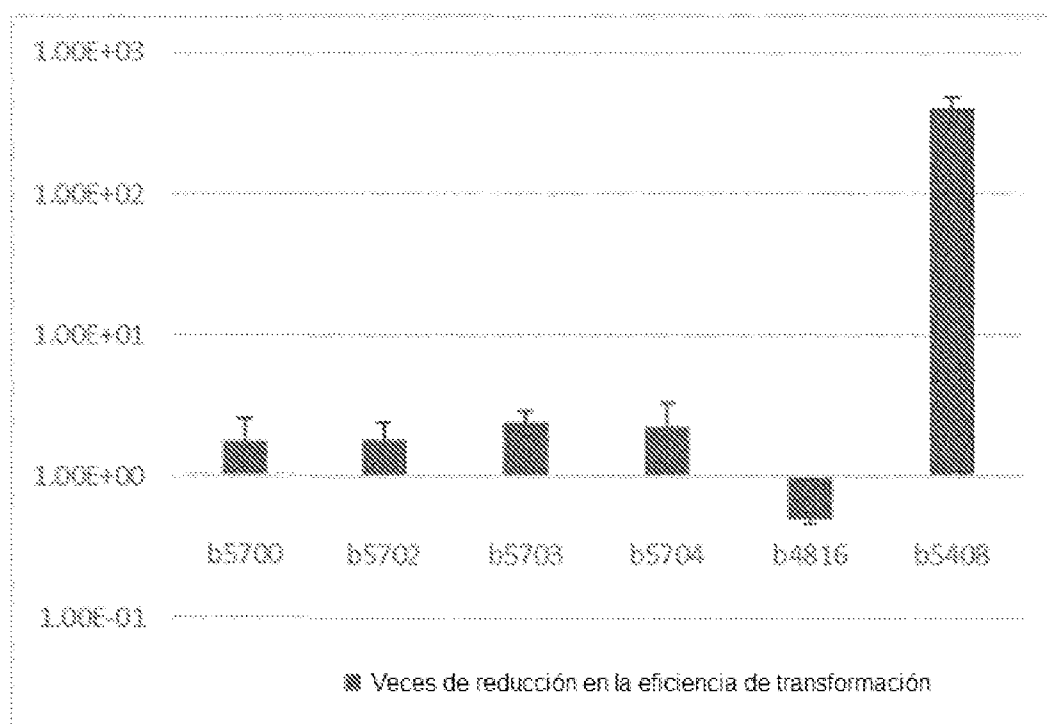


\*Construcciones que muestran actividad completa

Figura 3



**Figura 4**



**Figura 5**

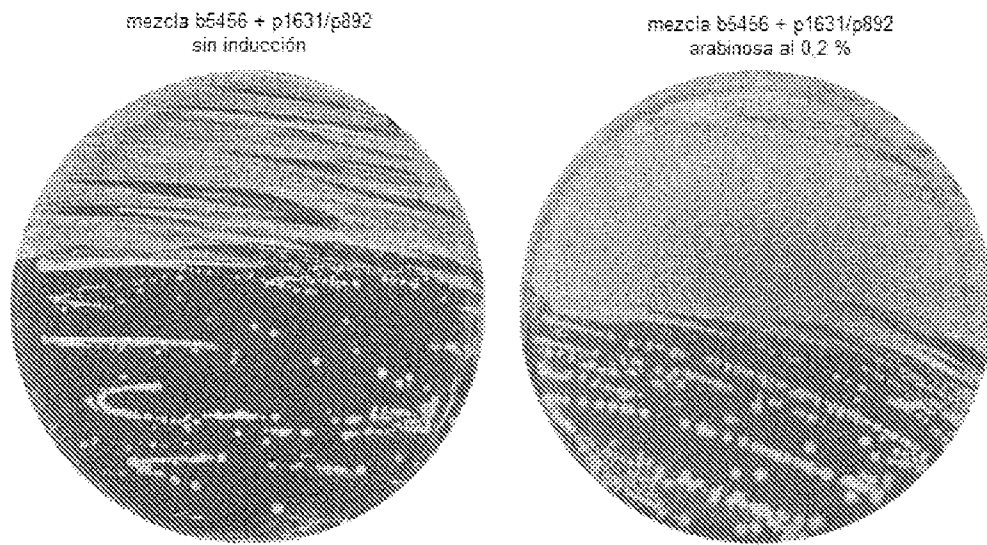
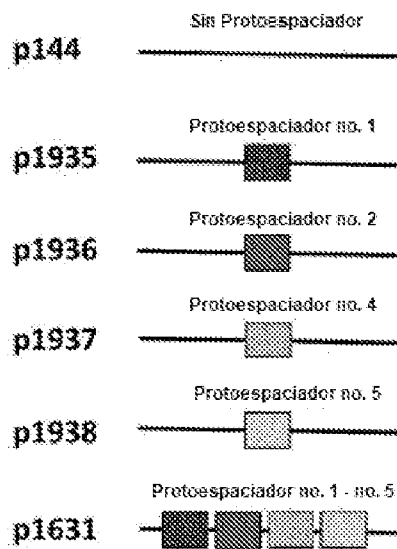


Figura 6

Plásmidos diana:



Dilución ( $10^4$ )

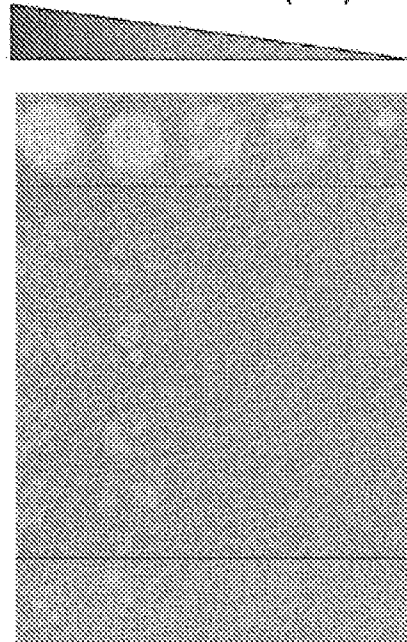
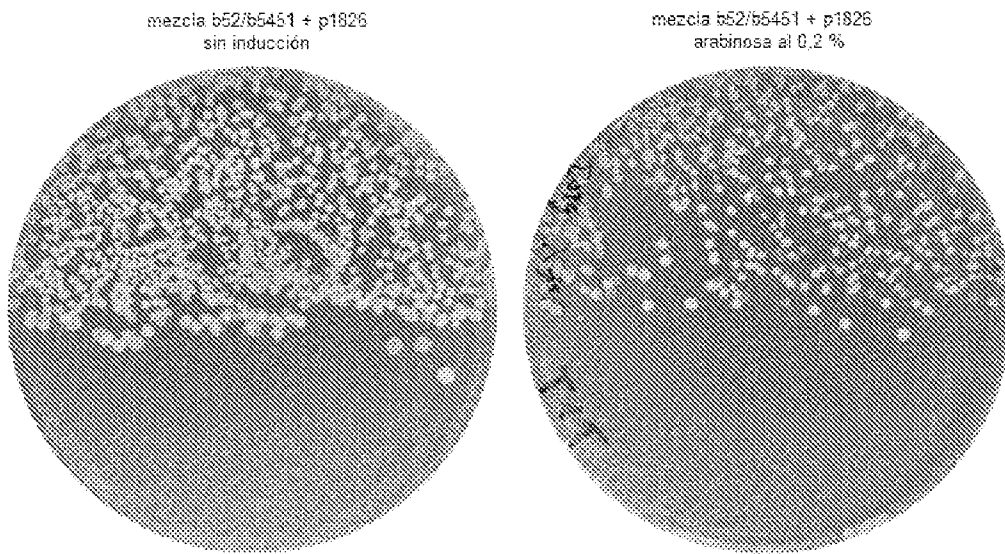
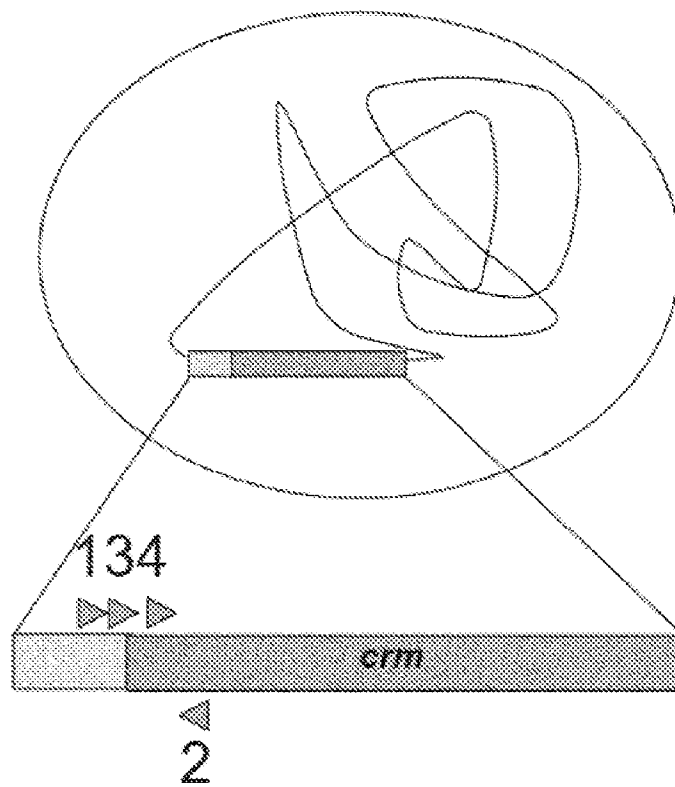


Figura 7



**Figura 8**  
**bSNP5810**



**Figura 9**

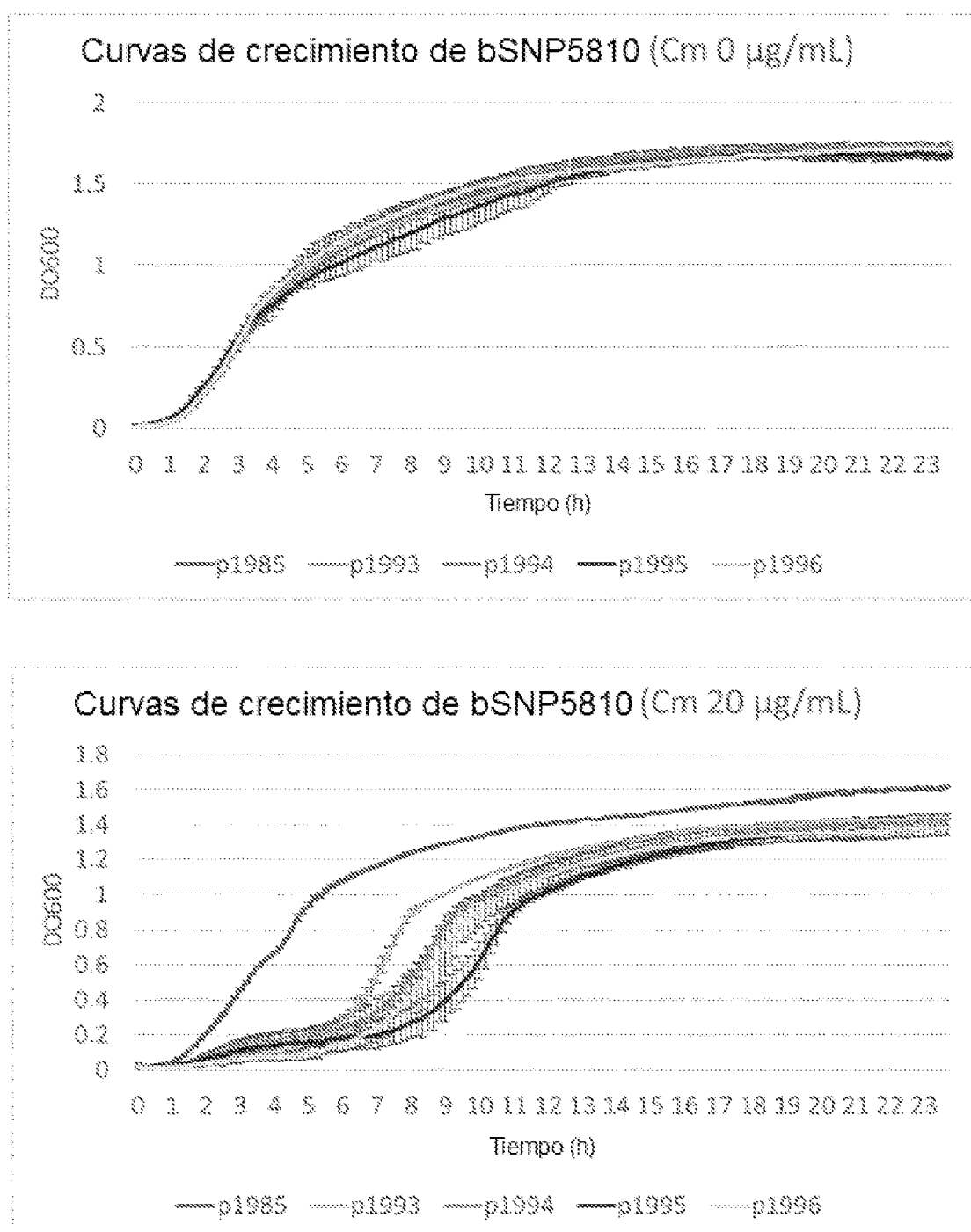


Figura 10

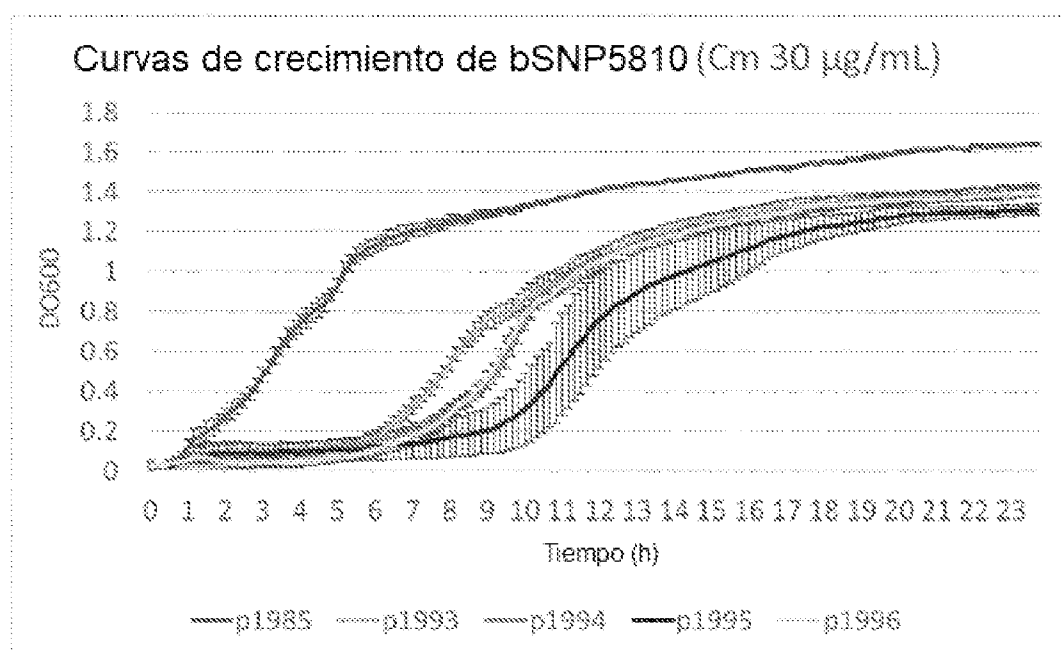
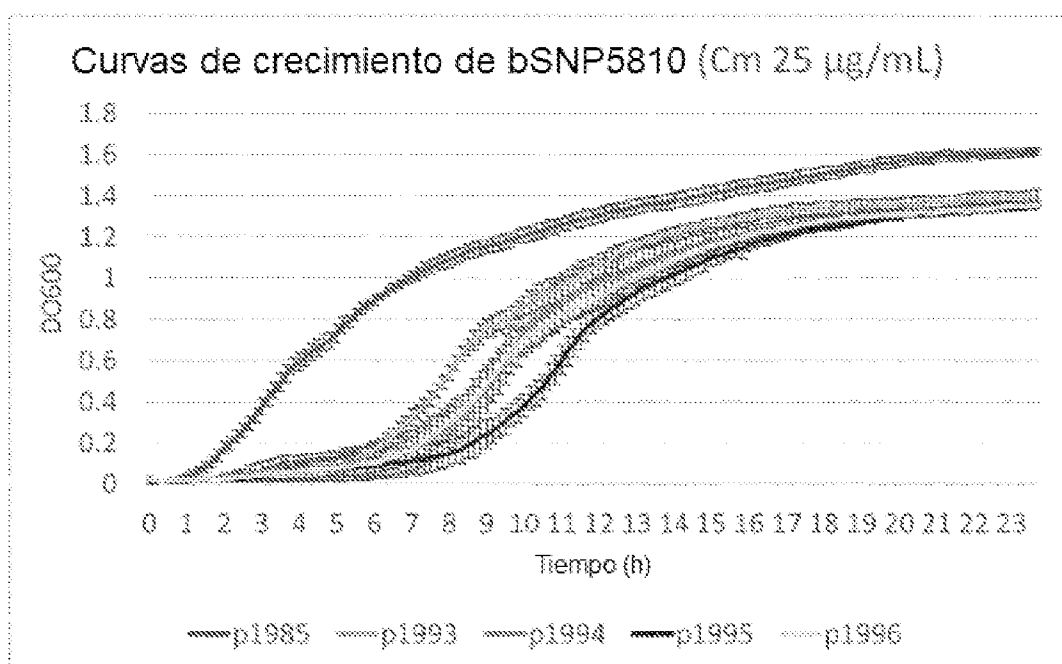
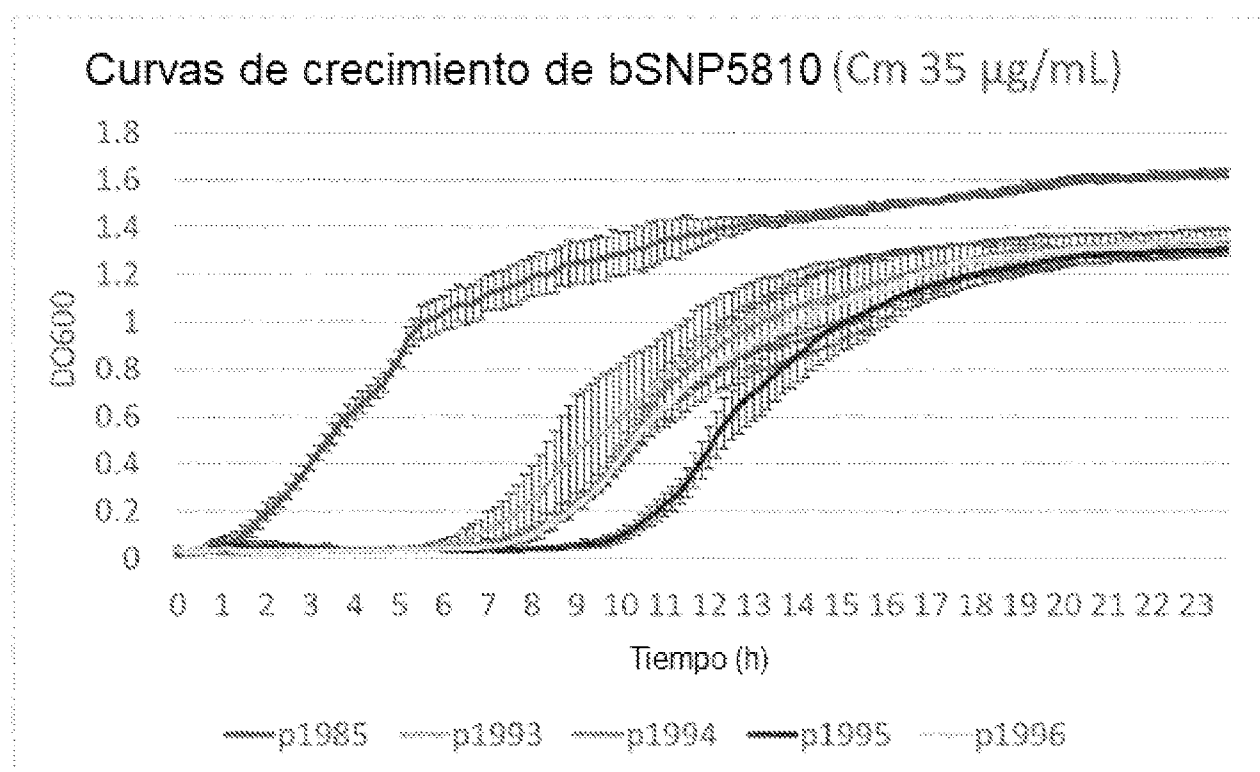


Figura 10 continuación

**Figura 10 continuación**

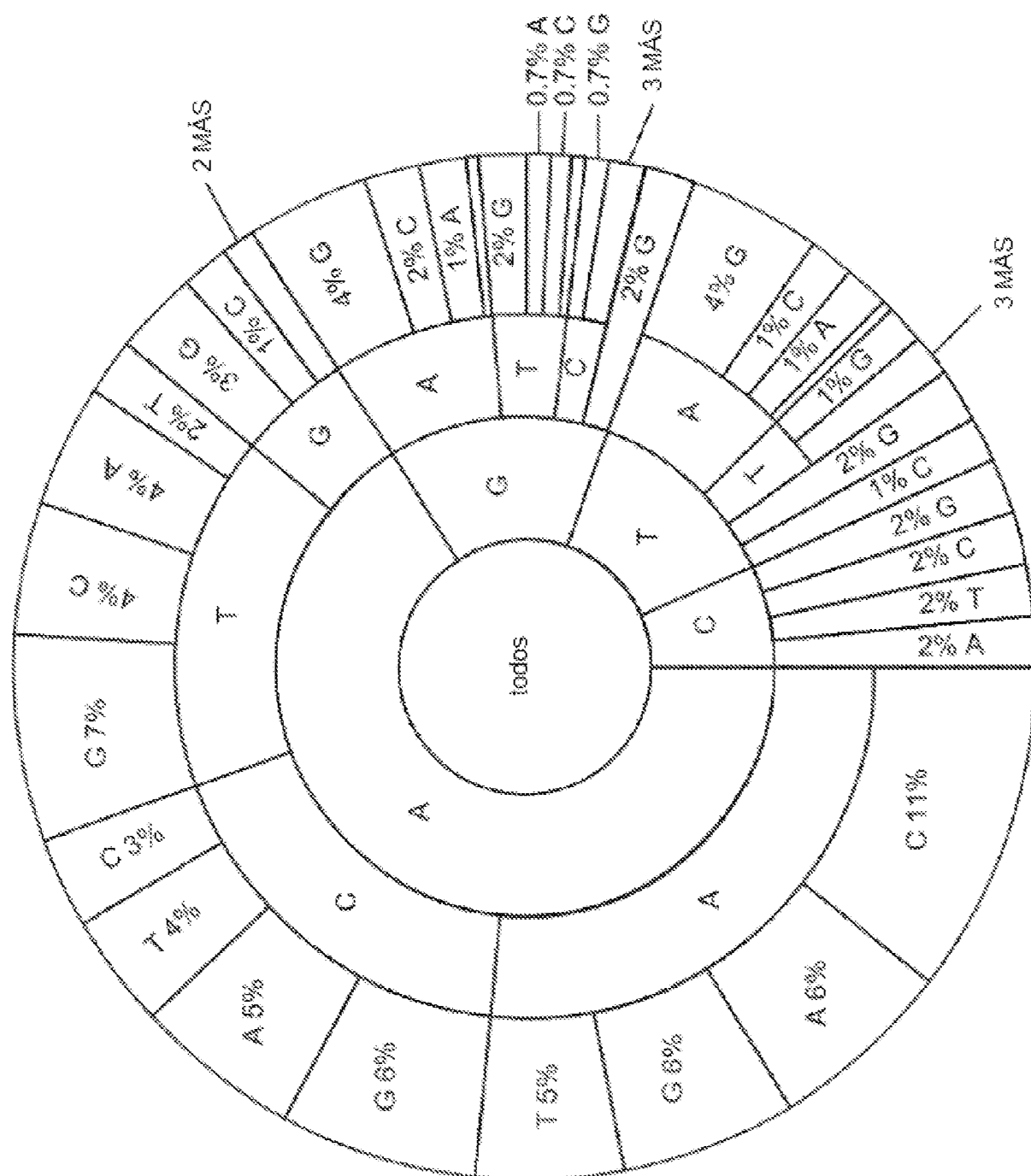


Figura 11

Calificaciones de los PAM

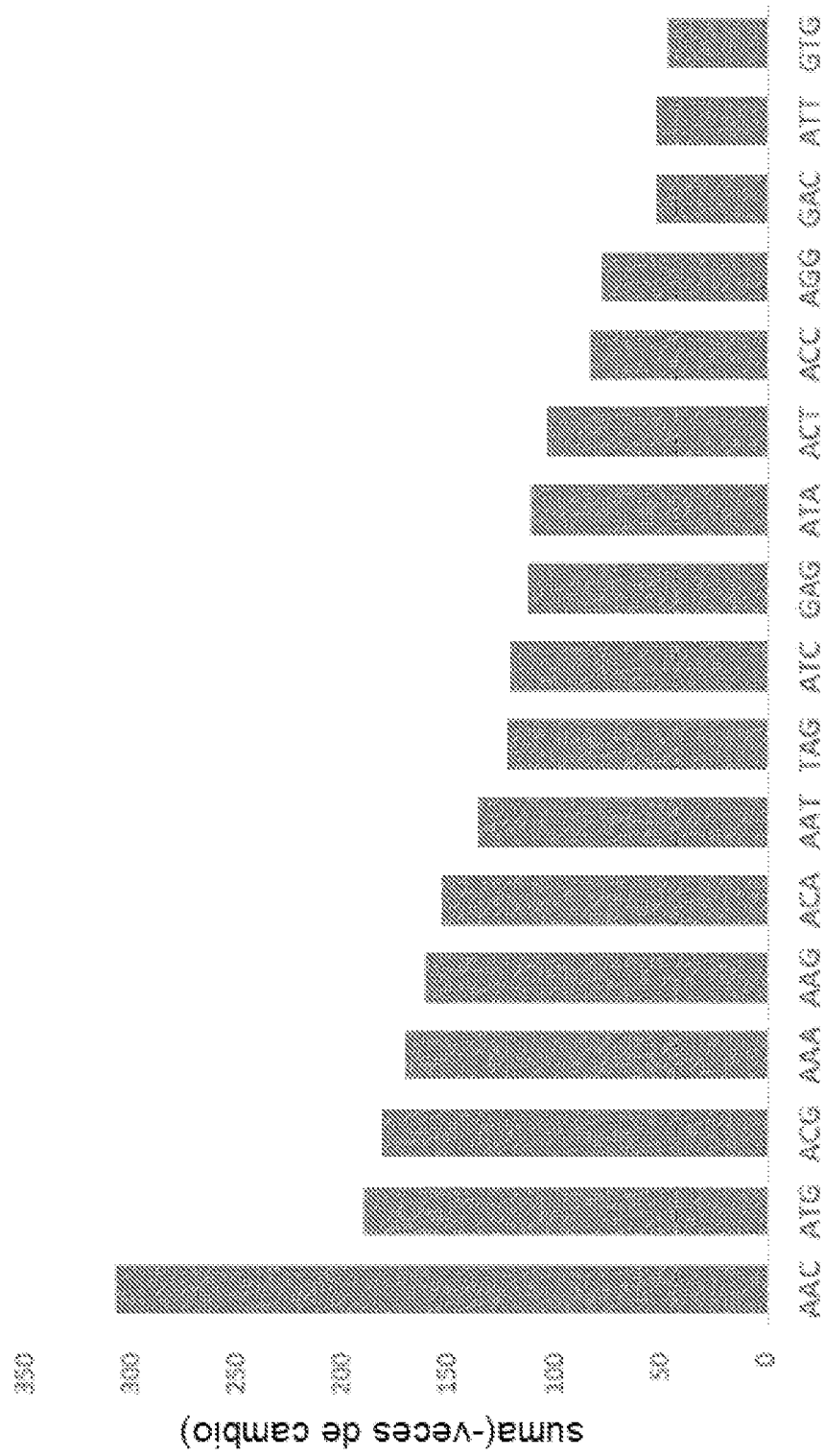
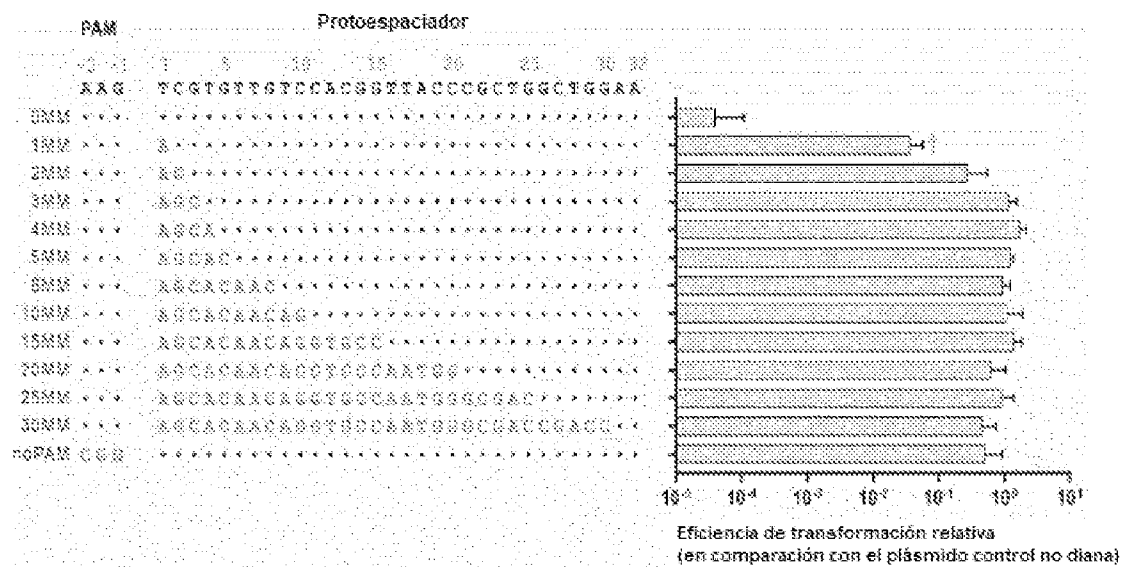
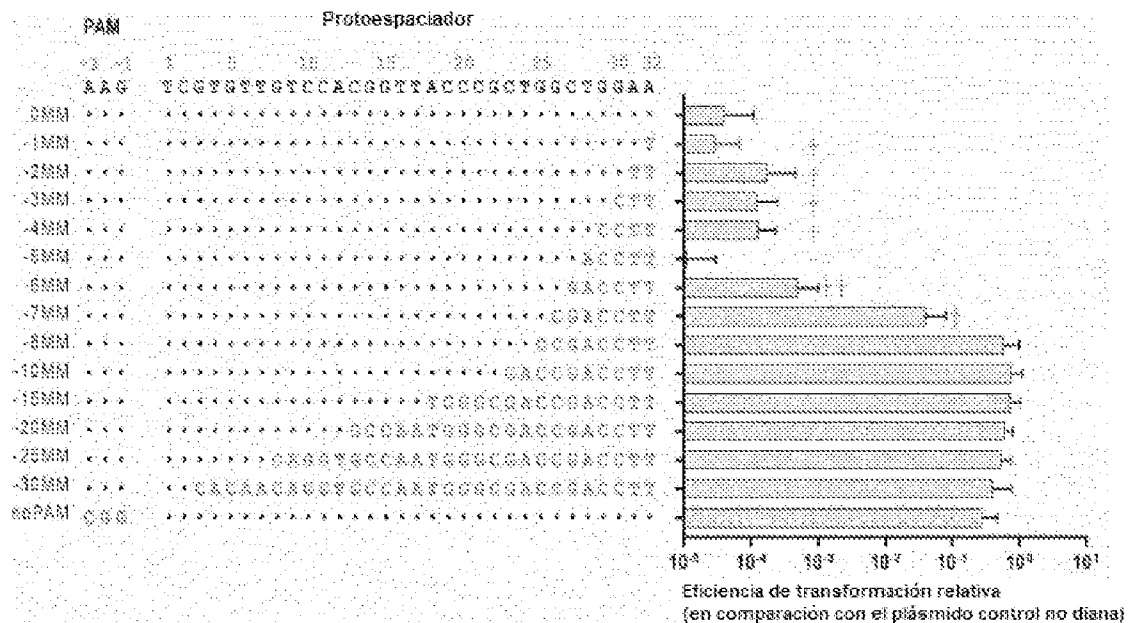


Figura 12



**Figura 13A:** errores de emparejamiento proximales a PAM (cadenas)



**Figura 13B:** errores de emparejamiento distales a PAM (cadenas)

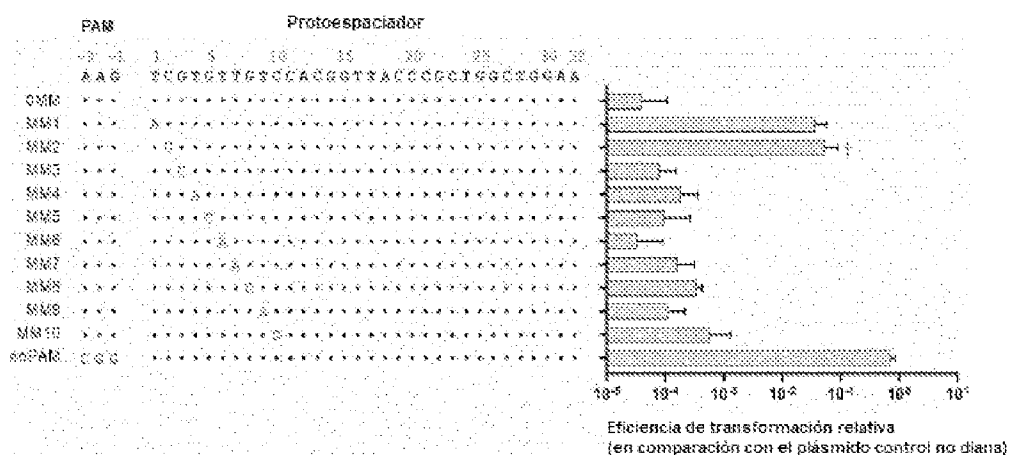


Figura 13C: errores de emparejamiento únicos proximales a PAM

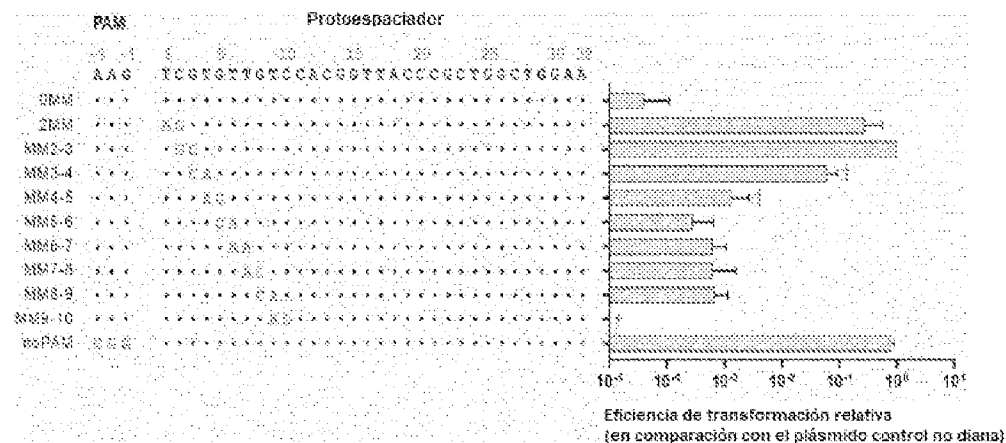


Figura 13D: errores de emparejamiento dobles proximales a PAM

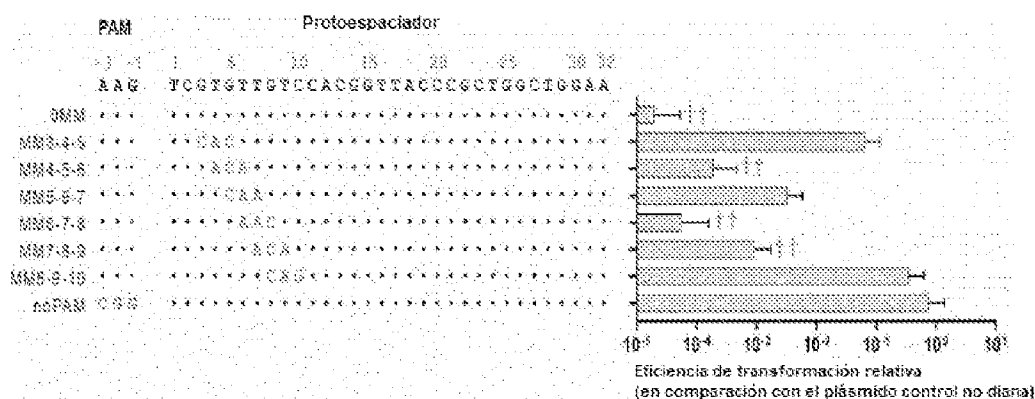
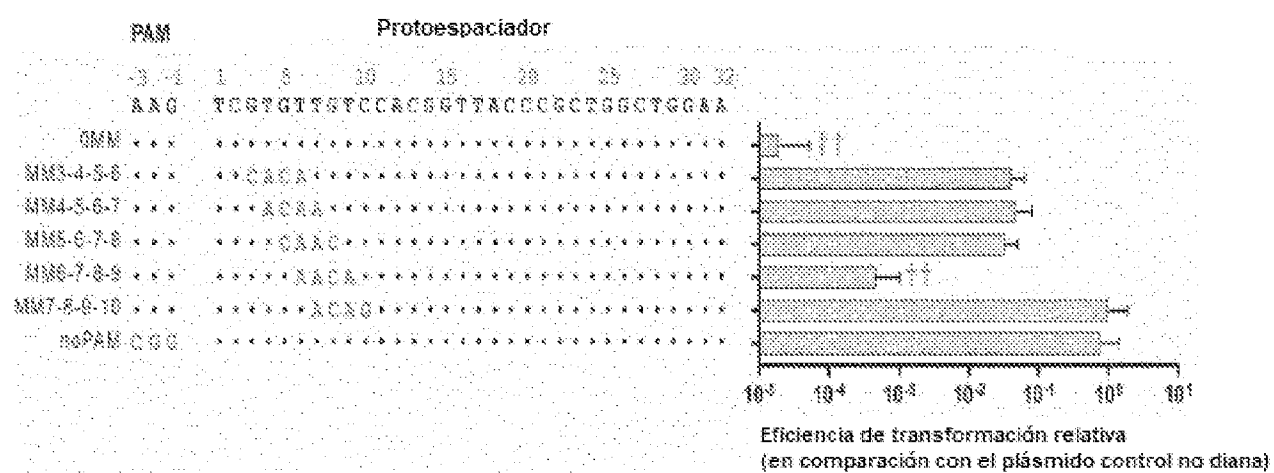


Figura 13E: errores de emparejamiento triples proximales a PAM



**Figura 13F:** errores de emparejamiento cuádruples proximales a PAM

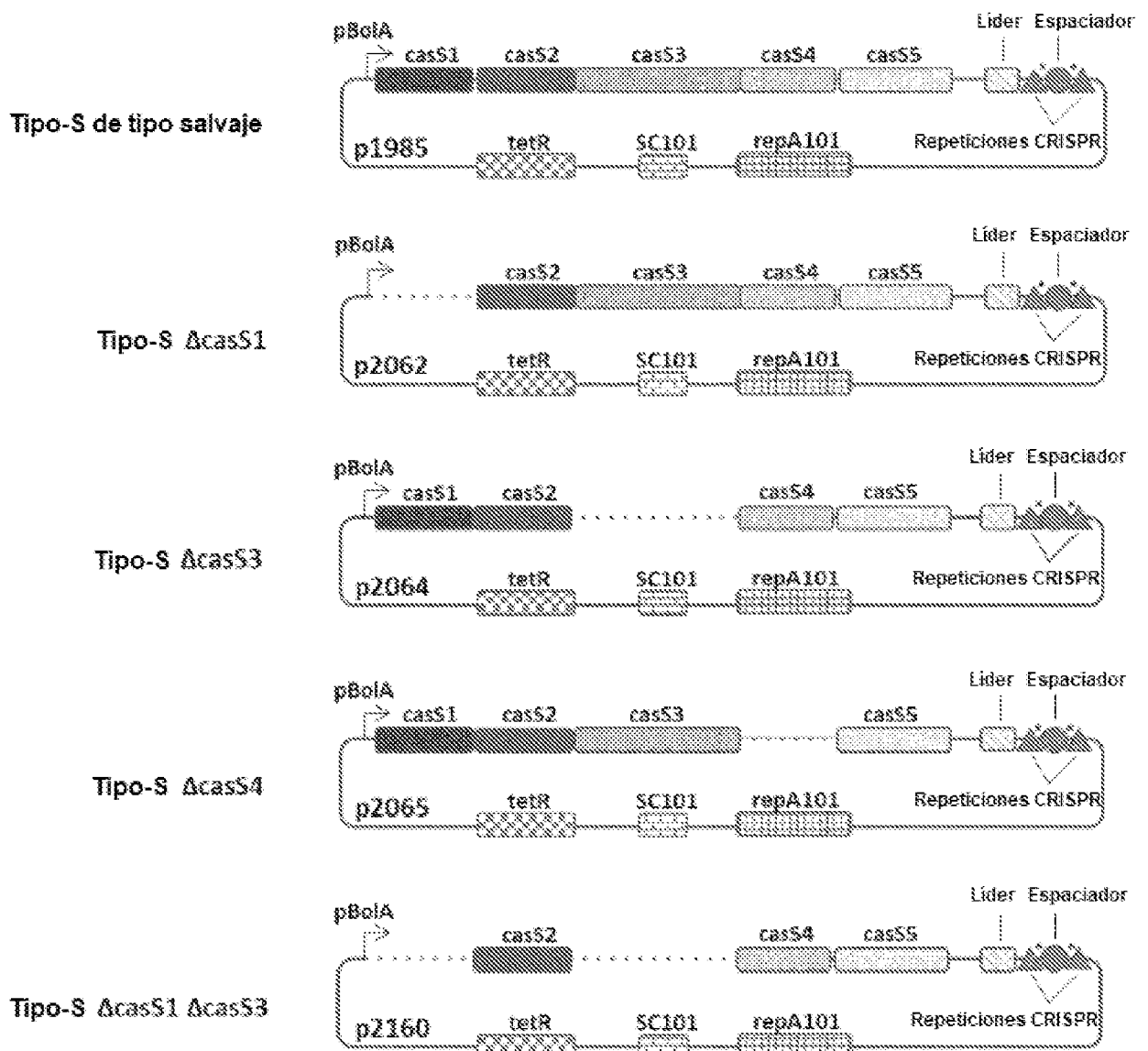


Figura 14

65815

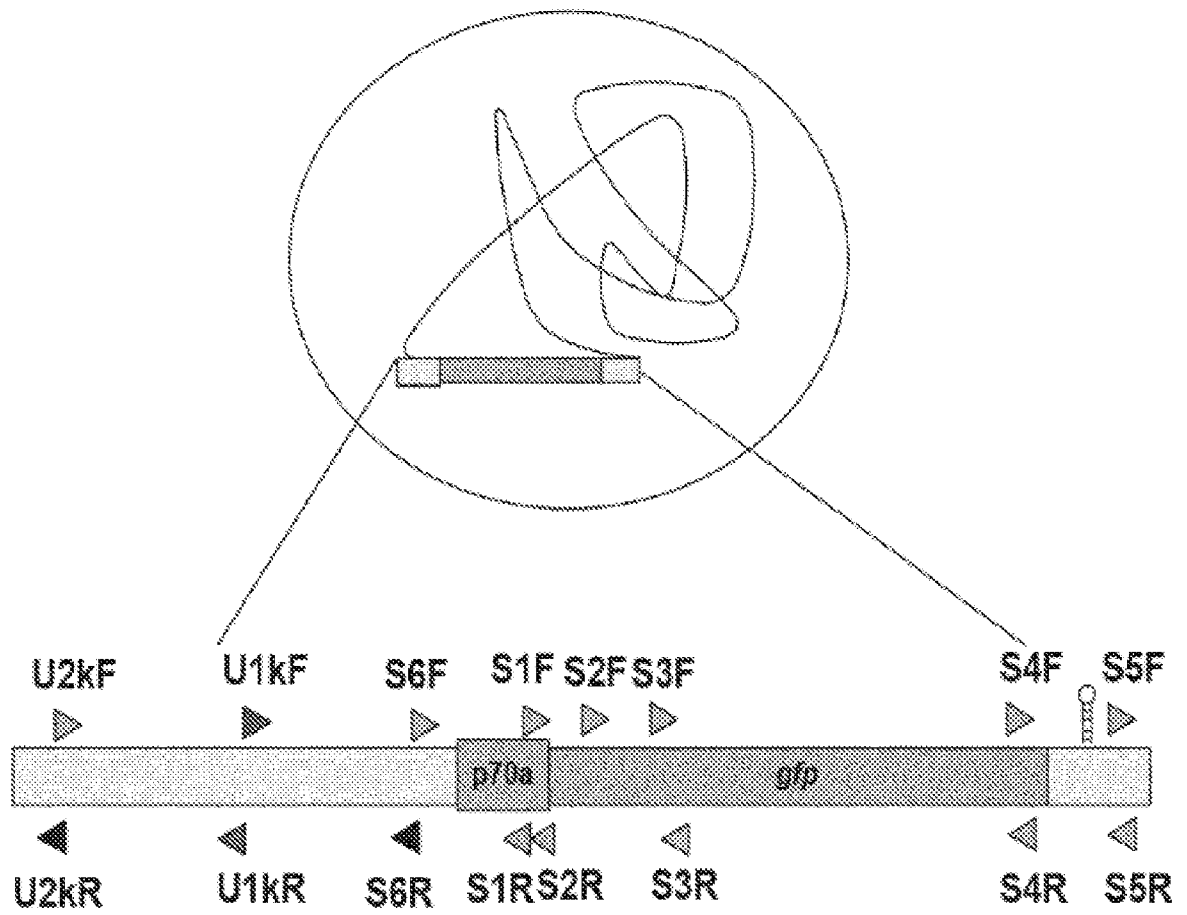


Figura 15

**CRISPRi Tipo-S de tipo salvaje\_cromosoma**

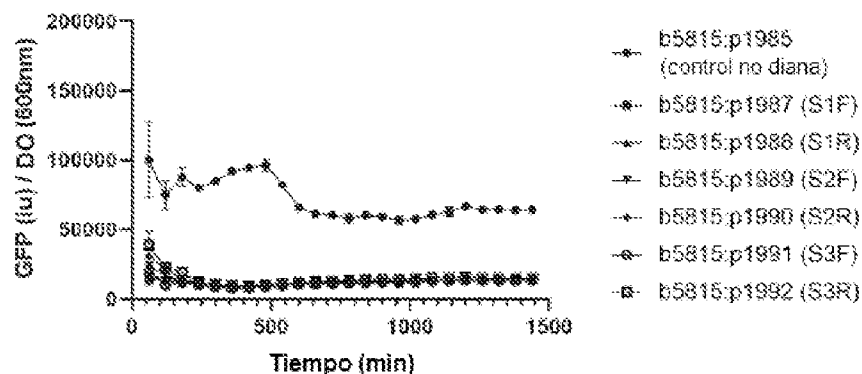


Figura 16A

**CRISPRi Tipo-S de tipo salvaje\_cromosoma**

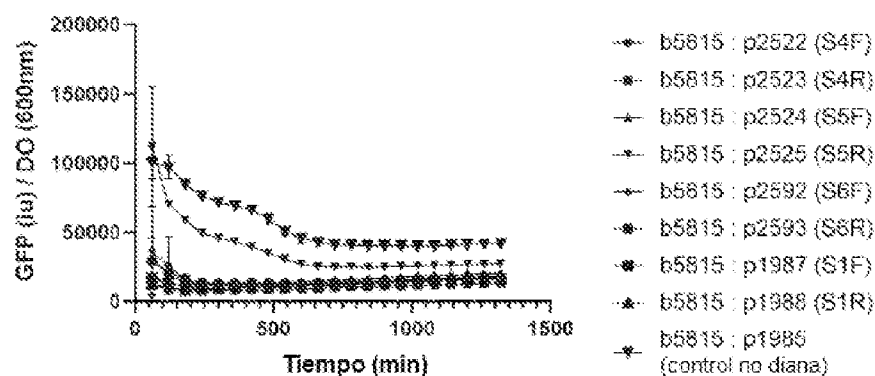


Figura 16B

**CRISPRi Tipo-S de tipo salvaje\_cromosoma**

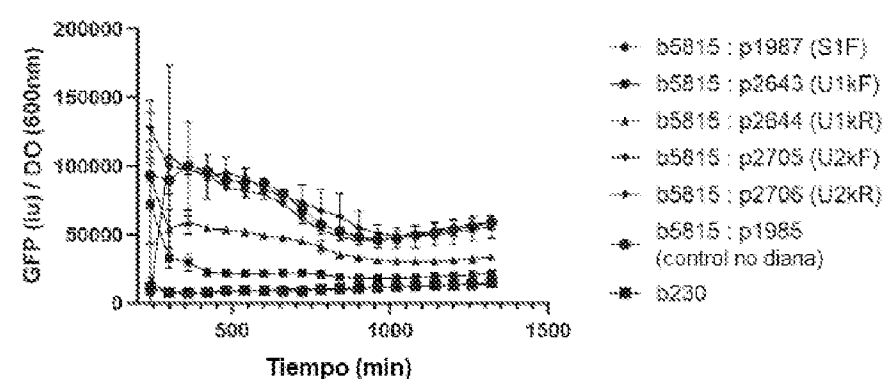


Figura 16C

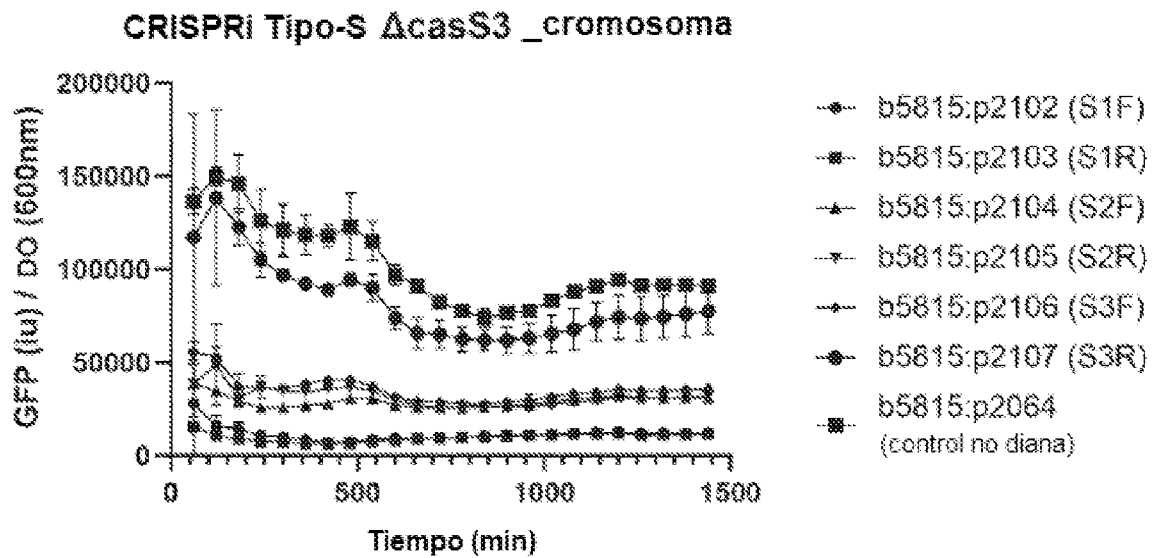


Figura 17A

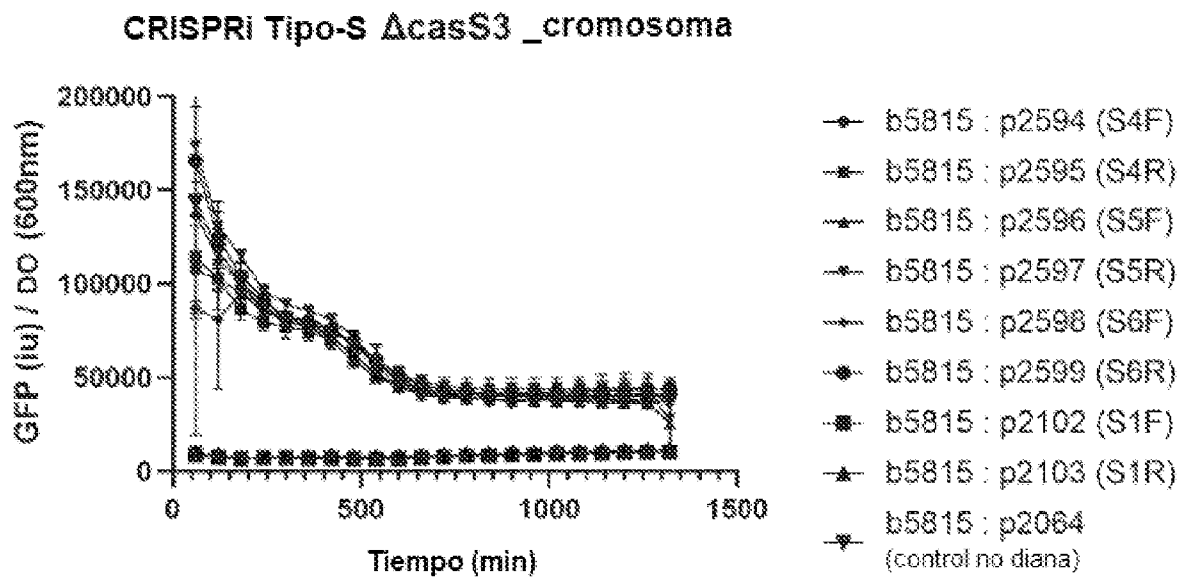


Figura 17B

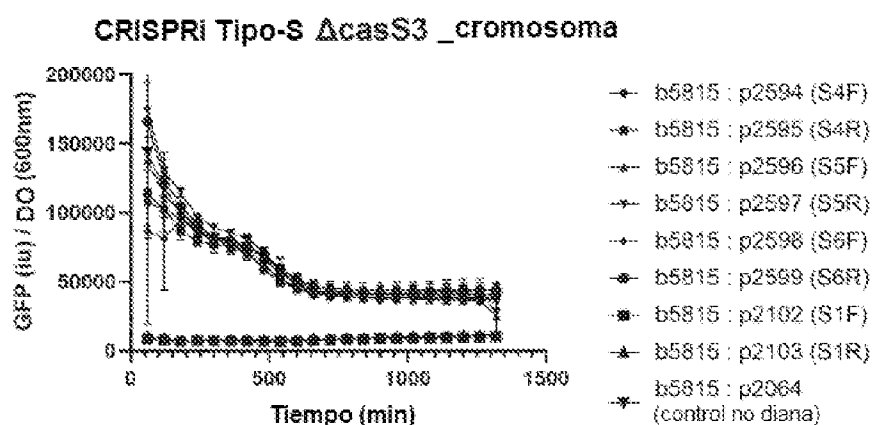


Figura 18A

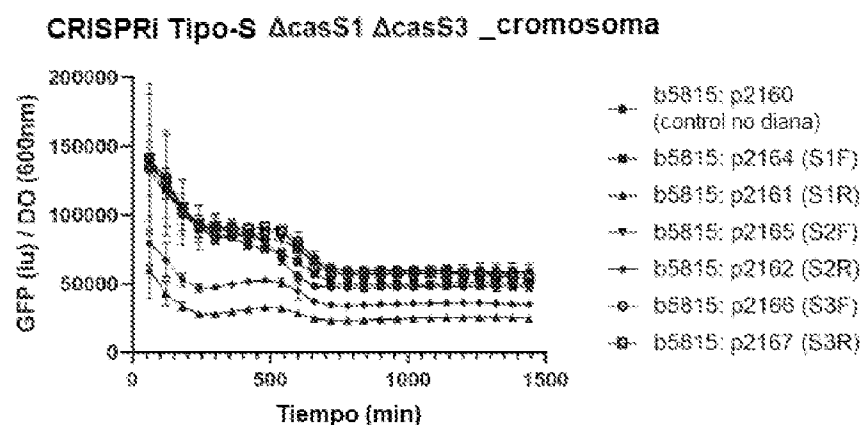


Figura 18B

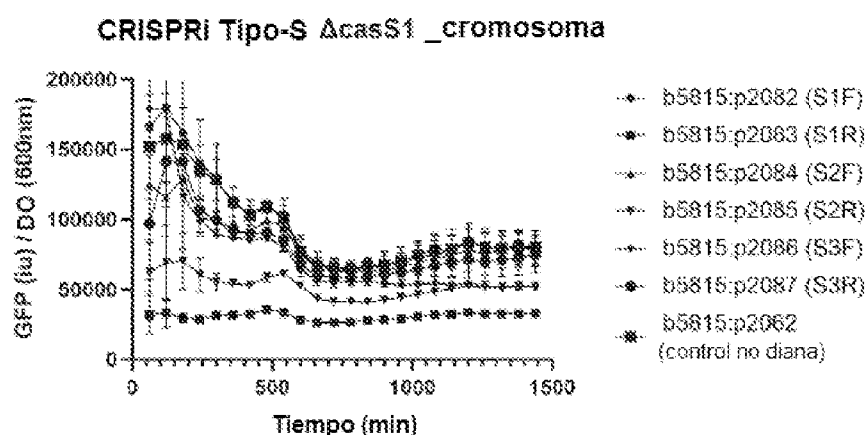


Figura 18C

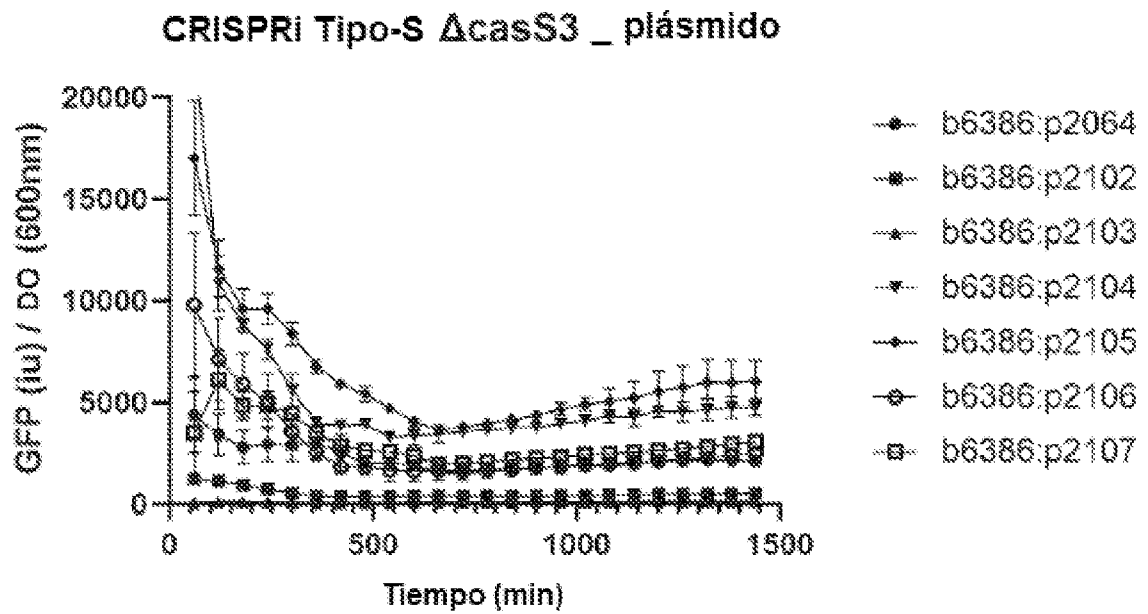


Figura 19A

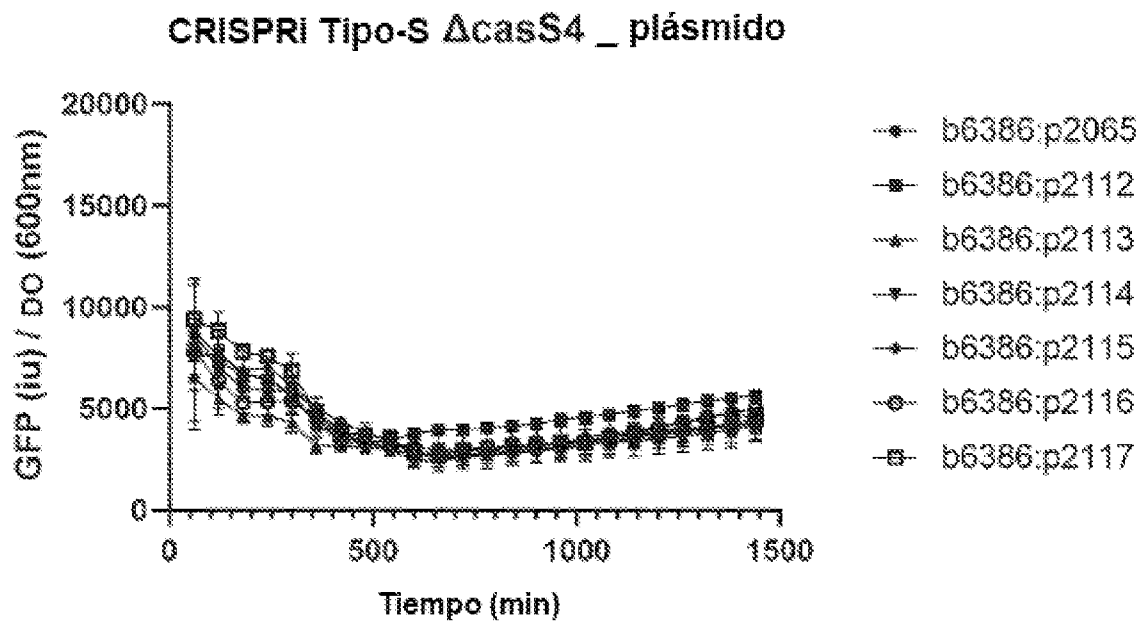


Figura 19B

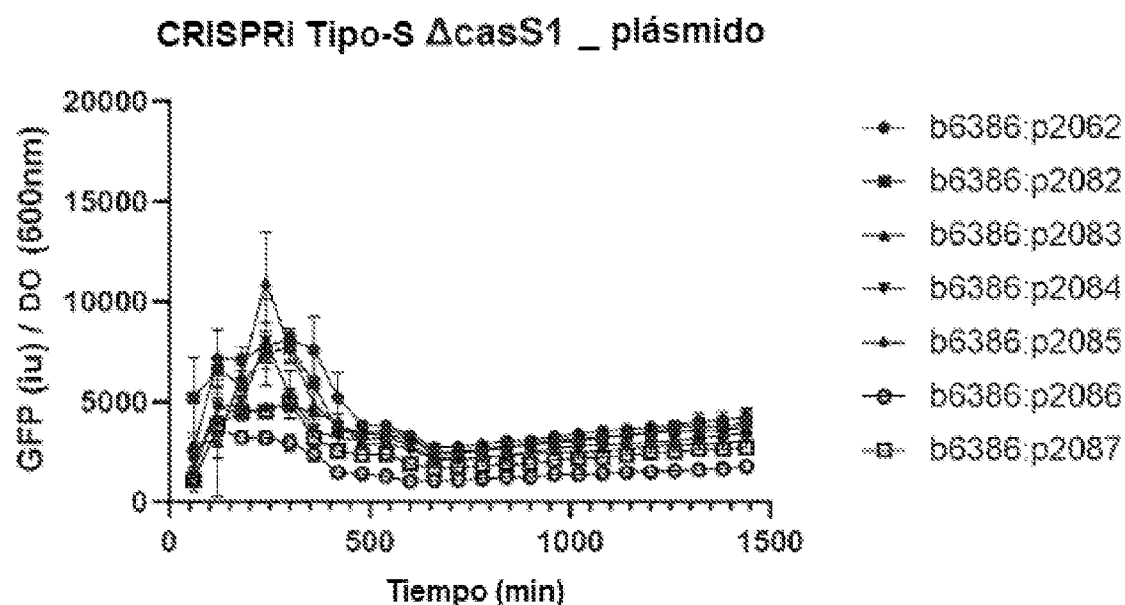


Figura 19C

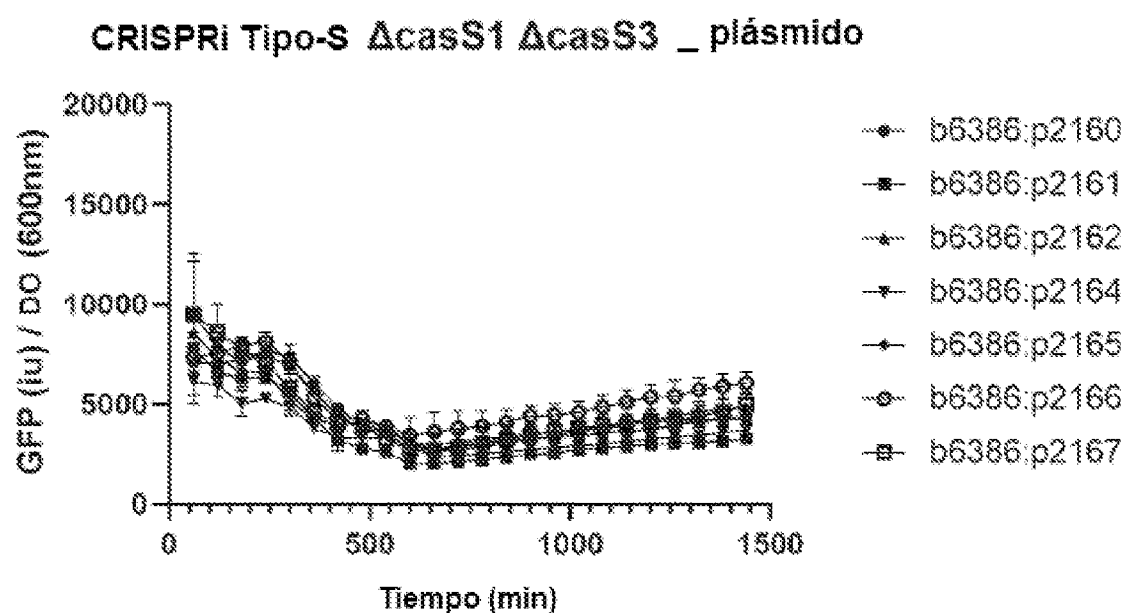


Figura 19D

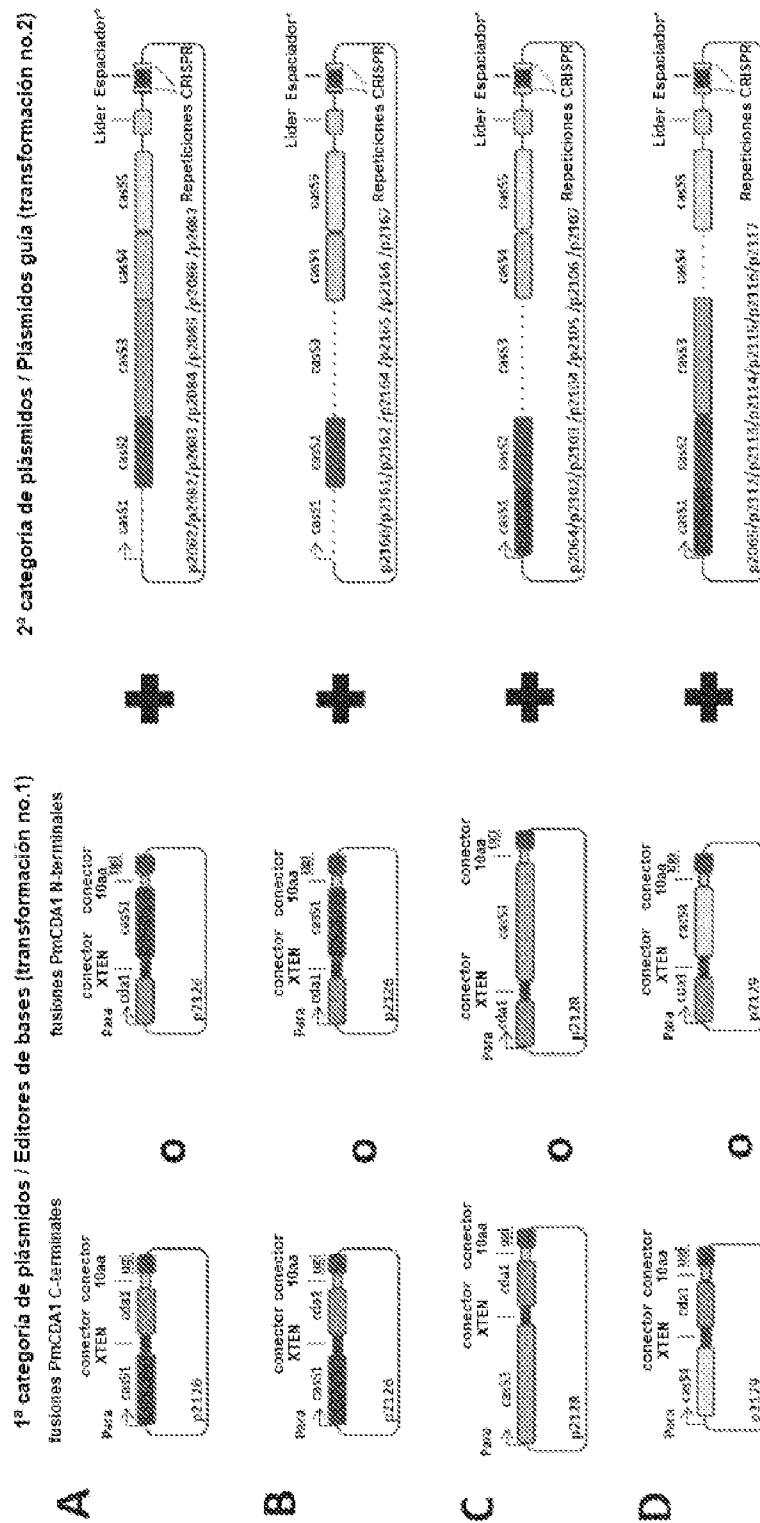




Figura 21A: PmCOA1-Cas-S1



Figura 21B: PmCDAl-Cas-S1 (Acas-53)

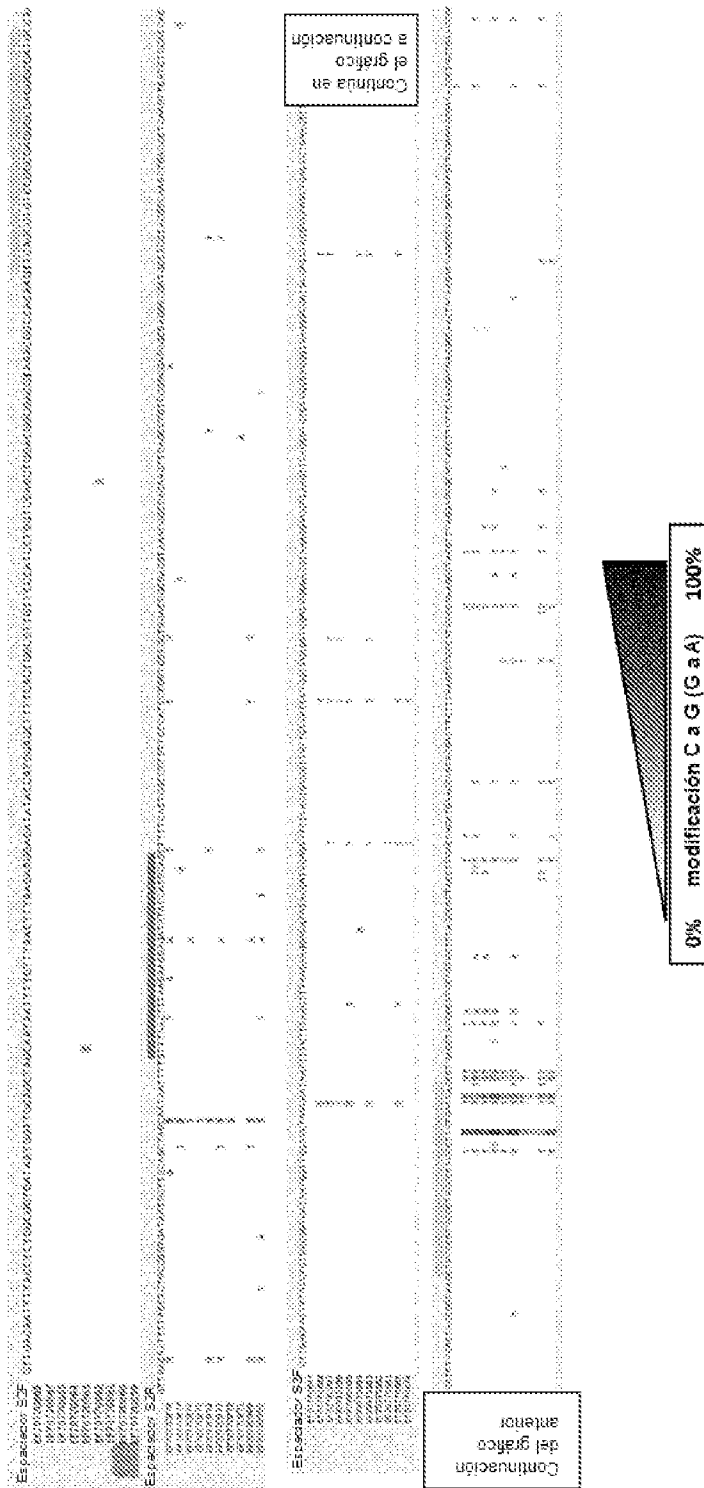


Figura 21C: PmCDA1-Cas-S3

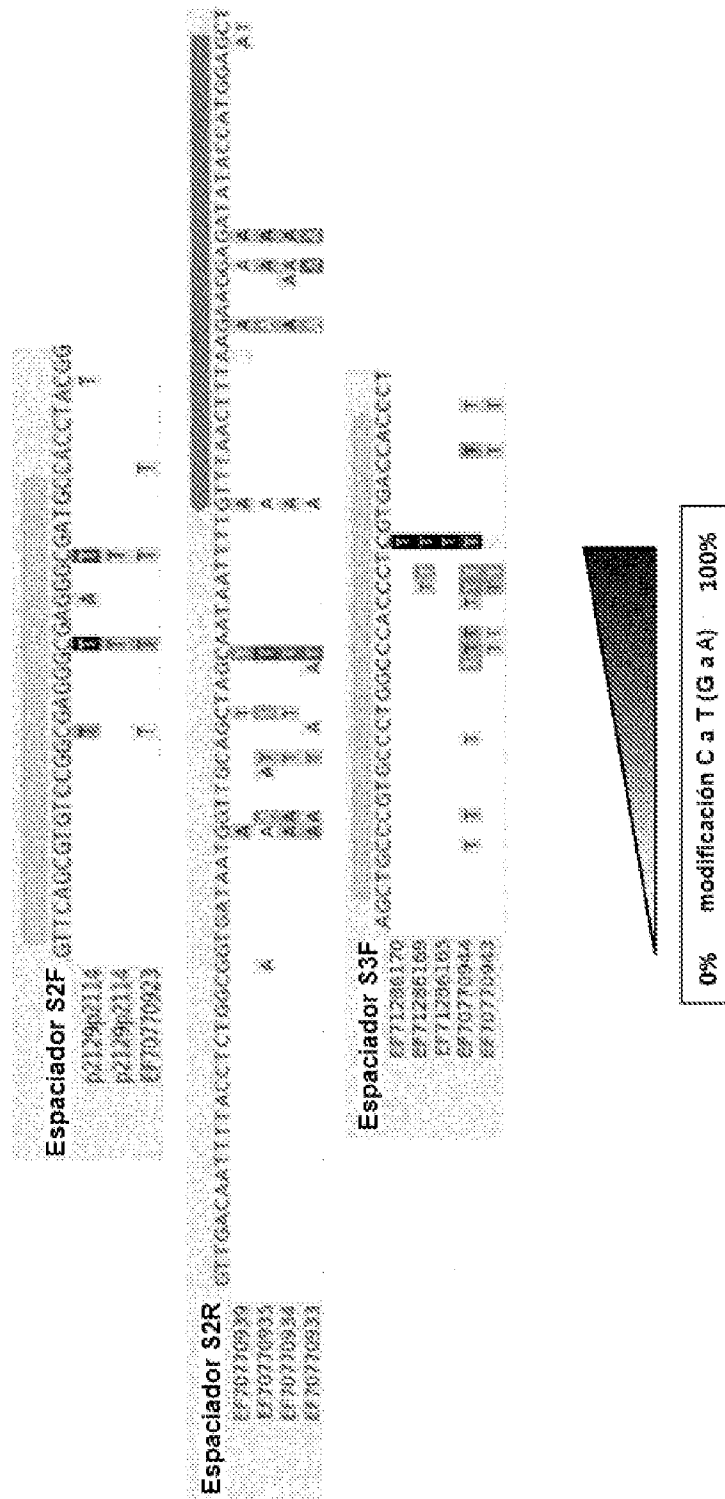


Figura 21D: PmCDAL-Cas-S4

2ª Transformación

1ª Transformación

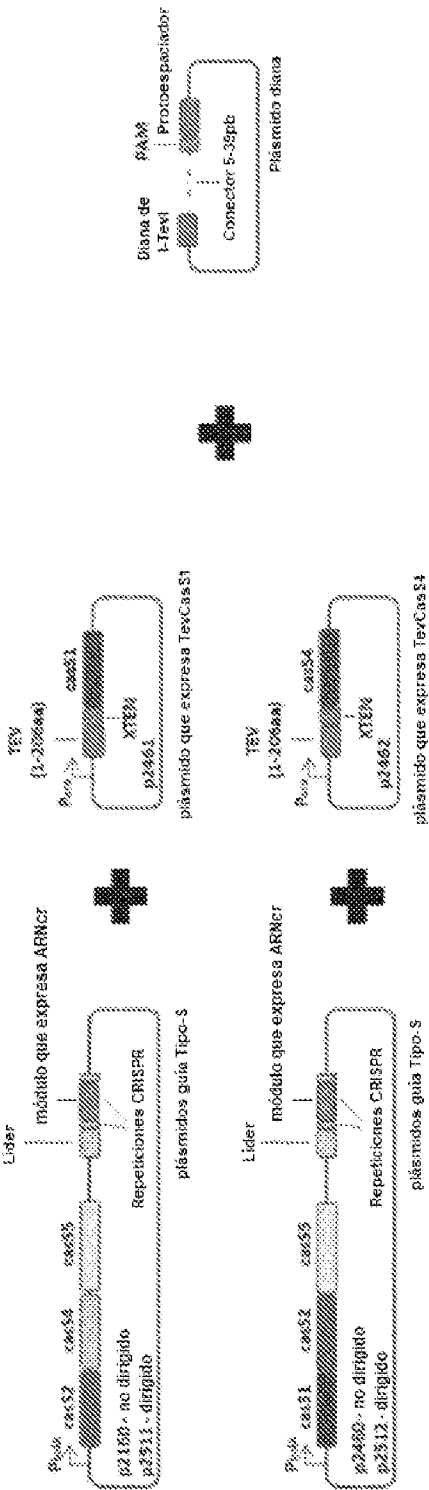


Figura 22

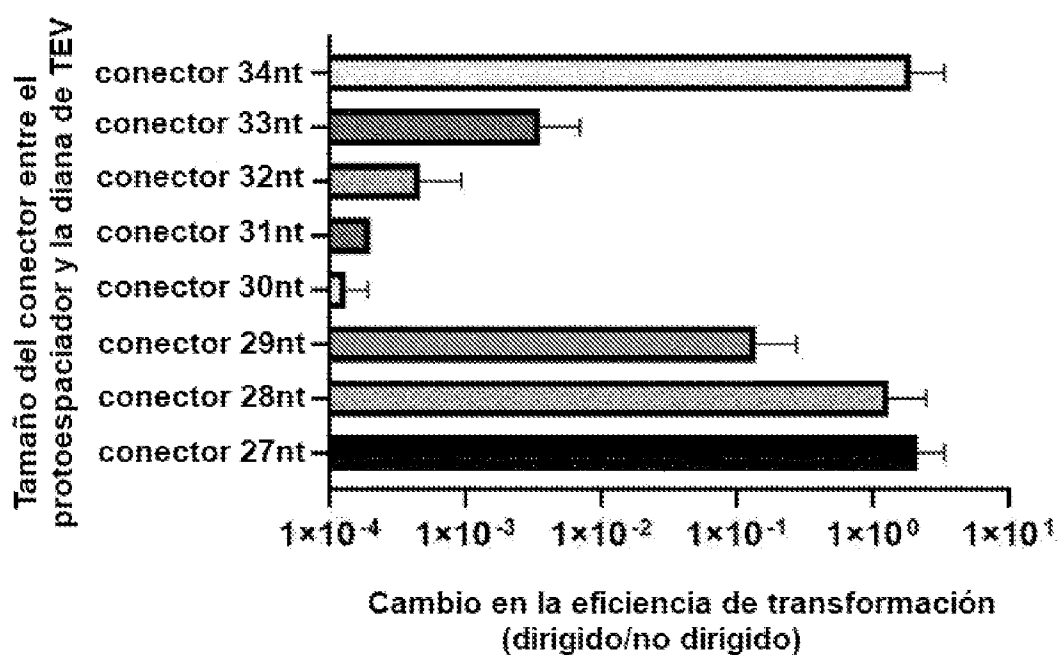


Figura 23

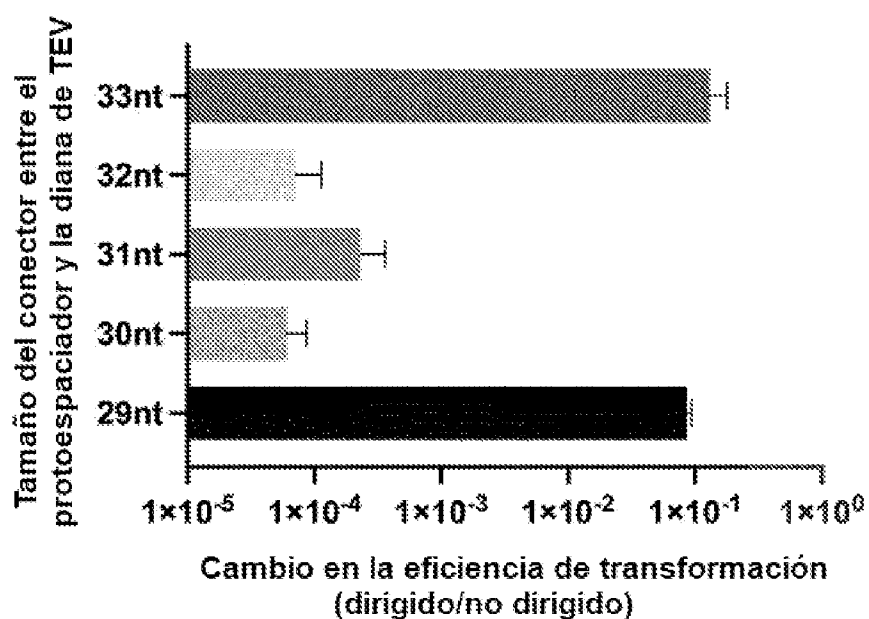


Figura 24

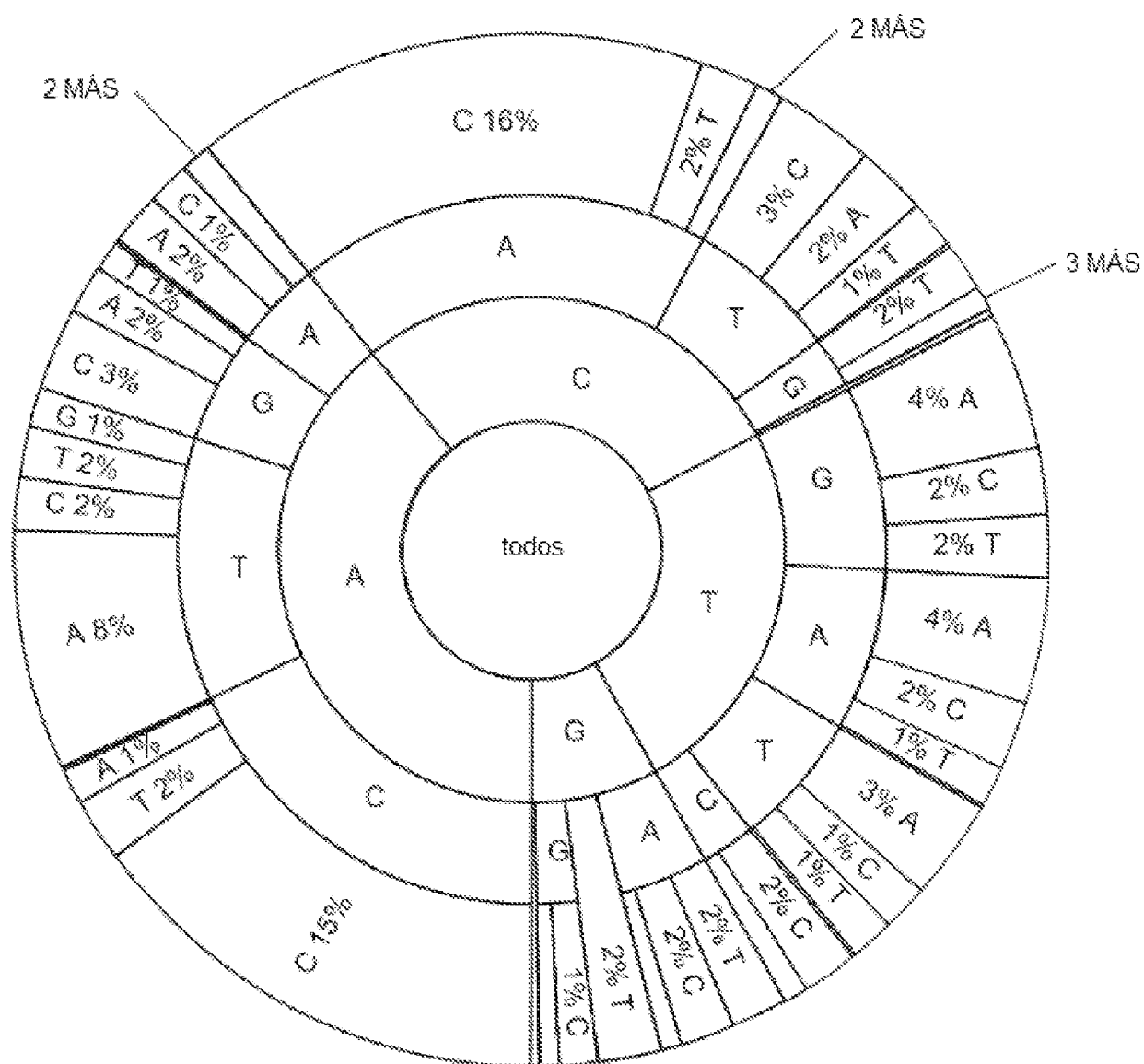


Figura 25