

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 774 677**

51 Int. Cl.:

C12N 9/22 (2006.01)

C07H 21/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.11.2014 PCT/US2014/066375**

87 Fecha y número de publicación internacional: **28.05.2015 WO15077318**

96 Fecha de presentación y número de la solicitud europea: **19.11.2014 E 14864613 (6)**

97 Fecha y número de publicación de la concesión europea: **08.01.2020 EP 3071694**

54 Título: **Proteínas CAS9 mutantes**

30 Prioridad:

19.11.2013 US 201361906374 P

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

22.07.2020

73 Titular/es:

**PRESIDENT AND FELLOWS OF HARVARD
COLLEGE (100.0%)
17 Quincy Street
Cambridge, MA 02138, US**

72 Inventor/es:

**CHAVEZ, ALEJANDRO;
POELWIJK, FRANCK y
CHURCH, GEORGE M.**

74 Agente/Representante:

DURAN-CORRETJER, S.L.P

ES 2 774 677 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

Proteínas CAS9 mutantes

5 SOLICITUD RELACIONADA

La presente solicitud reivindica prioridad a la Solicitud de Patente Provisional de los Estados Unidos N° 61/906,374, presentada el 19 de noviembre de 2013.

10 DECLARACIÓN DE INTERESES GUBERNAMENTALES

La presente invención se realizó con el apoyo del gobierno, bajo la Subvención No. P50 HG005550 otorgada por los Institutos Nacionales de Salud. El gobierno tiene ciertos derechos en la presente invención.

15 ESTADO DE LA TÉCNICA ANTERIOR

Los sistemas CRISPR-Cas (CRISP, de *Clustered Regularly Interspaced Short Palindromic Repeats*) bacterianos y de arqueas se basan en ARN guía cortos en complejo con proteínas Cas para dirigir la degradación de secuencias complementarias presentes dentro del ácido nucleico extraño invasor. Ver Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602-607 (2011); Gasiunas, G., Barrangou, R., Horvath, P. y Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579-2586 (2012); Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012); Sapranaukas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic acids research* 39, 9275-9282 (2011); y Bhaya, D., Davison, M. y Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics* 45, 273-297 (2011). Una reciente reconstitución *in vitro* del sistema CRISPR de *S. pyogenes* tipo II demostró que el ARNcr ("ARN CRISPR") fusionado a un ARNcrtra transcodificado normalmente ("ARN CRISPR transactivador") es suficiente para dirigir la proteína Cas9 para escindir secuencias de ADN diana de una manera específica de secuencia que coinciden con el ARNcr. Expresar un ARNg homólogo a un sitio diana da como resultado el reclutamiento de Cas9 y la degradación del ADN diana. Ver H. Deveau et al., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology* 190, 1390 (febrero de 2008).

Cas9 es una nucleasa de ADN que se puede programar para dirigirse a casi cualquier región de un genoma mediante la expresión de un ARN guía (ARNg) que contiene un motivo que recluta Cas9 y 20 pares de bases de complementariedad a una región del genoma a la que se desea el direccionamiento. Todos los miembros de la familia Cas9 caracterizados y putativos tienen un tamaño de diversas kilobases (> 3.000 pares de bases) con el miembro NM-Cas9 más pequeño funcionalmente validado (Cas9 de *Neisseria meningitidis*) con un tamaño de 3.249 pares de bases. El gran tamaño de esta proteína limita su potencial para aplicaciones biotecnológicas y terapéuticas debido a las dificultades de administración y manipulación.

CARACTERÍSTICAS

La presente invención se refiere a un procedimiento para fabricar una proteína Cas9 mutante, tal como se define en las reivindicaciones adjuntas.

Los aspectos de la presente divulgación se refieren a una proteína de unión al ADN guiada por ARN de un sistema CRISPR Tipo II que se une al ADN y está guiada por uno o más ARN, que se ha diseñado para omitir partes de la proteína mientras sigue funcionando como una nucleasa de unión al ADN guiada por ARN que puede unirse al ADN diana y crear una ruptura bicatenaria en el ADN diana. Según un aspecto, la proteína de unión al ADN guiada por ARN de un sistema CRISPR Tipo II es una proteína Cas9.

Los aspectos de la presente divulgación se refieren a una proteína de unión al ADN guiada por ARN de un sistema CRISPR Tipo II, que se ha diseñado para omitir partes de la proteína mientras sigue funcionando como una nickasa de unión al ADN guiada por ARN que puede unirse al ADN diana y crear una ruptura o mella en el ADN diana. Según un aspecto, la proteína de unión al ADN guiada por ARN de un sistema CRISPR Tipo II es una proteína Cas9.

Los aspectos de la presente divulgación se refieren a una proteína de unión al ADN guiada por ARN de un sistema CRISPR Tipo II que se ha diseñado para omitir partes de la proteína mientras sigue funcionando como una proteína de unión al ADN guiada por ARN que es una nucleasa nula, es decir, la proteína de unión al ADN guiada por ARN carece de actividad de nucleasa. Según un aspecto, la proteína de unión al ADN guiada por ARN de un sistema CRISPR Tipo II es una proteína Cas9.

Según un aspecto, las partes de una proteína de unión al ADN guiada por ARN se identifican para su eliminación mediante su identificación dentro de una población de especies de secuencias de proteínas de unión al ADN

guiadas por ARN, que no están bien conservadas o que de otra manera son muy divergentes dentro de una familia de proteínas y/o secuencias de proteínas de unión al ADN guiadas por ARN particulares entre límites, entre conservación baja y alta, denominados en la presente memoria descriptiva "bordes de conservación" dentro de una familia particular de proteínas de unión al ADN guiadas por ARN. Según este aspecto, las secuencias de aminoácidos dentro de una proteína de unión al ADN, tales como una proteína de unión al ADN guiada por ARN, tal como Cas9, se identifican como de alta conservación o de baja conservación utilizando los procedimientos descritos en la presente memoria descriptiva y tal como conocen los expertos en la materia. Según un aspecto, las secuencias de aminoácidos de alta conservación y las secuencias de aminoácidos de baja conservación son adyacentes, tal como inmediatamente adyacentes entre sí, dentro de la secuencia de proteínas de la proteína de unión al ADN en su conjunto. Las secuencias de aminoácidos de alta conservación y las secuencias de aminoácidos de baja conservación se distinguen por un aminoácido que separa una secuencia de aminoácidos de alta conservación de una secuencia de aminoácidos de baja conservación. De esta manera, en la presente memoria descriptiva, el aminoácido que separa una secuencia de aminoácidos de alta conservación de una secuencia de aminoácidos de baja conservación se denomina un "aminoácido de borde" o un "borde de conservación", en la medida en que está en un borde o parte terminal de una secuencia de aminoácidos de alta conservación o una secuencia de aminoácidos de baja conservación. Por consiguiente, los procedimientos de la presente divulgación contemplan identificar un aminoácido que separa una secuencia de aminoácidos de alta conservación de una secuencia de aminoácidos de baja conservación o que de otra manera, distingue una secuencia de aminoácidos de alta conservación de una secuencia de aminoácidos de baja conservación. En la presente memoria descriptiva, este aminoácido se denomina un "aminoácido de borde". Según este aspecto, un par de aminoácidos de borde pueden flanquear o unirse a cada extremo de una secuencia de aminoácidos de alta conservación. De la misma manera, un par de aminoácidos de borde pueden flanquear o unirse a cada extremo de una secuencia de aminoácidos de baja conservación. Aún según este aspecto, una realización de ejemplo se refiere a la identificación, dentro de la secuencia de proteínas de una proteína de unión al ADN en su conjunto, de secuencias de aminoácidos de alta conservación y secuencias de aminoácidos de baja conservación adyacentes. En particular, una realización de ejemplo se refiere a la identificación dentro de la secuencia de la proteína de una proteína de unión al ADN en su conjunto, de secuencias de alta conservación en tándem o en serie con secuencias de baja conservación y, en particular, una secuencia de alta conservación (HC, de *High Conservation*) delimitada en ambos los extremos por una secuencia de baja conservación (LC, de *Low Conservation*) o, alternativamente, una secuencia de baja conservación (LC) delimitada en ambos extremos por una secuencia de alta conservación (HC). De esta manera, ejemplos de secuencias en tándem o secuencias en serie identificadas por los procedimientos descritos en la presente memoria descriptiva pueden representarse esquemáticamente como LC-HC-LC o HC-LC-HC. Según este aspecto, una secuencia intermedia de alta conservación o baja conservación está delimitada por secuencias flanqueantes de baja conservación o alta conservación, respectivamente. En la secuencia tándem de ejemplo LC-HC-LC, un par de aminoácidos de borde distinguen o separan la secuencia de aminoácidos de alta conservación (HC) de las dos secuencias de aminoácidos flanqueantes de baja conservación (LC) que están unidas en ambos extremos o de cualquier otra manera a la secuencia de aminoácidos de alta conservación. En la secuencia tándem de ejemplo HC-LC-HC, un par de aminoácidos de borde distinguen o separan la secuencia de aminoácidos de baja conservación (LC) de las dos secuencias de aminoácidos flanqueantes de alta conservación (HC) que están unidas en ambos extremos o de cualquier otra manera a la secuencia de aminoácidos de baja conservación. Cuando se identifica una secuencia tándem de ejemplo de este tipo utilizando los procedimientos descritos en la presente memoria descriptiva, ya sea que la secuencia intermedia sea una secuencia de aminoácidos de alta conservación o una secuencia de aminoácidos de baja conservación, la secuencia intermedia se elimina para crear una proteína mutante de unión al ADN según los procedimientos descritos en la presente memoria descriptiva, que retiene la actividad de unión al ADN y que es de menor tamaño en comparación con la proteína de unión al ADN de tipo salvaje. Los aminoácidos del borde definen la secuencia intermedia que se eliminará flanqueando la secuencia intermedia o separando la secuencia intermedia de las secuencias adyacentes en serie para crear la proteína mutante de unión al ADN. Según ciertos aspectos de la presente divulgación, se puede eliminar una secuencia intermedia en las secuencias en tándem independientemente de si la secuencia intermedia es una secuencia de aminoácidos de alta conservación o una secuencia de aminoácidos de baja conservación, en la medida en que la proteína mutante de unión al ADN retenga la actividad proteica útil de unión al ADN.

Para identificar secuencias que no están bien conservadas o secuencias de proteínas entre los bordes de conservación dentro del contexto de la presente divulgación, se obtuvo una alineación de PFAM o se creó de otra manera a partir de una colección de secuencias que son el resultado de una búsqueda en la base de datos de homólogos de Cas9. Esta alineación estaba condicionada computacionalmente y se calculó la conservación como la entropía por posición (relativa) de las frecuencias de aminoácidos. Posteriormente, las secuencias se eliminan de la proteína de unión al ADN guiada por ARN para producir un mutante.

Según un aspecto, para identificar regiones dentro de la proteína Cas9 multidominio que pueden ser susceptibles de eliminación, se utiliza un enfoque bioinformático para identificar límites de dominio potenciales en las proteínas Cas9. Se crea una alineación de secuencia múltiple al realinear secuencias Cas9 en la base de datos PFAM (PF13395) utilizando MUSCLE, y la alineación está condicionada computacionalmente para la diversidad y secuencias de longitud completa. Se calcula la conservación de la secuencia como la entropía relativa de las frecuencias de aminoácidos observadas con respecto a las frecuencias promedio en todos los genes en *Escherichia coli*. Se aplica un filtro de borde de múltiples escalas (filtro de paso de banda de diferencia de gaussianas (DoG, de

difference of Gaussians)) al perfil de conservación para asignar límites potenciales de dominio de proteína, a los que se hace referencia en la presente memoria descriptiva como bordes de conservación. Se seleccionan las regiones entre los bordes de conservación para su eliminación en la primera iteración de mutantes de eliminación.

5 Según ciertos aspectos, la presente divulgación describe mutantes de eliminación NM-Cas9 sintéticos que son de menor tamaño pero que retienen la actividad de la proteína de tipo salvaje. Los mutantes de eliminación NM-Cas9 sintéticos se pueden utilizar para unirse al ADN como un complejo de colocalización con ARN guía en una célula y crear una ruptura bicatenaria, una ruptura monocatenaria o para localizar un grupo efector cerca del ADN diana de interés para realizar una función deseada.

10 Según ciertos aspectos, se da a conocer un procedimiento de detección de dominio basado en la alineación para identificar regiones de una proteína de unión al ADN, tal como Cas9, que son prescindibles para la unión al ADN, y que se pueden eliminar para formar una proteína mutante de unión al ADN que es de menor tamaño en comparación con la proteína de unión al ADN de tipo salvaje. Según los procedimientos descritos en la presente memoria descriptiva, se generan variantes de Cas9 minimizadas que muestran una actividad robusta en bacterias y células humanas. Según los aspectos descritos en la presente memoria descriptiva, se dan a conocer variantes de proteína funcionales mutantes de unión al ADN, tales como variantes de Cas9 funcionales mutantes, que son más pequeñas que las proteínas de unión al ADN de tipo salvaje.

15 Según ciertos aspectos, las proteínas de unión al ADN de ejemplo incluyen ortólogos de Cas9, tales como Cas9 de *Neisseria meningitidis* (NM, GI: 218767588) y Cas9 de *Streptococcus thermophilus* (ST1, GI: 116627542) que se ha demostrado que funcionan tanto en procariotas como en eucariotas superiores. Ver Hou, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. Proceedings of the National Academy of Sciences of the United States of America 110, 15644-15649, doi: 10.1073/pnas.1313587110 (2013).

20 Estos ortólogos de Cas9 de ejemplo son más pequeños en tamaño de gen en comparación con Cas9 de *Streptococcus pyogenes* (SP, GI: 13622193), es decir, aproximadamente, 3.200 frente a 4.100 pares de bases. Por lo tanto, los aspectos de la presente divulgación están dirigidos a reducir el tamaño de una proteína Cas9 de unión al ADN para aumentar la eficiencia con la que se puede administrar la proteína Cas9 de unión al ADN, particularmente utilizando tecnologías de empaquetamiento viral en las que la longitud del gen puede influir en gran medida en el título viral. Ver Kumar, M., Keller, B., Makalou, N. y Sutton, R. E. Systematic determination of the packaging limit of lentiviral vectors. Human gene therapy 12, 1893-1905, doi: 10.1089/104303401753153947 (2001); Wu, Z., Yang, H. y Colosi, P. Effect of genome size on AAV vector packaging. Molecular therapy: the journal of the American Society of Gene Therapy 18, 80-86, doi: 10.1038/mt.2009.255 (2010); y Gelinis, C. y Temin, H. M. Nondefective spleen necrosis virus-derived vectors define the upper size limit for packaging reticuloendotheliosis viruses. Proceedings of the National Academy of Sciences of the United States of America 83, 9211-9215 (1986).

25 La reducción sintética del tamaño de los genes Cas9 permite que sistemas reguladores más complejos y dominios funcionales se empaqueten dentro de vectores individuales. Según un aspecto adicional, se dan a conocer procedimientos para alterar sintéticamente la especificidad de PAM que permite la generación de variantes de Cas9 más pequeñas con un mayor potencial de direccionamiento.

30 Según ciertos aspectos, se dan a conocer procedimientos para producir Cas9 quiméricas intercambiando el dominio C-terminal de una primera especie de Cas9 con el dominio C-terminal de una segunda especie de Cas9. Según un aspecto, la presente divulgación da a conocer intercambio de dominio de Cas9 quimérica, tal como una quimera NM-ST1-Cas9 funcional, intercambiando el dominio C-terminal de NM con ST1. La proteína Cas9 quimérica exhibe especificidad a ARN guía ST1 y a PAM.

35 Según un aspecto, la célula es una célula procariota o una célula eucariota. Según un aspecto, la célula es una célula bacteriana, una célula de levadura, una célula vegetal o una célula animal. Según un aspecto, la célula es una célula de mamífero.

40 Según un aspecto, el ARN está entre, aproximadamente, 10 y, aproximadamente, 500 nucleótidos. Según un aspecto, el ARN está entre, aproximadamente, 20 y, aproximadamente, 100 nucleótidos.

45 Según un aspecto, el uno o más ARN es un ARN guía. Según un aspecto, el uno o más ARN es un ARNcr. Según un aspecto, el uno o más ARN es un ARNcrtra. Según un aspecto, el uno o más ARN es una fusión ARNcrtra-ARNcr.

50 Según un aspecto, el ADN diana es ADN genómico, ADN mitocondrial, ADN viral, elemento conjugable o ADN exógeno.

55 Según un aspecto, la proteína de unión al ADN guiada por ARN es de un sistema CRISPR Tipo II que se une al ADN y está guiada por uno o más ARN. Según un aspecto, la proteína de unión al ADN guiada por ARN es una proteína Cas9 que se une al ADN y está guiada por uno o más ARN.

Otras características y ventajas de ciertas realizaciones de la presente invención se harán más evidentes en la siguiente descripción de las realizaciones y dibujos de la misma, y a partir de las reivindicaciones.

DESCRIPCIÓN BREVE DE LOS DIBUJOS

5 El archivo de patente o solicitud contiene, como mínimo, un dibujo ejecutado en color. La Oficina proporcionará copias de esta publicación de patente o de solicitud de patente con el o los dibujos en color, previa solicitud y pago de la tarifa necesaria. Las características y ventajas anteriores y otras de la presente invención se entenderán más completamente a partir de la siguiente descripción detallada de realizaciones ilustrativas, tomadas conjuntamente con los dibujos adjuntos, en los que:

15 La figura 1 es una imagen en la que las células de *E. coli* contienen un informador YFP de la actividad de NM-Cas9 y se transforman con diversos genes nulos de nucleasa de NM-Cas9. En ausencia de NM-Cas9, las células son fluorescentes (cuadrante superior derecho-control negativo) y en presencia de nucleasa completa NM-Cas9 nulo las células no son fluorescentes (cuadrante superior izquierdo-control positivo). Se muestran dos de las eliminaciones de NM-Cas9 generadas, NM-Cas9- Δ 255-449 muestra niveles de represión casi de tipo salvaje (cuadrante inferior izquierdo) y NM-Cas9- Δ 874-922 muestra la carencia de la mayor parte de la capacidad de unión al ADN (cuadrante inferior derecho).

20 La figura 2 es un árbol filogenético, tal como se describe en Fonfara I, et al., (2014) Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems, *Nucleic Acids Res.* 42, 2577-90. Marcadas en rojo hay secuencias utilizadas como semilla inicial para la realineación de PFAM.

La figura 3 es un perfil de conservación de la alineación de Cas9 después del truncamiento a las posiciones de NM-Cas9.

25 La figura 4 es un perfil de conservación truncado a las posiciones en SP-Cas9.

La figura 5A es una gráfica de la conservación de aminoácidos de primer orden dentro de las proteínas Cas9. Las entropías relativas se calculan con respecto a la frecuencia promedio de aminoácidos en todos los genes en *Escherichia coli*. Las líneas verticales sobre el gráfico representan límites determinados por el algoritmo de detección de límites basado en la alineación, con líneas en negrita que representan los seis límites más significativos detectados. La figura 5B muestra las asignaciones de dominio basadas en Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic acid research* 42, 2577-2590, doi: 10.1093/nar/gkt1074 (2014). RuvCI-III son las partes que se pliegan en el dominio de nucleasa RuvC. HNH es el dominio de nucleasa de HNH. RRR es la región alfa-helicoidal rica en arginina. "Crosshatched" es la extensión de esta región basada en el tramo rico en arginina alrededor de la posición 140.

35 La figura 6A es un diseño esquemático que representa el informador transcripcional de Cas9 de NM. Se anota la ubicación del protoespaciador y el PAM específico de NM. La figura 6B es un gráfico de los datos del ensayo de represión transcripcional Cas9 de NM con diversos mutantes de NM de nucleasa nula ensayados en *E. coli*. Los datos representan el promedio \pm desviación estándar ($n = 4$). La figura 6C es un esquema que representa la estructura del dominio Cas9 de NM. Los cuadros blancos con contornos discontinuos dan la extensión de las regiones extirpadas más grandes de mutantes NM que provocan una alteración mínima en la actividad de unión al ADN.

40 La figura 7 se refiere al análisis de eliminación de Cas9 de NM y representa el diseño del informador transcripcional Cas9 de NM. Se anota la ubicación del protoespaciador y el PAM específico de NM. El ensayo de represión transcripcional Cas9 de NM con diversos mutantes NM de nucleasa nulos se ensayó en *E. coli*. Los datos representan el promedio \pm desviación estándar ($n = 5$).

45 La figura 8A representa el análisis de eliminación de Cas9 de ST1 y la validación funcional en *E. coli* y células humanas y, en particular, el diseño del informador transcripcional Cas9 de ST1. Se anota la ubicación del protoespaciador y la PAM específica de ST1. Ensayo represor transcripcional Cas9 de ST1 con mutantes de eliminación nula de nucleasa ST1 ensayados en *E. coli*. Los datos representan el promedio \pm desviación estándar ($n = 4$). La figura 8B es un esquema que describe la construcción del indicador para ensayar la activación de ST1 que contiene un promotor mínimo de CMV (CMV mínimo) en el sentido de 5' de un indicador tdTomate. Las proteínas de fusión Cas1-VP64 de la nucleasa nulas ST1 que se unen en el sentido de 5' del promotor CMV mínimo conducen a la activación transcripcional y la fluorescencia dentro de las células humanas. La figura 8C son imágenes de células transfectadas con activadores ST1 que incluyen mutantes de eliminación que se transfectaron junto con ARNgc y el indicador tdTomate y se visualizaron por microscopía de fluorescencia. La figura 8D es un gráfico que muestra la cuantificación de la activación de ST1 desde el panel C mediante citometría de flujo. Los datos representan el promedio \pm desviación estándar ($n = 3$).

50 La figura 9 se refiere al análisis de eliminación de Cas9 de TD y la validación funcional en *E. coli*. Los mutantes de eliminación de nucleasa nula TD se ensayaron utilizando un ensayo represor transcripcional. Los datos representan el promedio \pm desviación estándar ($n = 4$).

55 La figura 10A se refiere al análisis de intercambio de dominio de NM-ST1, según se determina mediante un ensayo de represión transcripcional y, en particular, el diseño de informadores transcripcionales NM y ST1 con la secuencia de la PAM específica de NM o ST1 ilustrada. La figura 10B es un esquema que muestra el esquema de Cas9 de NM y de ST1 con la ubicación de los puntos de intercambio de los aminoácidos indicados. Las figuras 10C-10F son gráficos de fluorescencia para mutantes de intercambio de dominio de NM-ST1 nucleasa

nula expresados en conjunción con ARN guía particulares a Cas9 de NM (figura 10C y figura 10D) o de ST1 (figura 10E y figura 10F), junto con informadores con secuencias PAM específicas para Cas9 de NM (figura 10C y figura 10F) o de ST1 (figura 10D y figura 10E). Los datos representan el promedio \pm desviación estándar ($n = 4$).

La figura 11A se refiere al análisis de intercambio de dominio de NM-ST1, según se determina mediante un ensayo de represión transcripcional, y en particular, el diseño del informador transcripcional ST1 con la secuencia de la PAM específica de ST1 ilustrada. La figura 11B es un esquema esquemático de Cas9 de NM y de ST1 con la ubicación de los puntos de intercambio de los aminoácidos indicados. Las figuras 11C-11D son gráficos de fluorescencia para mutantes de intercambio de dominio de NM-ST1 nucleasa nula expresados en conjunción con ARN guía particulares para Cas9 de NM (figura 11D) o de ST1 (figura 11C) junto con informadores con secuencias PAM específicas para Cas9 de ST1 (figura 11C y 11D). Los datos representan el promedio \pm desviación estándar ($n = 4$).

DESCRIPCIÓN DETALLADA

Las realizaciones de la presente invención se refieren a proteínas mutantes de unión al ADN guiadas por ARN del sistema CRISPR Tipo II. Dichos mutantes se crean mediante la eliminación de secuencias que no están bien conservadas o que son muy divergentes entre especies dentro de un género de proteínas de unión al ADN guiadas por ARN del sistema CRISPR Tipo II. Según un aspecto, se alinean las secuencias de especies dentro de una familia de proteínas de unión al ADN guiadas por ARN y se determinan secuencias de baja conservación o secuencias entre bordes de conservación. Estas secuencias se eliminan de una proteína de unión al ADN guiada por ARN particular. Entre las proteínas de unión al ADN guiadas por ARN de ejemplo se incluyen proteínas Cas9 presentes, por ejemplo, en los sistemas CRISPR Tipo II. Dichas proteínas Cas9 y sistemas CRISPR Tipo II están bien documentados en la técnica. Ver Makarova et al., Nature Reviews, Microbiology, vol. 9, junio de 2011, págs. 467-477. Las proteínas mutantes de unión al ADN descritas en la presente memoria descriptiva se pueden utilizar para realizar cortes bicatenarios en el ADN diana, cortes monocatenarios en el ADN diana o para unirse al ADN diana, de manera que se localice un grupo efector cerca del ADN diana de modo que ese grupo efector pueda interactuar con el ADN diana. Entre dichos grupos efectores se incluyen activadores, represores o modificadores epigenéticos conocidos por los expertos en la materia.

Las proteínas de unión al ADN de ejemplo que tienen actividad de nucleasa funcionan para cortar o mellar ADN bicatenario. Dicha actividad de nucleasa puede ser el resultado de que la proteína de unión al ADN tenga una o más secuencias de polipéptidos que exhiben actividad de nucleasa. Dichas proteínas de unión de ADN de ejemplo pueden tener dos dominios de nucleasa separados con cada dominio responsable de cortar o mellar una cadena particular del ADN bicatenario. Entre las secuencias de polipéptidos de ejemplo que tienen actividad de nucleasa, conocidas por los expertos en la materia, se incluyen el dominio relacionado con la nucleasa McrA-HNH y el dominio de nucleasa de tipo RuvC. Por consiguiente, las proteínas de unión al ADN de ejemplo son aquellas que en la naturaleza contienen uno o más del dominio relacionado con nucleasa McrA-NHH y el dominio de nucleasa de tipo RuvC.

Según un aspecto, una proteína de unión al ADN que tiene dos o más dominios de nucleasa se puede modificar o alterar para inactivar todos los dominios de nucleasa excepto uno. Se denomina a dicha proteína de unión al ADN modificada o alterada proteína nickasa de unión al ADN, en la medida en que la proteína corta o mella sólo ADN monocatenario o bicatenario. Cuando está guiada al ADN mediante ARN, la proteína nickasa de unión al ADN se denomina proteína nickasa de unión al ADN guiada por ARN.

Una proteína de unión al ADN de ejemplo es una proteína de unión al ADN guiada por ARN de un Sistema CRISPR Tipo II. Una proteína de unión al ADN de ejemplo es una proteína Cas9.

En *S. pyogenes*, Cas9 genera una ruptura bicatenaria de extremo romo 3 pb en el sentido de 5' del motivo adyacente protoespaciador (PAM, de *protospacer-adjacent motif*) mediante un proceso mediado por dos dominios catalíticos en la proteína: un dominio HNH que escinde la cadena complementaria del ADN y un dominio de tipo RuvC que escinde la cadena no complementaria. Ver Jinke et al., Science 337, 816-821 (2012). Se sabe que las proteínas Cas9 existen en muchos sistemas CRISPR Tipo II, incluidos los siguientes, tal como se identifica en la información complementaria de Makarova et al., Nature Reviews, Microbiology, vol. 9, junio de 2011, págs. 467-477: *Methanococcus maripaludis* C7; *Corynebacterium diphtheriae*; *Corynebacterium efficiens* YS-314; *Corynebacterium glutamicum* ATCC 13032 Kitasato; *Corynebacterium glutamicum* ATCC 13032 Bielefeld; *Corynebacterium glutamicum* R; *Corynebacterium kroppenstedtii* DSM 44385; *Mycobacterium abscessus* ATCC 19977; *Nocardia farcinica* IFM10152; *Rhodococcus erythropolis* PR4; *Rhodococcus jostii* RHA1; *Rhodococcus opacus* B4 uid36573; *Acidothermus cellulolyticus* 11B; *Arthrobacter chlorophenolicus* A6; *Kribbella flavida* DSM 17836 uid43465; *Thermomonospora curvata* DSM 43183; *Bifidobacterium dentium* Bd1; *Bifidobacterium longum* DJO10A; *Slackia heliotrinireducens* DSM 20476; *Persephonella marina* EX HI; *Bacteroides fragilis* NCTC 9434; *Capnocytophaga ochracea* DSM 7271; *Flavobacterium psychrophilum* JIP02 86; *Akkermansia muciniphila* ATCC BAA 835; *Roseiflexus castenholzii* DSM 13941; *Roseiflexus* RS1; *Synechocystis* PCC6803; *Elusimicrobium minutum* Pei191; bacteria filotipo Rs D17 de terminación no cultivada grupo 1; *Fibrobacter succinogenes* S85; *Bacillus cereus* ATCC 10987; *Listeria innocua*, *Lactobacillus casei*; *Lactobacillus rhamnosus* GG; *Lactobacillus salivarius* UCC118;

Streptococcus agalactiae A909; Streptococcus agalactiae NEM316; Streptococcus agalactiae 2603; Streptococcus dysgalactiae equisimilis GGS 124; Streptococcus equi zooepidemicus MGCS10565; Streptococcus gallolyticus UCN34 uid46061; Streptococcus gordonii Challis subst CHI; Streptococcus mutans NN2025 uid46353; Streptococcus mutans; Streptococcus pyogenes M1 GAS; Streptococcus pyogenes MGAS5005; Streptococcus pyogenes MGAS2096; Streptococcus pyogenes MGAS9429; Streptococcus pyogenes MGAS10270; Streptococcus pyogenes MGAS6180; Streptococcus pyogenes MGAS315; Streptococcus pyogenes SSI-1; Streptococcus pyogenes MGAS10750; Streptococcus pyogenes NZ131; Streptococcus thermophiles CNRZ1066; Streptococcus thermophiles LMD-9; Streptococcus thermophiles LMG 18311; Clostridium botulinum A3 Loch Maree; Clostridium botulinum B Eklund 17B; Clostridium botulinum Ba4 657; Clostridium botulinum F Langeland; Clostridium cellulolyticum H10; Finegoldia magna ATCC 29328; Eubacterium rectale ATCC 33656; Mycoplasma gallisepticum; Mycoplasma mobile 163K; Mycoplasma penetrans; Mycoplasma synoviae 53; Streptobacillus moniliformis DSM 12112; Bradyrhizobium BTAi1; Nitrobacter hamburgensis X14; Rhodopseudomonas palustris BisB18; Rhodopseudomonas palustris BisB5; Parvibaculum lavamentivorans DS-1; Dinoroseobacter shibae DFL 12; Gluconacetobacter diazotrophicus Pal 5 FAPERJ; Gluconacetobacter diazotrophicus Pal 5 JGI; Azospirillum B510 uid46085; Rhodospirillum rubrum ATCC 11170; Diaphorobacter TPSY uid29975; Verminephrobacter eiseniae EF01-2; Neisseria meningitidis 053442; Neisseria meningitidis alpha14; Neisseria meningitidis Z2491; Desulfovibrio salexigens DSM 2638; Campylobacter jejuni doylei 26997; Campylobacter jejuni 81116; Campylobacter jejuni; Campylobacter lari RM2100; Helicobacter hepaticus; Wolinella succinogenes; Tolumonas auensis DSM 9187; Pseudoalteromonas atlantica T6c; Shewanella pealeana ATCC 700345; Legionella pneumophila Paris; Actinobacillus succinogenes 130Z; Pasteurella multocida; Francisella tularensis novicida U112; Francisella tularensis holarctica; Francisella tularensis FSC 198; Francisella tularensis tularensis; Francisella tularensis WY96-3418; y Treponema denticola ATCC 35405. Por consiguiente, los aspectos de la presente divulgación se refieren a un mutante de una proteína Cas9 presente en un sistema CRISPR Tipo II, tal como cualquiera de las especies identificadas anteriormente. Una proteína Cas9 de ejemplo es la que se encuentra en Neisseria meningitidis, tal como Neisseria meningitidis 053442; Neisseria meningitidis alpha14; Neisseria meningitidis Z2491.

Entre las células, según la presente divulgación, se incluyen cualquier célula en la que los ácidos nucleicos extraños se pueden introducir y expresar tal como se describe en la presente memoria descriptiva. Debe entenderse que los conceptos básicos de la presente divulgación descritos en la presente memoria descriptiva no están limitados por el tipo de célula. Entre las células, según la presente divulgación, se incluyen células eucariotas, células procariotas, células animales, células vegetales, células fúngicas, células de arquea, células eubacterianas y similares. Entre las células se incluyen células eucariotas, tales como células de levadura, células vegetales y células animales. Entre las células particulares se incluyen células de mamífero. Entre las células particulares se incluyen células madre, tales como células madre pluripotentes, tales como células madre pluripotentes inducidas por humanos.

Entre los ácidos nucleicos diana se incluyen cualquier secuencia de ácido nucleico a la que una nucleasa proteica de unión al ADN guiada por ARN mutante pueda ser útil para cortar o mellar. Entre los ácidos nucleicos diana se incluyen genes. Para los propósitos de la presente descripción, el ADN, tal como el ADN bicatenario, puede incluir el ácido nucleico diana y un complejo de colocación se puede unir o colocalizar de otro modo con el ADN en el ácido nucleico diana o adyacente o cerca del mismo y de una manera en la que el complejo de colocación pueda tener un efecto deseado sobre el ácido nucleico diana. Entre estos ácidos nucleicos diana se pueden incluir ácidos nucleicos endógenos (o naturales) y ácidos nucleicos exógenos (o extraños). Un experto en la materia de la presente divulgación podrá identificar o diseñar fácilmente ARN guía y proteínas Cas9 mutantes que se colocalizan en un ADN que incluye un ácido nucleico diana. Entre los ADN se incluyen ADN genómico, ADN mitocondrial, ADN viral, un elemento conjugable o ADN exógeno.

Los ácidos nucleicos extraños (es decir, aquellos que no forman parte de la composición de ácido nucleico natural de una célula) se pueden introducir en una célula utilizando cualquier procedimiento conocido por los expertos en la materia para dicha introducción. Entre estos procedimientos se incluyen transfección, transducción, transducción viral, microinyección, lipofección, nucleofección, bombardeo de nanopartículas, transformación, conjugación y similares. Un experto en la materia comprenderá y adaptará fácilmente dichos procedimientos utilizando fuentes de bibliografía fácilmente identificables.

Según un aspecto, el material genético requerido para codificar una proteína Cas9 se reduce eliminando partes de la proteína Cas9 que no están bien conservadas o divergen de otro modo dentro de las especies dentro de una familia de Cas9 o están entre los bordes de conservación dentro de las especies dentro de una familia de Cas9. Al reducir el tamaño del ácido nucleico requerido para codificar una Cas9 funcional, se pueden incluir ácidos nucleicos adicionales con un vector diseñado para administrar la Cas9, tales como los ácidos nucleicos que codifican el ARN guía o elementos reguladores o dominios efectores. Si se utiliza el miembro de la familia Cas9 más pequeño, se requerirán, aproximadamente, 4.500 kilobases de ADN para codificar los elementos genéticos necesarios (proteína Cas9 y ARNg) para localizar adecuadamente Cas9 en el locus genómico deseado. A, aproximadamente, 4.500 pares de bases, Cas9 está cerca del límite de tamaño para el empaquetamiento dentro del vector viral basado en AAV (que es un vector viral aprobado por la regulación en Europa). Además, algunos de los primeros dominios efectores transcripcionales y epigenéticos que se fusionarán con proteínas de unión al ADN programables son mayores de 2.000 pares de bases y, de este modo, están fuera del límite de empaquetamiento para vectores AAV y

acercándose al límite de los sistemas de empaquetamiento lentivirales (aproximadamente, 8.000 pares de bases) una vez fusionados a Cas9.

5 Los siguientes ejemplos se exponen como representativos de la presente divulgación. Estos ejemplos no deben interpretarse como limitantes del alcance de la presente divulgación, ya que estas y otras realizaciones equivalentes serán evidentes a la vista de la presente divulgación, las figuras y las reivindicaciones adjuntas.

EJEMPLO I

10 Cas9 mutante

Para superar los problemas del gran tamaño del gen que codifica Cas9, se lleva a cabo una eliminación dirigida de diversas regiones dentro de uno de los miembros de la familia Cas9 caracterizados más pequeños, NM-Cas9 (Cas9 de *Neisseria meningitidis*). NM-Cas9 tiene un tamaño de 3.249 pb. Se determinan los requisitos para el direccionamiento al genoma y a los residuos involucrados en la actividad de nucleasa. Para generar versiones de NM-Cas9 que son de menor tamaño, se generó una alineación de proteínas Cas9 y se identificaron tramos contiguos de baja conservación o tramos entre los bordes de conservación para su eliminación. Se identificaron diversas regiones de interés y se eliminaron selectivamente de NM-Cas9, cuya función se evaluó posteriormente mediante la utilización de un ensayo represor de Cas9. En el ensayo, se utilizó una variante de NM-Cas9 que carece de actividad de nucleasa pero que puede ser dirigida a la región 5' de un gen informador. Si NM-Cas9 puede unirse al gen informador, reprimirá la transcripción y, en el caso de un informador fluorescente, las células aparecerán no fluorescentes.

25 Alineación de Cas9 y predicción de eliminación: Se obtuvieron secuencias completas de homólogos de Cas9 de la base de datos PFAM o de una búsqueda en la base de datos, tal como jackHMMER (R. D. Finn, J. Clements, S. R. Eddy, *Nucleic Acids Research* (2011) Edición del servidor web 39: W29-W37). En el caso de que la colección de secuencias no esté alineada, se crea una alineación utilizando un algoritmo de alineación tal como CLUSTALW (Sievers F, Wilm A, Dineen D. G., Gibson T. J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J. D., Higgins D. G. (2011), o equivalente. La alineación se cortó computacionalmente a las posiciones de la secuencia de interés y se acondicionó para disminuir el sesgo de alineación (por ejemplo, se eliminaron secuencias con una identidad por pares superior al 95 %). La conservación se calcula como la entropía o entropía relativa de las frecuencias de aminoácidos por posición, teniendo en cuenta la cantidad de aminoácidos y espacios en esa posición. Las eliminaciones se dirigen hacia regiones de baja conservación o entre bordes de conservación. En iteraciones con verificación experimental, las eliminaciones son expandidas o desplazadas.

35 Construcción y caracterización de las eliminaciones: Se generaron anteriormente plásmidos bacterianos que expresan NM-Cas9 nucleasa nula, tal como se describe en Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J. y Church, G. M. (2013) *Nat Methods* 10, 1116-1121. Para crear eliminaciones específicas dentro de NM-Cas9, se utilizó el ensamblaje Gibson, tal como se describe en Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3º y Smith, H. O. (2009) *Nat Methods* 6, 343-345. Se compraron y utilizaron en reacciones de PCR cebadores que contenían solapamiento complementario y que estaban diseñados para eliminar regiones diana dentro de NM-Cas9 junto con la inserción de un enlazador SGGGS. Los fragmentos de PCR se purificaron en gel, se ensamblaron in vitro utilizando el ensamblaje Gibson y se transformaron en *E. coli*. Se verificó la secuencia de los clones y se ensayaron utilizando una forma modificada de plásmidos de NM-Cas9 informadores anteriormente generados (ver (2013) *Nat Methods* 10, 1116-1121) en el que un solo plásmido (en lugar de dos) contiene el espaciador NM-Cas9, protoespaciador dirigido e informador de YFP para la actividad de NM-Cas9. Brevemente, en este ensayo, las células se transforman conjuntamente con variantes sintéticas de NM-Cas9 y el plásmido informador. Las células doblemente transformadas se hacen crecer a 37 °C, y la cantidad de fluorescencia YFP se mide utilizando un lector de placas de fluorescencia y se compara con las células que se transforman con un plásmido de control con NM-Cas9 nucleasa nula de tipo salvaje y el plásmido informador.

50 Las siguientes secuencias son para los dos mutantes de eliminación única NM-Cas9 más grandes que retuvieron niveles de actividad cercanos al tipo salvaje, según lo determinado por el ensayo de informador YFP. La secuencia del enlazador SGGGS que reemplaza las regiones eliminadas dentro de NM-Cas9 se muestra en MAYÚSCULAS.

55 NM-Cas9-Δ255-449

atggccgcttcaagcccaacccatcaactacatcctgggcctggccatcggcgcgcagcgtgggctgggccatggtggagatcgacgaggacgagaacccatctgcctgatcgacctgggtgtgcgcgtgttcgagcgcgctgaggtgcccagactggtgacagtctggctatggct

ES 2 774 677 T3

cgccggcttgcctctgttcggcgcctfactcggcggcgctcaccgccttctgcgcgctcggcgcctgctgaagcgcgaggggtgtgctg
caggctcccgacttcgacgagaacggcctgatcaagagcctgcccaactccttggcagctgcgcgctcggcctctggaccgcaagctga
ctcctctggagtggagcggcgtgctgctgcacctgatcaagcaccgggctacctgagccagcgaagaacgagggcgagaccgccgaca
aggagctgggtgctctgctgaaggcggtggccgacaacgcccacgccctgcagactggtgacttccgactcctgctgagctggccctgaa
caagttcgagaaggagagcggccacatccgaaccagcggcgactacagccacacctcagccgcaaggacctgcaggccgagctga
tcctgctgttcgagaagcagaaggagttcggcaacccccacgtgagcggcggcctgaaggaggcatcgagaccctgctgatgaccagc
gccccgccctgagcggcgacggcgtgcagaagatgTCCGGCGGCGGTTTCGggcgaccactacggcaagaagaacaccgag
gagaagatctacctgctctatccccggcgacgagatccgaacccctggtgctgctgcgcgcctgagccaggcccgcaaggtgatcaacg
gctggtgcccgtctacggcagccccggcgcacccatccacatcgagaccggcgagggggcaagagctcaaggaccgcaaggagatc
gagaagcggcaggaggagaaccgcaaggaccgcgagaaggccggcccaagttccgcgagtacttccccaaacttcgtgggcgagcccaa
gagcaaggacatctgaagctgcgcctgtacgagcagcagcagcgaagtgcctgtacagcggcaaggagatcaacctgggcccgcctgaa
cgagaagggtctacgtggagatcggcgtgcccctgcccctcagccgcacctgggacgacagcttcaacaacaaggctgctggtgctgggcagc
gaggctcagaacaagggaaccagacccccctacgagtacttcaacggcaaggacaacagccgcgagtgagcaggagttcaaggcccgcgt
ggagaccagccgcttccccgcagcaagaagcagcgcacctgctgagcaagttcagcaggagcggcttcaaggagcgcacctgaacga
cacccgctacgtgaaccgcttctgtgccagttcgtggccgaccgcatgcgctgaccggcaagggaagaagcgcgtgttcgccagcaac
ggccagatcaccaacctgctgcccggcttctggggcctgcgcaaggtgcgcggcggagaacgaccgccaccacgccctggacggcgtggt
ggtggcctgcagcaccgtggcctgagcagaagatcacccgcttctgctgcgctacaaggagatgaacgcttcgacggtaaaacctcagc
aaggagaccggcgaggtgctgcaccagaagaccacttccccagccctgggagttcttcgccaggaggtgatgatccgcgtgttcggca
agccccagcggcaagcccagttcaggaggccgacacccccgagaagctgcgcacctgctggccgagaagctgagcagccgccctga
ggcctgacaggtacgtgactcctctgttcgtgagccgcgcccccaaccgcaagatgagcggtcagggtcacatggagaccgtgaagagc
gccaagcgcctggacgagggcgtgagcgtgctgcgcgtgcccctgaccagctgaagctgaaggacctggagaagatggtgaaccgga
gcgcgagcccaagctgtacgaggccctgaaggcccgcctggaggcccaaggacgaccccgcaaggccttcgcccagcccttctaca
agtacgacaaggccggcaaccgacccagcaggtgaaggccgtgcgcgtggagcaggtcagaagaccggcgtgtgggtgcgcaacca
caacggcatcgccgacaacgccaccatggtgcgctggacgtgttcgagaaggcgacaagtaactacctggtgcccatctacagctggcag
gtggccaaggcatcctgcccgaccgcccgtggtgcagggcaaggacgaggaggactggcagctgatcgacgacagcttcaactcaag
ttcagcctgcacccaacgacctggtgaggtgatccaagaaggcccgcctgcttggctacttcgccagctgccaccggcaccggca
acatcaacatccgcatccacgacctggaccacaagatcggcaagaacggcatcctggaggcatcggcgtgaagaccgccctgagctcca
gaagttaccagatcgagagctgggcaaggagatccgccctgcccctgaagaagcgcctcctgtgcgctaa [SEQ ID NO:1]

NM-Cas9-Δ567-654

atggccgcttcaagcccaaccccatcaactacatcctgggctggccatcggcatcggcagcgtgggctgggcatggtggagatcgacg
aggacgagaaccccatctgcctgatcgacctgggtgtgcgcgtgttcgagcgcgctgaggtgcccaagactggtgacagtctggctatggct
cgccggcttgcctcgtctgttcggcgccttactcgccggcgcgctcaccgccttctgcgcgctcgccgctgctgaagcgcgagggtgtgctg
caggctgccgacttcgacgagaacggcctgatcaagagcctgcccaacactccttggcagctgcgcgctgccgctctggaccgcaagctga
ctcctctggagtggagcggcgtgctgctgcacctgatcaagcaccggctacctgagccagcgaagaacgagggcgagaccgccgaca
aggagctgggtgctctgctgaagggcgtggccgacaacgcccacgacctgcagactggtgacttccgactcctgctgagctggccctgaa
caagttcgagaaggagagcggccacatccgcaaccagcgcggcgactacagccacaccttcagccgcaaggacctgcaggccgagctga
tctctgttfcgagaagcagaaggagttcggcaacccccagctgagcggcgccctgaaggaggccatcgagacctgctgatgaccagc
gccccgacctgagcggcgacggcgtgcagaagatgctgggcaactgcaccttcgagccagccgagcccaaggccgcaagaacacctac
accgccgagcgttcatctggctgaccaagctgaacaacctgcgcaccttggagcagggcagcgcgccccctgaccgacaccgagcgc
gccacctgatggagagccctaccgcaagagcaagctgacctacgcccagggccgcaagctgctgggtctggaggacaccgcttcttca
aggcctgcgctacggcaaggacaacgccgagggcagcacctgatggagatgaaggcctaccacgccatcagccgcgcccctggagaag
gaggccctgaaggacaagaagatccttgaacctgagccccgagctgcaggacgagatcggcaccgcttcagcctgttcaagaccgac
gaggacatcaccggccgctgaaggaccgcatccagcccagatcctggaggccctgctgaagcacatcagcttcgacaagttcgtgcaga
tcagcctgaaggccctgcgcccgcacgtgccccctgatggagcagggcaagcgtacgacgagcctgcgcccagatctacggcgaccact
acggcaagaagaacaccgaggagaagatctacctgcctcctatccccgcgacgagatccgcaacccccgtggtgctgcgcccctgagcc
aggcccgcaaggtgatcaacggcgtggtgcgcccgtacggcagcccccccgcacccacatcgagaccgcccgcgaggtgggcaagag
cttcaaggaccgcaaggagatcgagaagcggcaggaggagaaccgcaaggaccgcgagaaggccgcccgaagtccgcgagcttcc
ccaacttctggggagcccaagagcaaggacatcctgaagctgcgctgtacgagcagcagcacggcaagtgcTCCGGCGGGC
GTTCGcagaagttcgacgaggacggcttcaaggagcgaacctgaacgacaccgctacgtgaaccgcttctgtgccagttctgtggcc
gaccgcatgcgctgaccggcaaggcaagaagcgcgtgttcgccagcaacggccagatcaccaacctgctgcgccccttctggggcctg
cgcaaggtgcgcccgagaacgaccgccaccacgcccctggagcggcgtggtggtggcctgcagcaccgtggccatgcagcagaagatcac
ccgcttcgtgcgctacaagagatgaacgccttcgacggtaaaaccatcgacaaggagaccggcgaggtgctgcaccagaagaccacttc
ccccagccctgggagttcttcgccaggaggtgatgatccgctgttccgcaagcccagcggcaagcccagttcgaggaggccgacacc
cccgagaagctgcgaccctgctggccgagaagctgagcagccgcccctgaggccgtgcacgagtacgtgactcctctgttcgtgagccgg
cccccaaccgcaagatgagcggctcagggtacatggagaccgtgaagagcggcaagcgcctggacgagggcgtgagcgtgctgcgcgtg
ccctgaccagctgaagctgaaggacctggagaagatggtgaaccgcgagcgcgagcccaagctgtacgaggccctgaaggcccgcct
ggaggcccacaaggacgaccccgcaaggccttcgcccagcccttctacaagtacgacaaggccggcaaccgcaccagcaggtgaagg
ccgtgcgctggagcaggtgcagaagaccggcgtgtgggtgcgaaccacaacggcatcggcacaacgccaccatggtgcgctggac
gttfcgagaaggcgacaagfactacctggtgcccactacagctggcaggtggcaaggccatcctgcccgaccgcgcccgtgtgaggg
gcaaggacgaggaggactggcagctgatcgacgacagcttcaacttcaagttcagcctgcaccccaacgacctggtggaggtgatcacia
gaaggcccgcatttggctacttcgccagctgccaccgcggcaccggcaacatcaacatccgcatccacgacctggaccacaagatcggc
aagaacggcactcctggagggcatcggcgtgaagaccgcccctgagcttccagaagfaccagatcgacgagctgggcaaggagatccgccc
tgcccctgaagaagcgcctcctgtgctgtaa [SEQ ID NO:2]

Plásmido informador YFP

agctctcgaacccagagtcgctcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcaatcgggagcggcgataccgta
 aagcacgaggaagcggcagccattcggcccaagctctcagcaatatacagggtagccaacgctatgctctgatagcggcggccacac
 ccagccggccacagtcgatgaatccagaaaagcggccatttccaccatgatattcggcaagcaggeatcgccatgggtcacgacgagatcc
 tcggcgtcgggcagtcgctgagcctggcgaacagttcggctggcgcgagccctgatgctctcgtccagatcctctgatcgacaaga
 ccggcttccatccgagtacgtgctgctgatgcgatgttctggtggcgaatgggcaggtagccggatcaagcgtatgcagccggcgc
 attgcatcagccatgatggatacttctcggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgccaatagcagccagtc
 ctccccgttcagtgacaacgtcgagcacagctgcgcaaggaaccccgtcgtggccagccacgatagccgctgcctcgtctgcagttc
 attcagggcaccggacaggtcggcttgacaaaaagaaccggggcggcctcgtgacagccggaacacggcggtcagagcagccga
 ttgctgtgtgcccagtcatagcccgaatagcctctccaccaagcggccggagaacctgctgcaatccatctgttcaatcatgcgaaacgat
 cctcatcctgtctctgatcagatcttgatccctgcgcatcagatcctggcggcaagaaagccatccagttactttgcagggcttccaaact
 taccagagggcggccagctggcaattccgacgtetaagaaaccattattatcatgacattaacctataaaaaataggcgatcacgagggccttt
 cgtcttcacctcaggggacaatgaaaacgttagtcatggcgcgcttgacggctagctcagtcctaggtacagtgctagcttaatgctcgac
 atagcagaactttaaaagtattcggcatgtgtgagctcccttctcatttcgagtgctacaatccggcgtatggccacgtagagcatacggaa
 aaaaaagtcaaaagcctccgaccggaggtcggccttacttctagcagagttgtagaaacgaaaaaggccatccgctcaggatggccttctg
 cttatttgatgctggcagttatggcggcgtcctgcccggaccctccgggcccgttgcctcgaacgttcaaatccgctcccggcgatttgt
 cctactcaggagagcgttaccgacaaacaacagataaaacgaaaggcccagctttcgaactgagcctttcgtttatgtatgctggcagttcc
 ctactctcgatgggagacccacactaccatggcgttacggcgttctacttctgagttcggcatggggcaggtgggaccaccgcgctac
 tggccaggcaaatctgtttatcagaccgcttctgctgttctgatttaactgtatcagggtgaaaatcttctcatccgcaaaaacagccccgt
 agaaaaaggacgttgatcggcacgtaagaggtccacgataaatatctaacaccgtcgtgttgactattttacctctggcggtgataatggtg
 catgtactagaattcttaactttaagaaggagatatacatatgaatccccatggcgaatacttttaagctcgtaaaggagaagaacttttactg
 gagtgtcccaattctgttgaattagatggatgtaatgggcacaaatftctgctagtgagaggggtgaagggtgatcaacatacggaaaact
 taccctaaatftatgactactggaaaactacctgttccatggccaacactgtcactactttcggttatggtctaaaatgctttgctagataccca
 gatcatatgaaacggcatgacttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatattttcaaatgatcgggaactacaa
 gacacgtgctgaagtcagttgaaggtgataccctgttaatagaatcgagttaaaaggtattgattttaaagaagatggaacattctggacac
 aaattggaatacaactataactcacacaatgtatacatatggcagacaaaagaatggaatcaaaagttactcaaaattagacacaacatt
 gaagatggaagcgttcaactagcagaccattatcaacaaaactccaattggcgtatggccctgtcctttaccagacaaccattacctgtctat
 caatctgcccttcgaaagatcccaacgaaaagagagaccacatggtcctcttgagttgtaacagctgctgggattacacatggcatggatga
 actatacaaaataagcttaaccgaagcgtttgatgtgatatcctttgctcggccgcaactagaggcatcaataaaacgaaaggctcagtcg
 aaagactggcctttcgtttatctgtttgttgctggtaacgctctcctgagtaggacaaatccggcgcctagacctagggtacgggtttgctg
 cccgcaaacgggctgttctggtgttgctagttgttatcagaatcgagatccggctcaggtttccggctgaaagcgtatttctccagaattg
 ccatgatttttccccacgggagggcgtcactggctcccgttctcggcagctttgattcgataagcagcatcgctgttccaggtgtctatgtg
 gactgttgagctgtaacaagttgtctcaggtgtcaattcatgttctagttgctttgtttactggttcacctgttctattaggtgttacatgctgttcatc

tggtacattgctgatctgtcatggtgaacagcttfaaatgcacaaaaactcgtaaaagctctgatgtatctatctttttfacaccgtttcatctgtgc
 atatggacagtttcccttgatataacgggtgaacagttgttctactttgtttgtagtcttgatgcttctactgatagatacaagagccataagaacc
 tcagatcctccgtatfttagccagatgttctctagtgtggttcgtgttttgcgtgagccatgagaacgaaccattgagatcatgcttactttgcatgt
 cactcaaaaatfttgcctcaaaactggtagctgaatfttgcagftaaagcctcgtgtagtgttttcttagtccgttacgtaggttaggaatctgatgt
 aatggttgtggtatfttgcaccattcattttatctggttgttctcaagttcgggttacgagatccattgtctatctagtccaactggaaaatcaacgta
 tcagtcggggcgccctcgttatcaaccaccaatftcatattgctgtaagtgtttaaacttttacttattggttcaaaaccattggftaagcctftttaa
 ctcatggtagttatfttcaagcattaacatgaactfaaattcatcaaggctaactctatatttgccttggtagtttctfttggtagttcttftaataacca
 ctcataaatcctcatagagtatttgtttcaaaagactfaacatgtccagattatattttatgaatftttaaactggaaaagataaggcaatatctctca
 ctaaaaactaattctaattttcgttggaaactggcatagttgtccactggaaaatcctaaagccttfaaccaaaggattcctgattccacagttc
 tcgctcatcagctctctggttgcctttagtaatacaccataagcattttccctactgatgttcatcatctgagcgtattggttataaggaacgataccgt
 ccgttcttctttaggtggtttcaatcgtggggtgagtagtgccacacagcataaaattagcttggttcatgctccgftaagtcatagcgactaat
 cgctagttcattgcttggaaaacaactaattcagacatacatctcaattggtctaggtgatttfaatcactataccaattgagatgggctagtcattg
 ataactactagtccttttccctttaggttgggtatctgtaaattctgctagaccttggctggaaaactgtaaattctgctagacctctgtaaattccgc
 tagaccttggtagttttttgtttatfttaagtggttataatftatagaataaagaaagaataaaaaagataaaaaagaatagatcccagccctgtgt
 ataactactactttagtcagttccgcagttattacaaaaggatgctgcgaaacgctgtttgctcctctacaaaacagaccttaaaaccctaaaggctt
 aagtagcaccctcgcaagctcgggcaaatcgtgaatattcctttgtctccgacctcaggcacctgagtcgctgtcttttcgtgacattcagttc
 gctgcgctcacggctctggcagtgatgggggtaaatggcaactacagggcctttatggattcatgcaaggaaactaccataatacaagaaa
 agcccgtcacgggcttctcagggcgtttatggcgggtctgctatgtggtgctatctgacttttggctgttcagcagttcctgccctctgattttccag
 tctgaccacttcggattatcccgtgacaggtcattcagactggctaatgcaccagtaaggcagcgggtatcatcaacaggcttacccttactg
 tccttagtcttggattctcaccataaaaaacgcccggcggaaccgagcgttctgaacaatccagatggagttctgaggtcattactggatc
 tataacaggagtccaageg [SEQ ID NO:3].

5 Según los procedimientos descritos en la presente memoria descriptiva, se han identificado diversas eliminaciones dentro de NM-Cas9, la mayor, NM-Cas9-Δ255-449, elimina 595 pares de bases y muestra solo una disminución del 16 % en la actividad medida por el ensayo del informador. Según ciertos aspectos, se dan a conocer proteínas Cas9 mutantes que tienen 1000 pares de bases menos o 900 pares de bases menos, en comparación con la Cas9 de tipo salvaje, tal como NM-Cas9 y retienen niveles de actividad cercanos a la de tipo salvaje.

10 EJEMPLO II

Direccionamiento a dominios de nucleasa de Cas9 para su eliminación

15 Junto con el direccionamiento a regiones de baja conservación de secuencia o entre bordes de conservación de secuencia, en el caso de que se desee una nickasa o un alelo nulo de la nucleasa de Cas9, se puede dirigir a los dominios de nucleasa de Cas9 junto con sus nucleótidos circundantes para eliminación. Utilizando este enfoque, se produjo un alelo funcional NM-Cas9 que carece del motivo HNH y los nucleótidos circundantes, NM-Cas9-Δ567-654, que retuvo una capacidad para unirse al ADN cercana a la del tipo salvaje, según lo determinado por el ensayo de informador YFP.

20 EJEMPLO III

Procedimientos para construir una biblioteca de eliminaciones de NM-Cas9 no sesgada

25 Además de adoptar un enfoque dirigido para generar eliminaciones de Cas9, los aspectos de la presente divulgación incluyen un enfoque de alto rendimiento para la creación de eliminaciones aleatorias y la detección de mutantes funcionales. Según un procedimiento de ejemplo, el ADN plasmídico que contiene el alelo Cas9 deseado se puede cortar utilizando una nucleasa promiscua, sonicando, pipeteando repetidamente la muestra u mediante otros medios

químicos, enzimáticos o ambientales. Una vez fragmentado, el ADN plasmídico se puede tratar con exonucleasas para eliminar los nucleótidos del gen Cas9. Después del tratamiento con exonucleasas, los extremos fragmentados se vuelven romos con enzimas tales como la nucleasa de frijol mungo o la polimerasa Klenow y se ligan para regenerar un plásmido Cas9 que contiene una eliminación aleatoria. Para insertar un dominio exógeno, tal como un motivo enlazador o efector dentro de la parte eliminada de Cas9, dichos dominios se pueden ligar al ADN fragmentado de extremos romos, y la posterior circularización del plásmido producirá una secuencia de codificación de Cas9 en la que se ha insertado el dominio exógeno dentro de la parte eliminada de Cas9. La biblioteca de moléculas circularizadas se transformará en *E. coli* y se extraerá el ADN plasmídico.

En este punto, la biblioteca se puede transformar en células que contienen un ensayo informador de actividad de Cas9 y se pueden identificar los miembros de la biblioteca que mantienen la actividad funcional. Alternativamente, para reducir el tamaño de la biblioteca a cribar, la secuencia de codificación para Cas9 de la biblioteca recién generada se puede aislar mediante digestión o PCR y se pueden seleccionar los fragmentos por tamaño para que sean más cortos que el gen Cas9 de tipo salvaje inicial. Estos miembros más pequeños se pueden ligar nuevamente al vector de partida y transformarse en células que contengan el informador de actividad de Cas9.

Además de la fragmentación de plásmidos, se puede generar una biblioteca de oligonucleótidos que tienen homología 3' con el gen Cas9 pero contienen homología 5' entre sí, en los que el extremo 3' de cada oligonucleótido se une a un tramo diferente de, aproximadamente, 30 pares de bases dentro de Cas9. Estos oligonucleótidos cubren las cadenas sentido y antisentido de la secuencia de codificación de Cas9. Posteriormente, se puede realizar la PCR con estos oligonucleótidos para generar una serie de fragmentos Cas9 con cada producto a partir de una reacción de PCR sentido dada, que tienen complementariedad con todos los demás productos de PCR antisentido y viceversa. Estos fragmentos se pueden hibridar conjuntamente utilizando procedimientos tales como el ensamblaje de Gibson o la PCR de extensión de solapamiento, seguido de la ligadura en un esqueleto de vector y transformarse en células, generando una biblioteca de variantes de Cas9 con extensiones aleatorias del gen Cas9 eliminadas. Para enlazadores más largos o para insertar un dominio efector dentro de las regiones eliminadas, los oligonucleótidos en sus extremos 5' deben contener complementariedad hacia el enlazador o dominio efector más largo y se debe incluir este dominio en la reacción de ensamblaje de Gibson o durante la PCR de extensión de solapamiento. Una vez que se ha generado una biblioteca, las variantes funcionales se pueden identificar utilizando un ensayo informador, tal como el sistema informador YFP descrito en la presente memoria descriptiva.

Ejemplo IV

Construcción del vector

Los plásmidos Cas9 de nucleasa nula fueron ST1 (Addgene # 48659) o se construyeron a partir de los plásmidos NM y TD (Addgene # 48646 y 48648, respectivamente) mediante la introducción de las siguientes mutaciones puntuales (NM: D16A D587A H588A N611A y TD: D13A D878A H879A N902A). Se generaron eliminaciones de Cas9 utilizando el ensamblaje Gibson. Las eliminaciones internas, cuando se realizaron, se unieron mediante un enlazador Ser-Gly-Gly-Gly-Ser de 5 aminoácidos, excepto NM Δ 566-620 que carece de un enlazador entre los fragmentos unidos. El dominio N-terminal intercambia los residuos fusionados 1-117 de ST1 sobre los residuos 118-1082 de NM. El dominio C-terminal intercambia los residuos fusionados 1-727 de NM sobre los residuos 743-1121 de ST1.

Ejemplo V

Construcciones de informador bacteriano

Las construcciones de informador utilizadas para el análisis de los mutantes de eliminación son similares a las publicadas anteriormente, excepto por que combinan el elemento espaciador y el indicador de YFP en un único esqueleto de plásmido SC101-kanR. Las construcciones de informador para el análisis de intercambio de dominio son idénticas a las utilizadas anteriormente. Ver Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* 10, 1116-1121, doi: 10.1038/nmeth.2681 (2013).

Ejemplo VI

Construcciones de informador de mamíferos

La construcción M-ST1n-VP64, el plásmido de ARN guía de ST1 y el informador transcripcional de mamífero específico de ST1 se habían publicado anteriormente en Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* 10, 1116-1121, doi: 10.1038/nmeth.2681 (2013) (Addgene # 48675, 48672 y 48678, respectivamente). Se produjeron mutantes por eliminación como en las construcciones bacterianas.

Ejemplo VII

Ensayos de represión

5 Los ensayos de represión de Cas9 se realizaron transformando conjuntamente células beta-NEB 10 (New England BioLabs) con la construcción espaciador/informador apropiada y el vector Cas9 a investigar. Las colonias de las transformaciones se recogieron y se hicieron crecer a 37 °C con agitación continua en placas de 96 pocillos. Las placas se leyeron al día siguiente utilizando un lector de microplacas Synergy Neo (BioTek), midiendo la fluorescencia a 495-528 nm y la absorbancia a 600 nm. Para los experimentos de intercambio, se ensayaron dos combinaciones diferentes de espaciador/protoespaciador anteriormente publicadas (A y B) (ver Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nature Methods 10, 1116-1121, doi: 10.1038/nmeth.2681 (2013)). Para todos los demás experimentos, solo se examinó la combinación B espaciador/protoespaciador.

15 Ejemplo VIII

Cultivo celular y transfecciones

20 Se mantuvieron células HEK 293T en medio Eagle modificado por Dulbecco (Invitrogen) con alta glucosa, complementada con FBS al 10 % (Invitrogen) y penicilina/estreptomina (Invitrogen). Las células se mantuvieron a 37 °C y el 5 % de CO₂ en una incubadora humidificada. Las células se transfectaron en placas de 24 pocillos sembradas con 50.000 células por pocillo. Se administraron 400 ng de activador Cas9, 100 ng de ARNg y 60 ng de plásmido informador a cada pocillo utilizando 2,5 µl de Lipofectamine 2000. Las células se cultivaron 36-48 horas adicionales antes de ser ensayadas utilizando inmunofluorescencia o FACS.

25 Ejemplo IX

Alineaciones de secuencia múltiple y filtro de borde

30 Se realizaron alineaciones de secuencia múltiple realineando secuencias de Cas9 en la base de datos PFAM (PF13395, 798 secuencias) en MUSCLE, y condicionando la alineación para la diversidad y secuencias completas utilizando un script MATLAB. Este procedimiento descrito con más detalle a continuación produjo 217 secuencias.

35 Para llegar a la realineación final de las secuencias de Cas9 en la base de datos PFAM (PF13395, 798 secuencias), se llevaron a cabo las siguientes etapas (todos los códigos de programa en MATLAB). Se obtuvo la alineación de Fonfara I, et al., (2014) Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems, Nucleic Acids Res. 42, 2577-90 y solo se incluyeron secuencias del grupo IIA e IIC, y desde este último solo hasta la rama que separa IIA e IIC (marcada en un recuadro rojo en la figura 2) que incluyó 49 secuencias totales.

40 Las secuencias se dividieron en dos grupos, uno con la inserción grande en, aproximadamente, la posición 150 y otro sin la misma (lo que distingue, por ejemplo, entre NM-Cas9 y ST-Cas9 por un lado y SP-Cas9 y TD-Cas9 por el otro). Estos grupos se alinean por separado utilizando MUSCLE (ver Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res 32, 1792-97, se realinean utilizando un enfoque en ventana (debido a la longitud de las secuencias), y posteriormente se alinean de nuevo perfil a perfil en alineación de una semilla.

45 Las secuencias en PF13395 se realinean con MUSCLE y utilizando la alineación de semilla. Todas las alineaciones utilizando una semilla se realizan alineando cada una de las secuencias diana una por una a la semilla. Esta alineación se utiliza para determinar la identidad de los mejores resultados entre las semillas y las secuencias diana, que se reordenan según la disminución de la identidad de los mejores resultados. A continuación, las secuencias diana se alinean nuevamente a la semilla una por una, ahora en orden de identidad decreciente. Se toma este enfoque de dos etapas para garantizar la solidez de la alineación. Además, estas secuencias se dividen dependiendo de si contienen la inserción o no, y los dos grupos separados se vuelven a alinear con la semilla como un perfil. Las secuencias cortas y las secuencias con grandes truncamientos se eliminan manualmente. Se eliminan las secuencias con más del 90 % de similitud por pares.

50 Las dos alineaciones resultantes se alinean perfil-perfil entre sí, lo que dio como resultado 217 secuencias totales. Las alineaciones resultantes se truncan a las posiciones del ortólogo de Cas9 de interés.

60 Se calculó la conservación de la secuencia como la entropía relativa con respecto a la frecuencia de fondo de los aminoácidos promediada sobre todos los genes en *Escherichia coli* 0157. La detección del límite de dominio se realizó aplicando un filtro de borde de diferencia de gaussianas (DoG) (ver Marr, D. & Hildreth, E. Theory of edge detection. Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society 207, 187-217 (1980)) al perfil de conservación resultante, promediando sobre múltiples escalas de longitud para proporcionar solidez a la elección de parámetros y la detección en diversas escalas de longitud.

En particular, se calculó la conservación de la alineación, siendo la entropía relativa (véase Cover, T. M. y Thomas, J. T. (2006) Elements of Information Theory; 2ª edición, Wiley-Interscience) de las frecuencias de aminoácidos con respecto a las frecuencias promedio para todos los genes en *Escherichia coli* O157:

5

$$D_i = \sum_{a=1}^{20} p_i^a \log \frac{p_i^a}{q^a}$$

en la que p_i^a es la frecuencia del aminoácido a en la posición i, y q^a es la frecuencia promedio del aminoácido a. La suma es sobre los 20 aminoácidos. El logaritmo es en base 2 y la entropía se da en bits. Las frecuencias q^a promedio son las siguientes:

10 {A C D E F G H I K L M N P Q R S T V W Y} = 0,094 0,012 0,052 0,058 0,038 0,073 0,022 0,059 0,045 0,104 0,027 0,041 0,044 0,044 0,057 0,060 0,055 0,070 0,015 0,029. La alineación de Cas9 produce el perfil de conservación en la figura 3, después del truncamiento a las posiciones de NM-Cas9.

15 El perfil de conservación se traza truncado a posiciones en SP-Cas9 (que es un ejemplo de las proteínas Cas9 más grandes de la subfamilia de tipo IIA). Se observan características similares en el extremo terminal N, aproximadamente, hasta la posición NM145 (SP170) y, posteriormente, nuevamente después de la posición NM200 (SP400), que es la inserción grande descrita anteriormente. Ver la figura 4.

20 Se identificaron los límites de dominio potenciales aplicando un filtro de borde de escala múltiple al perfil de conservación. Este filtro calcula la diferencia de gaussianas (DoG) (ver Marr, D y Hildreth, E (1980) Theory of Edge Detection, Proc R Soc Lond B Biol Sci 207, 187-217 para un intervalo de escalas y sumar los gráficos resultantes. Los extremos de esta curva se interpretan como los límites entre las regiones de baja y alta conservación en la proteína. Los dominios pueden mostrar diferentes niveles de conservación debido a su importancia funcional

25 potencialmente diferente que conduce a tasas diferenciales de divergencia evolutiva. La conservación diferencial puede ser característica de una proteína multidominio que es el resultado de la inserción del dominio sobre escalas de tiempo evolutivas. Los valores de los extremos se utilizan para ordenar los límites en términos de importancia.

30 Para la alineación de Cas9, restringida a las posiciones NM, se identificaron los siguientes límites o bordes de conservación o los aminoácidos de borde utilizando los procedimientos descritos en la presente memoria descriptiva:

Clasificación	Posición	Valor absoluto del extremo
1	736	0,0225
2	620	0,0148
3	554	0,0148
4	472	0,0144
5	288	0,0136
6	144	0,0131
7	87	0,0106
8	661	0,0101
9	825	0,0100
10	205	0,0099
11	512	0,0092
12	414	0,0092
13	108	0,0092
14	853	0,0067
15	438	0,0050
16	936	0,0041
17	1011	0,0038

Ejemplo X

35 Análisis computacional de miembros de la familia de Cas9

Para identificar las regiones dentro de la proteína Cas9 multidominio que pueden ser susceptibles de eliminación, se utilizó un enfoque bioinformático para identificar los límites potenciales entre dominios. Utilizando una alineación de semillas bien seleccionada, se realinearon las secuencias de Cas9 de longitud completa a partir de PFAM (PF13395) y se eliminaron las secuencias con alta identidad por pares, tal como se ha descrito anteriormente. La conservación de la secuencia de primer orden se calculó como la entropía relativa (véase Cover, T. M., Thomas, J. T. Elements of Information Theory, 2ª edición. (Wiley-Interscience, 2006)) de las frecuencias de aminoácidos observadas con respecto a las frecuencias promedio en todos los genes en *Escherichia coli*. Como se puede esperar que los dominios en una proteína multidominio exhiban niveles variables de conservación de secuencia, se puede

40

demandar la aplicación de un filtro de borde multiescala al perfil de conservación para identificar las ubicaciones de los límites del dominio.

5 Se realizó la detección de bordes en el perfil de conservación con un filtro de paso de banda de diferencia de gaussianas (DoG) que es sensible a un intervalo estrecho de frecuencias espaciales. Ver Marr, D. y Hildreth, E. Theory of edge detection. Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society 207, 187-217 (1980). Para permitir la detección a diversas escalas de longitud y hacer que el filtro sea insensible a una elección particular de parámetros, se realizó un promedio sobre escalas múltiples (5-50 aminoácidos). El filtro de paso de banda se aplicó posteriormente al perfil de conservación para la alineación de
10 Cas9. Las posiciones de límite potenciales identificadas para NM-Cas9 se muestran en la figura 5A, con los 6 límites de clasificación superior en líneas rojas más largas en negrita. Como se puede observar, el filtro identifica correctamente las estructuras de dominio HNH y RuvC conocidos que se asignaron anteriormente en Sapranaukas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic acid research 39, 9275-9282, doi: 10.1093/nar/gkr606 (2011) y Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic acid research 42, 2577-2590, doi: 10.1093/nar/gkt1074 (2014) (ver figura 5B), comenzando en el sentido de 5' de la posición 500 y variando hasta la posición 750. Los límites entre los dominios, alrededor de las posiciones 560 y 620, se predicen también correctamente. También se predice el límite de la primera hélice alfa rica en arginina en la posición 88. Uno de los resultados superiores de límite en el extremo N-terminal de la proteína, en la posición 144, no se había
20 identificado como un límite de dominio anteriormente en Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic acids research 42, 2577-2590, doi:10.1093/nar/gkt1074 (2014), pero representa una mejor delineación de la región alfa-helicoidal rica en Arg, que ahora incluye una segunda hélice rica en arginina conservada.

25 Ejemplo XI

Análisis de truncamiento y eliminación de Cas9

30 Para explorar dominios funcionales y potencialmente no funcionales dentro de Cas9, se generaron truncamientos experimentales en los extremos N-terminal y C-terminal, junto con un conjunto de eliminaciones internas modestas basadas en el análisis de detección de dominios descrito en la presente memoria descriptiva. Se analizaron los efectos de las eliminaciones utilizando un ensayo represor transcripcional en *Escherichia coli*, en el que una proteína Cas9 nucleasa nula funcional se unirá al extremo 5' de un informador YFP, disminuyendo de este modo su nivel de expresión (ver la figura 6A). Ver Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173-1183, doi:10.1016/j.cell. 2013.02.022 (2013); Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nature methods 10, 1116-1121, doi:10.1038/nmeth.2681 (2013) y Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic acids research 41, 7429-7437, doi:10.1093/nar/gkt520 (2013). Ninguno de los truncamientos del extremo N-terminal o C-terminal fue capaz de reprimir al informador, mientras que dos eliminaciones internas, NMΔ255-289 y NMΔ330-389, en el sentido de 3' y en el sentido de 5' del límite en la posición 288, junto con una eliminación NMΔ566-620 mostraron niveles de represión casi de tipo salvaje (figura 6B y figura 7). Se realizaron diversas rondas de análisis adicionales en las que las eliminaciones se expandieron iterativamente y se ensayó la capacidad de reprimir al informador (figura 6B). Se identificaron dos regiones grandes no superpuestas 254-449 y 567-654 (que comprendían el 18 % y el 8 % de la longitud total de la proteína, respectivamente) que podrían eliminarse con una pérdida insignificante en la actividad de NM (figura 6C). Tal como se observa en la alineación, las posiciones de NM 254-449 representan un tramo de conservación relativamente baja, en una región de la proteína que es específica de las proteínas Cas9. Las posiciones 567 a 654 representan el dominio HNH, un dominio que se sabe que es crítico en la catálisis de ADN de la Cas9, pero que se descubrió que era prescindible para la unión al ADN.
50

Para corroborar que las regiones eliminadas de NM-Cas9 no eran exclusivas de NM, sino que representan regiones generales que se pueden eliminar de otros miembros de la familia Cas9, se generaron las eliminaciones correspondientes dentro de las variantes de nucleasa nula de Cas9 de *Streptococcus thermophilus* (ST1) y Cas9 de *Treponema denticola* (TD, GI: 42525843) y se midió su función con el ensayo de represión transcripcional (figura 8A y figura 9). Los mutantes de eliminación correspondientes, tanto en ST1 como en TD, mostraron actividades similares a sus contrapartes de tipo salvaje, lo que sugiere que las regiones eliminadas son prescindibles para la unión al ADN de Cas9 en toda la filogenia de Cas9, incluso entre miembros más distantes dentro de la subfamilia de tipo II-A, tales como TD.
55

60 Se ensayó la actividad de las dos eliminaciones funcionales más grandes descritas en la presente memoria descriptiva dentro de ST1-Cas9 utilizando un ensayo activador transcripcional (figura 8B). De acuerdo con el análisis dentro de *E. coli*, ambos mutantes de eliminación de ST1 retuvieron una actividad comparable a la proteína de tipo salvaje cuando se fusionaron con el dominio de activación de VP64 y se dirigieron a un informador fluorescente en células humanas (figuras 8C-8D). El más grande de los dos mutantes de eliminación Δ255-450 (numeración ST1) genera un gen de Cas9 que tiene un tamaño de 2.793 pares de bases.
65

Ejemplo XII

Intercambio de dominios de Cas9

5 Los dominios N-terminal y C-terminal de Cas9 pueden desempeñar papeles críticos en la unión ARNcr:ARNcrtra y/o en la selectividad de PAM. Para analizar la actividad, se realizó una serie de mutantes de intercambio de dominio entre NM y ST1, reemplazando el extremo N-terminal y/o C-terminal de NM con la región homóloga de ST1. Se ensayaron posteriormente las proteínas quiméricas utilizando el ensayo de informador transcripcional descrito en la presente memoria descriptiva alterando el ARN guía y/o PAM específico de Cas9 dentro del informador para
10 determinar la influencia de los intercambios de dominio sobre la especificidad de la proteína (figura 10A). Las posiciones exactas para los intercambios de dominio se determinaron sobre la base del análisis de los límites de dominio: se seleccionaron las posiciones que estaban lo más cerca posible de los límites N-terminal y C-terminal más significativos identificados (figura 5A), que al mismo tiempo estaban casi completamente conservados dentro de la alineación (figura 10B). Ninguno de los intercambios de dominio N-terminal entre NM y ST1 dotó a NM de
15 propiedades novedosas, lo que sugiere que el extremo N-terminal de ST1 no es modular, sino que funciona en contexto con otras regiones de ST1 que no se transfirieron (figura 10C-10F). El intercambio C-terminal generó un híbrido NM-ST1 que era capaz de interactuar con el complejo ARNcr:ARNcrtra de ST1 y además podía suprimir un informador con un PAM específico de ST1 (figura 10E). Este resultado se validó adicionalmente con un informador específico ST1 adicional, tal como se muestra en las figuras 11(A)-11(D).
20

REIVINDICACIONES

- 5 1. Procedimiento para producir una proteína Cas9 mutante, que comprende identificar una o más secuencias de aminoácidos que no están altamente conservadas o tienen baja conservación dentro de una familia de proteínas Cas9, en el que la una o más secuencias de aminoácidos que no están altamente conservadas o tienen baja conservación se identifican obteniendo una alineación de secuencias de múltiples secuencias de aminoácidos de Cas9, e identificando tramos de baja conservación o tramos entre los bordes de conservación, en el que la conservación de dichas secuencias de aminoácidos se calcula como la entropía relativa de las frecuencias de aminoácidos por posición, identificar una secuencia de ácido nucleico que codifica la una o más secuencias de aminoácidos identificadas, generar una secuencia de ácido nucleico que codifica una proteína Cas9 diana que carece de la secuencia de ácido nucleico que codifica una o más secuencias de aminoácidos identificadas, y generar una proteína Cas9 mutante a partir de la secuencia de ácido nucleico generada, en el que la proteína Cas9 mutante retiene la actividad de unión al ADN.
- 15 2. Un procedimiento para producir una proteína Cas9 mutante, que comprende identificar una o más secuencias de aminoácidos dentro de una familia de proteínas Cas9 que son de alta conservación o de baja conservación, en el que cada una de las secuencias de alta o baja conservación está flanqueada por dos secuencias de aminoácidos flanqueantes de baja conservación o de alta conservación, respectivamente, en el que las secuencias de las proteínas de la familia Cas9 están alineadas, y la conservación de dichas secuencias de aminoácidos se calcula como la entropía relativa de las frecuencias de aminoácidos por posición, identificar una secuencia de ácido nucleico que codifica la una o más secuencias de aminoácidos, generar una secuencia de ácido nucleico que codifica una proteína Cas9 diana que carece de la secuencia de ácido nucleico que codifica la una o más secuencias de aminoácidos adicionales, y generar una proteína Cas9 mutante a partir de la secuencia de ácido nucleico generada,

20 en el que la proteína Cas9 mutante retiene la actividad de unión al ADN.
- 25 3. Procedimiento, según la reivindicación 1 o la reivindicación 2, en el que la proteína Cas9 diana tiene actividad de nucleasa.
- 30 4. Procedimiento, según la reivindicación 1 o la reivindicación 2, en el que la proteína Cas9 diana es una nickasa.
- 35 5. Procedimiento, según la reivindicación 1 o la reivindicación 2, en el que la proteína Cas9 diana es una nucleasa nula.
- 40 6. Procedimiento, según la reivindicación 1 o la reivindicación 2, en el que la secuencia de ácido nucleico que codifica la una o más secuencias de aminoácidos identificadas se elimina de la secuencia de ácido nucleico que codifica la proteína Cas9 diana.
- 45 7. Procedimiento, según la reivindicación 1 o la reivindicación 2, en el que la secuencia de ácido nucleico que codifica la una o más secuencias de aminoácidos identificadas se elimina de la secuencia de ácido nucleico que codifica la proteína Cas9 diana y se reemplaza con un enlazador.
8. Procedimiento, según la reivindicación 1 o la reivindicación 2, en el que la secuencia de ácido nucleico que codifica la una o más secuencias de aminoácidos identificadas se elimina de la secuencia de ácido nucleico que codifica la proteína Cas9 diana y se reemplaza con un enlazador SGGGS.
9. Proteína Cas9 mutante definida por la SEQ ID NO: 1.
- 50 10. Proteína Cas9 mutante definida por la SEQ ID NO: 2.

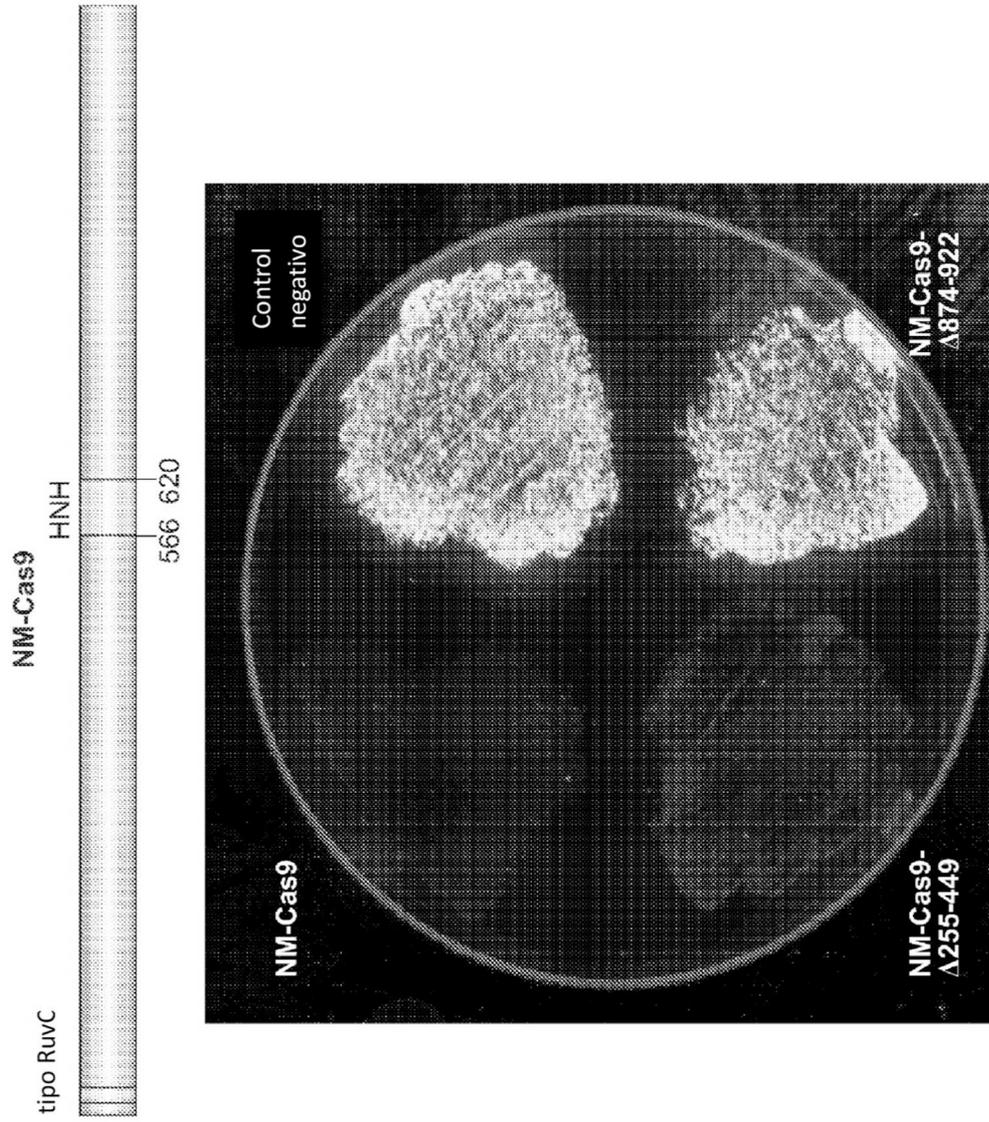


FIG. 1

FIG. 2A

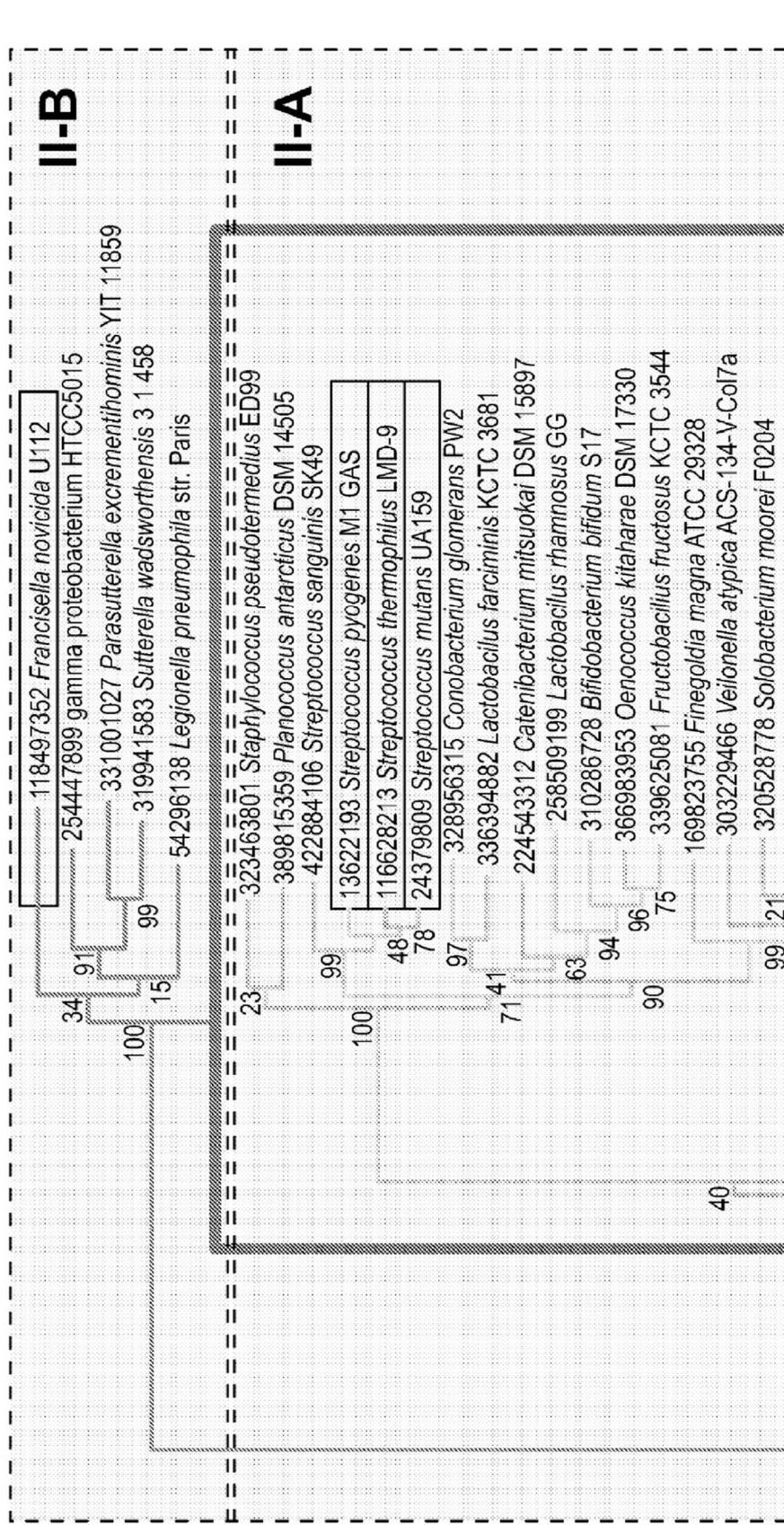


FIG. 2B

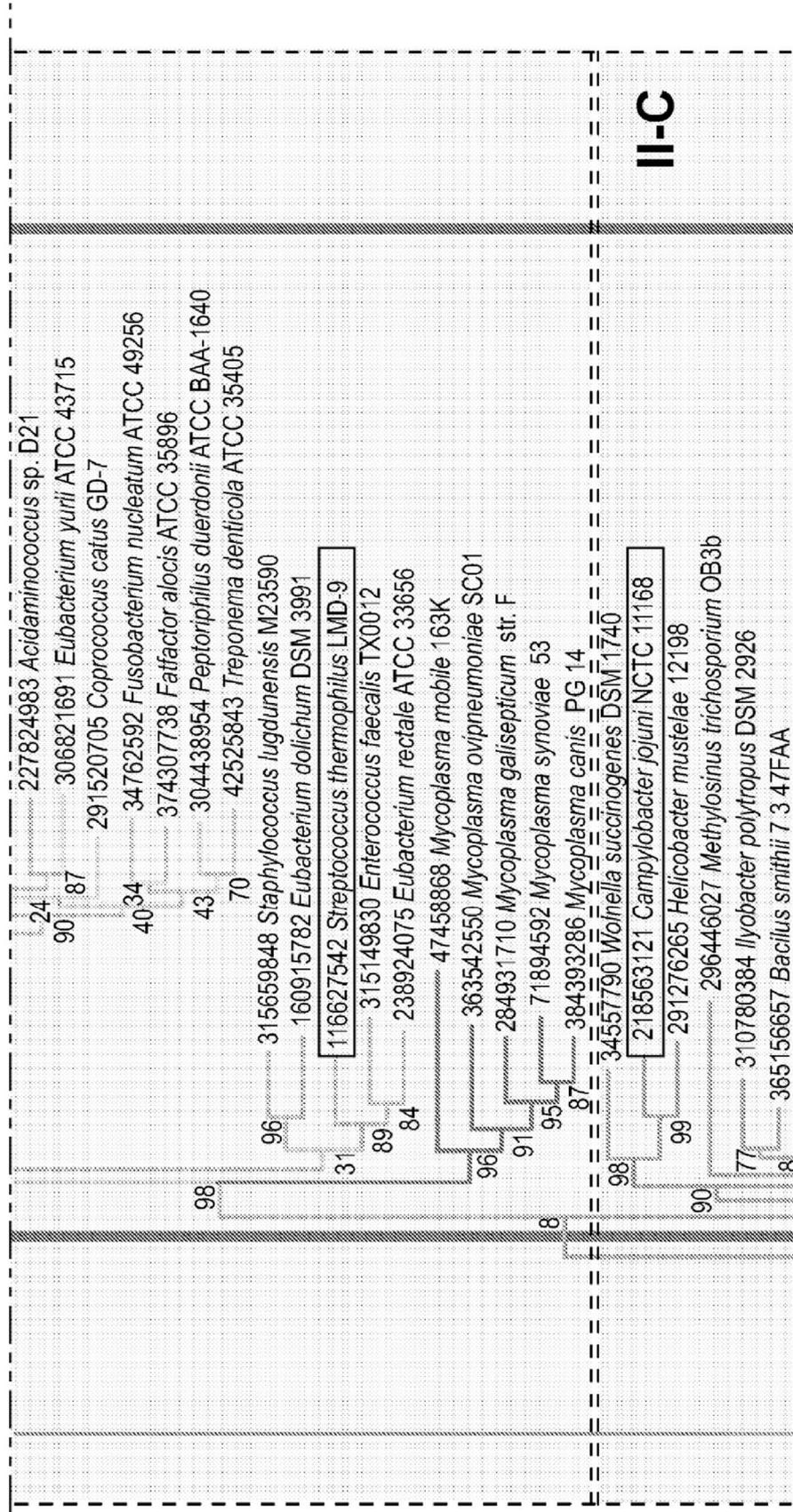


FIG. 2C

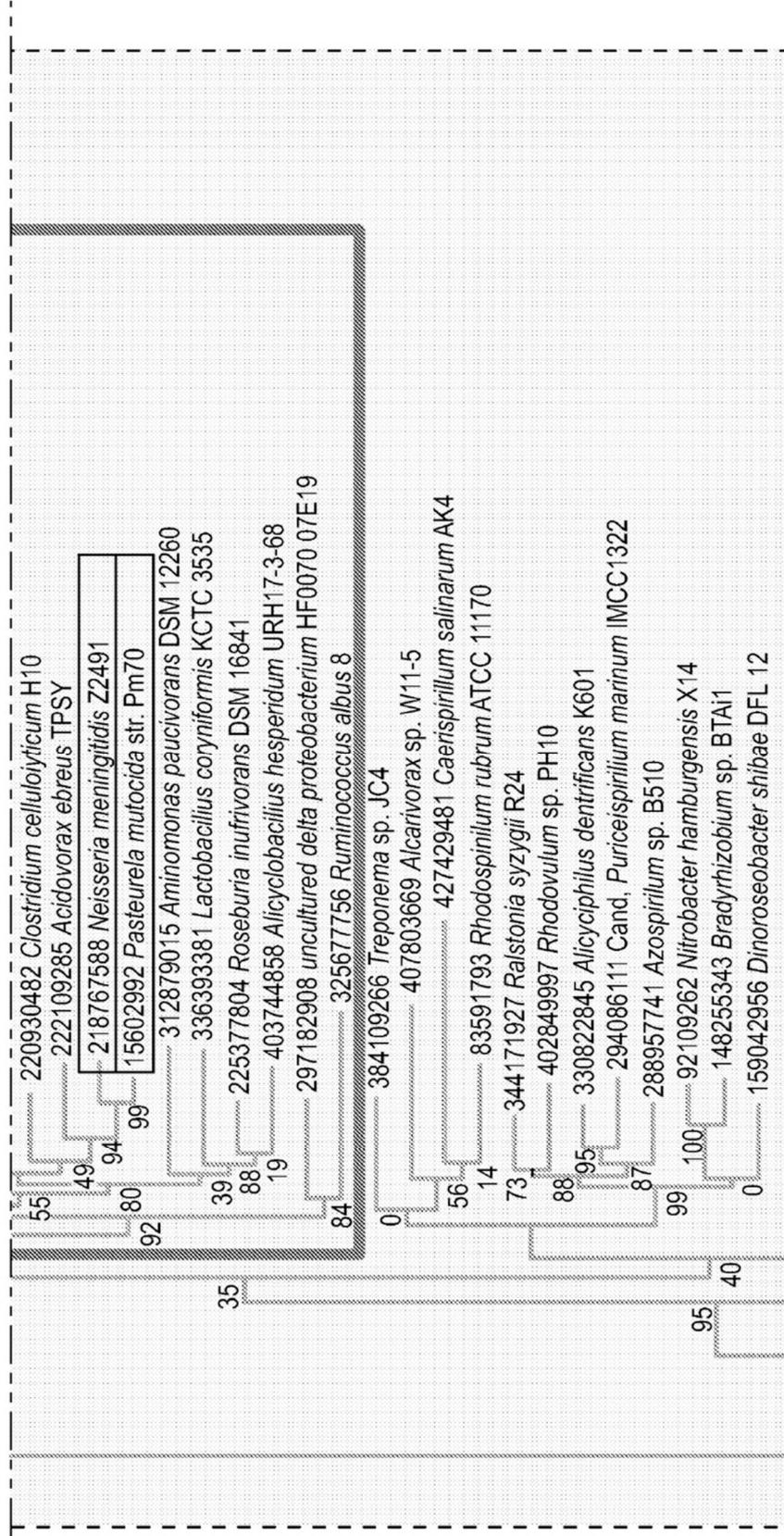


FIG. 2D

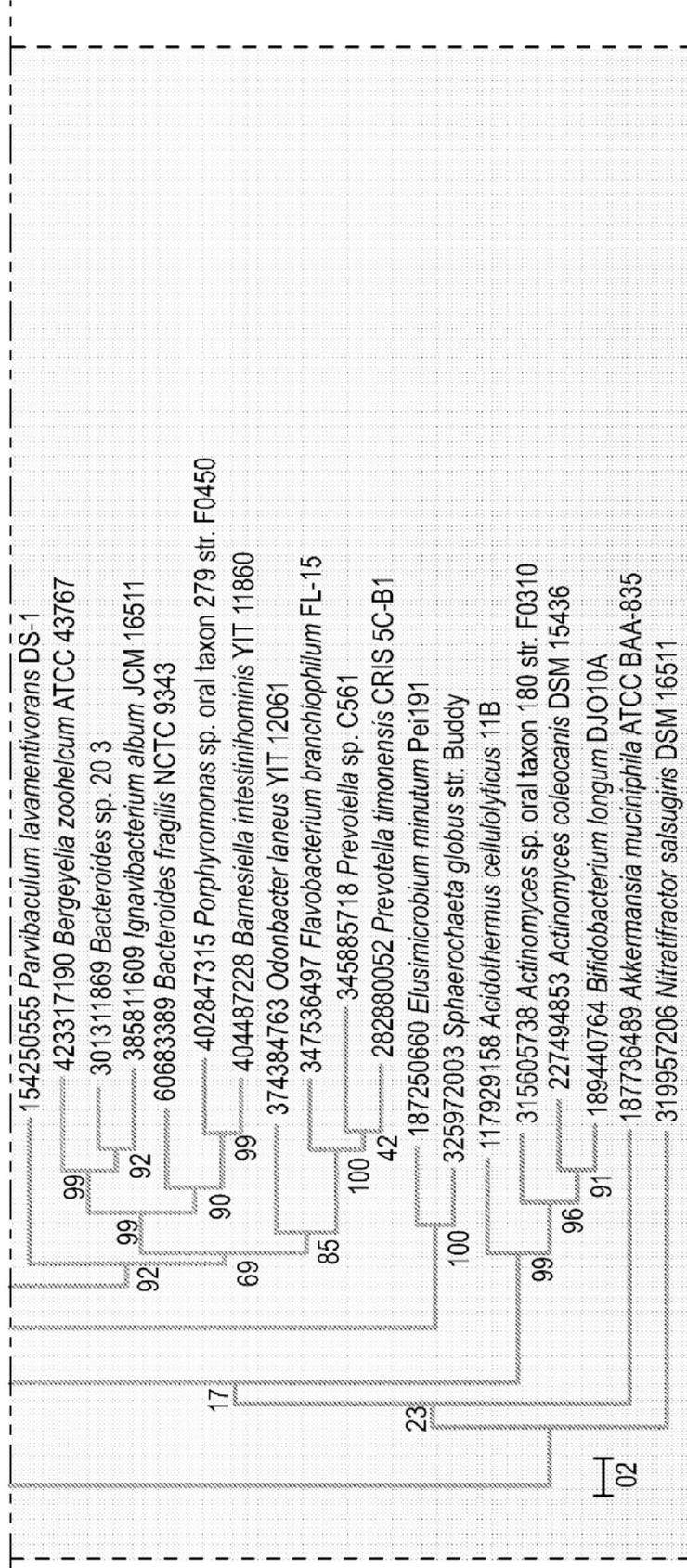


FIG. 3

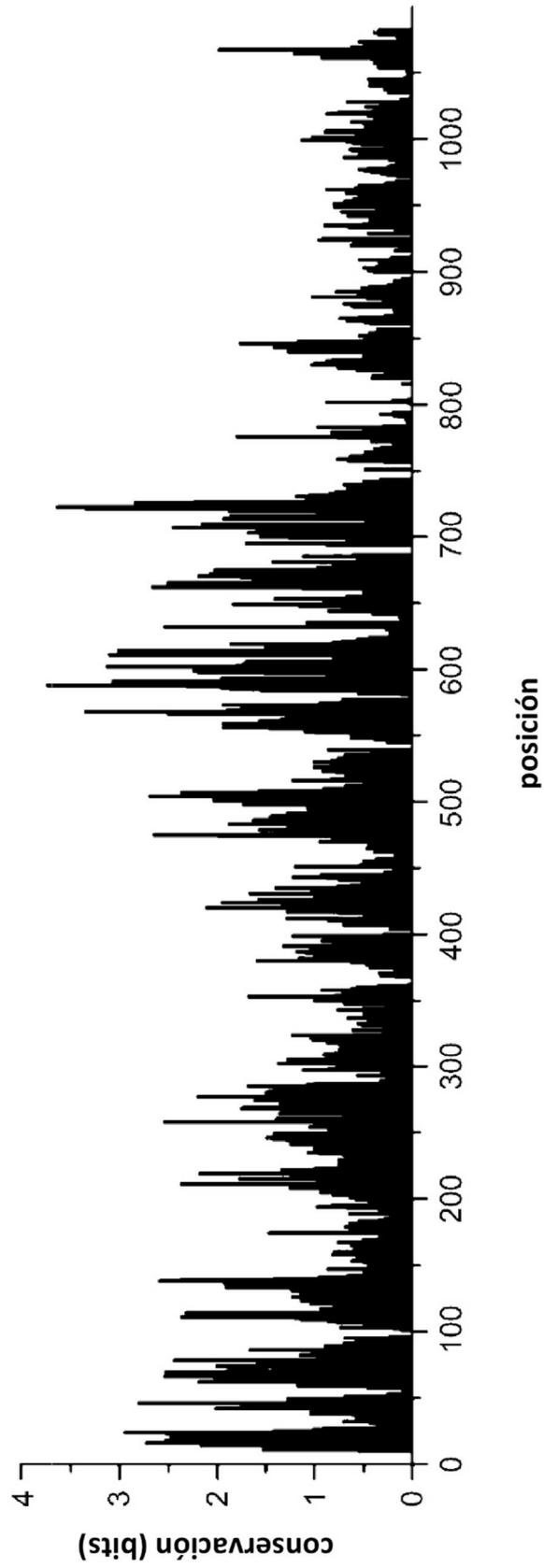


FIG. 4

Inserción
SP/TD-Cas9
↔

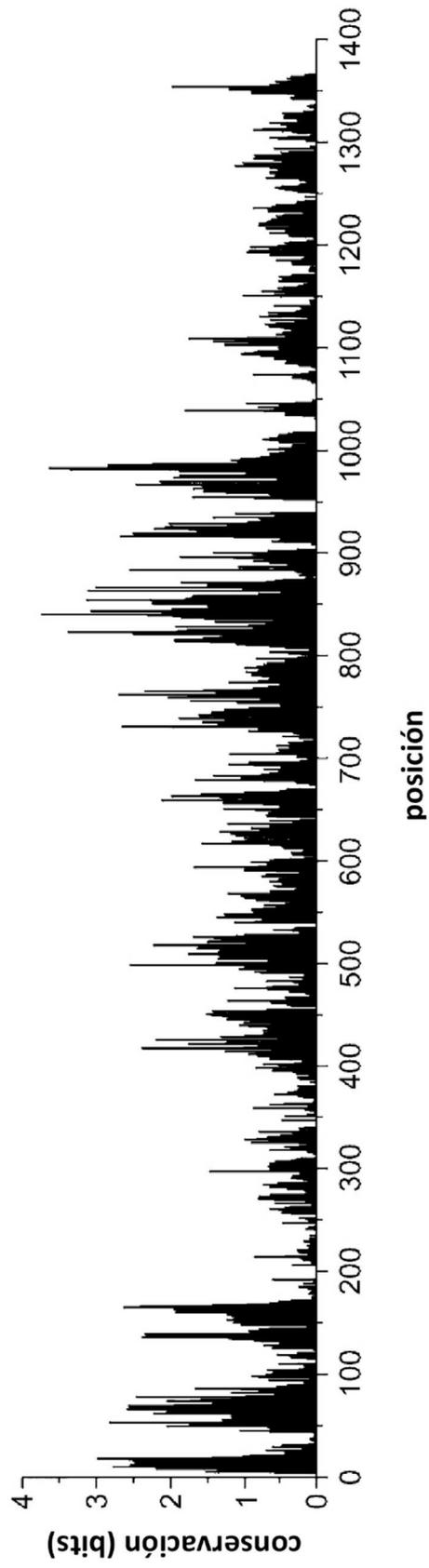


FIG. 5A

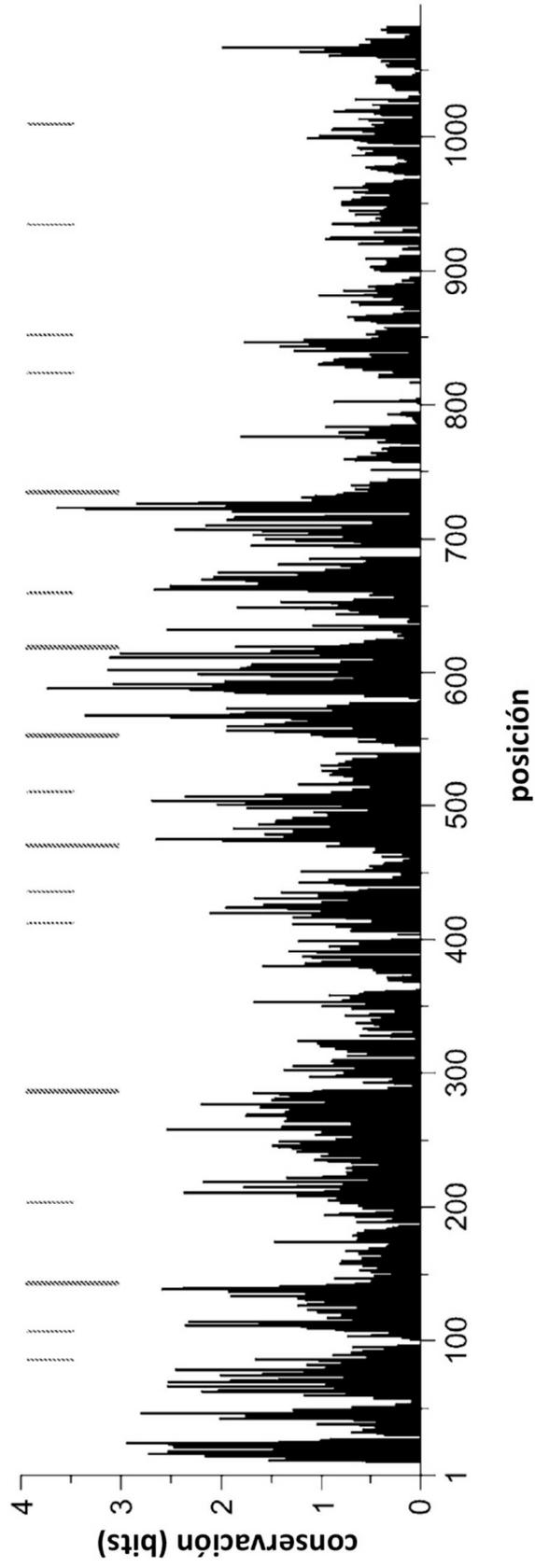
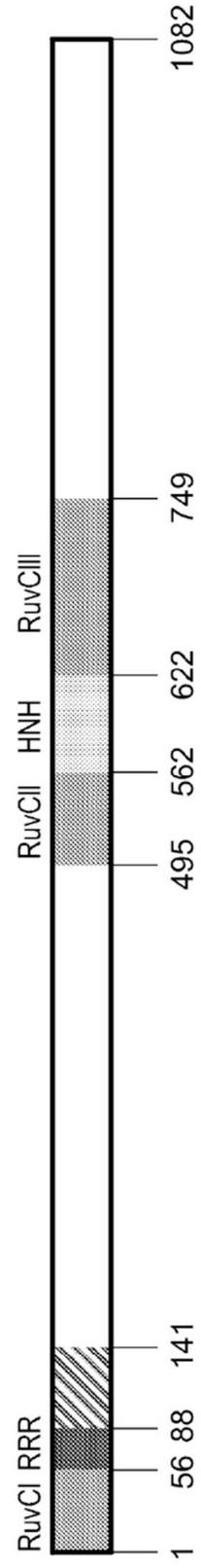
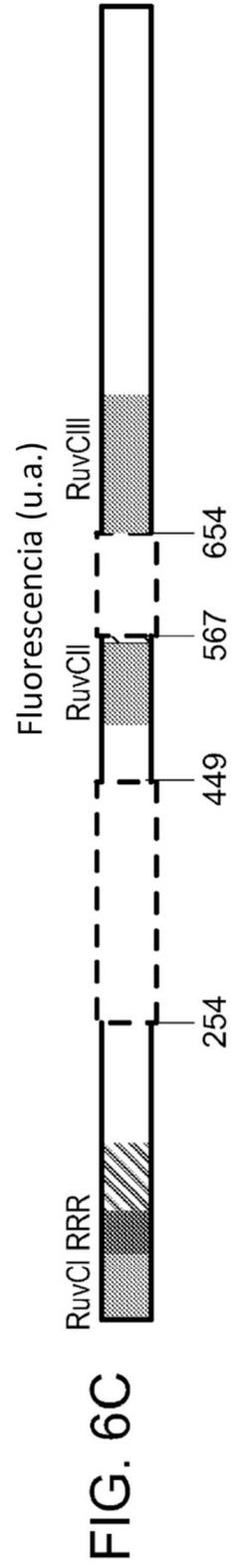
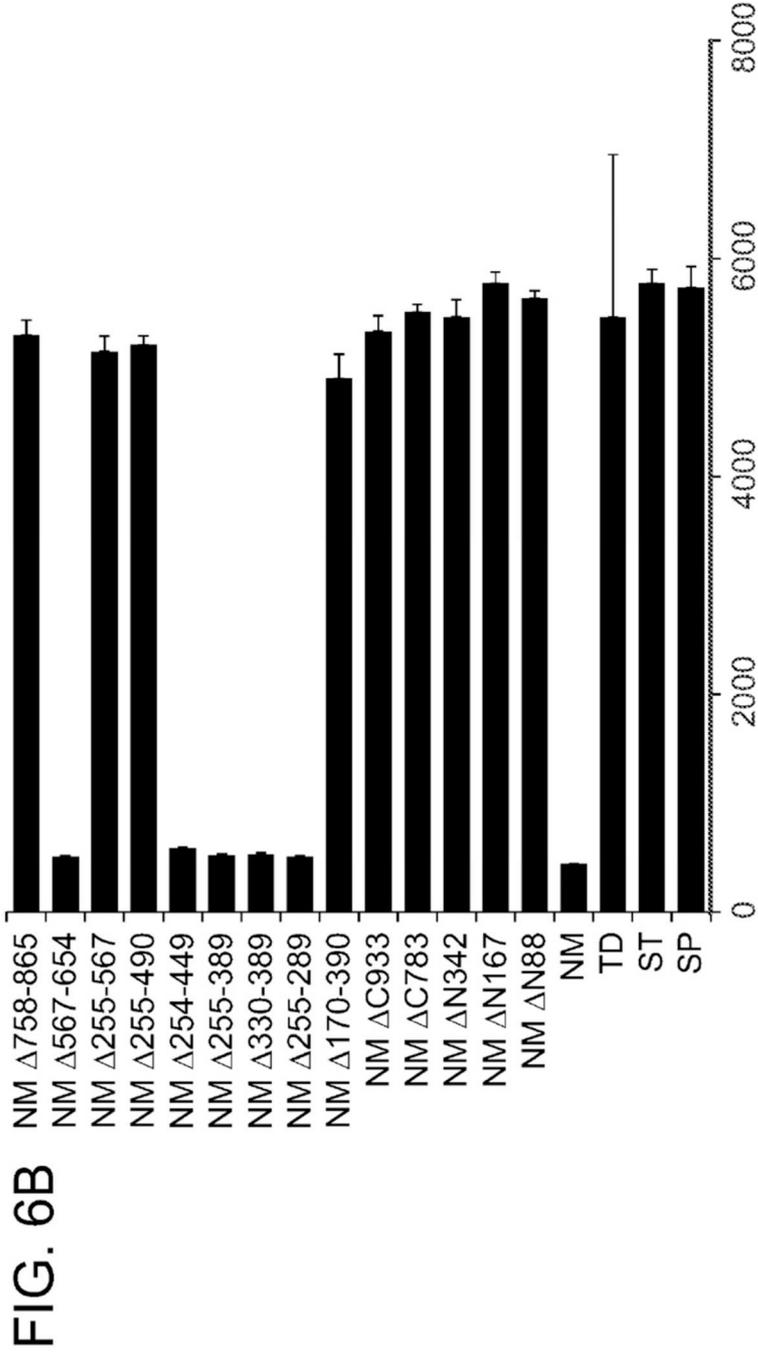
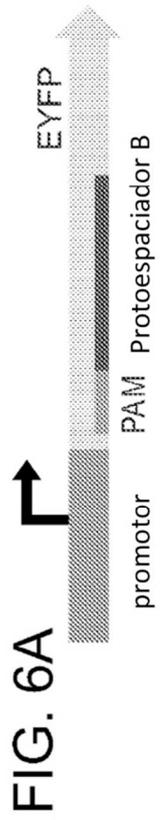


FIG. 5B





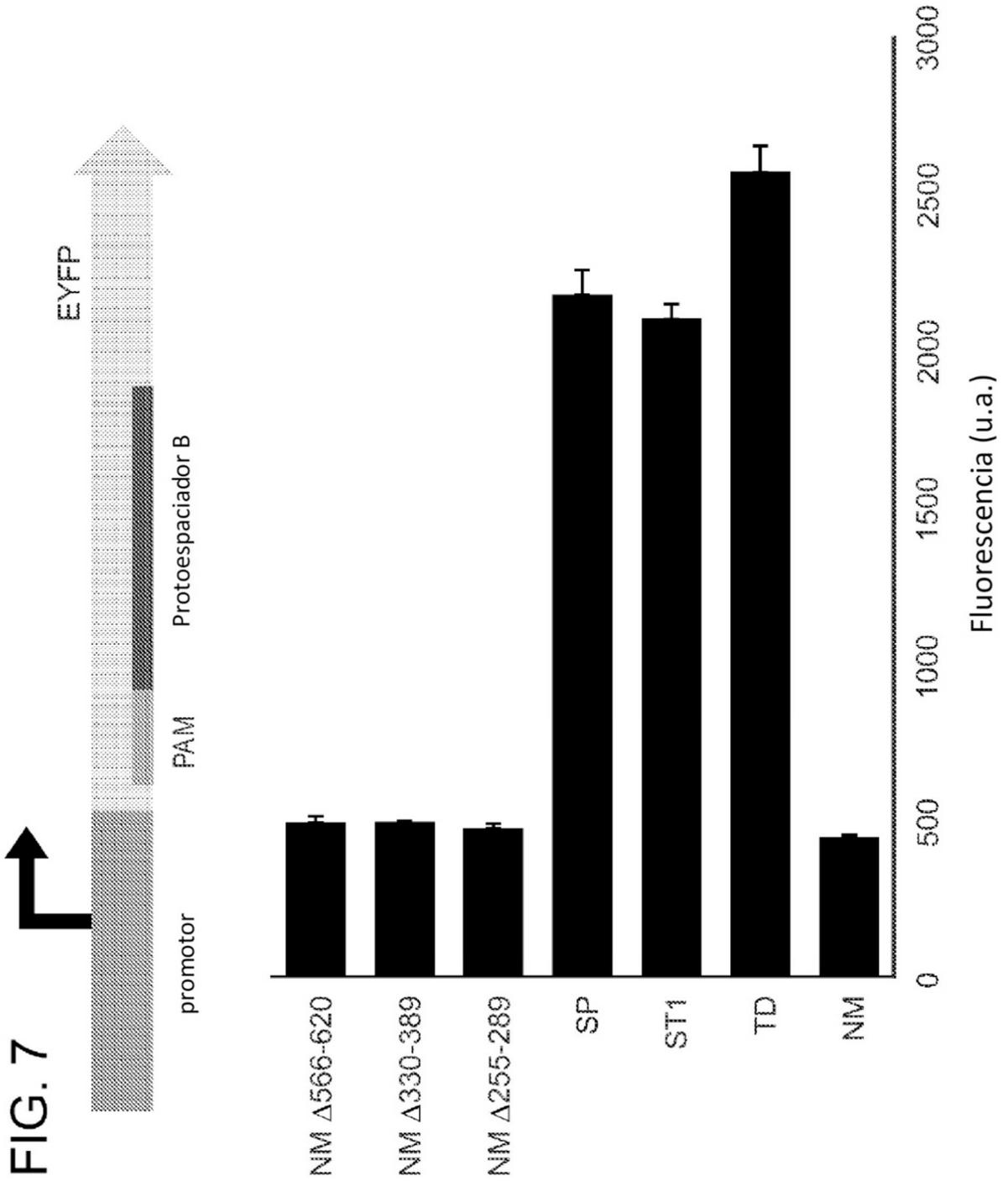


FIG. 8A

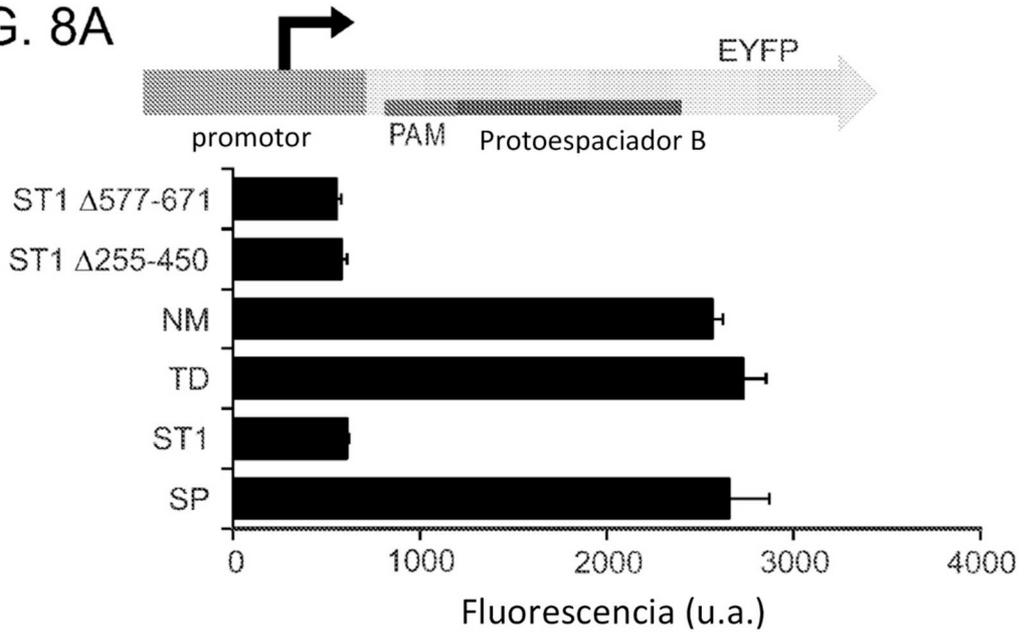


FIG. 8B



FIG. 8C

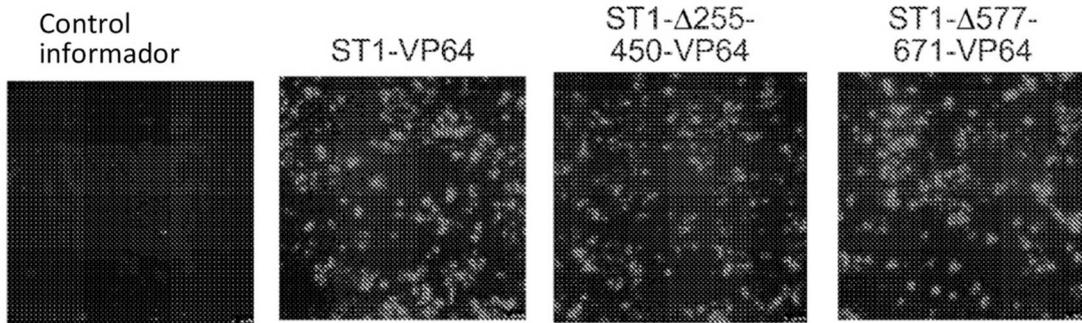
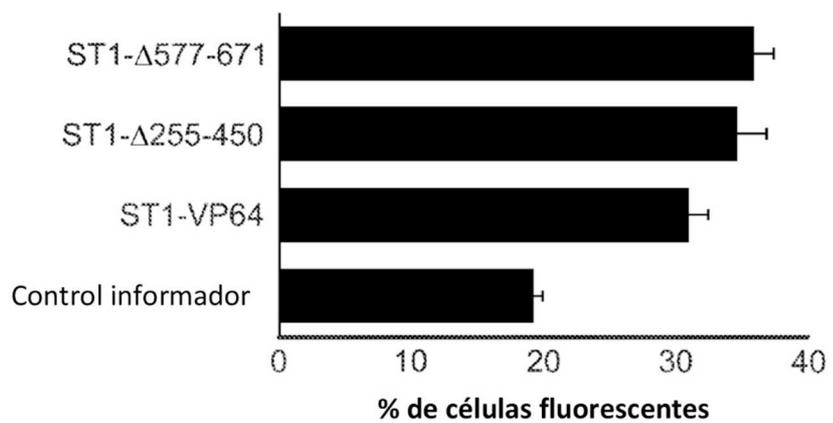


FIG. 8D



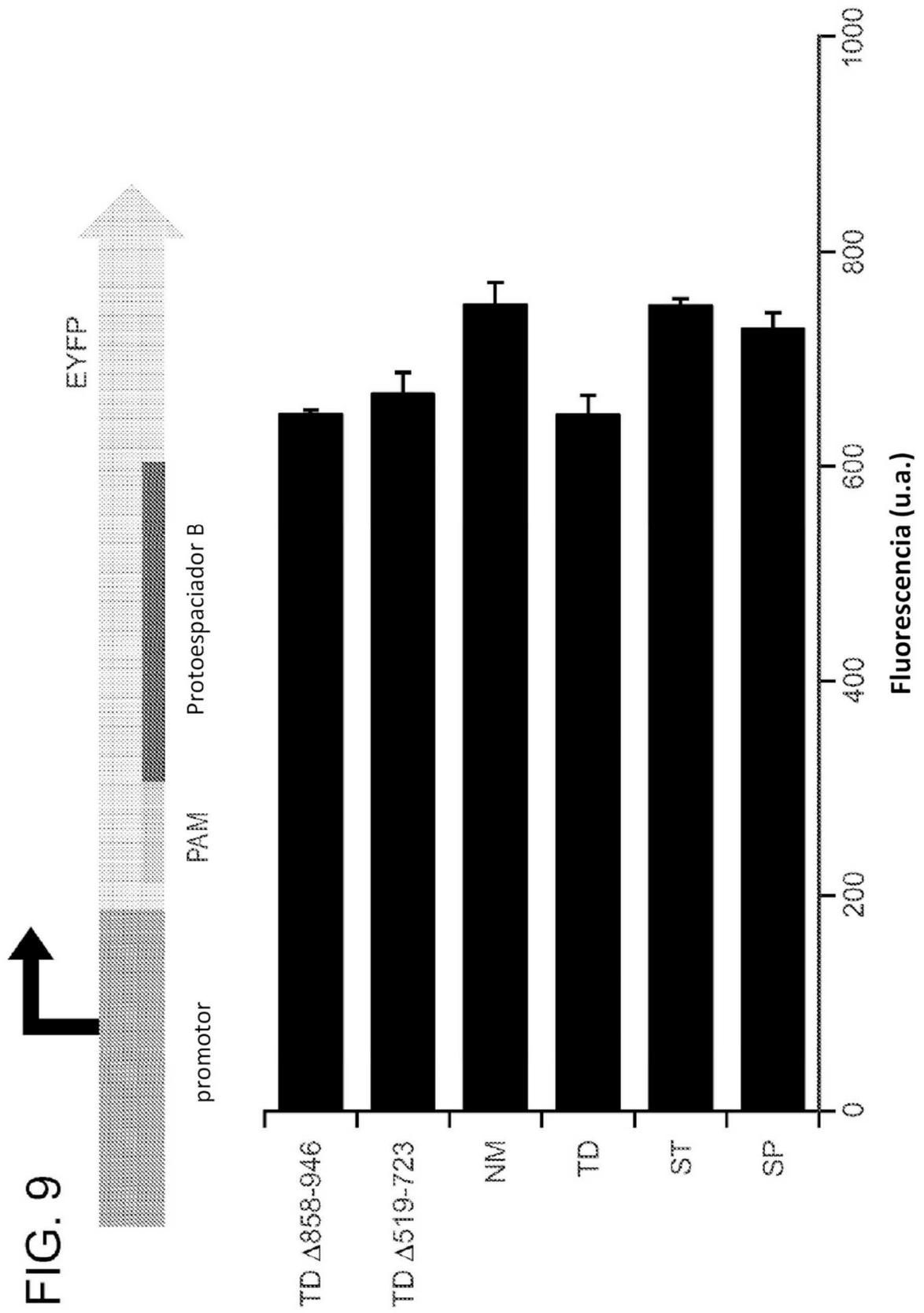


FIG. 10A

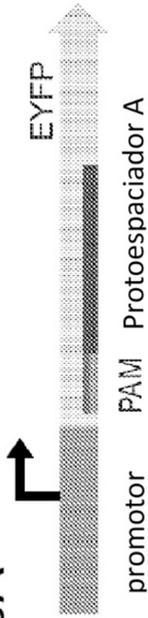


FIG. 10B

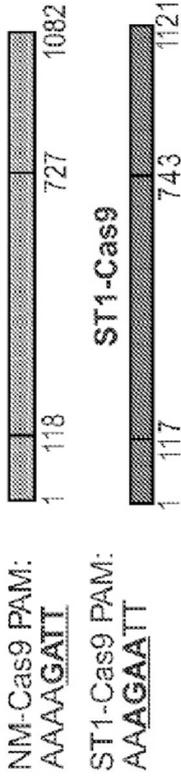


FIG. 10C ARNg NM con PAM informador NM

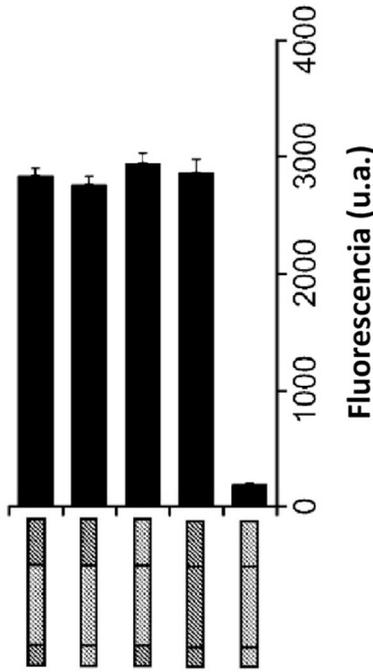


FIG. 10E ARNg ST1 con PAM informador ST1

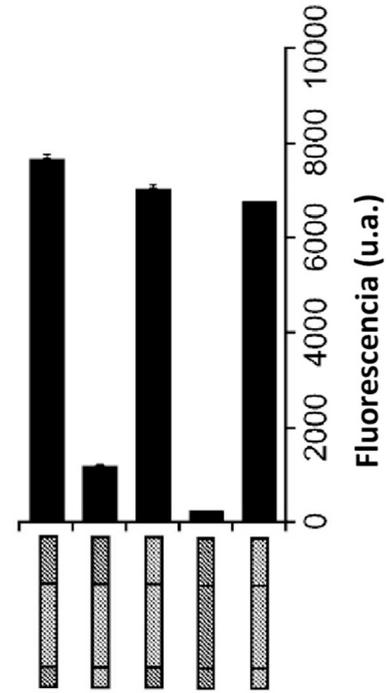


FIG. 10D ARNg NM con PAM informador ST1

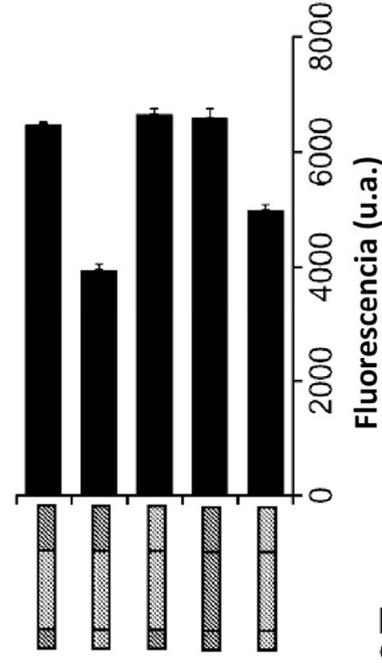


FIG. 10F ARNg ST1 con PAM informador NM

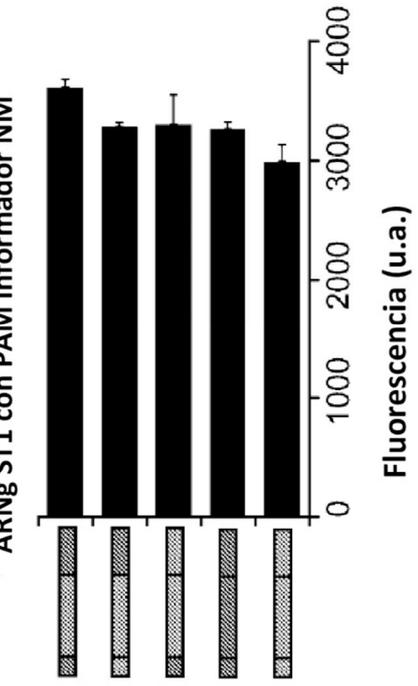


FIG. 11A

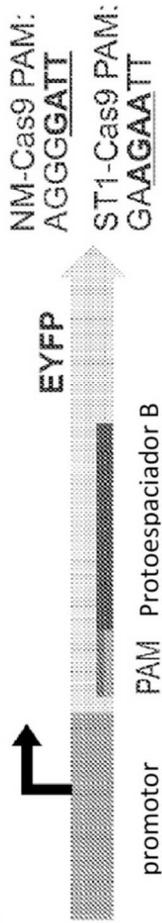


FIG. 11B



FIG. 11C

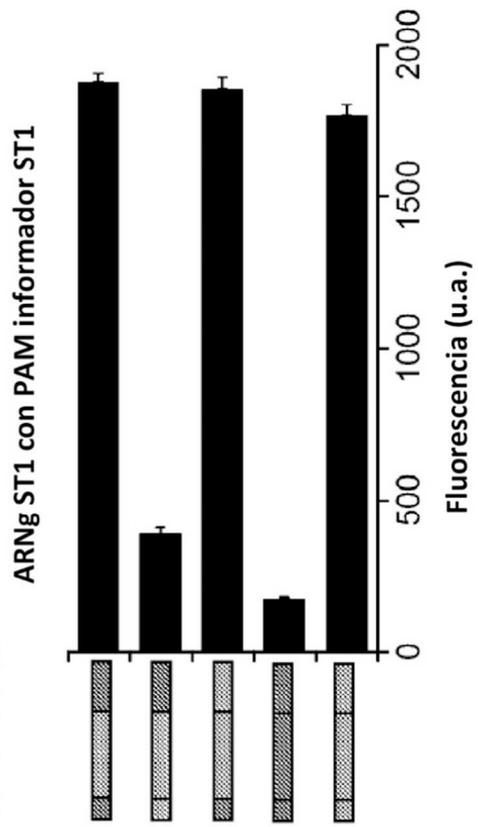
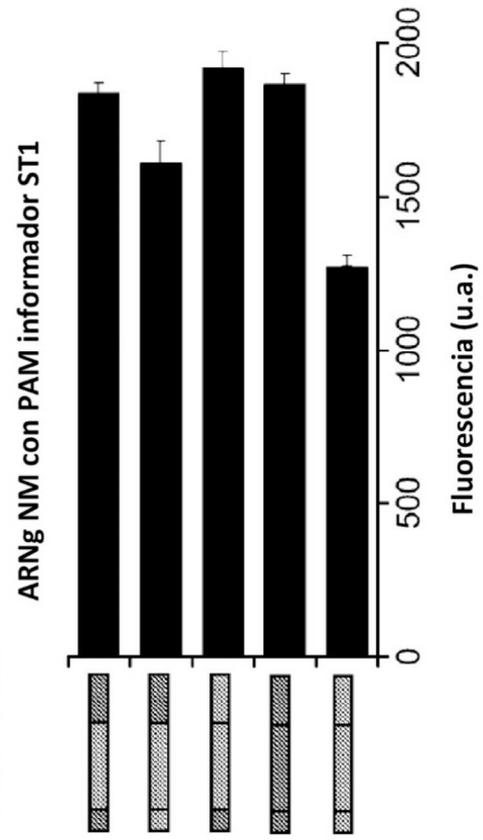


FIG. 11D



REFERENCIAS CITADAS EN LA DESCRIPCIÓN

Esta lista de referencias citada por el solicitante es únicamente para mayor comodidad del lector. No forman parte del documento de la Patente Europea. Incluso teniendo en cuenta que la compilación de las referencias se ha efectuado con gran cuidado, los errores u omisiones no pueden descartarse; la EPO se exime de toda responsabilidad al respecto.

Documentos de patentes citados en la descripción

- US 61906374

Literatura no patente citada en la descripción

- DELTCHEVA, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011, vol. 471, 602-607
- GASIUNAS, G.; BARRANGOU, R.; HORVATH, P.; SIKSNYS, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, vol. 109, E2579-2586
- JINEK, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, vol. 337, 816-821
- SAPRANAUSKAS, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic acids research*, 2011, vol. 39, 9275-9282
- BHAYA, D.; DAVISON, M.; BARRANGOU, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*, 2011, vol. 45, 273-297
- H. DEVEAU et al. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology*, February 2008, vol. 190, 1390
- HOU, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, vol. 110, 15644-15649
- KUMAR, M.; KELLER, B.; MAKALOU, N.; SUTTON, R. E. Systematic determination of the packaging limit of lentiviral vectors. *Human gene therapy*, 2001, vol. 12, 1893-1905
- WU, Z.; YANG, H.; COLOSI, P. Effect of genome size on AAV vector packaging. *Molecular therapy: the journal of the American Society of Gene Therapy*, 2010, vol. 18, 80-86
- GELINAS, C.; TEMIN, H. M. Nondefective spleen necrosis virus-derived vectors define the upper size limit for packaging reticuloendotheliosis viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 1986, vol. 83, 9211-9215
- FONFARA I et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.*, 2014, vol. 42, 2577-90
- FONFARA, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic acids research*, 2014, vol. 42, 2577-2590
- MAKAROVA et al. *Nature Reviews, Microbiology*, 09 June 2011, 467-477
- JINKE et al. *Science*, 2012, vol. 337, 816-821
- R.D. FINN ; J. CLEMENTS ; S.R. EDDY. *Nucleic Acids Research*, 2011
- ESVELT, K. M. ; MALI, P. ; BRAFF, J. L. ; MOOS-BURNER, M. ; YAUNG, S. J. ; CHURCH, G. M. *Nat Methods*, 2013, vol. 10, 1116-1121
- GIBSON, D. G. ; YOUNG, L. ; CHUANG, R. Y. ; VENTER, J. C. ; HUTCHISON, C. A., 3RD ; SMITH, H. O. *Nat Methods*, 2009, vol. 6, 343-345
- *Nat Methods*, 2013, vol. 10, 1116-1121
- ESVELT, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods*, 2013, vol. 10, 1116-1121
- EDGAR ; RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 2004, vol. 32, 1792-97
- MARR, D. ; HILDRETH, E. Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society*, 1980, vol. 207, 187-217
- COVER, TM ; THOMAS, JT. *Elements of Information Theory*. Wiley-Interscience, 2006
- SEE MARR, D ; HILDRETH, E. Theory of Edge Detection. *Proc R Soc Lond B Biol Sci*, 1980, vol. 207, 187-217
- COVER, T. M. ; THOMAS, J.T. *Elements of Information Theory*. Wiley-Interscience, 2006
- MARR, D. ; HILDRETH, E. Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Containing papers of a Biological character. Royal Society*, 1980, vol. 207, 187-217

- **QI, L. S. et al.** Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 2013, vol. 152, 1173-1183
- **BIKARD, D. et al.** Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research*, 2013, vol. 41, 7429-7437