

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 576 126**

51 Int. Cl.:

C12N 15/63 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **12.12.2013 E 13814364 (9)**

97 Fecha y número de publicación de la concesión europea: **09.03.2016 EP 2898075**

54 Título: **Modificación por tecnología genética y optimización de sistemas, métodos y composiciones enzimáticas mejorados para la manipulación de secuencias**

30 Prioridad:

12.12.2012 US 201261736527 P
02.01.2013 US 201361748427 P
30.01.2013 US 201361758468 P
25.02.2013 US 201361769046 P
15.03.2013 US 201361802174 P
15.03.2013 US 201361791409 P
28.03.2013 US 201361806375 P
20.04.2013 US 201361814263 P
06.05.2013 US 201361819803 P
28.05.2013 US 201361828130 P
17.06.2013 US 201361835931 P
17.06.2013 US 201361836101 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.07.2016

73 Titular/es:

THE BROAD INSTITUTE, INC. (33.3%)
415 Main Street
Cambridge, MA 02142, US;
MASSACHUSETTS INSTITUTE OF TECHNOLOGY (33.3%) y
PRESIDENT AND FELLOWS OF HARVARD COLLEGE (33.3%)

72 Inventor/es:

ZHANG, FENG;
RAN, FEI y
SHALEM, OPHIR

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 576 126 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

Modificación por tecnología genética y optimización de sistemas, métodos y composiciones enzimáticas mejorados para la manipulación de secuencias

SOLICITUDES RELACIONADAS

- 5 Esta solicitud reivindica prioridad de la solicitud de patente provisional de EE.UU. 61/836.101 titulada MODIFICACIÓN POR TECNOLOGÍA GENÉTICA Y OPTIMIZACIÓN DE SISTEMAS, MÉTODOS Y COMPOSICIONES ENZIMÁTICAS MEJORADOS PARA LA MANIPULACIÓN DE SECUENCIAS presentada el 17 de junio de 2013. Esta solicitud también reivindica prioridad de las solicitudes de patente provisionales de EE.UU. 61/758.468; 61/769.046; 61/802.174; 61/806.375; 61/814.263; 61/819.803 y 61/828.130, titulada MODIFICACIÓN
- 10 POR TECNOLOGÍA GENÉTICA Y OPTIMIZACIÓN DE SISTEMAS, MÉTODOS Y COMPOSICIONES PARA LA MANIPULACIÓN DE SECUENCIAS, presentadas el 30 de enero de 2013; 25 de febrero de 2013; 15 de de marzo de 2013; 28 de marzo de 2013; 20 de abril de 2013; 6 de mayo de 2013 y 28 de mayo de 2013, respectivamente. Esta solicitud también reivindica prioridad de las solicitudes de patente provisionales de EE.UU. 61/736.527 y 61/748.427, ambas con el título SISTEMAS, MÉTODOS Y COMPOSICIONES PARA LA MANIPULACIÓN DE
- 15 SECUENCIAS, presentadas el 12 de diciembre de 2012 y 2 de enero de 2013, respectivamente. La prioridad también se reivindica de las solicitudes de patente provisionales de EE.UU. 61/791.409 y 61/835.931, presentadas el 15 de marzo de 2013 y 17 de junio de 2013, respectivamente.

Se hace referencia a las solicitudes de patente provisionales de EE.UU. 61/836.127, 61/835.936, 61/836.080, 61/836.123 y 61/835.973, presentadas cada una el 17 de junio de 2013.

20 CAMPO DE LA INVENCIÓN

La presente invención se refiere, en general, a la modificación por tecnología genética y la optimización de sistemas, métodos y composiciones utilizados para el control de la expresión génica que implican la focalización de la secuencia, tales como la perturbación del genoma o la edición de genes, que se refieren a Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas (CRISPR) y componentes de los mismos.

25 DECLARACIÓN PATROCINADA POR EL GOBIERNO FEDERAL

Esta invención se hizo con el apoyo del Gobierno bajo la NIH Pioneer Award (1DP1MH100706) adjudicada por el National Institute of Health. El Gobierno tiene ciertos derechos en la invención.

ANTECEDENTES DE LA INVENCIÓN

- 30 Recientes progresos en las técnicas de secuenciación del genoma y métodos de análisis han acelerado significativamente la capacidad para catalogar y mapear factores genéticos asociados con una diversa gama de funciones biológicas y enfermedades. Se necesitan tecnologías de focalización del genoma, precisas para permitir una modificación por tecnología genética inversa sistemática de las variaciones genéticas causales, permitiendo una perturbación selectiva de elementos genéticos individuales, así como para avanzar en la biología de síntesis, biotecnológica y aplicaciones médicas. Aunque las técnicas de edición del genoma, tales como los dedos de zinc
- 35 diseñadores, efectores del tipo activador de la transcripción (TALEs), o meganucleasas migratorias están disponibles para la producción de perturbaciones del genoma focalizado, sigue habiendo una necesidad de nuevas tecnologías de la ingeniería del genoma que son asequibles, fáciles de instalar, expansibles y susceptibles de fijar como objetivo múltiples posiciones dentro del genoma eucariota.

SUMARIO DE LA INVENCIÓN

- 40 El CRISPR/Cas o el sistema CRISPR-Cas (ambos términos se utilizan indistintamente a lo largo de toda esta solicitud) no requieren la generación de proteínas modificadas para fijar como objetivo secuencias diana específicas, sino más bien una sola enzima Cas puede ser programada por una molécula de ARN corta para reconocer una diana de ADN específica, en otras palabras, la enzima Cas puede ser reclutada a una diana de ADN específica utilizando dicha molécula de ARN corta. Añadiendo el sistema CRISPR-Cas al repertorio de técnicas de

5 secuenciación del genoma y métodos de análisis puede simplificar significativamente la metodología y acelerar la capacidad de catalogar y mapear los factores genéticos asociados con una diversa gama de funciones biológicas y enfermedades. Para utilizar el sistema CRISPR-Cas de manera efectiva para la edición del genoma sin efectos perjudiciales, es fundamental entender aspectos de la modificación genética y la optimización de estas herramientas de modificación genética del genoma, que son aspectos de la presente descripción.

10 Por consiguiente, existe una necesidad urgente de sistemas y técnicas, alternativos y robustos, para la focalización de la secuencia de dirección con una amplia gama de aplicaciones. Aspectos de esta invención satisfacen esta necesidad y proporcionan ventajas relacionadas. Un complejo de CRISPR a modo de ejemplo comprende una enzima de CRISPR que forma complejo con una secuencia guía hibridada con una secuencia diana dentro del polinucleótido diana, en donde la enzima de CRISPR es un ortólogo Cas, p. ej., un ortólogo Cas9, de *Staphylococcus aureus* (SaCas9). La secuencia guía está enlazada a una secuencia de emparejamiento tracr que, a su vez, se hibrida a una secuencia tracr.

15 De acuerdo con la presente invención, se proporciona una composición que comprende componentes de un complejo de CRISPR que comprenden una secuencia o secuencias de polinucleótidos del sistema de CRISPR-Cas y una proteína Cas9 de Tipo II, en donde la proteína Cas9 de Tipo II es o comprende una Cas9 de *Staphylococcus aureus* (SaCas9) de acuerdo con la reivindicación 1. Aspectos adicionales de la invención se describen en las reivindicaciones dependientes. La invención también proporciona células, métodos y usos relacionados con la composición de acuerdo con las reivindicaciones independientes restantes. La invención es como se define en las reivindicaciones independientes, en particular la reivindicación 1. Cualesquiera aspectos que caigan fuera de las
20 reivindicaciones se proporcionan como descripción.

25 En un aspecto, la invención proporciona métodos para usar uno o más elementos de un sistema de CRISPR. El complejo de CRISPR de la invención proporciona un medio eficaz para modificar un polinucleótido diana. El complejo de CRISPR de la invención tiene una amplia diversidad de utilidades incluyendo modificar (p. ej., suprimir, insertar, translocar, inactivar, activar, reprimir, alterar la metilación, transferir restos específicos) un polinucleótido diana en una multiplicidad de tipos de células. Como tal, el complejo de CRISPR de la invención tiene un amplio espectro de aplicaciones en, p. ej., la edición del gen o del genoma, la regulación de genes, la terapia génica, el descubrimiento de fármacos, el rastreo de fármacos, el diagnóstico de enfermedades, y el pronóstico. El complejo de CRISPR comprende un ortólogo Cas9 de *Staphylococcus aureus* (SaCas9).

30 Aspectos de la invención se refieren a enzimas de CRISPR que tienen una función optimizada. Con respecto a la enzima de CRISPR que es una enzima Cas, realizaciones preferidas de la invención se refieren a ortólogos Cas9 que tienen especificidad diana mejorada en un sistema CRISPR-Cas9. Esto se puede realizar mediante enfoques que incluyen, pero no se limitan al diseño y preparación de ARNs guía que tiene una actividad óptima, la selección de enzimas Cas9 de una longitud específica, el truncamiento de la enzima Cas9 haciéndola más pequeña en longitud que la enzima Cas9 de tipo salvaje correspondiente, truncando las moléculas de ácido nucleico que
35 codifican la misma y generando enzimas Cas9 quiméricas en las que diferentes partes de la enzima se permutan o intercambian entre diferentes ortólogos para llegar a enzimas quiméricas que tienen especificidad adaptada. Aspectos de la invención se refieren también a métodos para mejorar la especificidad de la diana de una enzima ortólogo Cas9 o el diseño de un sistema de CRISPR-Cas9 que comprende diseñar o preparar ARNs guía que tiene actividad óptima y/o seleccionar o preparar una enzima ortólogo Cas9 que tiene un tamaño o una longitud más
40 pequeña que la correspondiente Cas9 de tipo salvaje, con lo cual se hace avanzar el empaquetamiento de un ácido nucleico que codifica la misma en un vector de suministro, ya que hay menos secuencia de codificación para la misma en el vector de suministro que para la correspondiente Cas9 de tipo salvaje y/o la generación de enzimas Cas9 quiméricas.

45 También se proporcionan usos en medicina de las presentes secuencias, vectores, enzimas o sistemas. También se proporcionan las mismas para uso en la edición de un gen o genoma. También se proporciona el uso de las mismas en la fabricación de un medicamento para la edición de un gen o genoma, por ejemplo el tratamiento mediante la edición del gen o genoma. También se proporcionan las presentes secuencias, vectores, enzimas o sistemas para uso en terapia.

50 En un aspecto adicional de la invención, una enzima de CRISPR, p. ej., una enzima Cas9 puede comprender una o más mutaciones y se pueden utilizar como una proteína de unión a ADN genérico con o sin fusión a, o estando operativamente enlazada a un dominio funcional. Las mutaciones pueden ser mutaciones introducidas artificialmente y pueden incluir, pero no se limitan a una o más mutaciones en un dominio catalítico. Ejemplos de dominios catalíticos con referencia a una enzima Cas9 pueden incluir, pero no se limitan a dominios RuvC I, RuvC II, RuvC III

y HNH. Ejemplos preferidos de mutaciones adecuadas son el o los residuos catalíticos en el extremo N del dominio RuvC I de Cas9 o el o los residuos catalíticos en el dominio HNH interno. En algunas realizaciones, la Cas9 es (o se deriva de) Cas9 de *Streptococcus pyogenes* (SpCas9). En tales realizaciones, las mutaciones preferidas son en cualquiera o en todas las posiciones 10, 762, 840, 854, 863 y/o 986 de SpCas9 o posiciones correspondientes en otros ortólogos Cas9 con referencia a la numeración de la posición de SpCas9 (que puede determinarse, por ejemplo, mediante herramientas de comparación de secuencias estándar, p. ej., ClustalW o MegAlign por Lasergene 10 suite). En particular, cualquiera o todas de las siguientes mutaciones son preferidos en SpCas9: D10A, E762A, H840A, N854A, N863A y/o D986A; así como también se prevé una sustitución conservativa para cualquiera de los aminoácidos de sustitución. También se prefieren las mismas mutaciones (o sustituciones conservativas de estas mutaciones) en posiciones correspondientes con referencia a la numeración de la posición de SpCas9 en otros ortólogos Cas9. Particularmente preferidos son SpCas9 D10 y H840. Sin embargo, en otras Cas9s, también se prefieren los residuos correspondientes a SpCas9 D10 y H840. Estos son ventajosos como cuando mutados individualmente proporcionan actividad nickase, y cuando ambas mutaciones están presentes la Cas9 se convierte en un mutante catalíticamente nulo que es útil para la unión de ADN genérico. Se han identificado y caracterizado mutaciones adicionales. Otros aspectos de la invención se refieren a la enzima Cas9 mutada que está fusionada o enlazada de forma operativa a los dominios que incluyen pero no están limitados a un activador transcripcional, represor transcripcional, una recombinasa, una transposasa, una remodelador de histona, una ADN metiltransferasa, una criptocromo, un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente.

Un aspecto adicional de la invención proporciona proteínas Cas9 quiméricas y métodos de generar proteínas Cas9 quiméricas. Proteínas Cas9 quiméricas son proteínas que comprenden fragmentos que proceden de diferentes ortólogos Cas9. Por ejemplo, el extremo N de un primer ortólogo Cas9 puede estar fusionado con el extremo C de un segundo ortólogo Cas9 para generar una proteína Cas9 quimérica resultante. Estas proteínas Cas9 quiméricas pueden tener una especificidad mayor o una mayor eficiencia que la especificidad o eficiencia original de cualquiera de las enzimas Cas9 individuales a partir de las cuales se generó la proteína quimérica. Estas proteínas quiméricas también pueden comprender una o más mutaciones o pueden estar enlazadas a uno o más dominios funcionales. Por lo tanto, aspectos de la invención se refieren a una enzima Cas quimérica, en donde la enzima comprende uno o más fragmentos de un primer ortólogo Cas y uno o más fragmentos de un segundo ortólogo Cas. En una realización de la invención, los uno o más fragmentos del primer o segundo ortólogo Cas son del extremo C o N del primer o segundo ortólogo Cas. En una realización adicional, el primer o segundo ortólogo Cas se selecciona de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*,

En una realización adicional, la invención proporciona métodos para generar componentes mutantes del complejo de CRISPR que comprende una enzima Cas, p. ej., un ortólogo Cas9. Los componentes mutantes pueden incluir, pero no se limitan a secuencias de tracrRNA y de emparejamiento tracr mutante o secuencias guías quiméricas mutantes que permiten mejorar el rendimiento de estos ARN en las células. También se proporciona el uso de la presente composición o la enzima en la preparación de un medicamento para la modificación de una secuencia diana.

La invención, en todavía otro aspecto, proporciona composiciones y métodos relacionados con una composición que se produce de forma no natural o modificada por ingeniería genética, que comprende:

A) - I. una secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema CRISPR-Cas, en donde la secuencia de polinucleótido comprende:

- (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota,
- (b) una secuencia de emparejamiento de tracr, y
- (c) una secuencia de tracr, y

II. una secuencia de polinucleótidos que codifica una enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear,

en donde (a), (b) y (c) están dispuestas en una orientación 5' a 3',

en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, y la secuencia de polinucleótidos que codifica una enzima CRISPR es ADN o ARN,

o

(B) I. polinucleótidos que comprenden:

- (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula procarionota, y

(b) al menos una o más secuencias de emparejamiento tracr,

II. una secuencia de polinucleótido que codifica una enzima de CRISPR, y

III. una secuencia de polinucleótidos que comprende una secuencia tracr,

5 en donde cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde el complejo de CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, y la secuencia de polinucleótido que codifica una enzima de CRISPR es ADN o ARN, y
 10 en donde la enzima CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*.

La invención, en un aspecto adicional, proporciona: (A) una composición que se produce de forma no natural o modificada por ingeniería genética, que comprende un sistema de vectores que comprende uno o más vectores que comprenden: I. un primer elemento regulador enlazado operativamente a una secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema CRISPR-Cas, en donde la secuencia de polinucleótidos comprende (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota, (b) una secuencia de emparejamiento tracr, y (c) una secuencia de tracr, y II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima CRISEPR que comprende al menos una o más
 20 secuencias de localización nuclear, en donde (a), (b) y (c) están dispuestos en una orientación 5' a 3', en donde los componentes I y II están situados en el mismo o en diferentes vectores del sistema, en donde, cuando se transcriben, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde el complejo CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, en donde la enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter* o (B) una composición que se produce de forma no natural o modificada por ingeniería genética que comprende un sistema de vectores que comprende uno o más
 30 vectores que comprende I. un primer elemento regulador enlazado operativamente a (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula procariota, y (b) por lo menos una o más secuencias de emparejamiento tracr II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR, y III. un tercer elemento regulador enlazado operativamente a una secuencia tracr, en el que los componentes I, II y III están situados en el mismo o diferentes vectores del sistema, en el que cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde el complejo de CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, en donde la enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*, y en donde se aplica al menos uno de los siguientes criterios.

45 Los criterios son los siguientes y se apreciará que se puede aplicar cualquier número de ellos, preferiblemente 1 o más, preferiblemente 2 o más, y preferiblemente 3 o más, 4 o más, o 5 o más, o todos se pueden aplicar:

- se selecciona la enzima de CRISPR que tiene un tamaño específico y tiene una longitud de al menos 500 aminoácidos, al menos 800-899 aminoácidos, al menos 900-999 aminoácidos, al menos 1000-1099 aminoácidos, al menos 1100-1119 aminoácidos, al menos 1200-1299 aminoácidos, al menos 1300-1399 aminoácidos, al menos
 50 1400-1499 aminoácidos, al menos 1500-1599 aminoácidos, al menos 1600-1699 aminoácidos o al menos 2000 aminoácidos;
- y/o la enzima de CRISPR está truncada en comparación con la enzima de CRISPR de tipo salvaje correspondiente;
- y/o la enzima de CRISPR es una nucleasa que dirige la escisión de las dos cadenas en la ubicación de la
 55 secuencia diana, o la enzima de CRISPR es una nickase que dirige la escisión de una cadena en el lugar de la secuencia diana;
- y/o la secuencia guía comprende al menos 10, al menos 15 o al menos 20 nucleótidos;
- y/o la enzima de CRISPR esta optimizada en codones o está optimizada en codones para la expresión en una célula eucariota;

- y/o la enzima de CRISPR comprende una o más mutaciones;
- y/o la enzima de CRISPR comprende una enzima de CRISPR quimérica;
- y/o la enzima de CRISPR tiene uno o más otros atributos comentados en esta memoria.

- 5 En algunas realizaciones, la enzima de CRISPR está truncada en comparación con una enzima de CRISPR de tipo salvaje o la enzima de CRISPR se compone de al menos 500 aminoácidos, al menos 800-899 aminoácidos, al menos 900-999 aminoácidos, al menos 1000-1099 aminoácidos, al menos 1100-1199 aminoácidos, al menos 1200-1299 aminoácidos, al menos 1300-1399 aminoácidos, al menos 1400-1499 aminoácidos, al menos 1500-1599 aminoácidos, al menos 1600-1699 aminoácidos o al menos 2000 aminoácidos. En formas de realización preferidas, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.
- 10 En algunas realizaciones, la enzima de CRISPR es una nucleasa que dirige la escisión de ambas cadenas en la ubicación de la secuencia diana, o la enzima de CRISPR es una nickase que dirige la escisión de una cadena en la ubicación de la secuencia diana. En realizaciones adicionales, la enzima de CRISPR es un mutante catalíticamente nulo que es una proteína de unión a ADN genérica. En realizaciones preferidas, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.
- 15 En algunas realizaciones, la secuencia guía comprende al menos quince nucleótidos. En algunas realizaciones, la enzima de CRISPR está optimizada en codones o está optimizada en codones para la expresión en una célula eucariota. En algunas realizaciones, la enzima de CRISPR comprende una o más mutaciones. En algunas realizaciones, la enzima de CRISPR comprende una enzima de CRISPR quimérica. En algunas realizaciones, la enzima de CRISPR tiene uno o más de otros atributos comentados en esta memoria. En realizaciones preferidas, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.
- 20 En determinadas realizaciones, la enzima de CRISPR comprende una o más mutaciones. La una o más mutaciones pueden estar en un dominio particular de la enzima. En una realización preferida, la una o más mutaciones pueden estar en un dominio catalítico. En una realización preferida adicional, el dominio catalítico es un dominio RuvC I, RuvC II, RuvC III o HNH. En una realización más preferida, la una o más mutaciones están en un dominio RuvCI o HNH de la enzima de CRISPR. En una realización preferida adicional, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9 y la mutación puede estar en una o en posiciones que incluyen pero no se limitan a las posiciones que corresponden a D10A, E762A, H840A, N854A, N863A o D986A con referencia a la numeración de la posición de SpCas9 y/o es una mutación como se comenta de otra manera en esta memoria. En algunas realizaciones, la enzima de CRISPR tiene una o más mutaciones en un dominio particular de la enzima, en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida a la secuencia de tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y en donde la enzima comprende, además, un dominio funcional. El dominio funcional puede incluir, pero no se limita a activador transcripcional, represor transcripcional, una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo, un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente.
- 25 En algunas realizaciones, el dominio funcional es el dominio activador transcripcional VP64. En algunas realizaciones, el dominio funcional es el dominio represor transcripcional KRAB. En algunas realizaciones, el dominio represor transcripcional es SID, o concatámeros de SID (es decir, SID4X). En algunas realizaciones, se proporciona una enzima de modificación epigenética, p. ej., una proteína de modificación histona o una proteína de modificación de cromatina epigenética. En algunas realizaciones, se proporciona un dominio activador, que puede ser el dominio activador P65.
- 30 Un aspecto adicional de la invención comprende métodos de modificar dos o más loci genómicos de interés. En una realización preferida de la invención, dos o más loci genómicos están diferencialmente modulados mediante la utilización de una o más enzimas de CRISPR, p. ej., dos o más ortólogos Cas9, estando enlazado operativamente cada uno de los ortólogos a uno o más dominios funcionales. En un aspecto, la invención proporciona un método para modificar dos o más loci genómicos en una célula eucariota. Por lo tanto, aspectos de la invención proporcionan un método para modular la expresión de dos o más loci genómicos de interés en un organismo, que comprende suministrar una composición que se produce de forma no natural o que se produce mediante ingeniería genética, que comprende un sistema de vector que comprende uno o más vectores que comprenden
- 35 I. un primer elemento regulador enlazado operativamente a una primera secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la primera secuencia de polinucleótidos comprende (i) una primera secuencia guía capaz de hibridarse con una primera secuencia diana en un primer locus genómico en una célula del organismo,

(ii) una primera secuencia de emparejamiento tracr, y
 (iii) una primera secuencia de tracr, y

II. un segundo elemento regulador enlazado operativamente a una segunda secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la segunda secuencia de polinucleótidos comprende

(i) una segunda secuencia guía capaz de hibridarse con una segunda secuencia diana en un segundo locus genómico en la célula del organismo,

(ii) una segunda secuencia de emparejamiento tracr, y

(iii) una segunda secuencia de tracr, y

III. un tercer elemento regulador enlazado operativamente a una secuencia codificadora de enzima que codifica una primera enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear y está enlazada operativamente a un primer dominio funcional,

IV. un cuarto elemento regulador enlazado operativamente a una secuencia codificadora de enzima que codifica una segunda enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear y está enlazada operativamente a un segundo dominio funcional, en donde (i), (ii) y (iii) en **I** y **II** están dispuestos en una orientación 5' a 3', en donde los componentes **I**, **II**, **III** y **IV** se encuentran en los mismos o en diferentes vectores del sistema, en donde, cuando se transcriben, cada una de las secuencias de emparejamiento tracr se hibrida con su secuencia tracr correspondiente y la primera y segunda secuencias guía dirigen la unión específica para la secuencia del primer y segundo complejos de CRISPR a la primera y segunda secuencia diana, en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, y en donde la expresión de la enzima de CRISPR proporciona la manipulación de la secuencia diana, en donde la primera y segunda enzimas de CRISPR comprenden cada una dos o más mutaciones, en donde la primera y segunda enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*, y en donde el primer locus genómico es modulado por la actividad del primer dominio funcional y el segundo locus genómico es modulado por la actividad del segundo dominio funcional. En una realización adicional, el primer dominio funcional se selecciona del grupo que consiste en un activador transcripcional, represor transcripcional, una recombinasa, una transposasa, una remodelador de histona, una ADN metiltransferasa, un criptocromo y un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente. En una realización adicional, el segundo dominio funcional se selecciona del grupo que consiste en un activador transcripcional, represor transcripcional, una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo y un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente. En realizaciones preferidas, la primera o segunda enzima de CRISPR es una Cas9 de *Sutterella wadsworthensis*, una Cas9 de *Filifactor aloicis*, una Cas9 de *Lactobacillus johnsonii*, una Cas9 de *Campylobacter lari*, una Cas9 de *Corynebacter diphtheriae*, una Cas9 de *Parvibaculum lavamentivorans*, una Cas9 de *Mycoplasma gallisepticum*, una Cas9 de *Staphylococcus aureus* subespecie *Aureus* Cas9, una Cas9 de *Legionella pneumophila* Paris, una Cas9 de *Treponema denticola*, una Cas9 de *Staphylococcus pseudintermedius*, una Cas9 de *Neisseria cinerea*.

En algunas realizaciones, la enzima de CRISPR es una enzima de CRISPR de tipo I, II o III, preferentemente una enzima CRISPR de tipo II. Esta enzima de CRISPR de tipo II puede ser cualquier enzima Cas. Una enzima Cas puede ser identificada como Cas9, ya que esto puede referirse a la clase general de enzimas que comparten homología con el mayor nucleasa con múltiples dominios nucleasa del sistema de CRISPR tipo II. Lo más preferiblemente, la enzima Cas9 es, o se deriva de SpCas9 o *Staphylococcus aureus* subespecie *Aureus* SaCas9. Por se deriva se entiende que la enzima derivada se basa en gran medida, en el sentido de tener un alto grado de homología de secuencia con, una enzima de tipo salvaje, pero que ha sido mutada (modificada) de alguna manera tal como se describe en esta memoria.

Se apreciará que las expresiones enzima Cas y CRISPR se utilizan en general en esta memoria de forma indistinta, a menos que de otro modo resulte evidente. Tal como se mencionó anteriormente, muchas de las numeraciones de residuos utilizadas en esta memoria se refieren a la enzima Cas9 del locus CRISPR de tipo II en *Streptococcus pyogenes*. Sin embargo, se apreciará que esta invención incluye muchas más Cas9s de otras especies de microbios tales como los pertenecientes al género *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* o *Campylobacter*, tal como SpCas9, SaCas9, St1Cas9, St3Cas9 y así sucesivamente, en donde St es *Streptococcus thermophilus*.

Un ejemplo de una secuencia optimizada en codones, en este caso optimizada para seres humanos (es decir, optimizada para la expresión en seres humanos) se proporciona en esta memoria, ver la secuencia optimizada en codones humana SaCas9. Si bien esto se prefiere, se apreciará que son posibles otros ejemplos y se conoce la optimización de codones para una especie huésped.

- 5 Aspectos adicionales de la invención se refieren a la especificidad de escisión mejorada, la secuencia tracr optimizada, el ARN guía quimérico optimizado, la estructura co-plegada de tracrRNA y la secuencia de emparejamiento tracr, la estabilización de estructuras secundarias de tracrRNA, tracrRNA con una región acortada de apareamiento de bases, tracrRNA con elementos de ARN fusionados, la clonación y el suministro simplificados, sistemas de toxicidad reducida y/o inducibles. Otro aspecto de la invención se refiere a la estabilización de ARN
- 10 quimérico, y/o secuencia guía y/o una parte del mismo de complejos de CRISPR, en donde la enzima de CRISPR es un ortólogo de CRISPR, en donde el ARN quimérico y/o la secuencia guía y/o una parte del mismo se estabiliza mediante nucleótidos sintéticos o modificados químicamente (p. ej., LNA/BNA; modificación de tiol, modificación de reticulación 2'/3'-OH), se modifica para ser resistentes a la degradación/hidrólisis y al que se han añadido elementos de estabilidad estructural.
- 15 La invención comprende, además, en determinadas realizaciones un método de modificar un organismo o un organismo no humano mediante la manipulación de una secuencia diana en un locus genómico de interés, que comprende suministrar una composición que se produce de forma no natural o que se produce mediante ingeniería genética que comprende un sistema de vectores que comprende uno o más vectores que codifican operativamente una composición comentada en esta memoria para la expresión de los mismos. Preferiblemente, el vector es un
- 20 vector viral, tal como vectores virales lenti- o baculo- o preferiblemente adeno-asociados, pero se conocen y se proporcionan otros medios de suministro (tales como sistemas de levadura, microvesículas, pistolas de genes/medios de fijación de vectores a nanopartículas de oro).

En esta memoria se describe diversos medios de suministro y se comentan en esta sección.

- 25 Suministro Viral: La enzima de CRISPR, por ejemplo una Cas9, y/o cualquiera de los presentes ARN, por ejemplo, un ARN guía, puede ser suministrada utilizando virus adeno-asociado (AAV), lentivirus, adenovirus u otros tipos de vectores virales, o combinaciones de los mismos. Cas9 y uno o más ARN guía pueden ser empaquetados en uno o más vectores virales. En algunas realizaciones, el vector viral es suministrado al tejido de interés, por ejemplo, mediante inyección intramuscular, mientras que otras veces el suministro viral es a través de la vía intravenosa, transdérmica, intranasal, oral, mucosal, u otros métodos de suministro. Una administración de este tipo puede ser o
- 30 bien a través de una sola dosis o de múltiples dosis. Un experto en la técnica entiende que la dosis real a suministrar en esta memoria pueden variar en gran medida dependiendo de una diversidad de factores tales como el vector elegido, la célula, organismo o tejido diana, el estado general del sujeto a tratar, el grado de transformación/modificación buscado, la vía de administración, el modo de administración, el tipo de de transformación/modificación buscado, etc.
- 35 Una dosificación de este tipo puede contener, además, por ejemplo, un soporte (agua, solución salina, etanol, glicerol, lactosa, sacarosa, fosfato de calcio, gelatina, dextrano, agar, pectina, aceite de cacahuete, aceite de sésamo, etc.), un diluyente, un soporte farmacéuticamente aceptable (p. ej., solución salina tamponada con fosfato), un excipiente farmacéuticamente aceptable, un adyuvante para potenciar la antigenicidad, un compuesto o molécula inmunoestimulador y/u otros compuestos conocidos en la técnica. El adyuvante en esta memoria puede contener
- 40 una suspensión de minerales (alumbre, hidróxido de aluminio, fosfato de aluminio) sobre los que se adsorbe el antígeno, o una emulsión de agua-en-aceite en la que la disolución de antígeno se emulsiona en aceite (MF-59, adyuvante incompleto de Freund), a veces con la inclusión de micobacterias muertas (adyuvante completo de Freund) para mejorar adicionalmente la antigenicidad (inhibe la degradación del antígeno y/o provoca la afluencia de macrófagos). Los adyuvantes también incluyen moléculas inmunoestimulantes tales como citoquinas, moléculas coestimulantes y, por ejemplo, moléculas de ADN o de ARN inmunoestimulantes tales como oligonucleótidos CpG. Una formulación de dosificación de este tipo es fácilmente determinable por un experto en la técnica. La dosificación puede contener, además, una o más sales farmacéuticamente aceptables tales como, por ejemplo, una sal de ácido mineral tal como un hidrocloreuro, un hidrobromuro, un fosfato, un sulfato, etc.; y las sales de ácidos orgánicos tales como acetatos, propionatos, malonatos, benzoatos, etc. Adicionalmente, también pueden estar presentes sustancias auxiliares tales como agentes humectantes o emulsionantes, sustancias tamponadoras del pH, geles o materiales
- 45 gelificantes, aromas, colorantes, microesferas, polímeros, agentes de suspensión, etc. Además, también pueden estar presentes uno o más de otros ingredientes farmacéuticos convencionales tales como conservantes, humectantes, agentes de suspensión, tensioactivos, antioxidantes, agentes antiapelmazantes, cargas, agentes quelantes, agentes de revestimiento, estabilizadores químicos, etc., especialmente si la forma de dosificación es una
- 50 forma reconstituible. Ingredientes a modo de ejemplo adecuados incluyen celulosa microcristalina,
- 55

carboximetilcelulosa de sodio, polisorbato 80, alcohol fenilético, clorobutanol, sorbato de potasio, ácido sórbico, dióxido de azufre, galato de propilo, los parabenos, vainillina de etilo, glicerol, fenol, paraclorofenol, gelatina, albúmina y una combinación de los mismos. Una discusión a fondo de excipientes farmacéuticamente aceptables está disponible en REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991).

5 En una realización en esta memoria, el suministro es a través de un adenovirus, que puede ser en una dosis de refuerzo única que contiene al menos 1×10^5 partículas (a la que también se alude como unidades de partículas, pu) de vector adenoviral. En una realización en esta memoria, la dosis es preferiblemente de al menos aproximadamente 1×10^6 partículas (por ejemplo, aproximadamente 1×10^6 - 1×10^{12} partículas), más preferiblemente al menos aproximadamente 1×10^7 partículas, más preferiblemente al menos aproximadamente 1×10^8 partículas (p. ej., aproximadamente 1×10^8 - 1×10^{11} partículas o aproximadamente 1×10^8 - 1×10^{12} partículas), y lo más preferiblemente al menos aproximadamente 1×10^9 partículas (p. ej., aproximadamente 1×10^9 - 1×10^{12} partículas), o incluso al menos aproximadamente 1×10^{10} partículas (p. ej., aproximadamente 1×10^{10} - 1×10^{12} partículas) del vector adenoviral. Alternativamente, la dosis comprende no más de aproximadamente 1×10^{14} partículas, preferiblemente no más de aproximadamente 1×10^{13} partículas, incluso más preferiblemente no más de aproximadamente 1×10^{12} partículas, incluso más preferiblemente no más de aproximadamente 1×10^{11} partículas, y lo más preferiblemente no más de aproximadamente 1×10^{10} partículas (p. ej., no más de aproximadamente 1×10^9 partículas). Por lo tanto, la dosis puede contener una sola dosis de vector adenoviral con, por ejemplo, aproximadamente 1×10^6 unidades de partículas (pu), aproximadamente 2×10^6 pu, aproximadamente 4×10^6 pu, aproximadamente 1×10^7 pu, aproximadamente 2×10^7 pu, aproximadamente 4×10^7 pu, aproximadamente 1×10^8 pu, aproximadamente 2×10^8 pu, aproximadamente 4×10^8 pu, aproximadamente 1×10^9 pu, aproximadamente 2×10^9 pu, aproximadamente 4×10^9 pu, aproximadamente 1×10^{10} pu, aproximadamente 2×10^{10} pu, aproximadamente 4×10^{10} pu, aproximadamente 1×10^{11} pu, aproximadamente 2×10^{11} pu, aproximadamente 4×10^{11} pu, aproximadamente 1×10^{12} pu, aproximadamente 2×10^{12} pu, aproximadamente 4×10^{12} pu, del vector adenoviral. Véanse, por ejemplo, los vectores adenovirales en la patente de EE.UU. N° 8.454.972 B2 expedida a Nabel, et. al., concedida el 4 de junio de 2013, y las dosis en la col. 29, líneas 36-58 de la misma. En una realización en esta memoria, el adenovirus se suministra a través de múltiples dosis.

En una realización en esta memoria, el suministro es a través de un AAV. Una dosis terapéuticamente eficaz para el suministro in vivo del AAV a un ser humano se cree que está en el intervalo de aproximadamente 20 a aproximadamente 50 ml de solución salina que contiene de aproximadamente 1×10^{10} a aproximadamente 1×10^{10} AAV funcional/ml de disolución. La dosificación puede ser ajustada para equilibrar el beneficio terapéutico contra efectos secundarios. En una realización en esta memoria, la dosis de AAV está generalmente en el intervalo de concentraciones de aproximadamente 1×10^5 hasta 1×10^{10} genomas de AAV, de aproximadamente 1×10^8 a 1×10^{20} genomas de AAV, de aproximadamente 1×10^{10} a aproximadamente 1×10^{16} genomas, o aproximadamente 1×10^{11} a aproximadamente 1×10^{16} genomas de AAV. Una dosificación humana puede ser de aproximadamente 1×10^{13} genomas de AAV. Tales concentraciones pueden ser suministradas en de aproximadamente 0,001 ml a aproximadamente 100 ml, DE aproximadamente 0,05 a aproximadamente 50 ml o de aproximadamente 10 a aproximadamente 25 ml de una disolución de soporte. Otras dosificaciones eficaces pueden establecerse fácilmente por un experto ordinario en la técnica a través de ensayos de rutina que establecen curvas de dosis-respuesta. Véase, por ejemplo, la patente de EE.UU. N° 8.404.658 B2 expedida a Hajjar, et al., concedida el 26 de marzo de 2013, en la col. 27, líneas 45-60.

En una realización en esta memoria, el suministro es a través de un plásmido. En tales composiciones de plásmidos, la dosis debe ser una cantidad suficiente de plásmido para inducir una respuesta. Por ejemplo, cantidades adecuadas de ADN de plásmido en composiciones de plásmidos puede ser de aproximadamente 0,1 a aproximadamente 2 mg, o de aproximadamente 1 µg a aproximadamente 10 µg.

45 Las dosis en esta memoria se basan en una media de un individuo de 70 kg. La frecuencia de administración está dentro del ámbito del profesional médico o veterinario (p. ej., médico, veterinario), o un científico experto en la técnica.

Los vectores virales pueden ser inyectados en el tejido de interés. Para la modificación del genoma específico para el tipo de célula, la expresión de Cas9 puede ser impulsada por un promotor específico de tipo celular. Por ejemplo, 50 la expresión específica para el hígado podría utilizar el promotor de albúmina y la expresión específica para las neuronas podría utilizar el promotor Sinapsina I.

Suministro de ARN: La enzima de CRISPR, por ejemplo, una Cas9, y/o cualquiera de los presentes ARN, por ejemplo un ARN guía, también se puede suministrar en forma de ARN. ARNm de Cas9 se pueden generar utilizando

la transcripción in vitro. Por ejemplo, ARNm de Cas9 se puede sintetizar utilizando una casete de PCR que contiene los siguientes elementos: secuencia T7_promotor-kozak (GCCACC)-Cas9-3' UTR de la cola beta globina-poliA (una cadena de 120 o más adeninas). La casete se puede utilizar para la transcripción por la T7polimerasa. ARNs guía también se pueden transcribir utilizando la transcripción in vitro a partir de una casete que contiene la secuencia T7_promotor-GG-ARN guía.

5

Para potenciar la expresión y reducir la toxicidad, la enzima CRISPR y/o el ARN guía se pueden modificar utilizando pseudo-U o 5-Metil-C.

ARNm de la enzima de CRISPR y ARN guía se pueden administrar simultáneamente utilizando nanopartículas o envueltas lipídicas.

10 Por ejemplo, Su X, Fricke I, Kavanagh DG, Irvine DJ ("In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm 6 de junio de 2011; 8(3):774-87 doi:10.1021/mp100390w Epub 1 de abril de 2011) describe nanopartículas de núcleo y envoltura estructuradas biodegradables con un núcleo poli (β -amino éster) (PBAE) envuelto por una envoltura bicapa de fosfolípidos. Estos fueron desarrollados para el suministro in vivo de ARNm. El componente PBAE sensible al pH fue escogido para fomentar la interrupción del endosoma, mientras que la capa de la superficie de lípidos se seleccionó para minimizar la toxicidad del núcleo de polimerización. Por lo tanto, éstos se prefieren para el suministro de ARN de la presente invención.

15

Además de ello, Michael S.D. Kormann et al. ("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: Nature Biotechnology, Volumen: 29, páginas: 154-157 (2011) publicado en línea el 9 de enero de 2011) describen el uso de envolturas de lípidos para suministrar ARN. El uso de envolturas de lípidos también se prefiere en la presente invención.

20

Métodos de suministro de ARNm son especialmente prometedores para suministrar actualmente al hígado.

ARNm de la enzima de CRISPR ARN guía también podrían ser suministrados por separado. ARNm de la enzima de CRISPR puede ser suministrado antes del ARN guía para dar tiempo a que la enzima de CRISPR se exprese. ARNm de la enzima de CRISPR puede ser administrada 1-12 horas (preferiblemente alrededor de 2-6 horas) antes de la administración de ARN guía.

25

Alternativamente, ARNm de la enzima de CRISPR y ARN guía se pueden administrar juntos. Ventajosamente, una segunda dosis de refuerzo de ARN guía se puede administrar 1 -12 horas (preferiblemente alrededor de 2-6 horas) después de la administración inicial de ARNm de la enzima de CRISPR + ARN guía.

Administraciones adicionales de ARNm de la enzima de CRISPR y/o ARN guía podrían ser útiles para alcanzar los niveles más eficientes de modificación del genoma.

30

Para la minimización de la toxicidad y los efectos fuera de objetivo, será importante controlar la concentración de ARNm de la enzima de CRISPR y el ARN guía suministrado. Concentraciones óptimas del ARNm de la enzima de CRISPR y ARN guía se pueden determinar mediante el ensayo de diferentes concentraciones en un modelo celular o animal y se puede utilizar la secuenciación profunda para analizar el grado de modificación en potenciales loci genómicos fuera de objetivo. Por ejemplo, para la secuencia guía que fija como objetivo 5'-GAGTCCGAGCAGAAGAAGAA-3' en el gen EMX1 del genoma humano, la secuenciación profunda se puede utilizar para evaluar el nivel de modificación en los dos loci siguientes fuera de objetivo, 1: 5'-GAGTCTAGCAGGAGAAGAA-3' y 2: 5'-GAGTCTAAGCAGAAGAAGAA-3'. La concentración que da el mayor nivel de modificación en objetivo al tiempo que minimiza el nivel de modificación fuera de objetivo debe ser elegido para el suministro in vivo.

35

40

Alternativamente, para minimizar el nivel de toxicidad y los efectos fuera de objetivo, la enzima de CRISPR nickase ARNm (por ejemplo Cas9 de *S. pyogenes* con la mutación D10A) puede ser suministrada con un par de ARNs guía que fijan como objetivo un sitio de interés. Los dos ARN guías tienen que estar separados de la siguiente manera. Secuencias guía en rojo (subrayado sencillo) y azul (subrayado doble), respectivamente (estos ejemplos se basan en el requisito del PAM para Cas9 de *Streptococcus pyogenes*).

45

ES 2 576 126 T3

Longitud del colgante (pb)	Diseño de ARN guía (secuencia guía y PAM codificados en color)
	5' -NNNNNNNNNNNNNNNNNNNNCCNGNNNNNNNNNNNNNGGNNNNNNNNNN-3'
14	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNGGNNNNNNNNNN-3'
13	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNGGNNNNNNNNNN-3'
12	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNnNGCNNNNNNNNNN-3'
11	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNGGNNNNNNNNNN-3
10	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNGGNNNNNNNNNN-3'
9	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNGGNNNNNNNNNN-3'
8	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNCCNNNNNNNNNN-5'

ES 2 576 126 T3

Longitud del colgante (pb)	Diseño de ARN guía (secuencia guía y PAM codificados en color)
	5'-NNNNNNNNNNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNNNNNN-3'
7	3'-NNNNNNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNN-5'
	5' NNNNNNNNNNNNNNNNNNN <u>CC</u> NNNNNNNNNN <u>GG</u> NNNNNNNNNNNNNNNNNN-3'
6	3'-NNNNNNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNN-3'
5	3' -NNNNNNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNN <u>GG</u> NNNNNNNNNNNNNNNNNN-3'
4	3' -NNNNNNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNN <u>CC</u> NNNNNNNNNN <u>GG</u> NNNNNNNNNNNNNNNNNN-3'
3	3' -NNNNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNNNNNNNNNN-3'
2	3' -NNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNNNNNNNNNN-3'
1	3' -NNNNNNNNNNNNNN <u>GG</u> NNNNNNNNNNNNNN <u>CC</u> NNNNNNNNNNNNNNNNNN-5'

ES 2 576 126 T3

Longitud del colgante (pb)	Diseño de ARN guía (secuencia guía y PAM codificados en color)
	5' -NNNNNNNNNNNNNNCCNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNN-3'
romo	
	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNN-3'
1	3' -NNNNNNNNNNNNNNNNNNNNNGGNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNN-3'
2	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNN-3'
3	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNN-3'
4	3' -NNNNNNNNNNNNNNNNNNNNGGNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3'
5	3' -NNNNNNNNNNNNNNNNNNNNGGNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'
	5' -NNNNNNNNNNNNNNNNNNNNCCNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3'
6	3' -NNNNNNNNNNNNNNNNNNNNGGNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNN-5'

ES 2 576 126 T3

Longitud del colgante (pb)	Diseño de ARN guía (secuencia guía y PAM codificados en color)
	5' - <u>NNNNNNNNNNNCCNGG</u> NN-3'
7	3' - <u>NNNNNNNNNNNNGGNCC</u> NN-5'
	5' - <u>NNNNNNNNNNNCCGG</u> NN-3'
8	3' - <u>NNNNNNNNNNNNNNGGCC</u> NN-5'
	5' - <u>NNNNNNNNNNNNNNNNNNNGG</u> NN- 3'
12	3' - <u>NNNNNNNNNNNNNNNNNNNCCGG</u> NN-5'
	5' - <u>NNNNNNNNNNNNNNNNNNNCCGG</u> NN-3'
13	3' - <u>NNNNNNNNNNNNNNNNNNNCCNGG</u> NN-5'
	5' - <u>NNNNNNNNNNNNNNNNNNNNC</u> GGNN-3'
14	3' - <u>NNNNNNNNNNNNNNNNNNNCCN</u> GGNN-5'
	5' - <u>NNNNNNNNNNNNNNNNNNNNC</u> GGNN-3'
15	3' - <u>NNNNNNNNNNNNNNNNNNNCCN</u> NGGNN-5'
	5' - <u>NNNNNNNNNNNNNNNNC</u> GGNN-3'
16	3' - <u>NNNNNNNNNNNNNNNNNNNCC</u> NNNNNGGNN-5'

Longitud del colgante (pb)	Diseño de ARN guía (secuencia guía y PAM codificados en color)
	5'-NNNNNNNNNNNNNNNNNNNNNN <u>CGG</u> NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN-3'
17	3'-NNNNNNNNNNNNNNNNNNNNNNNN <u>CC</u> NNNNNN <u>G</u> NNNNNNNNNNNNNNNNNNNNNN-5'

- La interrogación adicional del sistema ha dado a la solicitante evidencia del colgante 5' (véase, p. ej., Ran et al, Cell. 12 de septiembre de 2013;154(6):1380-9 y Solicitud de Patente Provisional de EE.UU. N° de Serie 61/871.301 presentada el 28 de agosto de 2013). La solicitante ha identificado, además, parámetros que se refieren a la escisión eficaz por parte del mutante Cas9 nickase cuando se combina con dos ARNs guía y estos parámetros incluyen, pero no se limitan a la longitud del colgante 5'. En realizaciones de la invención, el colgante 5' es a lo sumo de 200 pares de bases, preferiblemente a lo sumo de 100 pares de bases, o más preferiblemente a lo sumo de 50 pares de bases. En realizaciones de la invención, el colgante 5' es de al menos 26 pares de bases, preferiblemente de al menos 30 pares de bases o más preferiblemente de 34 a 50 pares de bases o de 1-34 pares de bases. En otros métodos preferidos de la invención, la primera secuencia guía que dirige la escisión de una cadena del dúplex de ADN cerca de la primera secuencia diana y la segunda secuencia de guía dirige la escisión de otra cadena cerca de la segunda secuencia diana da como resultado un corte como o un colgante 3'. En realizaciones de la invención, el colgante 3' es a lo sumo de 150, 100 ó 25 pares de bases o al menos de 15, 10 ó 1 pares de bases. En realizaciones preferidas, el colgante 3' es de 1-100 pares de bases.
- Aspectos de la invención se refieren a la expresión del producto génico que se encuentra disminuido o siendo introducido adicionalmente un polinucleótido molde en la molécula de ADN que codifica el producto génico o siendo escindida una secuencia intermedia precisamente al permitir que los dos colgantes 5' se reasocien o ligen o siendo alterada la actividad o función del producto génico, o siendo aumentada la expresión del producto génico. En una realización de la invención, el producto génico es una proteína.
- Sólo los pares de sgRNA que crean colgantes 5' con menos de 8 pb se solapan entre las secuencias guía (compensación mayor que -8 pb) fueron capaces de mediar en la formación de indel detectable. De manera importante, cada una de las guías utilizadas en estos ensayos es capaz de inducir de manera eficiente indeles cuando se combina con Cas9 de tipo salvaje, lo que indica que las posiciones relativas de los pares de guía son los parámetros más importantes en la predicción de la actividad de doble mella.
- Dado que Cas9n y Cas9H840A mellan cadenas opuestas de ADN, la sustitución de Cas9n con Cas9H840A con un par de sgRNA dados debería dar lugar a la inversión del tipo de colgante. Por ejemplo, un par de sgRNAs que generará un colgante 5' con Cas9n debería generar, en principio, el correspondiente colgante 3' en su lugar. Por lo tanto, los pares de sgRNA que conducen a la generación de un colgante 3' con Cas9n podrían utilizarse con Cas9H840A para generar un colgante 5'. Inesperadamente, la solicitante sometieron a ensayo Cas9H840A con un conjunto de pares de sgRNA diseñados para generar colgantes tanto 5' como 3' (intervalo de compensaciones de -278 a +58 pb), pero no fueron capaces de observar la formación de indeles. Puede ser necesario un trabajo adicional para identificar las reglas de diseño necesarias para el apareamiento de sgRNA para permitir un doble mellado por Cas9H840A.
- Opciones de suministro adicionales al cerebro incluyen la encapsulación de la enzima de CRISPR y ARN guía en forma de ADN o ARN en los liposomas y la conjugación a caballos de Troya moleculares para el suministro a la barrera hematoencefálica (BBB). Caballos de Troya moleculares han demostrado ser eficaces para el suministro de vectores de expresión B-gal en el cerebro de primates no humanos. El mismo enfoque se puede utilizar para el suministro de vectores que contienen enzima de CRISPR y ARN guía. Por ejemplo, Xia CF y Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology" Mol Pharm. mayo-junio de 2009; 6(3):747-51 doi : 10.1021/mp800194) describe cómo es posible la entrega de ARN corto de interferencia (siRNA) a células en cultivo, e in vivo, con el uso combinado de un anticuerpo monoclonal (mAb) específico para el receptor y la tecnología avidina-biotina. Los autores también informan que, debido a que la unión entre el mAb de fijación como objetivo y el siRNA es estable con la tecnología de avidina-biotina, y se observan efectos de RNAi en sitios distantes tales como el cerebro in vivo después de la administración intravenosa del siRNA fijado como objetivo.

Zhang Y, Schlachetzki F, Pardridge WM. ("Global non-viral gene transfer to the primate brain following intravenous administration" *Mol Ther.* enero de 2003; 7(1):11-8.) describen cómo la expresión de plásmidos que codifican informadores tales como luciferasa fueron encapsulados en el interior de un "virus artificial" constituido por inmunoliposoma 85 nm pegilado, que fue fijado como objetivo al cerebro de mono rhesus in vivo con un anticuerpo monoclonal (MAb) al receptor de la insulina huma (HIR). El HIRMAb permite que el liposoma que porta el gen exógeno se someta a transcitosis a través de la barrera sangre-cerebro y la endocitosis a través de la membrana plasmática neuronal después de la inyección intravenosa. El nivel de expresión del gen de luciferasa en el cerebro era 50 veces más alto en el mono rhesus en comparación con la rata. La expresión neuronal general del gen beta-galactosidasa en el cerebro de los primates fue demostrada tanto por histoquímica como por microscopía confocal. Los autores indican que este enfoque hace factibles transgénicos adultos reversibles en 24 horas. Por consiguiente, se prefiere el uso de inmunoliposoma. Éste se puede utilizar en unión con anticuerpos para fijar como objetivo tejidos diana específicos o proteínas de la superficie celular.

También se prefieren otros medios de suministro o de ARN tales como a través de nanopartículas (Clio, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R, y Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endotelial cells, *Advanced Functional Materials*, 19:3112-31 18, 2010) o exosomas (Schroeder, A., Levins, C., Cortez, C., Langer, R, y Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, *Journal of Internal Medicine*, 267: 9-21, 2010, PMID: 20059641). De hecho, los exosomas han demostrado ser particularmente útiles en el suministro de siRNA, un sistema con algunos paralelismos con el sistema de CRISPR. Por ejemplo, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA in vitro and in vivo" *Nat Protoc.* diciembre de 2012;7(12):2112-26 doi: 10.1038/nprot.2012.131 Epub 15 de noviembre de 2012) describen cómo los exosomas son herramientas prometedoras para el suministro de fármacos a través de diferentes barreras biológicas y pueden ser aprovechados para el suministro de siRNA in vitro e in vivo. Su enfoque es generar exosomas fijados como objetivo a través de transfección de un vector de expresión, que comprende una proteína exosomal fusionada con un ligando de péptido. Los exosomas son entonces purificados y caracterizados a partir del sobrenadante de células transfectadas y, a continuación, el siRNA se carga en los exosomas.

Un aspecto de la manipulación de una secuencia diana también se refiere a la manipulación epigenética de una secuencia diana. Esto puede ser del estado de cromatina de una secuencia diana tal como por modificación del estado de metilación de la secuencia diana (es decir, adición o eliminación de metilación o patrones de metilación o islas CpG), modificación de histona, el aumento o la reducción de la accesibilidad a la secuencia diana, o mediante la promoción o la reducción de plegado 3D.

Un vector puede significar no sólo un sistema viral o de levadura (por ejemplo, en los casos en los que los ácidos nucleicos de interés pueden estar operativamente enlazados a y bajo el control de (en términos de expresión, tal como para proporcionar en última instancia un ARN procesado) un promotor, sino que también un suministro directo de ácidos nucleicos a una célula huésped.

La invención comprende también, en determinadas realizaciones, un método para tratar o inhibir una afección provocada por un defecto en una secuencia diana en un locus genómico de interés en un sujeto o un sujeto no humano en necesidad del mismo, que comprende la modificación del sujeto o del sujeto no humano mediante la manipulación de la secuencia diana y en el que la afección es susceptible de tratamiento o la inhibición por manipulación de la secuencia diana que comprende proporcionar un tratamiento que comprende: suministrar una composición que se produce de forma no natural o modificada por ingeniería genética que comprende un sistema de vectores que comprende uno o más vectores que comprende codificar de manera operativa una composición de la presente memoria comentada para la expresión de los mismos, en el que la secuencia diana es manipulada por la composición cuando se expresa.

En determinadas realizaciones de los métodos de esta memoria, los métodos pueden incluir la expresión de la inducción, que puede ser inducir la expresión de la enzima de CRISPR y/o inducir la expresión de las secuencias guía, tracr o de emparejamiento tracr. En determinadas realizaciones de los métodos en esta memoria, el organismo o sujeto es un eucariota o un eucariota no humano. En determinadas realizaciones de los métodos en esta memoria, el organismo o sujeto es una planta. En determinadas realizaciones de los métodos en esta memoria, el organismo o sujeto es un mamífero o un mamífero no humano. En determinadas realizaciones de los métodos en esta memoria, el organismo o sujeto son algas.

Mientras que en los métodos de esta memoria el vector puede ser un vector viral y éste es ventajosamente un AAV, se pueden emplear otros vectores virales tal como se describen en esta memoria. Por ejemplo, se pueden utilizar baculovirus para la expresión en células de insectos. Estas células de insectos pueden, a su vez, ser útiles para

producir grandes cantidades de otros vectores, tales como vectores de AAV adaptados para el suministro de la presente invención.

5 Asimismo, se prevé un método para suministrar la presente enzima de CRISPR, que comprende suministrar a la célula un ARNm que codifican la enzima de CRISPR. Se apreciará que la enzima de CRISPR está truncada, es de un tamaño específico según se describe en esta memoria, es una nucleasa o nickase o una proteína de unión a ADN genérico, está optimizada en codones, comprende una o más mutaciones y/o comprende una enzima de CRISPR quimérica, o las otras opciones como se discuten en esta memoria.

Asimismo, se prevé un método para preparar un vector para el suministro de las composiciones o las presentes enzimas de CRISPR de la invención y para su uso en los presentes métodos.

10 Se prefieren los vectores virales AAV. Por lo tanto, en un aspecto adicional, se proporciona un método de preparar un vector viral AAV, que comprende transfectar un plásmido o plásmidos que contienen o que consisten esencialmente en una molécula o moléculas de ácido nucleico que codifican el AAV en células de infectadas con AAV, y suministrar AAV rep y/o cap obligatorio para la replicación y el empaquetamiento del AAV. A este respecto, se apreciará que la enzima de CRISPR está truncada, está compuesto por menos de mil aminoácidos o menos de
15 cuatro mil aminoácidos, es una nucleasa o nickase, está optimizada en codones, comprende una o más mutaciones y/o comprende una enzima de CRISPR quimérica tal como se comenta en esta memoria. En algunas realizaciones, el AAV rep y/o cap obligatorio para la replicación y el empaquetamiento del AAV se suministra mediante la transfección de las células con plásmido o plásmidos helper (cooperadores). En algunas realizaciones, el virus helper es un virus de la viruela, adenovirus, virus herpes o baculovirus. En algunas realizaciones, el virus de la viruela es un virus vacuna. En algunas realizaciones las células son células de mamíferos. Y en algunas
20 realizaciones las células son células de insectos y el virus helper es baculovirus.

La invención comprende además en determinadas realizaciones una enzima de CRISPR modificada. Diferencias con la enzima de CRISPR de tipo salvaje pueden comprender: la enzima de CRISPR modificada está truncada en comparación con una enzima de CRISPR de tipo salvaje, o la enzima de CRISPR es de un tamaño específico, p. ej.,
25 de al menos 500 aminoácidos, al menos 800-899 aminoácidos, al menos 900-999 aminoácidos, al menos 1000-1099 aminoácidos, al menos 1100-1199 aminoácidos, al menos 1200-1299 aminoácidos, al menos 1300-1399 aminoácidos, al menos 1400-1499 aminoácidos, al menos 1500-1599 aminoácidos, al menos 1600-1699 aminoácidos o al menos 2000 aminoácidos; y/o la enzima de CRISPR es una nucleasa que dirige la escisión de ambas cadenas en la ubicación de la secuencia diana, o la enzima de CRISPR es una nickase que dirige la escisión de una cadena en el lugar de la secuencia diana, o la enzima de CRISPR es un mutante catalítico nulo que funciona como una proteína de unión de ADN; y/o la enzima de CRISPR está optimizada en codones o está optimizada en codones para la expresión en una célula eucariota, y/o la enzima de CRISPR comprende una o más mutaciones y/o la enzima de CRISPR comprende una enzima de CRISPR quimérica y/o la enzima de CRISPR tiene uno o más de otros atributos comentados en la presente memoria. Por consiguiente, en determinadas realizaciones, la enzima de
35 CRISPR comprende una o más mutaciones en residuos catalíticos tales como spCa9 D10A, E762A, H840A, N854A, N863A o D986A o los correspondientes a ellos en otras enzimas Cas (tal como se describe en esta memoria) con referencia a la numeración de la posición de SpCas9, y/o tiene una o más mutaciones se encuentra en un dominio RuvC I, RuvC II, RuvC III, HNH u otro dominio descrito en esta memoria de la enzima de CRISPR y/o la enzima de CRISPR tiene una o más mutaciones en un dominio catalítico, en el que, cuando se transcribe, una secuencia de emparejamiento tracr se hibrida con una secuencia tracr y una secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a una secuencia diana, y en donde la enzima comprende, además, un dominio funcional. El dominio funcional puede incluir, pero no se limita a activador transcripcional, represor transcripcional, una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo, un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente. La enzima de CRISPR en determinadas realizaciones puede tener el dominio funcional que sea un dominio activador transcripcional, p. ej., VP64. En algunas realizaciones, un dominio represor transcripcional es KRAB. En algunas realizaciones, un dominio represor de la transcripción es SID, o concatámeros de SID (es decir SID4X). En algunas realizaciones, se proporciona una enzima modificadora epigenética.

En realizaciones preferidas, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9.

50 Aspectos de la invención también comprenden identificar nuevos ortólogos de enzimas de CRISPR. Los métodos para identificar nuevos ortólogos de enzimas CRISPR pueden implicar identificar secuencias tracr en los genomas de interés. La identificación de secuencias tracr puede estar relacionada con los siguientes pasos: Búsqueda de las repeticiones directas o secuencias de emparejamiento tracr en una base de datos para identificar una región

5 CRISPR que comprende una enzima de CRISPR, p. ej., Figura 18. Búsqueda de secuencias homólogas en la región de CRISPR que flanquean la enzima de CRISPR tanto en las direcciones sentido como antisentido. Búsqueda de terminadores de la transcripción y estructuras secundarias. Identificar cualquier secuencia que no sea una repetición directa o una secuencia de emparejamiento tracr, pero que tiene más de un 50% de identidad con la repetición directa o la secuencia de emparejamiento como una secuencia tracr potencial. Tomar la secuencia tracr potencial y analizar las secuencias de terminador de la transcripción asociados con la misma.

10 La invención comprende en determinadas realizaciones el uso de una composición como se describe en esta memoria o una enzima de CRISPR en la medicina. La invención comprende además, en determinadas realizaciones una composición o enzima de CRISPR de la invención utilizada en un método de la invención. La invención comprende también, en determinadas realizaciones, el uso de una composición o una enzima de CRISPR de la invención, preferiblemente ex vivo, en la edición de un gen o genoma o, preferiblemente un método ex vivo para la edición de un gen o genoma. La invención también comprende en determinadas realizaciones el uso de una composición de acuerdo con o una enzima de CRISPR de la invención en la fabricación de un medicamento para la edición ex vivo de un gen o genoma o para uso en un método tal como se comenta en esta memoria.

15 Además, la invención, en determinadas realizaciones, comprende una composición o una enzima de CRISPR de la invención, en donde la secuencia diana está flanqueada o es seguida, en su extremo 3', por un PAM adecuado para la enzima de CRISPR, típicamente una Cas y, en particular, una Cas9. Esta secuencia de PAM es específica para cada una de las Cas9, pero se puede determinar fácilmente por métodos descritos en esta memoria.

20 Por ejemplo, una PAM adecuada es 5'-NRG o 5'-NNGRR para las enzimas (o derivados de enzimas) SpCas9 o SaCas9, respectivamente. Se apreciará que la referencia hecha en esta memoria a *Staphylococcus aureus* incluye preferiblemente *Staphylococcus aureus subespecie aureus*.

25 Por consiguiente, es un objeto de la invención consiste no abarcar dentro de la invención alguno producto, proceso de fabricar el producto o método de utilizar el producto de forma conocida con anterioridad, de manera que la solicitante se reserva el derecho y, con ello, revela un descargo de responsabilidad de cualquier producto, proceso o método previamente conocido. Se señala, además, que la invención no pretende abarcar dentro del alcance de la invención cualquier producto, proceso o fabricación del producto o método de utilización del producto, que no cumpla con los requisitos de la descripción escrita y de habilitación de USPTO (35 USC § 112, párrafo primero) o la OEP (artículo 83 del CPE), de tal manera que la solicitante se reserva el derecho y con ello revela un descargo de responsabilidad de cualquier producto, proceso de fabricar el producto o el método de utilizar el producto descrito anteriormente.

35 Hay que señalar que en esta descripción y en particular en las reivindicaciones y/o párrafos, términos y expresiones tales como "comprende", "que comprende", "comprendiendo" y similares pueden tener el significado que se le atribuye en la ley de patentes de Estados Unidos; p. ej., puede significar "incluye", "incluido", "que incluye", y similares; y que términos tales como "que consiste esencialmente en" y "consiste esencialmente en" tienen el significado que se les atribuye en la ley de patentes de Estados Unidos, p. ej., permiten elementos no citados explícitamente, pero excluyen elementos que se encuentran en la técnica anterior o que afectan a una base o característica nueva de la invención.

Estas y otras realizaciones se describen o son evidentes a partir de y quedan abarcadas por la siguiente descripción detallada.

40 BREVE DESCRIPCIÓN DE LOS DIBUJOS

Las características nuevas de la invención se exponen con particularidad en las reivindicaciones adjuntas. Una mejor comprensión de las características y ventajas de la presente invención se obtendrá haciendo referencia a la siguiente descripción detallada que recoge realizaciones ilustrativas, en las que se utilizan los principios de la invención, y los dibujos adjuntos de los cuales:

45 La **Figura 1** muestra un esquema de nucleasa Cas9 guiada por ARN. La nucleasa Cas9 de *Streptococcus pyogenes* (amarillo) fija como objetivo el ADN genómico mediante un ARN guía sintético (SgRNA) que consiste en una secuencia guía de 20 nt y un armazón. Los pares de bases de la secuencia guía la diana de ADN, directamente aguas arriba de un motivo adyacente al protoespaciador 5'-NGG requerido (PAM) y Cas9 media en una rotura de doble cadena (DSB) ~ 3 pb aguas arriba del PAM (indicado por un triángulo).

- Las **Figuras 2A-2F** muestran un sistema CRISPR a modo de ejemplo y un posible mecanismo de acción (A), una adaptación de ejemplo para la expresión en células eucariotas, y los resultados de las pruebas que evalúan la localización nuclear y la actividad de CRISPR (B-F).
- 5 La **Figura 3** muestra una representación esquemática de plásmidos de suministro in vivo de AAV utilizando secuencias de repeticiones terminales invertidas (ITRs) y ARNs quiméricos guía para preferiblemente ayudar en el suministro por AAV o sistemas asociados a AAV.
- Las **Figuras 4A-D** muestran una representación circular del análisis filogenético revelando cinco familias de Cas9s, incluyendo tres grupos de Cas9s grandes (~ 1400 aminoácidos) y dos de Cas9 pequeñas (~ 1100 aminoácidos).
- 10 Las **Figuras 5 A-F** muestran una representación lineal del análisis filogenético que revelan cinco familias de Cas9s, incluyendo tres grupos de Cas9s grandes (~ 1400 aminoácidos) y dos de Cas9 pequeñas (~ 1100 aminoácidos).
- La **Figura 6** muestra un gráfico que representa la distribución de la longitud de ortólogos Cas9.
- La **Figura 7** muestra una representación de la logotipos de la secuencia de los PAMs de los ortólogos Cas9 como complementos inversos.
- 15 Las **Figuras 8 A-J** muestran 18 estructuras de ARN quimérico que conservaron la secuencia y estructuras secundarias del dúplex de secuencias emparejamiento tracr:tracr, al tiempo que se acorta la región de apareamiento basal y la fusión de los dos elementos de ARN a través de un bucle artificial.
- Las **Figuras 9 A-O** muestran una lista de secuencias Cas9 optimizadas en codones, humanas, a emparejarse con los ARNs guía quiméricos proporcionados en las Figuras 8 A-J.
- Las **Figuras 10 A-M** muestran secuencias en las que se ubican los puntos de mutación dentro del gen SpCas9.
- 20 La **Figura 11** muestra loci de CRISPR de Tipo II en diferentes organismos.
- La **Figura 12** muestra las secuencias guía ARN correspondientes a CRISPR loci en diferentes organismos.
- La **Figura 13 A-II** muestra una tabla que lista los ortólogos Cas9 y sus secuencias PAM correspondientes.
- La **Figura 14** muestra que el PAM para Staphylococcus aureus subespecie Aureus Cas9 es NNGRR.
- Las **Figuras 15A-D** muestran diseños de vectores sencillos y múltiples para SaCas9.
- 25 La **Figura 16** muestra un diseño del vector y las imágenes en gel de ortólogos Cas9 y utilizándose respectivos sgRNAs para escindir dos dianas candidatas presentes en un banco basado en pUC19.
- La **Figura 17** muestra la escisión in vitro mediante quimeras SpCas9, St3Cas9, Sp_St3 y quimer St3_Sp. Los PAM para Cas9 St3Cas9 y St3_Sp quiméricas son NGG.
- La **Figura 18** muestra una imagen del servidor de web de CRISPRs.
- 30 Las **Figuras 19 A-L** muestran un alineamiento de secuencias múltiple para 12 ortólogos Cas9. Dos residuos catalíticos están resaltados. El primer residuo resaltado es el residuo Asp catalítico en el dominio RuvCI, y el segundo residuo resaltado es el residuo catalítico His en el dominio HNH. La mutación de uno o de los otros residuos en Ala puede convertir Cas9 en una nickase. La mutación de ambos residuos convierte Cas9 en un mutante catalíticamente nulo - útil para la unión a ADN genérico.
- 35 La **Figura 20** muestra una transferencia Western que muestra que ortólogos Cas9 se expresan en células HEK 293FT; plásmidos de ADN que codifican ortólogos Cas9 son transfectados en células HEK 293 FT y los lisados de células se recogen para la transferencia Western. Nada de ADN se transfecta para el ortólogo Cas9 nº 8.

La **Figura 21** muestra la escisión in vitro de dianas candidato en el plásmido pUC19 por parte de 10 ortólogos Cas9. Los PAMs consenso se predicen por logos de secuencia (Figura 7), con base en los cuales se eligen dianas candidato en pUC19. 7 de 9 nuevos ortólogos Cas9 ensayados han escindido con éxito al menos una diana pUC19. SpCas9 también puede escindir NGA in vitro.

5 La **Figura 22** muestra un Gel Surveyor Representativo que muestra la escisión genómica por SaCas9.

La **Figura 23** muestra la Eficiencia de Escisión del Genoma de Secuencias PAM (Todas las dianas).

La **Figura 24** muestra la Eficiencia de Escisión del Genoma de Secuencias PAM (Dianas escindidas).

La **Figura 25** muestra la Eficiencia de Escisión del Genoma de Secuencias PAM (Todas las dianas, desechar las dianas de baja eficiencia y huérfanas).

10 La **Figura 26** muestra la Eficiencia de Escisión del Genoma de Secuencias PAM (Dianas escindidas, desechar las dianas de baja eficiencia y huérfanas).

La **Figura 27** muestra una Secuencia Logo para Tratar Separadores Escindidos y PAMs (Nuevo test del genoma endógeno que muestra que no se requiere T).

Las figuras en esta memoria son sólo para fines ilustrativos y no están necesariamente dibujadas a escala.

15 DESCRIPCIÓN DETALLADA DE LA INVENCION

La invención se refiere a la modificación por tecnología genética y la optimización de composiciones utilizadas para el control de la expresión génica que implican la focalización de la secuencia, tales como la perturbación del genoma o la edición de genes, que se refieren al sistema CRISPR/Cas y componentes de los mismos. En particular, la enzima de CRISPR es una enzima Cas, p. ej., un ortólogo Cas9 de *Staphylococcus aureus* (SaCas9).

20 La invención utiliza ácidos nucleicos para unir secuencias de ADN diana. Esto es ventajoso, ya que los ácidos nucleicos son mucho más fáciles y económicos de producir que, por ejemplo, péptidos, y la especificidad se puede variar de acuerdo con la longitud del tramo en el que se busca homología. Por ejemplo, no se requiere el posicionamiento en 3-D complejo de múltiples dedos.

25 Los términos y expresiones "polinucleótido", "nucleótido", "secuencia de nucleótidos", "ácido nucleico" y "oligonucleótido" se utilizan indistintamente. Se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sean desoxirribonucleótidos o ribonucleótidos, o análogos de los mismos. Los polinucleótidos pueden tener cualquier estructura tridimensional y pueden realizar cualquier función, conocida o desconocida. Los siguientes son ejemplos no limitativos de polinucleótidos: regiones codificadoras o no codificadoras de un gen o fragmento de gen, loci (locus) definidos a partir de análisis de ligamiento, exones, intrones, ARN mensajero (ARNm), ARN de transferencia, ARN ribosomal, ARN de interferencia corto (siRNA), ARN de horquilla corto (shRNA), micro-ARN (miRNA), ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácido nucleico y cebadores. El término también abarca estructuras de tipo ácido nucleico con cadenas principales sintéticas, véase, p. ej., Eckstein, 1991; Baserga et al., 1992; Miilligan, 1993; documentos WO97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; y Samstag, 1996. Un polinucleótido puede comprender uno o más nucleótidos modificados tales como nucleótidos metilados y análogos de nucleótidos. Si están presentes, modificaciones en la estructura de nucleótidos se pueden impartir antes o después del ensamblaje del polímero. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleótidos. Un polinucleótido puede modificarse adicionalmente después de la polimerización tal como por conjugación con un componente de marcaje.

40 En aspectos de la invención, las expresiones "ARN quimérico", "ARN guía quimérico", "ARN guía", "ARN guía sencillo" y "ARN guía sintético" se utilizan indistintamente y se refieren a la secuencia de polinucleótidos que comprende la secuencia guía, la secuencia tracr y la secuencia de emparejamiento tracr. La expresión "secuencia guía" se refiere a la secuencia de aproximadamente 20 pb dentro del ARN guía que especifica el sitio diana y puede utilizarse indistintamente con los términos "guía" o "espaciador". La expresión "secuencia de emparejamiento tracr" también se puede utilizar de forma indistinta con el término "repetición o repeticiones directas".

Tal como se utiliza en esta memoria, la expresión "tipo salvaje" es una expresión de la técnica entendida por las personas expertas y significa la forma típica de un organismo, cepa, gen o característica tal como se produce en la naturaleza que se distingue de las formas mutantes o variantes.

5 Tal como se utiliza en esta memoria, el término "variante" se deben tomar en el sentido de la exhibición de cualidades que tienen un patrón que se desvía de lo que ocurre en la naturaleza.

10 Las expresiones "que se produce de forma no natural" o "modificado por ingeniería genética" se utilizan indistintamente e indican la implicación de la mano del hombre. Las expresiones, cuando se refieren a moléculas de ácidos nucleicos o polipéptidos significan que la molécula de ácido nucleico o el polipéptido está al menos sustancialmente libre de al menos otro componente con el que están asociados de forma natural en la naturaleza y como se encuentra en la naturaleza.

15 "Complementariedad" se refiere a la capacidad de un ácido nucleico de formar enlace(s) hidrógeno con otra secuencia de ácido nucleico por apareamiento de bases de Watson-Crick tradicional u otros tipos no tradicionales. A porcentaje de complementariedad indica el porcentaje de residuos en una molécula de ácido nucleico que puede formar enlaces hidrógeno (p. ej., apareamiento de bases de Watson-Crick) con una segunda secuencia de ácido nucleico (p. ej., siendo 5, 6, 7, 8, 9, 10 de 10 50%, 60%, 70%, 80%, 90% y 100% complementarias). "Perfectamente complementarios" significa que todos los residuos contiguos de una secuencia de ácidos nucleicos se enlazarán por hidrógeno con el mismo número de residuos contiguos en una segunda secuencia de ácido nucleico. "Sustancialmente complementarios", tal como se utiliza en esta memoria, se refiere a un grado de complementariedad que es al menos 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% o 100% sobre una región de 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, o más nucleótidos, o se refiere a dos ácidos nucleicos que se hibridan en condiciones rigurosas.

25 Tal como se utiliza en esta memoria, "condiciones rigurosas" para la hibridación se refieren a condiciones en las que un ácido nucleico que tiene complementariedad con una secuencia diana se hibrida predominantemente con la secuencia diana, y sustancialmente no se hibrida con secuencias no diana. Las condiciones rigurosas son generalmente dependientes de la secuencia, y varían dependiendo de un cierto número de factores. En general, cuanto mayor sea la secuencia, más alta será la temperatura a la que la secuencia se hibrida específicamente con su secuencia diana. Ejemplos no limitantes de las condiciones restrictivas se describen en detalle en Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Parte I, Segundo Capítulo "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y. Cuando se haga referencia a una secuencia de polinucleótidos, entonces también se contemplan secuencias complementarias o parcialmente complementarias. Éstas son preferiblemente capaces de hibridarse a la secuencia de referencia en condiciones muy rigurosas. Generalmente, con el fin de maximizar la tasa de hibridación, se seleccionan condiciones de hibridación relativamente de baja rigurosidad: aproximadamente 20 a 25°C más bajas que el punto de fusión térmico (T_m). El T_m es la temperatura a la que el 50% de la secuencia diana específica se hibrida con una sonda perfectamente complementaria en solución a una fuerza iónica y pH definidos. Generalmente, con el fin de requerir al menos aproximadamente 85% de complementariedad de nucleótidos de las secuencias hibridadas, las condiciones de lavado altamente rigurosas se seleccionan para que sean aproximadamente 5 a 15°C inferiores al T_m. Con el fin de requerir al menos aproximadamente 70% de complementariedad de nucleótidos de las secuencias hibridadas, las condiciones de lavado moderadamente rigurosas se seleccionan para que sean aproximadamente 15 a 30°C inferiores al T_m. Condiciones de lavado altamente permisivas (rigurosidad muy baja) pueden ser tan bajas como 50°C inferiores al T_m, lo que permite un alto nivel de emparejamientos erróneos entre las secuencias hibridadas. Los expertos en la técnica reconocerán que otros parámetros físicos y químicos en las fases de hibridación y lavado también pueden alterarse para afectar el resultado de una señal de hibridación detectable de un nivel específico de homología entre las secuencias diana y sonda. Condiciones altamente rigurosas preferidas comprenden la incubación en formamida al 50%, 5xSSC y SDS al 1% a 42°C, o la incubación en 5xSSC y SDS al 45 1% a 65°C, con lavado en 0,2xSSC y SDS al 0,1% SDS a 65°C.

50 "Hibridación" se refiere a una reacción en la que uno o más polinucleótidos reaccionan para formar un complejo que se estabiliza a través de enlaces hidrógeno entre las bases de los residuos de nucleótidos. El enlace hidrógeno se puede producir por apareamiento de bases de Watson Crick, la unión Hoogstein, o en cualquier otra forma específica para la secuencia. El complejo puede comprender dos cadenas que forman una estructura dúplex, tres o más cadenas que forman un complejo multi-cadena, una sola cadena auto-hibridante, o cualquier combinación de éstas. Una reacción de hibridación puede constituir una etapa de un proceso más extenso tal como el inicio de la PCR, o la escisión de un polinucleótido por una enzima. A una secuencia capaz de hibridarse con una secuencia dada se la alude como el "complemento" de la secuencia dada.

Tal como se utiliza en esta memoria, la expresión "locus genómico" o "locus" (plural loci) es la ubicación específica de un gen o secuencia de ADN en un cromosoma. Un "gen" se refiere a tramos de ADN o ARN que codifican un polipéptido o una cadena de ARN que tiene que jugar un papel funcional en un organismo y, por lo tanto, es la unidad molecular hereditaria en los organismos vivos. Para los fines de esta invención, se puede considerar que los genes incluyen regiones que regulan la producción del producto génico, estén o no estas secuencias reguladoras adyacentes a las secuencias codificadoras y/o transcritas. Por consiguiente, un gen incluye, pero no se limita necesariamente a secuencias de promotor, terminadores, secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios de entrada al ribosoma interno, potenciadores, silenciadores, aisladores, elementos de borde, orígenes de replicación, sitios de unión a la matriz y regiones de control del locus.

Tal como se utiliza en esta memoria, "expresión de un locus genómico" o "expresión génica" es el proceso por el cual la información de un gen se utiliza en la síntesis de un producto génico funcional. Los productos de la expresión génica son a menudo proteínas, pero en genes no codificadores de proteínas tales como genes de ARNr o genes de ARNt, el producto es ARN funcional. El proceso de la expresión génica se utiliza por todos los eucariotas vivos conocidos (incluyendo organismos multicelulares), procariontes (bacterias y arqueas) y virus para generar productos funcionales para sobrevivir. Tal como se utiliza en esta memoria, "expresión" de un gen o ácido nucleico abarca no sólo la expresión del gen celular, sino también la transcripción y la traducción de ácido o ácidos nucleicos en los sistemas de clonación y en cualquier otro contexto. Tal como se utiliza en esta memoria, "expresión" también se refiere al proceso mediante el cual un polinucleótido se transcribe a partir de un molde de ADN (como en y ARNm u otro transcrito de ARN) y/o el proceso mediante el cual un ARNm transcrito se traduce posteriormente en péptidos, polipéptidos o proteínas. A transcritos y polipéptidos codificados se les puede aludir selectivamente como "producto génico". Si el polinucleótido se deriva de ADN genómico, la expresión puede incluir el corte y empalme de ARNm en una célula eucariota.

Los términos "polipéptido", "péptido" y "proteína" se utilizan indistintamente en esta memoria para referirse a polímeros de aminoácidos de cualquier longitud. El polímero puede ser lineal o ramificado, puede comprender aminoácidos modificados y puede estar interrumpido por no aminoácidos. Los términos también abarcan un polímero de aminoácidos que ha sido modificado; por ejemplo, formación de enlaces disulfuro, glicosilación, lipidación, acetilación, fosforilación o cualquier otra manipulación tal como conjugación con un componente de marcaje. Tal como se utiliza en esta memoria, el término "aminoácido" incluye aminoácidos naturales y/o no naturales o sintéticos, incluyendo glicina y tanto los isómeros ópticos D o L y análogos de aminoácidos y peptidomiméticos.

Tal como se utiliza en esta memoria, el término "dominio" o la expresión "dominio de proteína" se refiere a una parte de una secuencia de proteína que pueden existir y funcionar independientemente del resto de la cadena de proteína.

Tal como se describe en aspectos de la invención, la identidad de secuencia está relacionada con la homología de secuencia. Las comparaciones de homología pueden realizarse a simple vista, o más habitualmente, con la ayuda de programas de comparación de secuencias fácilmente disponibles. Estos programas informáticos disponibles comercialmente pueden calcular el porcentaje (%) de homología entre dos o más secuencias y también pueden calcular la identidad de secuencia compartida por dos o más secuencias de aminoácidos o de ácido nucleico. Las homologías de secuencia pueden generarse mediante cualquiera de un cierto número de programas de ordenador conocidos en la técnica, por ejemplo BLAST o FASTA, etc. Un programa informático adecuado para llevar a cabo dicha alineación es el paquete GCG Wisconsin Bestfit (Universidad de Wisconsin, EE.UU.; Devereux et al., 1984, Nucleic Acids Research 12:387). Ejemplos de otros software que pueden realizar comparaciones de secuencias incluyen, pero no se limitan a, el paquete BLAST (véase Ausubel et al., 1999 *ibid.* - Capítulo 18), FASTA (Atschul et al., 1990, J. Mol. Biol, 403-410) y el conjunto GENWORKS de herramientas de comparación. Tanto BLAST como FASTA están disponibles para la búsqueda fuera de línea y en línea (véase Ausubel et al., 1999 *ibid.*, páginas 7-58 a 7-60). Sin embargo, se prefiere utilizar el programa GCG Bestfit.

El % de homología puede calcularse sobre secuencias contiguas, es decir, una secuencia se alinea con la otra secuencia y cada uno de los aminoácidos o nucleótidos en una secuencia se compara directamente con el aminoácido o nucleótido correspondiente en la otra secuencia, un residuo cada vez. Esto se denomina alineamiento "sin huecos". Típicamente, estos alineamientos sin huecos se realizan sólo sobre un número relativamente corto de residuos.

Aunque este es un método muy simple y consistente, no tiene en consideración que, por ejemplo, en un par de secuencias de otro modo idénticas, una inserción o delección puede provocar que los siguientes restos de aminoácidos se coloquen fuera de alineamiento, por lo tanto, resultando potencialmente en una gran reducción en el

5 % de homología cuando se realiza un alineamiento global. En consecuencia, la mayoría de los métodos de comparación de secuencias se diseñan para producir alineamientos óptimos que tienen en cuenta posibles inserciones y deleciones sin penalizar indebidamente la puntuación de homología o identidad global. Esto se consigue insertando "huecos" en el alineamiento de secuencias para intentar maximizar la homología o identidad local.

10 Sin embargo, estos métodos más complejos asignan "penalizaciones por hueco" a cada uno de los huecos que se produce en el alineamiento, de modo que, para el mismo número de aminoácidos idénticos, un alineamiento de la secuencia con tan pocos huecos como sea posible - lo que refleja una mayor relatividad entre las dos secuencias comparadas - puede alcanzar una puntuación más alta que con muchos huecos. Se utilizan típicamente "costos de huecos de afinidad" que cargan un coste relativamente alto para la existencia de un hueco y una penalización más pequeña para cada uno de los residuos posteriores en el hueco. Este es el sistema de puntuación de huecos más comúnmente utilizado. Penalizaciones de huecos elevadas pueden producir, por supuesto, alineamientos optimizados con menos huecos. La mayoría de los programas de alineamiento permiten modificar las penalizaciones por hueco. Sin embargo, se prefiere utilizar los valores por defecto cuando se utiliza un software de este tipo para las comparaciones de secuencias. Por ejemplo, al utilizar el paquete GCG Wisconsin Bestfit la penalización por hueco por defecto para secuencias de aminoácidos es -12 para un hueco y -4 para cada extensión,

20 El cálculo del % máximo de homología, por tanto, requiere primero la producción de una alineación óptima, teniendo en consideración las penalizaciones por hueco. Un programa informático adecuado para llevar a cabo dicho alineamiento es el paquete GCG Wisconsin Bestfit (Devereux et al., 1984 Nuc. Acids Research 12 p 387). Ejemplos de otros software que pueden realizar comparaciones de secuencias incluyen, pero no se limitan a, el paquete BLAST (véase Ausubel et al., 1999 Short Protocols in Molecular Biology, 4ª Ed. - Capítulo 18), FASTA (Altschul et al., 1990, J. Mol. Biol. 403-410) y el conjunto GENWORKS de herramientas de comparación. Tanto BLAST como FASTA están disponibles para la búsqueda en línea y fuera de línea (véase Ausubel et al., 1999, Short Protocols in Molecular Biology, páginas 7-58 a 7-60). Sin embargo, para algunas aplicaciones, se prefiere utilizar el programa GCG Bestfit. Una nueva herramienta, denominada BLAST 2 Sequences, está también disponible para la comparación de secuencias de proteínas y nucleótidos (véase FEMS Microbiol Lett. 1999 174(2):247-50; FEMS Microbiol Lett. 1999 177(1):187-8 y el sitio web del Centro Nacional de información sobre Biotecnología en la página web de National Institutes for Health).

30 Aunque el % de homología final puede medirse en términos de identidad, el proceso de alineamiento en sí no se basa típicamente en una comparación de pares de todo o nada. En su lugar, se utiliza generalmente una matriz de puntuación de similitud a escala que asigna puntuaciones a cada comparación por pares basada en la similitud química o la distancia evolutiva. Un ejemplo de una matriz de este tipo comúnmente utilizada es la matriz BLOSUM62 - la matriz por defecto para el conjunto de programas BLAST. Los programas GCG Wisconsin generalmente utilizan cualquiera de los valores por defecto públicos o una tabla de comparación de símbolos personalizada, si se suministra (véase el manual del usuario para más detalles). Para algunas aplicaciones, se prefiere utilizar los valores por defecto públicos para el paquete GCG, o en el caso de otro software, la matriz por defecto, tal como BLOSUM62.

40 Alternativamente, las homologías porcentuales pueden calcularse utilizando la característica de alineamiento múltiple en DNASIS™ (Hitachi Software), basado en un algoritmo, análogo a CLUSTAL (Higgins DG y Sharp PM (1988), Gene 73 (1), 237-244). Una vez que el software ha producido un alineamiento óptimo, es posible calcular el % de homología, preferentemente el % de identidad de secuencia. El software hace normalmente esto como parte de la comparación de secuencias y genera un resultado numérico.

45 Las secuencias también pueden tener deleciones, inserciones o sustituciones de residuos de aminoácidos que producen un cambio silencioso y resultan en una sustancia funcionalmente equivalente. Sustituciones de aminoácidos deliberadas pueden realizarse sobre la base de la similitud en las propiedades de aminoácidos (tales como polaridad, carga, solubilidad, hidrofobicidad, hidrofiliicidad y/o la naturaleza anfipática de los residuos) y, por lo tanto, es útil para agrupar juntos aminoácidos en grupos funcionales. Los aminoácidos pueden agruparse basándose en las propiedades de sus cadenas laterales solas. Sin embargo, es más útil incluir los datos de mutación también. Los conjuntos de aminoácidos así obtenidos es probable que se conserven por razones estructurales. Estos conjuntos pueden describirse en forma de un diagrama de Venn (Livingstone CD y Barton GJ (1993) "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation" Comput. Appl. Biosci. 9: 745-756) (Taylor W.R. (1986) "The classification of amino acid conservation" J. Theor. Biol. 119; 205-218). Sustituciones conservadoras se pueden hacer, por ejemplo, de acuerdo con la tabla que figura a continuación que describe un diagrama de Venn generalmente aceptado de agrupación de aminoácidos.

Conjunto		Sub-conjunto	
Hidrofóbico	F W Y H K M I L V A G C	Aromático	F W Y H
		Alifático	I L V
Polar	W Y H K R E D C S T N Q	Cargado	H K R E D
		Positivamente cargado	H K R
		Negativamente cargado	E D
Pequeño	V C A G S P T N D	Minúsculo	A G S

- Realizaciones de la invención incluyen secuencias (tanto de polinucleótidos o polipéptidos) que puede comprender la sustitución homóloga (sustitución y reemplazo se utilizan ambos en esta memoria para dar a entender el intercambio de un residuo de aminoácido o nucleótido existente, con un residuo o nucleótido alternativo) que puede producirse, es decir, una sustitución comparable en el caso de aminoácidos tal como básico por básico, ácido por ácido, polar por polar, etc. También puede producirse una sustitución no homóloga, es decir, de una clase de residuo a otro o, alternativamente, que implica la inclusión de aminoácidos no naturales tales como ornitina (a la que se alude en lo sucesivo como Z), ornitina del ácido diaminobutírico (a la que se alude en lo sucesivo como B), ornitina norleucina (a la que se alude en lo sucesivo como O), pirilalanina, tienilalanina, naftilalanina y fenilglicina.
- Secuencias de aminoácidos variantes pueden incluir grupos espaciadores adecuados que pueden insertarse entre cualesquiera dos residuos de aminoácidos de la secuencia, incluyendo grupos alquilo tales como grupos metilo, etilo o propilo, además de espaciadores de aminoácidos tales como residuos glicina o β -alanina. Una forma adicional de variación, que implica la presencia de uno o más residuos aminoácidos en forma peptóide, pueda ser bien comprendida por los expertos en la técnica. Para evitar dudas, "la forma peptóide" se utiliza para referirse a residuos aminoácidos variantes, en donde el grupo de sustituyentes carbono α está en átomo de nitrógeno del residuo en lugar del carbono α . Procedimientos para preparar péptidos en forma peptóide son conocidos en la técnica, por ejemplo Simon RJ et al., *PNAS* (1992) 89(20), 9367-9371 y Horwell DC, *Trends Biotechnol.* (1995) 13(4), 132-134.
- La práctica de la presente invención emplea, a menos que se indique lo contrario, técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica y ADN recombinante, que están dentro de la experiencia de la técnica. Véase Sambrook, Fritsch y Maniatis, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2ª edición (1989); *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY* (F. M. Ausubel, et al. comps., (1987)); la serie *METHODS IN ENZYMOLOGY* (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames y G.R. Taylor comps. (1995)), Harlow y Lane, comps. (1988) *ANTIBODIES, A LABORATORY MANUAL*, y *ANIMAL CELL CULTURE* (R.I. Freshney, comp. (1987)).
- En un aspecto, la invención proporciona vectores que se utilizan en la ingeniería y optimización de los sistemas CRISPR/Cas.
- Tal como se utiliza en esta memoria, un "vector" es una herramienta que permite o facilita la transferencia de una entidad de un entorno a otro. Es un replicón, tal como un plásmido, fago, o cósmido, en el que se puede insertar otro segmento de ADN con el fin de llevar a cabo la replicación del segmento insertado. Generalmente, un vector es capaz de replicarse cuando se asocia con los elementos de control apropiados. En general, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido. Vectores incluyen, pero no se limitan a, moléculas de ácido nucleico que son de cadena sencilla, de cadena doble, o parcialmente de doble cadena; moléculas de ácido nucleico que comprenden uno o más extremos libres, sin extremos libres (p. ej., circulares); moléculas de ácido nucleico que comprenden ADN, ARN, o ambos; y otras variedades de polinucleótidos conocidos en la técnica. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN de doble cadena circular en el que se pueden insertar segmentos de ADN adicionales tal como por técnicas de clonación molecular estándares. Otro tipo de vector es un vector viral, en el que las secuencias de ADN o ARN derivadas de virus están presentes en el vector para el empaquetamiento en un virus (p. ej., retrovirus, retrovirus defectuosos en la replicación, adenovirus, adenovirus defectuosos en la replicación, y virus adeno-asociados). Los vectores virales también incluyen polinucleótidos portados por un virus para la transfección en una célula huésped. Determinados vectores son capaces de replicación autónoma en una célula huésped en la que se introducen (p. ej., vectores bacterianos que tienen un origen bacteriano de replicación y vectores episomales de mamífero). Otros vectores (p. ej., vectores no episomales de mamífero) se integran en el genoma de una célula huésped tras la

introducción en la célula huésped y, con ello, se replican junto con el genoma del huésped. Además, determinados vectores son capaces de dirigir la expresión de genes a los que están enlazados operativamente. Tales vectores se denominan en esta memoria "vectores de expresión". Vectores de expresión comunes de utilidad en técnicas de ADN recombinante están a menudo en forma de plásmidos. En esta memoria se proporciona una discusión adicional de vectores.

Los vectores de expresión recombinantes pueden comprender un ácido nucleico de la invención en una forma adecuada para la expresión del ácido nucleico en una célula huésped, lo que significa que los vectores de expresión recombinantes incluyen uno o más elementos reguladores, que pueden seleccionarse sobre la base de las células huésped a utilizar para la expresión, que están operativamente enlazadas a la secuencia de ácido nucleico a expresar. Dentro de un vector de expresión recombinante, "enlazada operativamente" pretende significar que la secuencia de nucleótidos de interés está enlazada al o a los elementos reguladores de una manera que permite la expresión de la secuencia de nucleótidos (p. ej., en un sistema de transcripción/traducción in vitro o en una célula huésped cuando el vector se introduce en la célula huésped). Con respecto a métodos de recombinación y de clonación, se hace mención a la solicitud de patente de EE.UU. 10/815.730, publicada el 2 de septiembre de 2004 como US 2004-0171156 A1.

Aspectos y realizaciones de la invención se refieren a vectores bicistrónicos para ARN quimérico y Cas9 o un ortólogo Cas9. En algunas realizaciones, la Cas9 es impulsada por el promotor promotor CBh. En algunas realizaciones, el ARN quimérico es impulsado por un promotor U6. Preferiblemente, el CBh y U6 se utilizan juntos, en el sentido de que Cas9 es impulsada por el promotor CBh y el ARN quimérico es impulsado por un promotor U6. En algunas realizaciones, el ARN guía quimérico consiste en una secuencia guía de 20 pb (Ns) unidos a la secuencia tracr (que va desde la primera "U" de la cadena inferior al extremo del transcrito), que está truncada en diversas posiciones tal como se indica. Las secuencias guía y tracr están separadas preferiblemente por la secuencia de emparejamiento tracr. Un ejemplo preferido de una secuencia de emparejamiento tracr es GUUUUAGAGCUA. Esto es seguido preferiblemente por una secuencia de bucle. El bucle es preferiblemente GAAA, pero no se limita a esta secuencia o, de hecho, sólo es de 4 pb de longitud. De hecho, secuencias formadoras de bucle preferidas para su uso en estructuras de horquilla son de cuatro nucleótidos de longitud, y lo más preferiblemente tienen la secuencia GAAA. Sin embargo, se pueden utilizar secuencias de bucle más largas o más cortas, al igual que secuencias alternativas. Las secuencias incluyen preferiblemente un triplete de nucleótidos (por ejemplo, AAA), y un nucleótido adicional (por ejemplo C o G). Ejemplos de secuencias formadoras de bucle incluyen CAAA y AAAG. A lo largo de esta solicitud, ARN quimérico también se puede denominar ARN guía sencillo o guía sintético (sgRNA).

Por ejemplo, tal como se describe en detalle específico en el Ejemplo 2, ARNs guía quiméricos pueden ser diseñados tal como se muestra en la Figura 8. Los loci de CRISPR en algunas de estas familias se representan en la Figura 11. Las correspondientes secuencias de ARN guía se muestran en la Figura 12. Los autores de la invención analizaron la secuencia de ADN genómico dentro ~ 2kb de las proteínas Cas9 e identificaron repeticiones directas que van desde 35 pb a 50 pb, con separadores intermedios que van desde 29 pb a 35 pb. Sobre la base de la secuencia de repetición directa, realizaron búsquedas de secuencias candidatas de tracrRNA con los siguientes criterios: fuera de la matriz de crRNA, pero que contienen un alto grado de homología con repeticiones directas (tal como se requiere para el apareamiento de bases repetición directa:tracrRNA; el análisis personalizado de cálculo), fuera de las regiones codificantes de los componentes de la proteína, que contienen señales de terminación de la transcripción de ~ 60 pb-120 pb aguas abajo de la región de homología con repeticiones directas y el co-plegamiento con repetición directa para formar un dúplex, seguido de dos o más estructuras de horquilla en el extremo distal de la secuencia de tracrRNA. Sobre la base de estos criterios de predicción, los autores de la invención seleccionaron un conjunto inicial de 18 proteínas Cas9 y sus repeticiones y tracrRNAs directos asociados de forma única distribuidos a través de las cinco familias Cas9. La solicitante generó, además, un conjunto de 18 estructuras de ARN quiméricas que conservaron la secuencia y las estructuras secundarias del dúplex repetición directa:tracrRNA nativo, al tiempo que acortan la región de apareamiento de bases y fusionan los dos elementos de ARN a través de un bucle artificial (Figuras 6 A-J).

La expresión "elemento regulador" pretende incluir promotores, potenciadores, sitios de entrada al ribosoma interno (IRES), y otros elementos de control de la expresión (p. ej., señales de terminación de la transcripción tales como señales de poliadenilación y secuencias poli-U). Se describen estos elementos reguladores, por ejemplo, en Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Elementos reguladores incluyen los que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huéspedes y los que dirigen la expresión directa de la secuencia de nucleótidos sólo en determinadas células huéspedes (p. ej., secuencias reguladoras específicas para el tejido). Un promotor específico para el tejido puede dirigir la expresión principalmente en un tejido deseado de interés tal como

el músculo, neurona, hueso, piel, sangre, órganos específicos (p. ej., hígado, páncreas), o tipos de células particulares (p. ej., linfocitos). Los elementos reguladores también pueden dirigir la expresión de una manera dependiente del tiempo tal como de una manera dependiente del ciclo celular o dependiente de la fase de desarrollo, que puede o puede no ser también específico para el tejido o tipo de célula. En algunas realizaciones, un vector puede comprender uno o más promotores pol III (p. ej., 1, 2, 3, 4, 5 o más promotores pol III), uno o más promotores pol II (p. ej., 1, 2, 3, 4, 5 o más promotores pol II), uno o más promotores pol I (p. ej., 1, 2, 3, 4, 5 o más promotores pol I), o combinaciones de los mismos. Ejemplos de promotores pol III incluyen, pero no se limitan a promotores U6 y H1. Ejemplos de promotores pol II incluyen, pero no se limitan al promotor del virus del sarcoma de Rous retroviral (RSV) LTR (opcionalmente con el potenciador RSV), el promotor de citomegalovirus (CMV) (opcionalmente con el potenciador CMV) [véase, p. ej., Boshart et al., *Cell*, 41:521-530 (1985)], el promotor SV40, el promotor de dihidrofolato reductasa, el promotor β -actina, el promotor fosfoglicerol quinasa (PGK) y el promotor EF1 α . También quedan abarcados por la expresión "elemento regulador" elementos potenciadores tales como WPRE; potenciadores de CMV; el segmento R-U5' en LTR de HTLV-I (*Mol. Cell. Biol.*, Vol 8(1), págs. 466-472, 1988); el potenciador SV40; y la secuencia de intrón entre los exones 2 y 3 de β -globina de conejo (*Proc. Natl. Acad. Sci. USA*, Vol, 78(3), págs. 1527-31, 1981). Se apreciará por los expertos en la técnica que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped a transformar, el nivel de expresión deseado, etc. Un vector puede ser introducido en células huéspedes para con ello producir transcritos, proteínas o péptidos, incluyendo proteínas o péptidos de fusión, codificados por ácidos nucleicos tal como se describe en esta memoria (p. ej., repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR), transcritos, proteínas, enzimas, formas mutantes de los mismos, proteínas de fusión de los mismos, etc.). Con respecto a secuencias reguladoras, se hace mención de la publicación de la solicitud de patente de EE.UU. 10/491.026. Con respecto a los promotores, se hace mención a la publicación PCT WO 2011/028929 y la solicitud de patente de EE.UU. 12/511.940.

Los vectores pueden ser diseñados para la expresión de transcritos de CRISPR (p. ej., transcritos de ácidos nucleicos, proteínas o enzimas) en células procariotas o eucariotas. Por ejemplo, los transcritos de CRISPR pueden expresarse en células bacterianas tales como *Escherichia coli*, células de insectos (utilizando vectores de expresión de baculovirus), células de levaduras o células de mamífero. Células huéspedes adecuadas se discuten en Goeddel, *GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY* 185, Academic Press, San Diego, California (1990). Alternativamente, el vector de expresión recombinante puede transcribirse y traducirse in vitro, por ejemplo utilizando secuencias reguladoras del promotor T7 y T7 polimerasa.

Los vectores se pueden introducir y propagar en una célula procariota o procariótica. En algunas realizaciones, se utiliza un procariota para amplificar copias de un vector que se ha de introducir en una célula eucariota o como un vector intermedio en la producción de un vector a ser introducido en una célula eucariota (por ejemplo, una ampliación de un plásmido como parte de un sistema de empaquetado de un vector viral). En algunas realizaciones, se utiliza un procariota para amplificar copias de un vector y expresar uno o más ácidos nucleicos tal como para proporcionar una fuente de una o más proteínas para el suministro a una célula huésped u organismo huésped. La expresión de proteínas en procariotas se lleva a cabo más a menudo en *Escherichia coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas de fusión o de no fusión. Los vectores de fusión añaden un cierto número de aminoácidos a una proteína codificada en el mismo tal como al extremo amino de la proteína recombinante. Tales vectores de fusión pueden cumplir uno o más fines tales como: (i) aumentar la expresión de la proteína recombinante; (ii) aumentar la solubilidad de la proteína recombinante; y (iii) ayudar en la purificación de la proteína recombinante, actuando como un ligando en purificación por afinidad. A menudo, en vectores de expresión de fusión, un sitio de escisión proteolítica se introduce en la unión del resto de fusión y la proteína recombinante para permitir la separación de la proteína recombinante del resto de fusión después de la purificación de la proteína de fusión. Tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen Factor Xa, trombina y enteroquinasa. Ejemplos de vectores de expresión de fusión incluyen pGEX (Pharmacia Biotech Inc; Smith y Johnson, 1988. *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) y pRIT 5 (Pharmacia, Piscataway, N.J.) que fusionan la glutatión S transferasa (GST), proteína de unión a maltosa, o proteína A, respectivamente, a la proteína recombinante diana.

Ejemplos de vectores de expresión de *E. coli* sin fusión inducibles adecuados incluyen pTrc (Amrann et al., (1988) *Gene* 69: 301-315) y pET 11d (Studier et al., *GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY* 185, Academic Press, San Diego, Calif. (1990) 60-89).

En algunas realizaciones, un vector es un vector de expresión de levadura. Ejemplos de vectores para la expresión en la levadura *Saccharomyces cerevisiae* incluyen pYepSec1 (Baldari, et al., 1987. *EMBO J.* 6:229-234), pMFa (Kuijan y Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz et al., 1987. *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) y picZ (Invitrogen Corp, San Diego, Calif).

En algunas realizaciones, un vector impulsa la expresión de proteínas en células de insecto utilizando vectores de expresión de baculovirus. Vectores de baculovirus disponibles para la expresión de proteínas en células de insecto cultivadas (p. ej., células SF9) incluyen la serie pAc (Smith, et al., 1983. *Mol. Cell. Biol.* 3:2156-2165) y la serie pVL (Lucklow y Summers, 1989. *Virology* 170:31-39).

5 En algunas realizaciones, un vector es capaz de impulsar la expresión de una o más secuencias en células de mamífero utilizando un vector de expresión de mamífero. Ejemplos de vectores de expresión de mamífero incluyen pCDM8 (Seed, 1987. *Nature* 329:840) y pMT2PC (Kaufman, et al., 1987. *EMBO J.* 6:187-195). Cuando se utiliza en
10 células de mamífero, las funciones de control del vector de expresión son proporcionadas típicamente por uno o más elementos reguladores. Por ejemplo, promotores comúnmente utilizados se derivan de poliovirus, adenovirus 2, citomegalovirus, virus de simio 40, y otros descritos en esta memoria y conocidos en la técnica. Para otros sistemas de expresión adecuados tanto para células procariotas como eucariotas véanse, p. ej., los Capítulos 16 y 17 de Sambrook, et al., *MOLECULAR CLONING: A LABORATORY MANUAL*. 2ª ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989.

15 En algunas realizaciones, el vector de expresión de mamífero recombinante es capaz de dirigir la expresión del ácido nucleico preferentemente en un tipo de célula particular (p. ej., elementos reguladores específicos para tejidos se utilizan para expresar el ácido nucleico). Elementos reguladores específicos para tejidos son conocidos en la técnica. Ejemplos no limitantes de promotores específicos para tejidos, adecuados, incluyen el promotor de albúmina (específico para el hígado; Pinkert, et al., 1987. *Genes Dev.* 1: 268-277), promotores linfocitos-específicos (Calame y Eaton, 1988. *Adv. Immunol.* 43:235-275), en particular promotores de receptores de células T (Winoto y Baltimore, 1989. *EMBO J.* 8:729-733) e inmunoglobulinas (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen y Baltimore, 1983 *Cell* 33: 741-748), promotores específicos para neuronas (p. ej., el promotor para neurofilamentos; Byrne y Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), promotores específicos para páncreas (Edlund, et al., 1985. *Science* 230:912-916) y promotores específicos para glándulas mamarias (p. ej., promotor de suero lácteo; patente de EE.UU. Nº. 4.873.316 y la Publicación de Solicitud Europea Nº 264.166.). Promotores regulados en el desarrollo
20 también están comprendidos, p. ej., los promotores *hox* murinos (Kessel y Gruss, 1990. *Science* 249:374-379) y el promotor de α -fetoproteína (Campes y Tilghman, 1989. *Genes Dev.* 3: 537-546). Con respecto a estos vectores procarióticos y eucarióticos, se hace mención a la patente de EE.UU. 6.750.059. Otras realizaciones de la invención pueden estar relacionadas con el uso de vectores virales, en lo que respecta a los que se hace mención en la solicitud de patente de EE.UU. 13/092.085. Elementos reguladores específicos para tejidos son conocidos en la
25 técnica y, en este sentido, se hace mención a la patente de EE.UU. 7.776.321.

En algunas realizaciones, un elemento regulador está enlazado operativamente a uno o más elementos de un sistema de CRISPR con el fin de impulsar la expresión de los uno o más elementos del sistema de CRISPR. En general, las CRISPRs (Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas), también conocidas como SPIDRs (Repeticiones Directas Intercaladas Separadas), constituyen una familia de loci de ADN
35 que habitualmente son específicos para una especie bacteriana particular. El locus de CRISPR puede comprender una clase distinta de Repeticiones de Secuencia Corta Intercaladas (SSRs) que fueron reconocidas en *E. coli* (Ishino et al. *J. Bacteriol.*, 169:5429-5433 [1987]; y Nakata et al. *J. Bacteriol.*, 171: 3553-3556 [1989]), y genes asociados. SSRs intercaladas similares han sido identificadas en *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena* y *Mycobacterium tuberculosis* (Véase, Groenen et al. *Mol. Microbiol.*, 10:1057-1065 [1993]; Hoe et al., *Emerg. Infect. Dis.*, 5:254-263 [1999]; Masepohl et al., *Biochim. Biophys. Acta* 1307:26-30 [1996]; y Mojica et al., *Mol. Microbiol.*, 17:85-93 [1995]). Los loci CRISPR difieren típicamente de otros SSRs por la estructura de las repeticiones, que se han denominado repeticiones cortas regularmente espaciadas (SRSRs) (Janssen et al., *OMICS J. Integ. Biol.*, 6:23-33 [2002]; y Mojica et al., *Mol. Microbiol.*, 36:244-246 [2000]). En general, las repeticiones son elementos cortos que se producen en grupos que están espaciados regularmente por secuencias intermedias únicas con una longitud sustancialmente constante (Mojica et al., [2000], *supra*). Aunque las secuencias de repetición están altamente conservadas entre las cepas, el número de repeticiones intercaladas y las secuencias de las regiones del espaciador difieren típicamente de una cepa a otra (van Embden et al., *J. Bacteriol.*, 182:2393-2401 [2000]). Loci CRISPR han sido identificados en más de 40 procariotas (Véase, p. ej., Jansen et al. *Mol. Microbiol.*, 43:1565-1575 [2002]; y Mojica et al., [2005]), incluyendo, pero no limitado a *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*,
45 *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*,
50 *Xanthomonas*, *Yersinia*, *Treponema* y *Thermotoga*.

En general, "sistema de CRISPR" se refiere colectivamente a transcritos y otros elementos implicados en la expresión de o dirigir la actividad de genes CRISPR-asociados ("Cas"), incluyendo secuencias que codifican un gen Cas, una secuencia tracr (CRISPR trans-activante) (p. ej., tracrRNA o un tracrRNA parcial activo), una secuencia de emparejamiento tracr (que abarca una "repetición directa" y una repetición directa parcial procesada por tracrRNA en el contexto de un sistema de CRISPR endógeno), una secuencia guía (a la que también se alude como un "espaciador" en el contexto de un sistema de CRISPR endógeno), u otras secuencias y transcritos de un locus CRISPR. En algunas realizaciones, uno o más elementos de un sistema de CRISPR se derivan de un sistema de CRISPR de tipo I, tipo II o de tipo III. En algunas realizaciones, uno o más elementos de un sistema de CRISPR se derivan de un organismo particular que puede comprender un sistema de CRISPR endógeno tal como *Streptococcus pyogenes*. En general, un sistema de CRISPR se caracteriza por elementos que fomentan la formación de un complejo de CRISPR en el sitio de una secuencia diana (al que también se alude como un protoespaciador en el contexto de un sistema de CRISPR endógeno). En el contexto de la formación de un complejo de CRISPR, "secuencia diana" se refiere a una secuencia a la que una secuencia guía está diseñada para que tenga complementariedad, en donde la hibridación entre una secuencia diana y una secuencia guía fomenta la formación de un complejo de CRISPR. Una secuencia diana puede comprender cualquier polinucleótido tal como polinucleótidos de ADN o ARN. En algunas realizaciones, una secuencia diana se encuentra en el núcleo o citoplasma de una célula.

En realizaciones preferidas de la invención, el sistema de CRISPR es un sistema de CRISPR de tipo II y la enzima Cas es Cas9, que cataliza la escisión de ADN. La acción enzimática por Cas9 derivada de *Streptococcus pyogenes* o cualquier Cas9 estrechamente relacionada genera roturas de doble cadena en las secuencias del sitio diana que se hibridan a 20 nucleótidos de la secuencia guía y que tienen una secuencia del motivo adyacente al protoespaciador (PAM) NGG/NRG (por ejemplo, tal como se discute en otra parte, un PAM adecuado es 5'-NRG o 5'-NNGRR para SpCas9 o enzimas (o enzimas derivadas) SaCas9, respectivamente) después de los 20 nucleótidos de la secuencia diana. La actividad de CRISPR a través de Cas9 para el reconocimiento de ADN específico para el sitio y la escisión se definen por la secuencia guía, la secuencia tracr que se hibrida en parte a la secuencia guía y la secuencia de PAM. Se describen más aspectos del sistema de CRISPR en Karginov y Hannon. The CRISPR system: small RNA-guided defense in bacteria and archaea, *Mole Cell* 2010, 15 de enero; 37(1): 7.

El locus de CRISPR de tipo II de *Streptococcus pyogenes* SF370 contiene un racimo de cuatro genes Cas9, Cas1, Cas2 y Csn1, así como dos elementos de ARN no codificantes, tracrRNA y una matriz característica de secuencias repetitivas (repeticiones directas) interespaciadas por tramos cortos de secuencias no repetitivas (espaciadores, aproximadamente de 30 pb cada uno). En este sistema, la rotura de doble cadena de ADN (DSB) fijada como objetivo se genera en cuatro etapas secuenciales (Figura 2A). En primer lugar, dos ARN no codificantes, la matriz pre-crRNA y tracrRNA se transcriben a partir del locus de CRISPR. En segundo lugar, tracrRNA se hibrida a las repeticiones directas de pre-crRNA, que se procesa luego en crRNA maduros que contienen secuencias de espaciador individuales. En tercer lugar, el complejo crRNA:tracrRNA maduro dirige Cas9 a la diana de ADN que consiste en el protoespaciador y el PAM correspondiente a través de la formación de heteroduplex entre la región del espaciador del crRNA y el ADN del protoespaciador. Por último, Cas9 media en la escisión del ADN diana aguas arriba del PAM para crear una DSB dentro del protoespaciador (Figura 2A). La Figura 2B muestra la ubicación nuclear de Cas9 optimizado en codones. Para fomentar la iniciación de la transcripción precisa, se seleccionó el promotor U6 basado en ARN polimerasa III para impulsar la expresión de tracrRNA (Figura 2C). De manera similar, se desarrolló una construcción basada en el promotor U6 para expresar una matriz de pre-crRNA que consiste en un único espaciador flanqueado por dos repeticiones directas (DRs, también abarcados por la expresión "secuencias de emparejamiento de tracr"; Figura 2C). El espaciador inicial se diseñó para fijar como objetivo un sitio diana de 33 pares de bases (pb) (protoespaciador de 30 pb más una secuencia del motivo de CRISPR (PAM) de 3 pb que satisface el motivo de reconocimiento de NGG de Cas9) en el locus EMX1 humano (Figura 2C), un gen clave en el desarrollo de la corteza cerebral.

La Figura 16 muestra ortólogos Cas9 y respectivos sgRNAs se utilizan para escindir dos dianas candidatas presentes en un banco basado en pUC19. La diana 1 es seguida por un PAM aleatorio que contiene 7 bases degeneradas (5'-NNNNNNN -3'), y la diana 1', que contiene la misma secuencia diana que la diana 1, es seguida por un PAM fijo (5'-TGGAGAAT-3). El sgRNA de cada uno de los ortólogos Cas9 contiene la secuencia guía contra la diana 1 o diana 1'. Imágenes en gel muestran la escisión con éxito por 20 ortólogos Cas9, lo que indica que estos diseños de sgRNA son funcionales con sus respectivas enzimas Cas9.

En algunas realizaciones, repeticiones directas o secuencias de emparejamiento tracr se descargan de la base de datos CRISPRs o se identifican in silico mediante el rastreo de motivos repetitivos que 1. se encuentran en una ventana de 2 kb de secuencia genómica flanqueante del locus CRISPR de tipo II, 2. abarcan de 20 a 50 pb, y 3.

están separadas por 20 a 50 pb. En algunas realizaciones, se pueden utilizar 2 de estos criterios, por ejemplo, 1 y 2, 2 y 3 ó 1 y 3. En algunas realizaciones, se pueden utilizar los 3 criterios.

5 En algunas realizaciones tracrRNA candidato se predice posteriormente por 1. homología de la secuencia con repeticiones directas (búsqueda con motivos en Geneious con apareamientos erróneos de hasta 18 pb), 2. presencia de un terminador de la transcripción independiente de Rho predicho en la dirección de la transcripción, y 3. estructura secundaria de horquilla estable entre tracrRNA y la repetición directa. En algunas realizaciones, se pueden utilizar 2 de estos criterios, por ejemplo 1 y 2, 2 y 3, ó 1 y 3. En algunas realizaciones, se pueden utilizar los 3 criterios. En algunas realizaciones, diseños de ARNs guía quiméricos sintéticos (sgRNAs) incorporan al menos 8 pb de la estructura dúplex entre la repetición directa y tracrRNA.

10 Varios aspectos del sistema de CRISPR pueden mejorarse aún más para aumentar la eficiencia y versatilidad de la fijación como objetivo de CRISPR. Una actividad de Cas9 óptima puede depender de la disponibilidad de Mg²⁺ libre a niveles más altos que los presentes en el núcleo de mamíferos (véase, p. ej., Jinek et al., 2012, Science, 337:816), y la preferencia por un motivo NGG/NRG inmediatamente aguas abajo de la protoespaciador restringe la capacidad de orientar por término medio cada 12 pb en el genoma humano.

15 Típicamente, en el contexto de un sistema de CRISPR endógeno, la formación de un complejo de CRISPR (que comprende una secuencia guía hibridada a una secuencia diana y formando complejo con una o más proteínas Cas) resulta en la escisión de una o las dos cadenas en o cerca de (p. ej., dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, o más pares de bases) de la secuencia diana. Sin desear estar limitados por la teoría, la secuencia tracr, que puede comprender o consistir en la totalidad o una porción de una secuencia tracr de tipo salvaje (p. ej., aproximadamente 20, 26, 32, 45, 48, 54, 63, 67, 85, o más nucleótidos de una secuencia tracr de tipo salvaje), también pueden formar parte de un complejo de CRISPR, tal como mediante la hibridación a lo largo de al menos una porción de la secuencia de tracr a la totalidad o una parte de una secuencia de emparejamiento tracr que está enlazada operativamente a la secuencia guía. En algunas realizaciones, uno o más vectores que impulsan la expresión de uno o más elementos de un sistema de CRISPR se introducen en una célula huésped de manera que la expresión de los elementos del sistema de CRISPR dirigen la formación de un complejo de CRISPR en uno o más sitios diana. Por ejemplo, una enzima Cas, una secuencia guía enlazada a una secuencia tracr-emparejamiento, y una secuencia de tracr podrían ser unidas operativamente en cada caso para separar elementos reguladores en vectores separados. Alternativamente, dos o más de los elementos expresados a partir de los mismos o diferentes elementos reguladores, se pueden combinar en un solo vector, con uno o más vectores adicionales que proporcionan cualesquiera componentes del sistema de CRISPR no incluidos en el primer vector. Elementos del sistema de CRISPR que se combinan en un único vector pueden estar dispuestos en cualquier orientación adecuada, tal como un elemento situado en 5' con respecto a ("aguas arriba" de) o 3' con respecto a ("aguas abajo" de) un segundo elemento. La secuencia codificadora de un elemento puede estar situada en la misma o en una cadena diferente de la secuencia codificadora de un segundo elemento, y estar orientada en la misma o en dirección opuesta. En algunas realizaciones, un único promotor dirige la expresión de un transcrito que codifica una enzima de CRISPR y una o más de la secuencia guía, secuencia de emparejamiento tracr (opcionalmente enlazada operativamente a la secuencia de guía) y una secuencia de tracr incrustada dentro de una o más secuencias de intrones (p. ej., cada una en un intrón diferente, dos o más en al menos un intrón, o todas en un solo intrón). En algunas realizaciones, la enzima de CRISPR, la secuencia guía, la secuencia de emparejamiento tracr y la secuencia tracr están enlazadas operativamente a y son expresada a partir del mismo promotor.

45 En algunas realizaciones, un vector comprende uno o más sitios de inserción, tales como una secuencia de reconocimiento de endonucleasa de restricción (al que también se alude como un "sitio de clonación"). En algunas realizaciones, uno o más sitios de inserción (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más sitios de inserción) están situados aguas arriba y/o aguas abajo de uno o más elementos de secuencia de uno o más vectores. En algunas realizaciones, un vector comprende un sitio de inserción aguas arriba de una secuencia de emparejamiento tracr y, opcionalmente, aguas abajo de un elemento regulador enlazado operativamente a la secuencia de emparejamiento tracr, de tal manera que después de la inserción de una secuencia guía en el sitio de inserción y tras la expresión de la secuencia guía dirige la unión específica para la secuencia de un complejo CRISPR a una secuencia diana en una célula eucariota. En algunas realizaciones, un vector comprende dos o más sitios de inserción, estando situado cada uno de los sitios de inserción entre dos secuencias de emparejamiento tracr con el fin de permitir la inserción de una secuencia guía en cada uno de los sitios. En una disposición de este tipo, las dos o más secuencias guía pueden comprender dos o más copias de una única secuencia guía, dos o más secuencias guía diferentes, o combinaciones de éstas. Cuando se utilizan múltiples secuencias guía diferentes, se puede utilizar una única construcción de expresión para fijar como objetivo la actividad de CRISPR a múltiples secuencias diana diferentes, correspondiente, dentro de una célula. Por ejemplo, un único vector puede comprender aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20,

o más secuencias guía. En algunas realizaciones, pueden proporcionarse aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más de dichos vectores que contienen la secuencia guía, y opcionalmente suministrarse a una célula.

5 En algunas realizaciones, un vector comprende un elemento regulador enlazado operativamente a una secuencia codificadora de la enzima que codifica una enzima de CRISPR tal como una proteína Cas. Ejemplos no limitantes de proteínas Cas incluyen Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (también conocida como Csn1 y Csx12), Cas 10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homólogos de las mismas, o versiones modificadas de las mismas. En algunas 10 realizaciones, la enzima de CRISPR no modificada tiene actividad de escisión de ADN, tal como Cas9. En algunas realizaciones, la enzima de CRISPR dirige la escisión de una o ambas cadenas en la ubicación de una secuencia diana, tal como dentro de la secuencia diana y/o dentro del complemento de la secuencia diana. En algunas realizaciones, la enzima de CRISPR dirige la escisión de una o ambas cadenas dentro de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, o más pares de bases desde el primer o el último nucleótidos de una secuencia diana. En algunas realizaciones, un vector codifica una enzima de CRISPR que está mutada con respecto a una enzima de tipo salvaje correspondiente, de tal manera que la enzima de CRISPR mutada carece de la capacidad de escindir una o ambas cadenas de un polinucleótido diana que contiene una secuencia diana. Por ejemplo, una sustitución-aspartato-a alanina (D10A) en el dominio catalítico RuvC I de Cas9 de *S. pyogenes* convierte Cas9 de una nucleasa que escinde ambas hebras en una nicasa (escinde una sola cadena). Otros 20 ejemplos de mutaciones que hacen de Cas9 una nicasa incluyen, sin limitación, H840A, N854A y N863A en SpCas9. Tal como se comenta en esta memoria, posiciones correspondientes puede estar conservadas en otras Cas9s, es decir, en ortólogos Cas9 de o derivados de otras especies bacterianas, con referencia a la numeración de la posición de SpCas9. (Figura 19) muestra un alineamiento de secuencias múltiples de 12 ortólogos Cas9 e indica el residuo Asp catalítico conservado en el dominio RuvC I y el residuo His catalítico conservado en el dominio HNH. La mutación de uno u otro residuo en Ala pueden convertir el ortólogo Cas9 en una nickase. La mutación de los dos 25 residuos puede convertir el ortólogo Cas9 en un mutante catalíticamente nulo - útil para la unión a ADN genérico. Como un ejemplo adicional, dos o más dominios catalíticos de Cas9 (RuvC I, RuvC II y RuvC III o el dominio HNH) pueden mutarse para producir un ortólogo Cas9 mutado que carece sustancialmente de toda actividad de escisión de ADN. En algunas realizaciones, una mutación D10A se combina con uno o más de mutaciones H840A, N854A o 30 N863A para producir una enzima Cas9 que carece sustancialmente de toda actividad de escisión de ADN. En algunas realizaciones, una enzima de CRISPR se considera que carece sustancialmente de toda actividad de escisión de ADN cuando la actividad de escisión de ADN de la enzima mutada es menor que aproximadamente 25%, 10%, 5%, 1%, 0,1%, 0,01%, o inferior con respecto a su forma no mutada.

35 Una sustitución aspartato-a-alanina (D10A) en el dominio catalítico RuvC I de SpCas9 fue diseñada para convertir la nucleasa en una nickase (SpCas9n) (véase, p. ej., Sapranaukas et al., 2011, *Nucleic Acid Research*, 39: 9275; Gasiunas et al., 2012, *Proc. Natl. Acad. Sci. USA*, 109:E2579), de manera que el ADN genómico mellado sufre la reparación dirigida a homología de alta fidelidad (HDR). La solicitante utilizó el ensayo SURVEYOR para confirmar que SpCas9n no genera indeles en la diana protoespaciador EMX1. Se observó que la co-expresión de crRNA quimérico que fija como objetivo EMX1 (que tiene también el componente tracrRNA) con SpCas9 produjo indeles en el sitio diana, mientras que la co-expresión con SpCas9n no lo hizo ($n = 3$). Además de ello, la secuenciación de 327 amplicones no detectó indeles inducidos por SpCas9n. El mismo locus fue seleccionado para testar HR mediada por CRISPR al co-transfectar células HEK 293FT con EMX1, hSpCas9 o hSpCas9n que fija como objetivo ARN quimérico, así como un molde de HR para introducir un par de sitios de restricción (HindIII y NheI) cerca del protoespaciador. De hecho, SpCas9 y SpCas9n catalizaban la integración del molde HR en el locus EMX1. La 45 amplificación por PCR de la región diana, seguido de la digestión por restricción con HindIII reveló productos de escisión que corresponden a tamaños de los fragmentos esperados, mediando SpCas9 y SpCas9n niveles similares de las eficiencias de HR. La solicitante verificó, además, HR utilizando la secuenciación de Sanger de amplicones genómicos y demostró la utilidad de CRISPR para facilitar la inserción de genes fijada como objetivo en el genoma de mamíferos. Dada la especificidad diana de 14 pb (12 pb desde el espaciador y 2-pb desde el PAM) de la SpCas9 de tipo salvaje, la disponibilidad de un nickase puede reducir significativamente la probabilidad de modificaciones fuera de objetivo, ya que roturas de una cadena sencilla no son sustratos para la vía NHEJ propenso a errores. La Figura 10A-M proporciona un esquema que indica las posiciones de las mutaciones en SpCas9 y ortólogos Cas9 50 comparten típicamente la organización general de los dominios 3-4 RuvC y un dominio HNH. El dominio RuvC más próximo a 5' escinde la cadena no complementaria, y el dominio HNH escinde la cadena complementaria.

55 El residuo catalítico en el dominio RuvC 5' se identifica a través de la comparación de homología de la Cas9 de interés con otros ortólogos Cas9 (de locus de CRISPR de tipo II de *S. pyogenes*, locus 1 de CRISPR de *S. thermophilus*, locus 3 de CRISPR de *S. thermophilus* y locus de CRISPR de tipo II de *Franciscilla novicida*), y el

residuo Asp conservada está mutado a alanina para convertir Cas9 en una enzima de mellado de cadena complementaria. De manera similar, los residuos His y Asn conservados en los dominios HNH están mutados a Alanina para convertir Cas9 en una enzima de mellado de cadena no complementaria.

5 En algunas realizaciones, una secuencia codificadora de la enzima que codifica una enzima CRISPR está optimizada en codones para la expresión en células particulares, tales como células eucariotas. Las células eucariotas pueden ser las de o derivadas de un organismo particular, tal como un mamífero, incluyendo, pero no limitado a, ser humano, ratón, rata, conejo, perro o mamífero no humano, incluyendo primate no humano. Se excluyen procesos para modificar la identidad genética de la línea germinal de seres humanos y/o procesos para modificar la identidad genética de animales que es probable que les provoque sufrimiento sin beneficio médico sustancial alguno para el hombre o los animales, y también animales que resultan de estos procesos.

15 En general, la optimización de codones se refiere a un proceso de modificar una secuencia de ácido nucleico para la expresión potenciada en las células huéspedes de interés mediante el reemplazo de al menos un codón (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, o más codones) de la secuencia nativa con codones que son con más frecuencia o lo más frecuentemente utilizados en los genes de esa célula huésped, al tiempo que se mantiene la secuencia de aminoácidos nativa. Diversas especies exhiben una desviación particular para ciertos codones de un aminoácido particular. Desviación del codón (diferencias en el uso de codones entre organismos) se correlaciona a menudo con la eficacia de la traducción de ARN mensajero (ARNm) que, es a su vez, se cree que depende, entre otros, de las propiedades de los codones que están siendo traducidos y la disponibilidad de moléculas de ARN de transferencia (ARNt) particulares. El predominio de los ARNt seleccionados en una célula es, por lo general, un reflejo de los codones más frecuentemente utilizados en la síntesis de péptidos. Por consiguiente, los genes pueden adaptarse para la expresión génica óptima en un organismo dado sobre la base de la optimización de codones. Tablas de uso de codones están fácilmente disponibles, por ejemplo, en la base de datos "Codon Usage Database", disponible en www.kazusa.or.jp/codon/ (visitada el 9 de julio de 2002), y estas tablas pueden ser adaptados en un cierto número de maneras. Véase Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence database status for the year 2000" Nucl. Acids Res. 28:292 (2000). También están disponibles algoritmos informáticos para la optimización de codones de una secuencia particular para la expresión en una célula huésped particular, tal como Gene Forge (Aptagen; Jacobus, PA). En algunas realizaciones, uno o más codones (p. ej., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, o más, o todos los codones) en una secuencia que codifica una enzima de CRISPR corresponden al codón utilizado con mayor frecuencia para un aminoácido particular.

30 En algunas realizaciones, un vector codifica una enzima de CRISPR que comprende una o más secuencias de localización nuclear (NLSs), tal como aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más NLSs. En algunas realizaciones, la enzima de CRISPR puede comprender aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más NLSs en o cerca del extremo amino, aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más NLSs en o cerca del extremo carboxi, o una combinación de éstas (p. ej., una o más NLS en el extremo amino y una o más NLS en el extremo carboxi). Cuando está presente más de una NLS, cada una se puede seleccionar independientemente de las otras, de manera que una sola NLS puede estar presente en más de una copia y/o en combinación con una o más de otras NLSs presentes en una o más copias. En una realización preferida de la invención, la enzima de CRISPR comprende a lo sumo 6 NLSs. En algunas realizaciones, una NLS se considera cerca del extremo N o C cuando el aminoácido más cercano de la NLS está dentro de aproximadamente 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50 o más aminoácidos a lo largo de la cadena de polipéptido desde el extremo N o C. Ejemplos no limitantes de NLSs incluyen una secuencia NLS derivada de: la NLS del antígeno T grande del virus SV40, que tiene la secuencia de aminoácidos PKKKRKV; la NLS del nucleoplasma (p. ej., la NLS bipartita nucleoplasmica con la secuencia KRPAATKKAGQAKKKK); las NLS c-myc que tienen la secuencia de aminoácidos PAAKRVKLD o RQRRNELKRSP; la NLS hRNPAI M9 que tiene la secuencia NQSSNFGPMKGGNFGGRSSGPGYGGGGQYFAKPRNQGGY; la secuencia RMRIZFKNKGKDTAELRRRVEVSVELRKAKKDEQILKRRNV del dominio IBB de importina-alfa; las secuencias VSRKRPRP y PPKKARED de la proteína T mioma; la secuencia POPKKKPL de p53 humana; la secuencia SALIKKKKMAP de ratón c-ab1 IV; las secuencias DRLRR y PQKKRK del virus NS1 de la gripe; la secuencia RKLKKKIKKL del antígeno delta del virus de la hepatitis; la secuencia REKKKFLKRR de la proteína Mx1 de ratón; la secuencia KRKGDEVDGVDEVAKKSKK de la poli(ADP-ribosa)polimerasa humana; y la secuencia RKCLQAGMNLARKTKK del glucocorticoide de los receptores de hormonas esteroides (humanas).

55 En general, la una o más NLSs son de una fuerza suficiente para impulsar la acumulación de la enzima de CRISPR en una cantidad detectable en el núcleo de una célula eucariota. En general, la fuerza de la actividad de localización nuclear puede derivar del número de NLSs en la enzima de CRISPR, la o las NLSs particulares utilizadas, o una combinación de estos factores. La detección de la acumulación en el núcleo puede ser realizada por cualquier técnica adecuada. Por ejemplo, un marcador detectable puede estar fusionado a la enzima CRISPR, de manera que

- 5 puede ser visualizada la ubicación dentro de una célula, tal como en combinación con medios para detectar la ubicación del núcleo (p. ej., un colorante específico para el núcleo, tal como DAPI). Los núcleos de las células también pueden aislarse de las células, cuyos contenidos pueden luego ser analizados mediante cualquier procedimiento adecuado para la detección de proteínas, tal como inmunohistoquímica, transferencia Western o un ensayo de actividad enzimática. La acumulación en el núcleo también puede determinarse indirectamente tal como mediante un ensayo para el efecto de la formación de complejos CRISPR (p. ej., ensayo para la escisión o mutación de ADN en la secuencia diana, o ensayo para la actividad de la expresión génica alterada, afectada por la formación de complejos de CRISPR y/o actividad de la enzima de CRISPR), en comparación con un control no expuesto a la enzima o complejo CRISPR, o expuesto a una enzima de CRISPR que carece de la una o más NLSs.
- 10 En general, una secuencia guía es cualquier secuencia de polinucleótidos que tiene suficiente complementariedad con una secuencia de polinucleótidos diana para hibridarse con la secuencia diana y dirigir la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana. En algunas realizaciones, el grado de complementariedad entre una secuencia guía y su secuencia diana correspondiente, cuando se alinean óptimamente utilizando un algoritmo de alineación adecuado, se trata de o más de aproximadamente 50%, 60%, 15 75%, 80%, 85%, 90%, 95%, 97,5%, 99%, o más. Un alineamiento óptimo puede ser determinado con el uso de cualquier algoritmo adecuado para el alineamiento de secuencias, ejemplos no limitativos de los cuales incluyen el algoritmo de Smith-Waterman, el algoritmo de Needleman-Wunsch, algoritmos basados en la Transformada de Burrows-Wheeler (p. ej., el Alineador de Burrows Wheeler), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; disponible en www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (disponible en soap.genomics.org.cn) y Maq (disponible en maq.sourceforge.net). En algunas realizaciones, una secuencia guía está relacionada en aproximadamente o más de aproximadamente 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75 o más nucleótidos de longitud. En algunas realizaciones, una secuencia guía es menor que aproximadamente 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, o menos nucleótidos de longitud. La capacidad de una secuencia de guía para dirigir unión específica para la secuencia de un complejo 20 CRISPR a una secuencia diana se puede determinar mediante cualquier ensayo adecuado. Por ejemplo, los componentes de un sistema de CRISPR, suficientes para formar un complejo de CRISPR, incluyendo la secuencia guía a ensayar, se pueden proporcionar a una célula huésped que tiene la secuencia diana correspondiente, tal como por transfección con vectores que codifican los componentes de la secuencia CRISPR, seguido de un estudio de escisión preferencial dentro de la secuencia diana tal como el ensayo Surveyor como se describe en esta memoria. De manera similar, la escisión de una secuencia de polinucleótido diana puede ser evaluada en un tubo de ensayo proporcionando la secuencia diana, componentes de un complejo de CRISPR, que incluye la secuencia guía a ensayar y una secuencia guía de control diferente de la secuencia guía de ensayo, y comparando la unión o la tasa de escisión en la secuencia diana entre las reacciones de ensayo y de la secuencia guía de control. Son posibles otros ensayos, y se les ocurrirán a los expertos en la técnica.
- 25 30
- 35 Nickase multiplezda: Aspectos de optimización y las enseñanzas de Cas9 detallada en esta solicitud también se pueden utilizar para generar nickases Cas9. Nickases Cas9 se pueden utilizar ventajosamente en combinación con pares de ARNs guía para generar roturas de la doble cadena de ADN con colgantes definidos. Cuando se utilizan dos pares de ARNs guía, es posible escindir un fragmento de ADN intermedio. Si un trozo exógeno de ADN es escindido por los dos pares de ARNs guía para generar colgantes compatibles con el ADN genómico, entonces el fragmento de ADN exógeno se puede ligar en el ADN genómico para reemplazar el fragmento escindido. Por ejemplo, esto se puede utilizar para eliminar la expansión de repetición de trinucleótidos en el gen huntingtina (HIT para el tratamiento de la enfermedad de Huntington.
- 40
- Cas9 y su ARN guía quimérico, o combinación de tracrRNA y crRNA, se pueden suministrar como ADN o ARN. El suministro de Cas9 y ARN guía ambos como moléculas de ARN (normal o con un contenido de bases o 45 modificaciones de la cadena principal) se puede utilizar para reducir la cantidad de tiempo que la proteína Cas9 persiste en la célula. Esto puede reducir el nivel de la actividad de escisión fuera de objetivo en la célula diana. Dado que el suministro de Cas9 como ARNm requiere tiempo para ser traducido en proteína, podría ser ventajoso suministrar el ARN guía varias horas después del suministro de ARNm de Cas9, para maximizar el nivel de ARN guía disponible para la interacción con la proteína Cas9.
- 50 En situaciones en las que la cantidad de ARN guía es limitante, puede ser deseable introducir Cas9 como ARNm y ARN guía en forma de un casete de expresión de ADN con un promotor que impulsa la expresión del ARN guía. De esta manera la cantidad de ARN guía disponible será amplificada a través de la transcripción.
- 55 Se puede introducir una diversidad de sistemas de suministro para introducir Cas9 (ADN o ARN) y ARN guía (ADN o ARN) en la célula huésped. Éstos incluyen el uso de liposomas, vectores virales, electroporación, nanopartículas, nanohilos (Shalek et al., Nano Letters, 2012), exosomas. Liposomas de caballos de Troya moleculares (Pardridge et

al., Cold Spring Harb Protoc; 2010; doi: 10.1101/pdb.prot5407) se pueden utilizar para suministrar Cas9 y ARN guía a través de la barrera hematoencefálica.

Discusión de ARNs guía para ortólogos: En algunas realizaciones, se selecciona una secuencia guía para reducir el grado de estructura secundaria dentro del ARN guía. En algunas realizaciones, aproximadamente o menos de aproximadamente 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, o menos de los nucleótidos del ARN guía participa en el apareamiento de bases auto-complementario cuando se pliegan de manera óptima. El plegamiento óptimo se puede determinar por cualquier algoritmo de plegamiento de polinucleótidos adecuado. Algunos programas se basan en el cálculo de la energía libre de Gibbs mínima. Un ejemplo de un algoritmo de este tipo es mFold, tal como se describe por Zuker y Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Otro ejemplo de algoritmo de plegamiento es el servidor web en línea RNAfold, desarrollado en el Instituto de Química Teórica de la Universidad de Viena, utilizando el algoritmo de predicción de estructura centroide (véase, p. ej. A.R. Gruber et al., 2008, Cell 106(1):23-24, y PA Carr y GM Church 2009, Nature Biotechnology 27(12): 1151-62). Un método de optimización del ARN guía de un ortólogo Cas9 comprende romper tramos poliU en el ARN guía. Tramos poliU que pueden romperse puede comprender una serie de 4, 5, 6, 7, 8, 9 ó 10 Us.

En general, una secuencia de emparejamiento tracr incluye cualquier secuencia que tiene una suficiente complementariedad con una secuencia tracr para fomentar uno o más de: (1) la escisión de una secuencia guía flanqueada por secuencias de emparejamiento tracr en una célula que contiene la secuencia tracr correspondiente; y (2) la formación de un complejo de CRISPR en una secuencia diana, en donde el complejo CRISPR comprende la secuencia de emparejamiento tracr hibridada a la secuencia de tracr. En general, el grado de complementariedad es con referencia al alineamiento óptimo de la secuencia de emparejamiento tracr y la secuencia tracr, a lo largo de la longitud de la más corta de las dos secuencias. El alineamiento óptimo puede ser determinado por cualquier algoritmo de alineamiento adecuado, y puede representar, además, estructuras secundarias tales como la auto-complementariedad dentro de la secuencia de tracr o la secuencia de emparejamiento tracr. En algunas realizaciones, el grado de complementariedad entre la secuencia tracr y la secuencia de emparejamiento tracr a lo largo de la longitud de la más corta de las dos cuando se alinean óptimamente es de aproximadamente o más de aproximadamente 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97,5%, 99%, o superior. En algunas realizaciones, la secuencia de tracr es aproximadamente o más de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, o más nucleótidos de longitud. En algunas realizaciones, la secuencia tracr y la secuencia de emparejamiento tracr están contenidas dentro de un único transcrito, de modo que la hibridación entre los dos produce un transcrito que tiene una estructura secundaria, tal como una horquilla. En una realización de la invención, el transcrito o secuencia de polinucleótido transcrita tiene al menos dos o más horquillas. En realizaciones preferidas, el transcrito tiene dos, tres, cuatro o cinco horquillas. En una realización adicional de la invención, el transcrito tiene a lo sumo cinco horquillas. En una estructura de horquilla, la parte de la secuencia 5' del final "N" y aguas arriba del bucle corresponde a la secuencia de emparejamiento de tracr, y la parte de la secuencia 3' del bucle corresponde a la secuencia de tracr. Ejemplos no limitantes adicionales de polinucleótidos sencillos que comprenden una secuencia guía, una secuencia de emparejamiento de tracr y una secuencia de tracr son como sigue (listados de 5' a 3'), en donde "N" representa una base de una secuencia guía, el primer bloque de letras en minúscula representa la secuencia de emparejamiento de tracr y el segundo bloque de letras en minúscula representa la secuencia tracr, y la secuencia poli-T final representa el terminador de la transcripción: 1)

NNNNNNNNNNNNNNNNNNNNgTTTTgtactctcaagatttGAAAtaaactctgcagaagctacaaagataaggctt
catgccgaaatcaacacctgtcattttatggcagggtgttttcgatttaaTTTTTT; (2)

NNNNNNNNNNNNNNNNNNNNgTTTTgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca
acacctgtcattttatggcagggtgttttcgatttaaTTTTTT; (3)

NNNNNNNNNNNNNNNNNNNNgTTTTgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatca
acacctgtcattttatggcagggtgtTTTTTT; (4)

NNNNNNNNNNNNNNNNNNNNgTTTTtagagctaGAAAtagcaagttaaataaggctagtcggtatcaactgaaaa
agtggcaccgagtcggtgcTTTTTT; (5)

NNNNNNNNNNNNNNNNNNNNgTTTTtagagctaGAAATAGcaagttaaataaggctagtcggtatcaactgaa
aaagtTTTTTTT; (6)

NNNNNNNNNNNNNNNNNNNNgTTTTtagagctagAAATAGcaagttaaataaggctagtcggtatcaTTTTT TTT. En algunas realizaciones, las secuencias (1) a (3) se utilizan en combinación con Cas9 de *S. thermophilus* CRISPR1. En algunas realizaciones, las secuencias (4) a (6) se utilizan en combinación con Cas9 de *S. pyogenes*. En algunas realizaciones, la secuencia de tracr es un transcrito separado de un transcrito que puede comprender la secuencia de emparejamiento de tracr.

Discusión de los emparejamientos de tracr para ortólogos: En algunas realizaciones, también se proporciona un molde de recombinación. Un molde de recombinación puede ser un componente de otro vector tal como se describe en esta memoria, contenido en un vector separado, o proporcionado como un polinucleótido separado. En algunas

realizaciones, un molde de recombinación está diseñado para servir como un molde en la recombinación homóloga tal como dentro o cerca de una secuencia diana mellada o escindida por una enzima de CRISPR como parte de un complejo de CRISPR. Un polinucleótido molde puede ser de cualquier longitud adecuada, tal como aproximadamente o más de aproximadamente 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, o más nucleótidos de longitud. En algunas realizaciones, el polinucleótido molde es complementario a una parte de un polinucleótido que puede comprender la secuencia diana. Cuando se alinean óptimamente, un polinucleótido molde podría solaparse con uno o más nucleótidos de una secuencia diana (p. ej., aproximadamente o más de aproximadamente 1, 5, 10, 15, 20, o más nucleótidos). En algunas realizaciones, cuando una secuencia de un molde y un polinucleótido, que comprende una secuencia diana, se alinean de manera óptima, el nucleótido más próximo del polinucleótido del molde está dentro de aproximadamente 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, o más nucleótidos de la secuencia diana.

En algunas realizaciones, la enzima de CRISPR es parte de una proteína de fusión que comprende uno o más dominios de proteínas heterólogos (p. ej., aproximadamente o más de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, o más dominios, además de la enzima de CRISPR). Una proteína de fusión enzima de CRISPR puede comprender cualquier secuencia de la proteína adicional, y opcionalmente una secuencia de unión entre cualquiera de dos dominios. Ejemplos de dominios de proteínas que pueden fusionarse con una enzima de CRISPR incluyen, sin limitación, macadores de epítomos, secuencias de gen indicador y dominios de proteínas que tienen una o más de las siguientes actividades: actividad metilasa, actividad desmetilasa, actividad de activación de la transcripción, actividad de represión de la transcripción, actividad del factor de liberación de la transcripción, actividad de la modificación de histonas, actividad de escisión del ARN y actividad de unión a ácido nucleico. Ejemplos no limitantes de marcadores de epítomos incluyen marcadores de histidina (His), marcadores V5, marcadores FLAG, marcadores de hemaglutinina de la gripe (HA), marcadores Myc, marcadores VSV-G y marcadores de tiorredoxina (Trx). Ejemplos de genes informadores incluyen, pero no se limitan a, glutatión-S-transferasa (GST), peroxidasa de rábano picante (HRP), cloranfenicol acetiltransferasa (CAT), beta-galactosidasa, beta-glucuronidasa, luciferasa, proteína fluorescente verde (GFP), HcRed, DsRed, proteína fluorescente cian (CFP), proteína amarilla fluorescente (YFP) y proteínas autofluorescentes incluyendo la proteína fluorescente azul (BFP). Una enzima de CRISPR puede fusionarse a una secuencia de gen que codifica una proteína o un fragmento de una proteína que une moléculas de ADN o une otras moléculas celulares, incluyendo pero no limitada a la proteína de unión a maltosa (MBP), S-tag, fusiones del dominio de unión a ADN (DBD) de Lex A, fusiones del dominio de unión a ADN de GAL4 y fusiones de la proteína BP16 del virus herpes simplex (HSV). Dominios adicionales que pueden formar parte de una proteína de fusión que puede comprender una enzima de CRISPR se describen en el documento US20110059502. En algunas realizaciones, se utiliza una enzima de CRISPR marcada para identificar la ubicación de una secuencia diana.

En algunas realizaciones, una enzima de CRISPR puede formar un componente de un sistema inducible. La naturaleza inducible del sistema permitiría un control de espacio-temporal de la edición gen o la expresión génica utilizando una forma de energía. La forma de energía puede incluir, pero no se limita a la radiación electromagnética, energía sonora, energía química y energía térmica. Ejemplos de sistema inducible incluyen promotores de tetraciclina inducibles (Tet-On o Tet-Off), sistemas de activaciones de transcripción de dos híbridos de moléculas pequeñas (FKBP, ABA, etc.), o sistemas inducibles por la luz (fitocromo, dominios LOV, o criptocromo). En una realización, la enzima CRISPR puede ser una parte de un Efecto Transcripcional Inducible por la Luz (LITE) para dirigir cambios en la actividad transcripcional de una manera específica para la secuencia. Los componentes de una luz pueden incluir una enzima de CRISPR, un heterodímero citocromo sensible a la luz (p. ej., de *Arabidopsis thaliana*), y un dominio de activación/represión transcripcional.

En algunos aspectos, la invención proporciona métodos que comprenden suministrar uno o más polinucleótidos, tales como o uno o más vectores tal como se describen en esta memoria, uno o más transcritos de los mismos, y/o una o más proteínas transcritas de los mismos, a una célula huésped. En algunos aspectos, la invención proporciona, además, células producidas por tales métodos, y animales que comprenden o son producidos a partir de tales células. En algunas realizaciones, una enzima de CRISPR en combinación con (y opcionalmente formando complejo con) una secuencia guía se suministra a una célula. Métodos de transferencia génica basados en virus y no virales convencionales pueden utilizarse para introducir ácidos nucleicos en células de mamífero o tejidos diana. Tales métodos se pueden utilizar para administrar ácidos nucleicos que codifican componentes de un sistema de CRISPR para células en cultivo, o en un organismo huésped. Sistemas de suministro de vectores no virales incluyen plásmidos de ADN, ARN (p. ej., un transcrito de un vector descrito en esta memoria), ácido nucleico desnudo y ácido nucleico formando complejo con un vehículo de suministro, tal como un liposoma. Sistemas de suministro de vectores virales incluyen virus de ADN y ARN, que tienen genomas episomales o integrados después dl suministro a la célula. Para una revisión de procedimientos de terapia génica, véase Anderson, *Science* 256: 808- 813 (1992); Nabel y Felgner, *TIBTECH* 11: 211-217 (1993); Mitani y Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357: 455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne,

Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer y Perricaudet. *British Medical Bulletin* 51(1):31-44 (1995); Haddada et al., en *Current Topics in Microbiology and Immunology* Doerfler y Böhm (comps.) (1995); y Yu et al, *Gene Therapy* 1:13-26 (1994).

5 Métodos de suministro no viral de ácidos nucleicos incluyen lipofección, microinyección, biolística, virosomas, liposomas, inmunoliposomas, policación o conjugados de lípido:ácido nucleico, ADN desnudo, viriones artificiales y absorción potenciada por agente de ADN. La lipofección se describe, p. ej., en las patentes de EE.UU. N^os. 5.049.386, 4.946.787; y 4.897.355) y reactivos de lipofección se venden comercialmente (p. ej., TransfectamTM y LipofectinTM). Lípidos catiónicos y neutros que son adecuados para una lipofección de reconocimiento de receptores eficiente de polinucleótidos incluyen los de Felgner, documento WO 91/17424; documento WO 91/16024. El
10 suministro puede ser a células (p. ej., administración in vitro o ex vivo) o tejidos diana (p. ej., la administración in vivo).

La preparación de complejos de lípido:ácido nucleico, incluyendo liposomas fijados como objetivo tales como complejos de inmunolípidos, es bien conocida para un experto en la técnica (véase, p. ej., *Crystal, Science* 270: 404-410 (1995); Blaese et al, *Cancer Gene Ther.* 2: 291-297 (1995); Behr et al, *Bioconjugate Chem.* 5: 382-389 (1994); Remy et al, *Bioconjugate Chem* 5: 647-654 (1994); Gao et al., *Gene Therapy* 2:710-722 (1995); Ahmad et al., *Cancer Res.* 52: 4817-4820 (1992); patentes de EE.UU. N^os 4.186.183, 4.217.344, 4.235.871, 4.261.975, 4.485.054, 4.501.728, 4.774.085, 4.837.028 y 4.946.787).

El uso de sistemas basados en virus de ARN o ADN para el suministro de ácidos nucleicos aprovecha los procesos altamente evolucionados para fijar como objetivo un virus a células específicas en el cuerpo y el tráfico de la carga útil viral al núcleo. Vectores virales pueden administrarse directamente a los pacientes (in vivo) o pueden utilizarse para tratar células in vitro y las células modificadas se pueden opcionalmente administrar a pacientes (ex vivo). Sistemas basados en virus convencionales podrían incluir vectores retroviral, de lentivirus, de virus adeno-asociados y de virus del herpes simplex para la transferencia de genes. La integración en el genoma huésped es posible con métodos de transferencia de genes de retrovirus, lentivirus y virus adeno-asociados, a menudo resulta en la expresión a largo plazo del transgén insertado. Adicionalmente, las elevadas eficiencias de la transducción se han observado en muchos tipos de células y tejidos diana diferentes.

El tropismo de un retrovirus puede alterarse por incorporación de proteínas de la envoltura extraña y expansión de la población diana potencial de células diana. Los vectores lentivirales son vectores retrovirales que son capaces de transducir o infectar células que no se dividen y que producen típicamente altos títulos virales. Por lo tanto, la selección de un sistema de transferencia de genes retroviral dependería del tejido diana. Los vectores retrovirales se componen de repeticiones cis-terminales de larga acción con capacidad de empaquetamiento para hasta 6-10 kb de secuencia extraña. LTRs que actúan en cis mínimos son suficientes para la replicación y el empaquetamiento de los vectores, que se utilizan luego para integrar el gen terapéutico en la célula objetivo para proporcionar una expresión del transgen permanente. Vectores retrovirales ampliamente utilizados incluyen los basados en virus de la leucemia murina (MuLV), virus de la leucemia del mono gibón (GaLV), virus de la inmunodeficiencia de simios (SIV), virus de la inmunodeficiencia humana (VIH) y combinaciones de los mismos (véase, p. ej., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al, *J. Virol.* 66: 1635-1640 (1992); Sommerfelt et al, *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63: 2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); documento PCT/US94/05700).

En aplicaciones en las que se prefiere la expresión transitoria, se pueden utilizar sistemas basados en adenovirus. Los vectores basados en adenovirus son capaces de una muy alta eficiencia de transducción en muchos tipos de células y no requieren la división celular. Con este tipo de vectores se han obtenido un título y niveles de expresión altos. Este vector puede producirse en grandes cantidades en un sistema relativamente simple.

Vectores de virus adeno-asociados ("AAV") también pueden utilizarse para transducir células con ácidos nucleicos diana, p. ej., en la producción in vitro de ácidos nucleicos y péptidos, y para procedimientos de terapia génica in vivo y ex vivo (véase, p. ej., West et al., *Virology* 160:38-47 (1987); patente de EE.UU. N^o. 4.797.368; documento WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). La construcción de vectores AAV recombinantes se describe en un cierto número de publicaciones, incluyendo la patente de EE.UU. N^o. 5.173.414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat y Muzyczka, *PNAS* 81:6466-6470 (1984); y Samulski et al., *J. Virol.* 63: 03822-3828 (1989).

Células de empaquetamiento se utilizan típicamente para formar partículas de virus que son capaces de infectar una célula huésped. Tales células incluyen células 293, que empaquetan adenovirus, y células ψ 2 o células PA317, que empaquetan retrovirus. Los vectores virales utilizados en terapia génica se generan habitualmente por el productor

de una línea celular que empaqueta un vector de ácido nucleico en una partícula viral. Los vectores contienen típicamente las secuencias virales mínimas requeridas para el empaquetamiento y la posterior integración en un huésped, siendo reemplazadas otras secuencias virales por una casete de expresión para el o los polinucleótidos a expresar. Las funciones virales que faltan se suministran normalmente en trans por la línea celular de empaquetamiento. Por ejemplo, los vectores AAV utilizados en terapia génica poseen típicamente sólo secuencias ITR del genoma de AAV que se requieren para el empaquetamiento y la integración en el genoma huésped. El ADN viral se empaqueta en una línea celular, que contiene un plásmido auxiliar que codifica los otros genes AAV, a saber rep y cap, pero que carece de secuencias ITR. La línea celular también puede estar infectada con adenovirus como un helper (cooperador). El virus helper fomenta la replicación del vector AAV y la expresión de genes AAV a partir del plásmido helper. El plásmido helper no está empaquetado en cantidades significativas debido a una carencia de secuencias ITR. La contaminación con adenovirus puede reducirse mediante, p. ej., tratamiento térmico al que el adenovirus es más sensible que AAV. Métodos adicionales para el suministro de ácidos nucleicos a las células son conocidos por los expertos en la técnica. Véase, por ejemplo, el documento US20030087817.

Por consiguiente, el AAV se considera un candidato ideal para su uso como un vector de transducción. Tales vectores de transducción AAV pueden comprender suficientes funciones que actúan en cis para replicar en presencia de adenovirus o virus herpes o virus de la viruela (p. ej., virus vacuna) funciones helper proporcionadas en trans. AAV recombinante (rAAV) puede utilizarse para transportar genes exógenos a células de una diversidad de linajes. En estos vectores, los genes cap y/o rep de AAV se eliminan del genoma viral y se reemplazan por un segmento de ADN de elección. Vectores AAV actuales pueden alojar hasta 4300 bases de ADN insertado.

Existe un cierto número de maneras de producir rAAV, y la invención proporciona rAAV y métodos para preparar rAAV. Por ejemplo, un plásmido o plásmidos que contienen o consisten esencialmente en la construcción viral deseada se transfectan en células infectadas con AAV. Además, un segundo o adicional plásmido helper se co-transfecta en estas células para proporcionar los genes cap y/o rep de AAV que son obligatorias para replicación y el empaquetamiento de la construcción viral recombinante. Bajo estas condiciones, las proteínas rep y/o cap de AAV actúan en trans para estimular la replicación y el empaquetamiento de la construcción de rAAV. Dos a tres días después de la transfección, se recoge el rAAV. Tradicionalmente el rAAV se recoge de las células junto con adenovirus. El adenovirus contaminante se inactiva luego por tratamiento térmico. En la presente invención, el rAAV se recoge ventajosamente no de las propias células, sino a partir de sobrenadante celular. Por consiguiente, en un aspecto inicial, la invención proporciona preparar rAAV, y además de lo anterior, rAAV se puede preparar por un método que comprende o consiste esencialmente en: infectar células susceptibles con un ADN exógeno que contiene rAAV, incluyendo ADN para la expresión, y virus helper (p. ej., adenovirus, virus herpes, virus de la viruela tales como el virus vacuna) en el que el rAAV carece de cap y/o rep funcionales (y el virus helper (p. ej., adenovirus, virus herpes, virus de la viruela tales como el virus vacuna) proporciona la función cap y/o rep de la que carece el rAAV); o infectar células susceptibles con un rAAV que contiene ADN exógeno incluyendo ADN para la expresión, en donde el recombinante carece de cap y/o rep funcional, y transfectar dichas células con un plásmido que suministra la función cap y/o rep de la que carece el rAAV; o infectar células susceptibles con un rAAV que contiene ADN exógeno incluyendo ADN para la expresión, en donde el recombinante carece de cap y/o rep funcional, en donde dichas células suministran la función cap y/o rep de la que carece el recombinante carece; o transfectar las células susceptibles con un AAV que carecen de cap y/o rep funcional y plásmidos para la inserción de ADN exógeno en el recombinante, de modo que el ADN exógeno se expresa por el recombinante y para suministrar las funciones rep y/o cap, con lo que la transfección resulta en un rAAV que contiene el ADN exógeno incluyendo ADN para la expresión que carece de cap y/o rep funcional.

El rAAV puede ser de un AAV tal como se describe en esta memoria, y ventajosamente puede ser un rAAV1, rAAV2, AAV5 o rAAV que tiene un híbrido o una cápside que puede comprender AAV1, AAV2, AAV5 o cualquier combinación de los mismos. Se puede seleccionar el AAV del rAAV con respecto a las células a ser fijadas como objetivo por el rAAV; p. ej., se puede seleccionar AAV serotipos 1, 2, 5 o un híbrido o la cápside de AAV 1, AAV2, AAV5 o cualquier combinación de los mismos para fijar como objetivo células del cerebro o neuronales; y se puede seleccionar AAV4 para fijar como objetivo tejido cardíaco, AAV8 para fijar como objetivo tejido del hígado.

Además de las células 293, en la práctica de la invención se pueden utilizar otras células y la infectividad relativa de determinados serotipos de AAV in vitro en cuanto a estas células (véase Grimm, D. et al, J. Virol 82.: 5887-5911 (2008)) es la siguiente:

ES 2 576 126 T3

Línea Celular	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2,5	0,0	0,1	10	0,7	0,0
HEK293	25	100	2,5	0,1	0,1	5	0,7	0,1
HeLa	3	100	2,0	0,1	6,7	1	0,2	0,1
HepG2	3	100	16,7	0,3	1,7	5	0,3	ND
Hep1A	20	100	0,2	1,0	0,1	1	0,2	0,0
911	17	100	11	0,2	0,1	17	0,1	ND
CHO	100	100	14	1,4	333	50	10	1,0
COS	33	100	33	3,3	5,0	14	2,0	0,5
MeWo	10	100	20	0,3	6,7	10	1,0	0,2
NIH3T3	10	100	2,9	2,9	0,3	10	0,3	ND
A549	14	100	20	ND	0,5	10	0,5	0,1
HT1180	20	100	10	0,1	0,3	33	0,5	0,1
Monocitos	1111	100	ND	ND	125	1429	ND	ND
DC inmadura	2500	100	ND	ND	222	2857	ND	ND
DC madura	2222	100	ND	ND	333	3333	ND	ND

5 La invención proporciona rAAV que contiene o consiste esencialmente en una molécula de ácido nucleico exógeno que codifica un sistema de CRISPR (Repeticiones Palindrómicas Cortas Agrupadas y Regularmente Interespaciadas), p. ej., una pluralidad de casetes que comprenden o consisten en una primera casete que comprende o consiste esencialmente en un promotor, una molécula de ácido nucleico que codifica una proteína CRISPR-asociada (Cas) (proteínas nucleasa o helicasa putativas), p. ej., Cas9 y un terminador, y dos, o más, de manera ventajosa hasta el límite de tamaño de empaquetamiento del vector, p. ej., en total (incluyendo la primera casete) de cinco casetes que comprenden o consisten esencialmente en un promotor, una molécula de ácido nucleico que codifica ARN guía (gRNA) y un terminador (p. ej., cada una de las casetes representada esquemáticamente como Promotor-gRNA1-terminador, Promotor-gRNA2-terminador.... Promotor-gRNA(N)-terminador (en donde N es un número que puede ser insertado que se encuentra en un límite superior del límite de tamaño de empaquetamiento del vector), o dos o más rAAVs individuales, cada uno de los cuales contiene una o más de una casete de un sistema de CRISPR, p. ej., un primer rAAV que contiene la primera casete que comprende o que consiste esencialmente en un promotor, una molécula de ácido nucleico que codifica Cas, p. ej., Cas9 y un terminador y un segundo rAAV que contiene una pluralidad, cuatro, casetes que comprende o consiste

10

15

esencialmente en un promotor, molécula de ácido nucleico que codifica ARN guía (gRNA) y un terminador (p. ej., cada una de las casetes representada esquemáticamente como Promotor-gRNA1-terminador, Promotor-gRNA2-terminador, Promotor-gRNA(N)-terminador (en donde N es un número que puede ser insertado que se encuentra en un límite superior del límite de tamaño de empaquetamiento del vector). Dado que el rAAV es un virus de ADN, las moléculas de ácido nucleico en la discusión en esta memoria en relación con AAV o rAAV son ventajosamente ADN. El promotor es en algunas realizaciones ventajosamente promotor de Sinapsina 1 humana (hSyn).

5 Se prefieren dos formas de empaquetar moléculas de ácido nucleico codificadoras de Cas9, p. ej., ADN, en vectores virales para mediar en la modificación del genoma in vivo:
Para lograr una inactivación del mediada por NHEJ:

10 Vector de virus sencillo:
Vector que contiene dos o más casetes de expresión:
Promotor-molécula de ácido nucleico que codifica Cas9-terminador
Promotor-gRNA1-terminador
Promotor-gRNA2-terminador
15 Promotor-gRNA(N)-terminador (hasta el límite de tamaño del vector)

Vector de virus doble:
Vector 1 que contiene una casete de expresión para impulsar la expresión de Cas9
Promotor-molécula de ácido nucleico que codifica Cas9-terminador
Vector 2 que contiene uno o más casetes de expresión para impulsar la expresión de uno o más RNAs guía
20 Promotor-gRNA1-terminador
Promotor-gRNA(N)-terminador (hasta el límite de tamaño del vector)

Mediar en la reparación directa de la homología. Además de los enfoques de vector de virus sencillo y doble descritos anteriormente, se utiliza un vector adicional para suministrar un molde de reparación directa de la homología.

25 El promotor utilizado para impulsar la expresión de la molécula de ácido nucleico que codifica Cas9 puede incluir:
AAV ITR puede servir como un promotor: esto es ventajoso para eliminar la necesidad de un elemento promotor adicional (que puede ocupar espacio en el vector). El espacio adicional liberado se puede utilizar para impulsar la expresión de elementos adicionales (gRNA, etc.). Además, la actividad ITR es relativamente más débil, por lo que se puede utilizar para reducir la toxicidad debido a la sobre-expresión de Cas9.

30 Para la expresión ubicua, se pueden utilizar promotores: CMV, CAG, CBh, PGK, SV40, cadenas pesada o ligera de ferritina, etc.

Para la expresión en el cerebro, se pueden utilizar promotores: Sinapsina I para todas las neuronas, CaMKIIalpha para las neuronas excitadoras, GAD67 o GAD65 o VGAT para las neuronas GABAérgicas, etc.

35 Para la expresión en el hígado, se puede utilizar promotor de albúmina
Para la expresión en pulmón, se puede utilizar SP-B
Para células endoteliales, se puede utilizar ICAM
Para células hematopoyéticas, se puede utilizar IFNbeta o CD45,
Para osteoblastos, se puede utilizar OG-2
El promotor utilizado para impulsar ARN guía puede incluir:

40 Promotores Pol III tales como U6 o H1
Uso de promotor Pol II y casetes intrónicas para expresar gRNA

En cuanto a AAV, el AAV puede ser AAV1, AAV2, AAV5 o cualquier combinación de los mismos. Se puede seleccionar el AAV del AAV en relación con las células a fijar como objetivo; p. ej., se puede seleccionar AAV serotipos 1, 2, 5 o un AAV1, AAV2, AAV5 híbrido o de la cápside o cualquier combinación de los mismos para fijar como objetivo el cerebro o células neuronales; y se puede seleccionar AAV4 para fijar como objetivo tejido cardíaco. AAV8 es útil para el suministro al hígado. Los promotores y vectores anteriores se prefieren individualmente.

En algunas realizaciones, una célula huésped es transfectada de forma transitoria o no transitoria con uno o más vectores descritos en esta memoria. En algunas realizaciones, una célula se transfecta tal como se produce de forma natural en un sujeto. En algunas realizaciones, una célula que se transfecta se toma de un sujeto. En algunas realizaciones, la célula se deriva de células tomadas de un sujeto tal como una línea celular. En la técnica se conoce una amplia diversidad de líneas celulares para el cultivo de tejidos. Ejemplos de líneas celulares incluyen, pero no se limitan a, C8161 CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEK293T,

HEKa, MiaPaCell, Panel, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 epitelial del riñón de mono, fibroblastos de embrión de ratón BALB/ 3T3, 3T3 Swiss, 3T3-L1, fibroblastos
 5 fetales humanos 132-d5; fibroblastos de ratón 10.1, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, células BCP-1, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-
 10 29, Jurkat, células JY, células K562, Ku812, KCL22, KG1, KEYO1, LNCap, Ma-ME1 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, células Saos-2, Sf-9, SkBr3, T2, T-47D, T84, línea celular THP1, U373, U87, U937, VCaP, células Vero, WM39, WT-49, X63, YAC-1, YAR, y variedades transgénicas de las mismas.
 15 Las líneas celulares están disponibles de una diversidad de fuentes conocidas por los expertos en la técnica (véase, p. ej., American Type Culture Collection (ATCC) (Manassus, Va.)). En algunas realizaciones, una célula transfectada con uno o más vectores descritos en esta memoria se utiliza para establecer una nueva línea celular que puede comprender una o más secuencias derivadas del vector. En algunas realizaciones, una célula transfectadas transitoriamente con los componentes de un sistema de CRISPR, tal como se describe en esta memoria (por
 20 ejemplo, por transfección transitoria de uno o más vectores, o la transfección con ARN), y modificadas a través de la actividad de un complejo de CRISPR, se utiliza para establecer una nueva línea celular que puede comprender células que contienen la modificación, pero que carecen de cualquier otra secuencia exógena. En algunas realizaciones, las células transfectadas de forma transitoria o no transitoria con uno o más vectores descritos en esta memoria, o líneas celulares derivadas de estas células se utilizan en la evaluación de uno o más compuestos de
 25 ensayo.

En algunas realizaciones, uno o más vectores descritos en esta memoria se utilizan para producir un animal transgénico no humano o planta transgénica. En algunas realizaciones, el animal transgénico es un mamífero, tal como un ratón, rata o conejo. Métodos para producir plantas y animales transgénicos son conocidos en la técnica, y generalmente comienzan con un método de transfección celular tal como se describe en esta memoria.

30 Con los recientes progresos en genómica de cultivos, la capacidad de utilizar los sistemas de CRISPR-Cas para llevar a cabo la edición y la manipulación de genes eficiente y rentable permitirá la rápida selección y la comparación de manipulaciones genéticas individuales y multiplexadas para transformar dichos genomas para una producción mejorada y rasgos potenciados. A este respecto se hace referencia a las patentes y publicaciones de EE.UU. Patente de EE.UU. Nº 6.603.061 - Agrobacterium-Mediated Plant Transformation Method; patente de EE.UU. Nº.
 35 7.868.149 - Plant Genome Sequences and Uses Thereof y de EE.UU. 2009/0100536 – Transgenic Plants with Enhanced Agronomic Traits. En la práctica de la invención también se mencionan el contenidos y la descripción de Morrell et al "Crop genomics: advances and applications" Nat Rev Genet. 29 dic. 2011;13(2):85-96 En una forma de realización ventajosa de la invención, el sistema CRISPR/Cas9 se utiliza para modificar genéticamente microalgas (Ejemplo 7). Por consiguiente, referencia en esta memoria a células animales también se puede aplicar, mutatis
 40 mutandis, a células vegetales, a menos que de otro modo resulte evidente.

En un aspecto, la invención proporciona métodos de modificar un polinucleótido diana en una célula eucariota, que puede ser in vivo, ex vivo o in vitro. En algunas realizaciones, el método comprende el muestreo de una célula o población de células de un animal humano o no humano o planta (incluyendo micro-algas), y la modificación de la
 45 célula o células. El cultivo puede producirse en cualquier fase ex vivo. La célula o las células pueden incluso ser re-introducidos en el animal no humano o la planta (incluyendo micro-algas). Para las células reintroducidas se prefiere particularmente que las células sean células madre.

En algunas realizaciones, el método comprende permitir que un complejo de CRISPR se una al polinucleótido diana para efectuar la escisión de dicho polinucleótido diana, modificando así el polinucleótido diana, en el que el complejo de CRISPR puede comprender una enzima de CRISPR formando complejo con una secuencia guía hibridada a una
 50 secuencia diana dentro de dicho polinucleótido diana, en el que dicha secuencia guía está unida a una secuencia de emparejamiento tracr que a su vez se hibrida a una secuencia de tracr.

En un aspecto, la invención proporciona un método de modificar la expresión de un polinucleótido en una célula eucariota. En algunas realizaciones, el método comprende permitir que un complejo de CRISPR se una al polinucleótido de manera que dicha unión resulte en una expresión incrementada o reducida de dicho polinucleótido;
 55 en el que el complejo de CRISPR comprende una enzima de CRISPR formando complejo con una secuencia guía

hibridada a una secuencia diana dentro de dicho polinucleótido, en el que dicha secuencia guía está enlazada a una secuencia de emparejamiento tracr que a su vez se hibrida a una secuencia de tracr. Consideraciones similares se aplican como antes a métodos para modificar un polinucleótido diana. De hecho, estas opciones de muestreo, cultivo y re-introducción se aplican a lo largo de los aspectos de la presente invención.

5 En un aspecto, la invención proporciona kits que contienen uno o más de los elementos descritos en los métodos y las composiciones anteriores. Los elementos pueden proporcionarse individualmente o en combinaciones, y se pueden proporcionar en cualquier recipiente adecuado, tal como un vial, una botella, o un tubo. En algunas realizaciones, el kit incluye instrucciones en uno o más idiomas, por ejemplo, en más de un idioma.

10 En algunas realizaciones, un kit comprende uno o más reactivos para su uso en un proceso que utiliza uno o más de los elementos descritos en esta memoria. Los reactivos pueden proporcionarse en cualquier recipiente adecuado. Por ejemplo, un kit puede proporcionar uno o más tampones de reacción o de almacenamiento. Los reactivos se pueden proporcionar en una forma que sea utilizable en un ensayo particular, o en una forma que requiere la adición de uno o más de otros componentes antes de su uso (p. ej., en forma de concentrado o liofilizada). Un tampón puede ser cualquier tampón, incluyendo, pero no limitado a un tampón carbonato de sodio, un tampón bicarbonato de sodio, un tampón borato, un tampón Tris, un tampón MOPS, un tampón HEPES, y combinaciones de los mismos. En algunas realizaciones, el tampón es alcalino. En algunas realizaciones, el tampón tiene un pH de aproximadamente 7 a aproximadamente 10. En algunas realizaciones, el kit comprende uno o más oligonucleótidos correspondientes a una secuencia guía para la inserción en un vector con el fin de enlazar operativamente la secuencia guía y un elemento regulador. En algunas realizaciones, el kit comprende un polinucleótido molde de recombinação homóloga.

15 En un aspecto, la invención proporciona métodos para usar uno o más elementos de un sistema de CRISPR. El complejo de CRISPR de la invención proporciona un medio eficaz para la modificación de un polinucleótido diana. El complejo de CRISPR de la invención tiene una amplia variedad de utilidad, incluyendo modificar (por ejemplo, suprimir, insertar, translocar, inactivar, activar) un polinucleótido diana en una multiplicidad de tipos de células. Como tal, el complejo de CRISPR de la invención tiene un amplio espectro de aplicaciones en, p. ej., la terapia génica, el rastreo de fármacos, el diagnóstico de enfermedades y el pronóstico. Un complejo de CRISPR a modo de ejemplo puede comprender una enzima de CRISPR formando complejo con una secuencia guía hibridada a una secuencia diana dentro del polinucleótido diana. La secuencia guía está enlazada a una secuencia de emparejamiento tracr que, a su vez, se hibrida a una secuencia de tracr.

25 En una realización, esta invención proporciona un método para escindir un polinucleótido diana. El método puede comprender modificar un polinucleótido diana utilizando un complejo de CRISPR que se une al polinucleótido diana y efectúa la escisión de dicho polinucleótido diana. Típicamente, el complejo de CRISPR de la invención, cuando se introduce en una célula, crea una ruptura (p. ej., una rotura de la cadena sencilla o doble) en la secuencia del genoma. Por ejemplo, el método puede utilizarse para escindir un gen de la enfermedad en una célula.

30 La ruptura creada por el complejo de CRISPR puede ser reparada por unos procesos de reparación tales como la vía de unión de extremos no homólogos (NHEJ) propensa a errores o la reparación dirigida a la homología de alta fidelidad (HDR). Durante estos procesos de reparación, un molde de polinucleótido exógeno puede ser introducido en la secuencia del genoma. En algunos métodos, el proceso de HDR se utiliza para modificar la secuencia del genoma. Por ejemplo, un molde de polinucleótido exógeno que puede comprender una secuencia que se ha de integrar flanqueada por una secuencia aguas arriba y una secuencia aguas abajo se introduce en una célula. Las secuencias aguas arriba y aguas abajo comparten similitud de secuencia con cualquiera de los lados del sitio de integración en el cromosoma.

35 Cuando se desee, un polinucleótido donante puede ser ADN, p. ej., un plásmido de ADN, un cromosoma bacteriano artificial (BAC), un cromosoma artificial de levadura (YAC), un vector viral, un trozo lineal de ADN, un fragmento de la PCR, un ácido nucleico desnudo, o un ácido nucleico formando complejo con un vehículo de suministro tal como un liposoma o poloxámero.

40 El molde de polinucleótido exógeno comprende una secuencia para ser integrada (p. ej., un gen mutado). La secuencia de integración puede ser una secuencia endógena o exógena a la célula. Ejemplos de una secuencia a ser integrada incluyen polinucleótidos que codifican una proteína o un ARN no codificante (p. ej., un microARN). Por lo tanto, la secuencia de la integración puede estar operativamente enlazada a una secuencia o secuencias de control apropiadas. Alternativamente, la secuencia a ser integrada puede proporcionar una función reguladora.

50

Las secuencias de aguas arriba y aguas abajo del molde de polinucleótido exógeno se seleccionan para fomentar la recombinación entre la secuencia cromosómica de interés y el polinucleótido donante. La secuencia de aguas arriba es una secuencia de ácido nucleico que comparte similitud de secuencia con la secuencia del genoma de aguas arriba del sitio diana para la integración. De manera similar, la secuencia de aguas abajo es una secuencia de ácido nucleico que comparte similitud de secuencia con la secuencia cromosómica aguas abajo del sitio diana de la integración. Las secuencias aguas arriba y aguas abajo en el molde de polinucleótido exógeno puede tener 75%, 80%, 85%, 90%, 95% o 100% de identidad de secuencia con la secuencia del genoma de destino. Preferiblemente, las secuencias de aguas arriba y aguas abajo en la plantilla de polinucleótido exógeno tienen aproximadamente 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con la secuencia del genoma fijada como objetivo. En algunos métodos, las secuencias de aguas arriba y aguas abajo en el molde de polinucleótido exógeno tienen aproximadamente el 99% o el 100% de identidad de la secuencia con la secuencia del genoma fijada como objetivo.

Una secuencia de aguas arriba o aguas abajo puede comprender de aproximadamente 20 pb a aproximadamente 2500 pb, por ejemplo, aproximadamente 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400 ó 2500 pb. En algunos métodos, la secuencia aguas arriba o aguas abajo a modo de ejemplo tiene aproximadamente 200 pb a aproximadamente 2000 pb, aproximadamente 600 pb a aproximadamente 1000 pb, o más particularmente aproximadamente 700 pb a aproximadamente 1000 pb.

En algunos métodos, el molde de polinucleótido exógeno puede comprender, además, un marcador. Este marcador puede hacer más fácil rastrear integraciones fijadas como objetivo. Ejemplos de marcadores adecuados incluyen sitios de restricción, proteínas fluorescentes o marcadores seleccionables. El molde de polinucleótido exógeno de la invención puede construirse utilizando técnicas recombinantes (véase, por ejemplo, Sambrook et al., 2001 y Ausubel et al., 1996).

En un método a modo de ejemplo, para la modificación de un polinucleótido diana mediante la integración de un molde de polinucleótido exógeno, se introduce en la secuencia del genoma una rotura de la cadena doble por el complejo de CRISPR, la ruptura se repara mediante recombinación homóloga de un molde de polinucleótido exógeno de tal forma que el molde se integra en el genoma. La presencia de una rotura de cadena doble facilita la integración del molde.

En otras realizaciones, esta invención proporciona un método para modificar la expresión de un polinucleótido en una célula eucariota. El método puede comprender aumentar o disminuir la expresión de un polinucleótido diana utilizando un complejo de CRISPR que se une al polinucleótido,

En algunos métodos, un polinucleótido diana puede ser inactivado para efectuar la modificación de la expresión en una célula. Por ejemplo, tras la unión de un complejo de CRISPR a una secuencia diana en una célula, el polinucleótido diana se inactiva de manera que la secuencia no se transcribe, la proteína codificada no se produce o la secuencia no funciona como la secuencia de tipo salvaje. Por ejemplo, una proteína o secuencia codificadora de microARN puede inactivarse de manera que no se produzca la proteína.

En algunos métodos, una secuencia de control puede ser inactivada de manera que ya no funcione como una secuencia de control. Tal como se utiliza en esta memoria, "secuencia de control" se refiere a cualquier secuencia de ácido nucleico que efectúa la transcripción, traducción o la accesibilidad de una secuencia de ácido nucleico. Ejemplos de una secuencia de control incluyen un promotor, un terminador de la transcripción, y un promotor son secuencias de control.

La secuencia diana inactivada puede incluir una mutación por delección (es decir, la delección de uno o más nucleótidos), una mutación por inserción (es decir, la inserción de uno o más nucleótidos), o una mutación sin sentido (es decir, la sustitución de un único nucleótido con otro nucleótido de modo que se introduce un codón de parada). En algunos métodos, la inactivación de una secuencia diana da como resultado la "inactivación" de la secuencia diana.

Una expresión alterada de una o más secuencias del genoma asociadas con una vía de señalización bioquímica puede ser determinada por ensayo de una diferencia en los niveles de ARNm de los genes correspondientes entre la célula modelo de ensayo y una célula control, cuando están en contacto con un agente candidato. Alternativamente, la expresión diferencial de las secuencias asociadas con una vía de señalización bioquímica se determina mediante la detección de una diferencia en el nivel del producto génico o polipéptido codificado.

Para ensayar para una alteración inducida por el agente en el nivel de transcritos de ARNm o polinucleótidos correspondientes, ácido nucleico contenido en una muestra se extrae primero de acuerdo con métodos estándares en la técnica. Por ejemplo, el ARNm puede aislarse utilizando diversas enzimas líticas o disoluciones químicas de acuerdo con los procedimientos establecidos en Sambrook et al. (1989), o puede extraerse mediante resinas de unión a ácido nucleico siguiendo las instrucciones adjuntas proporcionadas por los fabricantes. El ARNm contenido en la muestra de ácido nucleico extraído se detecta a continuación mediante procedimientos de amplificación o ensayos de hibridación convencionales (p. ej., análisis de transferencia Northern) de acuerdo con métodos ampliamente conocidos en la técnica o en base a los métodos ejemplificados en esta memoria.

Para los fines de esta invención, amplificación significa cualquier método que emplea un cebador y una polimerasa capaz de replicar una secuencia diana con fidelidad razonable. La amplificación puede ser portada por el ADN-polimerasas naturales o recombinantes tales como TaqGold™, T7 ADN polimerasa, el fragmento Klenow de la ADN polimerasa de E. coli y transcriptasa inversa. Un método de amplificación preferido es la PCR. En particular, el ARN aislado se puede someter a un ensayo de transcripción inversa que está acoplado con una reacción en cadena de la polimerasa cuantitativa (RT-PCR) con el fin de cuantificar el nivel de expresión de una secuencia asociada con una vía de señalización bioquímica.

La detección del nivel de expresión génica puede llevarse a cabo en tiempo real en un ensayo de amplificación. En un aspecto, los productos amplificados pueden ser directamente visualizados con agentes de unión a ADN fluorescentes, incluyendo, pero no limitados a, intercaladores de ADN y aglomerantes de de la muesca de ADN. Debido a que la cantidad de los intercaladores incorporados en las moléculas de ADN de doble cadena es típicamente proporcional a la cantidad de los productos de ADN amplificados, se puede determinar convenientemente la cantidad de los productos amplificados mediante la cuantificación de la fluorescencia del colorante intercalado utilizando sistemas ópticos convencionales en la técnica. Colorante de unión a ADN adecuado para esta aplicación incluyen verde SYBR, azul SYBR, DAPI, yoduro de propidio, Hoeste, oro SYBR, bromuro de etidio, acridinas, proflavina, naranja de acridina, acriflavina, fluorcoumanina, elipticina, daunomicina, cloroquina, distamicina D, cromomicina, homidio, mitramicina, polipiridilos rutenio, antramycin, y similares.

En otro aspecto, otros marcadores fluorescentes tales como sondas específicas para la secuencia se pueden emplear en la reacción de amplificación para facilitar la detección y cuantificación de los productos amplificados. La amplificación cuantitativa basada en la sonda se basa en la detección específica para la secuencia de un producto amplificado deseado. Utiliza sondas específicas para la diana, fluorescentes (p. ej., sondas TaqMan®) resultando en una mayor especificidad y sensibilidad. Métodos para realizar la amplificación cuantitativa basada en la sonda están bien establecidos en la técnica y se enseña en la Patente de EE.UU. N° 5.210.015.

Todavía en otro aspecto, se pueden realizar ensayos de hibridación convencionales utilizando sondas de hibridación que comparten homología de secuencia con las secuencias asociadas con una vía de señalización bioquímica. Típicamente, se deja que las sondas formen complejos estables con las secuencias asociadas con una vía de señalización bioquímica contenida dentro de la muestra biológica derivada del sujeto de ensayo en una reacción de hibridación. Se apreciará por un experto en la técnica que, cuando se utilice antisentido como el ácido nucleico sonda, los polinucleótidos diana proporcionados en la muestra se eligen para que sean complementarios a las secuencias de los ácidos nucleicos antisentido. En cambio, cuando la sonda de nucleótidos es un ácido nucleico sentido, el polinucleótido diana se selecciona para que sea complementario a las secuencias del ácido nucleico sentido.

La hibridación puede realizarse en condiciones de diferentes rigurosidad, por ejemplo tal como se describen en esta memoria. Condiciones de hibridación adecuadas para la práctica de la presente invención son tales que la interacción de reconocimiento entre la sonda y secuencias asociadas con una vía de señalización bioquímica es a la vez lo suficientemente específica y lo suficientemente estable. Condiciones que aumentan la rigurosidad de una reacción de hibridación son ampliamente conocidas y están publicadas en la técnica. Véase, por ejemplo, (Sambrook, et al., (1989); Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, segunda edición). El ensayo de hibridación se puede formar utilizando sondas inmovilizadas sobre cualquier soporte sólido, incluyendo, pero no limitadas a nitrocelulosa, vidrio, silicio, y una diversidad de matrices de genes. Un ensayo de hibridación preferido se lleva a cabo en chips de genes de alta densidad tal como se describe en la Patente de EE.UU. N° 5.445.934.

Para una detección conveniente de los complejos de sonda-diana que se forman durante el ensayo de hibridación, las sondas de nucleótidos se conjugan a un marcador detectable. Marcadores detectables adecuados para su uso en la presente invención incluyen cualquier composición detectable por medios fotoquímicos, bioquímicos,

5 espectroscópicos, inmunoquímicos, eléctricos, ópticos o químicos. Una amplia variedad de marcadores detectables adecuados son conocidos en la técnica, que incluyen marcadores fluorescentes o quimioluminiscentes, marcadores de isótopos radiactivos, enzimáticos u otros ligandos. En realizaciones preferidas, probablemente se deseará emplear un marcador fluorescente o una etiqueta de enzima tal como digoxigenina, β -galactosidasa, ureasa, fosfatasa alcalina o peroxidasa, complejo de avidina/biotina.

10 Los métodos de detección utilizados para detectar o cuantificar la intensidad de hibridación dependerán típicamente del marcador arriba seleccionado. Por ejemplo, los marcadores radiactivos se pueden detectar utilizando una película fotográfica o un phosphoimager. Los marcadores fluorescentes pueden ser detectados y cuantificados utilizando un fotodetector para detectar la luz emitida. Los marcadores enzimáticos se detectan típicamente
 10 proveyendo a la enzima de un sustrato y midiendo el producto de reacción producido por la acción de la enzima sobre el sustrato; y finalmente marcadores colorimétricos son detectadas simplemente visualizando el marcador coloreado.

15 Un cambio inducido por agente en la expresión de secuencias asociadas con una vía de señalización bioquímica también se puede determinar mediante el examen de los productos génicos correspondientes. La determinación del nivel de proteína implica típicamente a) poner en contacto la proteína contenida en una muestra biológica con un agente que se une específicamente a una proteína asociada con una vía bioquímica de señalización; y (b) identificar cualquier complejo de agente:proteína así formado. En un aspecto de esta realización, el agente que une específicamente una proteína asociada con una vía de señalización bioquímica es un anticuerpo, preferiblemente un anticuerpo monoclonal.

20 La reacción se lleva a cabo poniendo en contacto el agente con una muestra de las proteínas asociadas con una vía de señalización bioquímica derivada de las muestras de ensayo en condiciones que permitan que se forme un complejo entre el agente y las proteínas asociadas con una vía de señalización bioquímica. La formación del complejo se puede detectar directa o indirectamente de acuerdo con procedimientos estándares en la técnica. En el método de detección directo, los agentes se suministran con un marcador detectable y agentes que no han reaccionado pueden separarse del complejo; la cantidad de marcador restante indicando con ello la cantidad de complejo formado. Para un método de este tipo, es preferible seleccionar los marcadores que permanecer unidos a los agentes, incluso durante condiciones de lavado rigurosas. Es preferible que el marcador no interfiera con la reacción de unión. En la alternativa, un procedimiento de detección indirecta puede utilizar un agente que contiene un marcador introducido química o enzimáticamente. Un marcador deseable no interfiere generalmente con la unión o la estabilidad del complejo agente:polipéptido resultante. Sin embargo, el marcador se diseña típicamente para
 25 que sea accesible a un anticuerpo para una unión eficaz y, por lo tanto, para generar una señal detectable.

Una amplia diversidad de marcadores adecuados para la detección de los niveles de proteína son conocidos en la técnica. Ejemplos no limitantes incluyen radioisótopos, enzimas, metales coloidales, compuestos fluorescentes, compuestos bioluminiscentes y compuestos quimioluminiscentes.

35 La cantidad de complejos agente:polipéptido formados durante la reacción de unión se puede cuantificar mediante ensayos cuantitativos estándares. Tal como se ilustra anteriormente, la formación de complejo agente:polipéptido se puede medir directamente por la cantidad de marcador permaneció en el sitio de unión. En una alternativa, la proteína asociada con una vía de señalización bioquímica se testa en cuanto a su capacidad para competir con un análogo marcado por sitios de unión en el agente específico. En este ensayo competitivo, la cantidad de marcador capturado es inversamente proporcional a la cantidad de secuencias de proteínas asociadas con una vía de señalización bioquímica presente en una muestra de ensayo.

40 En la técnica está disponible un cierto número de técnicas para el análisis de proteínas sobre la base de los principios generales esbozados anteriormente. Estos incluyen, pero no se limitan a radioinmunoensayos, ELISA (ensayos inmunoradiométricos ligados a enzima), inmunoensayos "sándwich", ensayos inmunoradiométricos, inmunoensayos in situ (utilizando, p. ej., oro coloidal, enzimas o radioisótopos etiquetas), análisis de transferencia Western, ensayos de inmunoprecipitación, ensayos de inmunofluorescencia y SDS-PAGE.

50 Anticuerpos que reconocen específicamente o se unen a proteínas relacionadas con una vía de señalización bioquímica son preferibles para la realización de los análisis de proteínas mencionados anteriormente. En los casos en los que se desee, se pueden utilizar los anticuerpos que reconocen un tipo específico de modificaciones post-traduccionales (p. ej., modificaciones inducibles en la vía bioquímica de señalización). Modificaciones post-traduccionales incluyen, pero no se limitan a glicosilación, lipidación, acetilación y fosforilación. Estos anticuerpos pueden adquirirse de proveedores comerciales. Por ejemplo, los anticuerpos anti-fosfotirosina que reconocen

específicamente las proteínas tirosina-fosforiladas están disponibles de un número de proveedores, incluyendo Invitrogen y Perkin Elmer. Anticuerpos anti-fosfotirosina son particularmente útiles en la detección de proteínas que son fosforiladas diferencialmente en sus residuos de tirosina en respuesta a un estrés ER. Tales proteínas incluyen, pero no se limitan al factor 2 alfa de iniciación de la traducción eucariótica (eIF-2 α). Alternativamente, estos anticuerpos pueden generarse utilizando tecnologías de anticuerpos policlonales o monoclonales convencionales mediante la inmunización de un animal huésped o una célula productora de anticuerpos con una proteína diana que exhibe la modificación post-traducciona

En la práctica del método en cuestión, puede ser deseable discernir el patrón de expresión de una proteína asociada con una vía de señalización bioquímica en diferentes tejidos corporales, en diferentes tipos de células y/o en diferentes estructuras subcelulares. Estos estudios se pueden realizar con el uso de anticuerpos específicos para el tejido, específicos para la célula o específicos para la estructura subcelular, capaces de unirse a marcadores de proteínas que se expresan preferentemente en determinados tejidos, tipos de célula o estructuras subcelulares.

Una expresión alterada de un gen asociado con una vía de señalización bioquímica también se puede determinar examinando un cambio en la actividad del producto génico con relación a una célula de control. El ensayo para un cambio inducido por agente en la actividad de una proteína asociada con una vía de señalización bioquímica dependerá de la actividad biológica y/o la vía de transducción de señales que está bajo investigación. Por ejemplo, en los casos en los que la proteína es una quinasa, un cambio en su capacidad para fosforilar el o los sustratos aguas abajo puede ser determinado por una diversidad de ensayos conocidos en la técnica. Ensayos representativos incluyen, pero no se limitan a inmunotransferencia e inmunoprecipitación con anticuerpos tales como anticuerpos anti-fosfotirosina que reconocen las proteínas fosforiladas. Además, la actividad quinasa se puede detectar por ensayos quimioluminiscentes de alto rendimiento tales como los ensayos AlphaScrea™ (disponible de Perkin Elmer) y eTag™ (Chan-Hui, et al. (2003) Clinical Immunology 111:162-174).

En los casos en los que la proteína asociada con una vía de señalización bioquímica es parte de una cascada de señalización que conduce a una fluctuación de la condición de pH intracelular, moléculas sensibles al pH tales como colorantes de pH fluorescentes se pueden utilizar como moléculas informadoras. En otro ejemplo en el que la proteína asociada a una vía de señalización bioquímica es un canal de iones, se pueden vigilar las fluctuaciones en el potencial de membrana y/o en la concentración intracelular de iones. Un cierto número de kits comerciales y dispositivos de alto rendimiento son particularmente adecuados para un rastreo rápido y robusto para moduladores de canales de iones. Instrumentos representativos incluyen FLIPRTM (Molecular Devices, Inc.) y VIPR (Aurora Biosciences). Estos instrumentos son capaces de detectar reacciones en pocillos de más de 1000 muestras de una microplaca de forma simultánea, y proporcionan una medición en tiempo real y datos funcionales en el espacio de un segundo o incluso un milisegundo.

En la práctica de cualquiera de los métodos descritos en esta memoria, un vector adecuado se puede introducir en una célula o un embrión a través de uno o más métodos conocidos en la técnica, incluyendo, sin limitación, microinyección, electroporación, sonoporación, biolística, transfección mediada por fosfato de calcio, transfección catiónica, transfección de liposomas, transfección de dendrímeros, transfección de choque térmico, transfección por nucleofección, magnetofección, lipofección, impalefacción, transfección óptica, captación de propiedad reforzada por el agente de ácidos nucleicos y el suministro a través de liposomas, inmunoliposomas, virosomas o viriones artificiales. En algunos métodos, el vector se introduce en un embrión mediante microinyección. El vector o los vectores se pueden microinyectar en el núcleo o en el citoplasma del embrión. En algunos métodos, el vector o los vectores se pueden introducir en una célula por nucleofección.

El polinucleótido diana de un complejo de CRISPR puede ser cualquier polinucleótido endógeno o exógeno a la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que reside en el núcleo de la célula eucariota. El polinucleótido diana puede ser una secuencia codificadora un producto génico (p. ej., una proteína) o una secuencia no codificadora (p. ej., un polinucleótido regulador o un ADN basura).

Ejemplos de polinucleótidos diana incluyen una secuencia asociada con una vía de señalización bioquímica, p. ej., un gen o polinucleótido asociado a la vía de señalización bioquímica. Ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado a la enfermedad. Un gen o polinucleótido "asociado a la enfermedad" se refiere a cualquier gen o polinucleótido que está proporcionando productos de transcripción o traducción en un nivel anormal o en una forma anormal en células derivadas de tejidos afectados por la enfermedad en comparación con los tejidos o células de un control no enfermo. Puede ser un gen que se expresa a un nivel anormalmente alto; puede ser un gen que se expresa a un nivel anormalmente bajo, en donde la expresión alterada se correlaciona con la aparición y/o el progreso de la enfermedad. Un gen asociado a la enfermedad también se refiere a un gen que posee

mutación o mutaciones o una variación genética que es directamente responsable o está en desequilibrio de enlace con un gen o genes que es responsable de la etiología de una enfermedad. Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar en un nivel normal o anormal.

5 El polinucleótido diana de un complejo de CRISPR puede ser cualquier polinucleótido endógeno o exógeno a la célula eucariota. Por ejemplo, el polinucleótido diana puede ser un polinucleótido que reside en el núcleo de la célula eucariota. El polinucleótido diana puede ser una secuencia codificadora de un producto génico (p. ej., una proteína) o una secuencia no codificadora (p. ej., un polinucleótido regulador o un ADN basura).

El polinucleótido diana de un complejo de CRISPR puede incluir un cierto número de genes y polinucleótidos asociados a la enfermedad, así como genes y polinucleótidos asociados a la vía de señalización bioquímica.

10 Ejemplos de polinucleótidos diana incluyen una secuencia asociada con una vía de señalización bioquímica, p. ej., un gen o polinucleótido asociado a la vía de señalización bioquímica. Ejemplos de polinucleótidos diana incluyen un gen o polinucleótido asociado a la enfermedad. Un gen o polinucleótido "asociado a la enfermedad" se refiere a cualquier gen o polinucleótido que está proporcionando productos de la transcripción o traducción en un nivel anormal o en una forma anormal a células derivadas de tejidos afectados por la enfermedad en comparación con los tejidos o células de un control no enfermo. Puede ser un gen que se expresa a un nivel anormalmente alto; puede ser un gen que se expresa a un nivel anormalmente bajo, en donde la expresión alterada se correlaciona con la aparición y/o progreso de la enfermedad. Un gen asociado a la enfermedad también se refiere a un gen que posee la o las mutaciones o variación genética que es directamente responsable o está en desequilibrio de enlace con un gen o genes que son responsables de la etiología de una enfermedad. Los productos transcritos o traducidos pueden ser conocidos o desconocidos, y pueden estar en un nivel normal o anormal.

Ejemplos de genes y polinucleótidos asociados a una enfermedad están disponibles en McKusick-Nathans Institute for Genetic Medicine, Universidad Johns Hopkins (Baltimore, Md.) y el Centro Nacional de Información sobre Biotecnología, Biblioteca Nacional de Medicina (Bethesda, Md.), disponible en la World Wide Web.

25 Ejemplos de genes y polinucleótidos asociados con la enfermedad se enumeran en las Tablas A y B. Información específica de la enfermedad está disponible en McKusick-Nathans Institute for Genetic Medicine, Universidad Johns Hopkins (Baltimore, Md.) y el Centro Nacional de Información sobre Biotecnología, Biblioteca Nacional de Medicina (Bethesda, Md.), disponible en la World Wide Web. Ejemplos de señalización de genes y polinucleótidos asociados a la vía de señalización bioquímica se enumeran en la Tabla C.

30 Mutaciones en estos genes y estas vías pueden resultar en la producción de proteínas inadecuadas o proteínas en cantidades inadecuadas que afectan a la función. Genes, proteínas y vías de este tipo pueden ser el polinucleótido diana de un complejo de CRISPR.

Tabla A

ENFERMEDAD/TRASTORNOS	GENE(S)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alfa; PPAR
	gamma; WT1 (Tumor de Wilms); miembros de la Familia de receptores FGF
	(5 miembros: 1, 2, 3, 4, 5); CDKN2a; APC; RB
	(retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR

ES 2 576 126 T3

ENFERMEDAD/TRASTORNOS	GENE(S)
	(Receptor de Andr6geno); TSG101; IGF; Receptor de IGF1; Igf1 (4
	variantes); Igf2 (3 variantes); Receptor de Igf1; Receptor de Igf2;
	Bax; Bcl2; familia de caspasas (9 miembros:
	1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Macular relacionada con la edad	Abcr; Ccl2; Cc2; cp (ceruloplasmina); Timp3; catepsinaD;
Degeneraci6n	Vldlr; Ccr2
Esquizofrenia	Neuregulina1 (Nrg1); Erb4 (receptor para Neuregulina);
	Complexina1 (Cplx1); Tph1 Triptofanohidroxilasa; Tph2
Tript6fano	hidroxilasa 2; Neurexina 1; GSK3; GSK3a;
	GSK3b
Trastornos	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA;
	DTNBP1; Dao (Dao1)
Repetici6n de Trinucle6tidos	HTT (Enfermedad de Huntington); SBMA/SMAX1/AR (Enfermedad de Kennedy);
Trastornos	FXN/X25 (Ataxia de Friedrich); ATX3 (Enfermedad de Machado-
	Joseph); ATXN1 and ATXN2 (ataxias espinocerebelares);
	DMPK (distrofia miot6nica); Atrofin-1 y Atnl
	(Enfermedad DRPLA); CBP (Creb-BP – inestabilidad global); VLDLR
	(Alzheimer); Atxn7; Atxn10
S6ndrome X Fr6gil	FMR2; FXR1; FXR2; mGLUR5
Relacionado con Secretasa	APH-1 (alfa y beta); Presenilina (Psen1); nicastrina
Trastornos	(Ncstn); PEN-2

ES 2 576 126 T3

ENFERMEDAD/TRASTORNOS	GENE(S)
Otros	Nos1; Parp1; Nat1; Nat2
Trastornos relacionados con Priones	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c)
Adicción a fármacos	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)
Autismo	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexina 1; Frágil X (FMR2 (AFF2); FXR1; FXR2; Mglur5)
Enfermedad de Alzheimer	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterina; PS1; SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1, Aquaporina 1); Uchl1; Uchl3; APP
Inflamación	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a(CTLA8); IL-17b; IL-17c; IL,-17d; IL-17f); IL-23; Cx3cr1; ptpn22; TNFa; OD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3c11
Enfermedad de Parkinson	x-Sinucleína; DJ-1; LRRK2; Parkin; PINK1

Tabla B:

Enfermedades y trastornos de la sangre y coagulación	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); síndrome de linfocito desnudo (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Trastornos de hemorragia (TBXA2R, P2RX1, P2X1); Factor H y tipo factor H 1 (HF 1, CFH, HUS); Factor V y factor VIII (MCFD2); Deficiencia de factor VII (F7); Deficiencia de factor X (F10); Deficiencia de factor XI (F11); Deficiencia de factor XII (F 12, HAF); Deficiencia de factor XIIIa (F13A1, F13A); Deficiencia de factor XIIIb (F13B); anemia de Fanconi (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCI, PHF9, FANCL, FANCM, KIAA1596); Trastornos de linfocitosis hemofagocítica (PRF1, HPLH2,
--	--

ES 2 576 126 T3

	UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemofilia A (F8, F8C, HEMA); Hemofilia B (F9, HEMB), Trastornos hemorrágicos (PI, ATT, F5); Deficiencias y trastornos de leucocitos (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); anemia de células de Sickle (HBB); Talasemia (HBA2, HBB, HBD, LCRB, HBA1).
Enfermedades y trastornos de disregulación celular y oncología	Linfoma no-Hodgkin B-cel1 (BCL7A, BCL7); Leucemia (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL 11, ARLTS 1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
Enfermedades y trastornos relacionados con la inflamación y el sistema inmune	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Síndrome linfoproliferativo autoinmune (TNFRSF6, APT1, FAS, CD95, ALPS1A); Inmunodeficiencia combinada, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), Susceptibilidad o infección por VIH (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCR5 (CCR5)); Inmunodeficiencias (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflamación (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), IL-23, Cx3crl, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Inmunodeficiencias combinadas graves (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Enfermedades y trastornos metabólicos, del hígado, riñon y proteínas	Neuropatía amiloide (TTR, PALB); Amiloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Fibrosis quística (CFTR, ABCC7, CF, MRP7); Enfermedades de almacenamiento de glucógeno (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Adenoma hepático, 142330 (TCF1, HNF1A, MODY3), Insuficiencia hepática, brote temprano y trastorno neurológico (SCOD1, SCO1), Deficiencia de lipasa hepática (LIPC), Hepatoblastoma, cáncer y carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Enfermedad de riñón quístico medular (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); fenilcetonuria (PAH, PKU1, QDPR, DHPR, PTS); Riñón poliquístico y enfermedad hepática (FCYT, PKHD 1, ARPKD, PKD 1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Enfermedades y trastornos Musculares/ del Esqueleto	Distrofia muscular de Becker (DMD, BMD, MYF6), Distrofia muscular de Duchenne (DMD, BMD); Distrofia muscular de Emery-Dreifuss (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Distrofia muscular facioescapulohumeral (FSHMD1A, FSHD1A); Distrofia muscular (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G,

	CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT 1, CAV3, LGMD1C, SEPN 1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Atrofia muscular (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
Enfermedades y trastornos neurológicos y neuronales	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); enfermedad de Alzheimer (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAUI, URK, ACE, DCP 1, ACE 1, MPO, PACIP 1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autismo (Mecp2, BZRAP1, MDGA2, Sema5A, Neurexina 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Síndrome X frágil (FMR2, FXR1, FXR2, mGLUR55); Enfermedad y trastornos tipo enfermedad de Huntington (HD, IT 15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Enfermedad de Parkinson (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Síndrome de Rett (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Sinucleína, DJ-1); Esquizofrenia (Neuregulina1 (Nrg1), Erb4 (receptor para Neuregulina), Complexina1 (Cplx1), Tph1 Triptófano hidroxilasa, Tph2, Triptófano hidroxilasa 2, Neurexina 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Trastornos relacionados con secretasa (APH-1 (alfa y beta), Presenilina (Psen1), nicastrina, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trastornos de la repetición de trinucleótidos (HTT (Enfermedad de Huntington), SBMA/SMAX1/AR (Enfermedad de Kennedy), FXN/X25 (Ataxia de Friedrich), ATX3 (Enfermedad de Machado-Joseph), ATXN1 y ATXN2 (ataxias espinocerebelares), DMPK (distrofia miotónica), Atrofia-1 y Atn1 (DRPLA Dx), CBP (Creb-BP – inestabilidad global), VLDLR (Alzheimer), Atxn7, Atxn10).
Enfermedades y trastornos oculares	Degeneración macular relacionada con la edad (Abcr, Cc12, Cc2, cp (ceruloplasmina), Timp3, catepsinaD, Vldlr, Ccr2); Cataratas (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Opacidad y distrofia de la córnea (APOA1, TGFB1, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Córnea plana congénita (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Amaurosis congénita del hígado (CRB1, RP 12, CRX, CORD2, CRD, RPGRIP 1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Distrofia macular (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

Tabla C:

FUNCIÓN CELULAR	GENES
Señalización PI3K/AKT	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; FPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2;
	ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3;
	IPRCAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7;
	YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;
	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;
	HSP90AA1; RPS6KB1
Señalización ERK/MAPK	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2;
	EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6;
	MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1;
	PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN;
	EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ;
	PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1;
	PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1;
	CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Señalización del Receptor de Glucocorticoides	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13;
	RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1;
	PIK3C2A; CDKN1C; TRAF2; SERPINA1; NCOA3;
	MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP;
	CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;
	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1
Señalización de Guía Axonal	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12;
	IGF 1; RAC 1; RAP1A; EIF4E; PRKCZ; NRP 1; NTRK2;
	ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2;
	CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11;
	PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;
	FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1;
	GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3;
	CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
Señalización del Receptor Efrina	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2;
	PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2;
	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;
	AKT3; SGK
Señalización del Citoesqueleto de Actina	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
	IPRCAA2; FIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8;
	PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; 1TGA1; KRAS; RHOA; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7;
	PPP1CC; PXN; VIL2; RAF 1; GSN; DYRK1A; ITGB 1;
	MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3;
	ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
Señalización de la Enfermedad de Huntington	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1;
	GNB2L1; BCL2L1; CAPN1; MAPK3; CLASP8; HDAC2;
	HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A;
	HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1;
	PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Señalización de la Apoptosis	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;
	BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF;
	RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2;
	CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2;
	BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK;
	CASP3; BIRC3; CARP1
Señalización del Receptor de Células B	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11;
	AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A;
	IMAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1;
	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Señalización de la Extravasación de Leucocitos	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	P1K3C3; MAPK; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MUMP9
Señalización de Integrina	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	IRLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;
	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
Señalización de la Respuesta de Fase Aguda	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
	AKT2; IKBKB; PIK3GA; FOS; NFKB2; MAP3K14;
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;
	AKT3; IL 1R1; IL6
Señalización de PTEN	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;
	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CLASP3; RPS6KB1
Señalización de p53	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIQC3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP 1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
Señalización del Receptor de Aril Hidrocarbano	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
Señalización del Metabolismo Xenobiótico	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB 1; KEAP 1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP94AA1
Señalización de SAPK/JNK	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;
	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPS1;
	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;
	CRKL; BRAF; SGK
Señalización de PPAR/RXR	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	TGFB1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	ADIPOQ
Señalización de NF-KB	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAY6;
	TBKL1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2;
	KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF;
	INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL 1R1
Señalización de Neuregulina	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1;
	MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI;
	CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS;
	PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2;
	ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL;
	AKT3; PRKCA; HSP90AA1; RPS6KB1
Señalización de Wnt y Beta catenina	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
	AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A;
	WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK;
	LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFB1; CCND1;
	GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B;
	AKT3; SOX2
Señalización del Receptor de Insulina	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1;
	PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3;
	MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1;
	SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN;
	AP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1;
	GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK;
	RPS6KB1
Señalización de IL-6	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3;
	MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1;
	MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG;
	RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3;
	MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
Colestasis Hepática	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA;
	RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8;
	PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABC1;
	TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8;
	CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4;
	JUN; IL1R1; PRKCA; IL6
Señalización de IGF-1	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2;
	PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;
	IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;
	FOXO1; SRF; CTGF; RPS6KB1
Respuesta de Estrés Oxidativo mediada por NRF2	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
	NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP;
	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;
	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Fibrosis Hepática/Activación de Células Estelares	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFB1; SMAD4; VEGFA; BAX;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
Señalización de PPAR	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Señalización de Fc Epsilon RI	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
Señalización del Receptor Acoplado a Proteína G	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; REB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;
	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
Metabolismo de Inositol Fosfato	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
	MAPK8; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
MAPK8	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	PIM1; PIK3G2A; DYRK1A; MAP2K2; PIP5KIA; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
Señalización de PDGF	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB;
	PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC;
	PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2;
	PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
Señalización de VEGF	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;
	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3;
	BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN;
	RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN;
	VEGFA; AKT3; FOXO1; PRKCA
Señalización de Células Asesinas Naturales	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11;
	KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB;
	PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6;
	P1K3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1;
	PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Ciclo Celular: G1/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
Regulación del punto de verificación	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	IGSK3B; RBL1; HDAC6

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
Señalización del Receptor de Células T	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS;
	NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN;
	MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10;
	JUN; VAV3
Señalización del Receptor de Muerte	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD;
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPS1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3
Señalización de FGF	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
Señalización de GM-CSF	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;
	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
Señalización de la Esclerosis Lateral Amiotrófica	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGF; BIRC2; BAX; AKT3; CASP3; BIRC3
Señalización de JAK/Stat	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1
Metabolismo de Nicotinato y Nicotinamida	PRKCE; IRAK; PRKAA2; EIF2AK2; GRK6; MAPK1;
	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAF2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Señalización de Quimioquinas	CXCR4; ROCK2; MAPS1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
Señalización de IL-2	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;
	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Depresión Sináptica a Largo Plazo	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Señalización del Receptor de Estrógenos	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Vía de la Ubiquitinación de Proteínas	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
Señalización de IL-10	TRAF6; CCR1; ELK1; IKKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;
	IKBK; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6
Activación de VDR/RXR	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;
	LRP5; CEBPB; FOXO1; PRKCA
Señalización de TGF-beta	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGEBR1; SMAD4; JUN; SMAD5
Señalización del Receptor de tipo Toll	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;
	NFKB1; TLR2; JUN
Señalización de p38 MAPK	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K.7; TGFBR1; MYC; ATF4; IL 1R1;
	SRF; STAT1
Señalización de Neurotrofina/TRK	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A;
	RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	CDC42; JUN; ATF4
Activación de FXR/RXR	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8;
	APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Potenciación Sináptica a Largo Plazo	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
Señalización de Calcio	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;
	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
Señalización de EGF	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Señalización de Hipoxia en el	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;
Sistema Cardiovascular	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
Inhibición Mediada por LPS/IL-1	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1;
de la Función RXR	MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2;
	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
Activación de LXR/RXR	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;
	SREBF1; IL1R1; CCL2; IL6; MMP9
Procesamiento Amiloide	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
Señalización de IL-4	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
Ciclo Celular: G2/M Regulación del Punto	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
de Verificación de Daño a ADN	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
	PRKDC; ATM; SFN; CDKN2A
Señalización de Óxido Nítrico en el	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
Sistema Cardiovascular	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
Metabolismo de Purina	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD 1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
Señalización mediada por cAMP	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Disfunción Mitocondrial	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Señalización de Notch	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4
Vía del Estrés del Retículo Endoplásmico	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
	EIF2AK3; CASP3
Metabolismo de Pirimidina	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;
	NT5E; POLD1; NME1
Señalización de Parkinson	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7;
	PARK2; CASP3
Señalización Cardíaca y Beta Adrenérgica	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC;
	PPP2R5C
Glicolisis/Gluconeogénesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Señalización de Interferón	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Señalización Sonic Hedgehog	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
Metabolismo de Glicerofosfolípidos	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Degradación de Fosfolípidos	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Metabolismo de triptófano	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Degradación de Lisina	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Vía de la Reparación de la Escisión de Nucleótidos	ERCC5; ERCC4; XPA; XPC; ERCC1
Metabolismo de Almidón y Sacarosa	UCHL1; HK2; GCK; GPI; HK1
Metabolismo de Aminoazúcares	NQO1; HK2; GCK; HK1
Metabolismo del Ácido Araquidónico	PRDX6; GRN; YWHAZ; CYP1B1
Señalización del Ritmo Circadiano	CSNK1E; CREB1; ATF4; NR1D1
Sistema de Coagulación	BDKRB1; F2R; SERPINE1; F3
Señalización del Receptor de Dopamina	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Metabolismo de Glutation	IDH2; GSTP1; ANPEP; IDH1
Metabolismo de Glicerolípidos	ALDH1A1; GPAM; SPHK1; SPHK2
Metabolismo del Ácido Linoleico	PRDX6; GRN; YWHAZ; CYP1B1
Metabolismo de Metionina	DNMT1; DNMT3B; AHCY; DNMT3A
Metabolismo de Piruvato	GLO1; ALDH1A1; PKM2; LDHA

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
Metabolismo de Arginina y Prolina	ALDH1A1; NOS3; NOS2A
Señalización de Eicosanoides	PRDX6; GRN; YWHAZ
Metabolismo de Fructosa y Manosa	HK2; GCK; HK1
Metabolismo de Galactosa	HK2; GCK; HK1
Biosíntesis de Estilbeno, Coumarina y	PRDX6; PRDX1; TYR
Lignina	
Vía de la Presentación de Antígenos	CALR; B2M
Biosíntesis de Esteroides	NQO1; DHCR7
Metabolismo de Butanoato	ALDH1A1; NLGN1
Ciclo del Citrato	IDH2; IDH1
Metabolismo de Ácidos Grasos	ALDH1A1; CYP1B1
Metabolismo de Glicerofosfolípidos	PRDX6; CHKA
Metabolismo de Histidina	PRMT5; ALDH1A1
Metabolismo de Inositol	ERO1L; APEX1
Metabolismo de Xenobióticos	GSTP1; CYP1B1
por el Citocromo p450	
Metabolismo del Metano	PRDX6; PRDX1
Metabolismo de Fenilalanina	PRDX6; PRDX1

ES 2 576 126 T3

FUNCIÓN CELULAR	GENES
Metabolismo de Propanoato	LDH1A1; LDHA
Metabolismo de Selenoaminoácido	PRMT5; AHCY
Metabolismo de Esfingolípidos	SPHK1; SPHK2
Metabolismo de Aminofosfonato	PRMT5
Metabolismo de Andrógenos y Estrógenos	PRMT5
Metabolismo de Ascorbato y Aldarato	ALDH1A1
Biosíntesis de Ácidos Biliares	ALDH1A1
Metabolismo de Cisteína	LDHA
Biosíntesis de Ácidos Grasos	FASN
Señalización del Receptor de Glutamato	GNB2L1
Respuesta al Estrés Oxidativo medida por NRF2	PRDX1
Vía del Fosfato de Pentosa	GPI
Interconversiones de Pentosa y Glucuronato	UCHL1
Metabolism de Retinol	ALDH1A1

FUNCIÓN CELULAR	GENES
Metabolismo de Riboflavina	TYR
Metabolismo de Tirosina	PRMT5, TYR
Biosíntesis de Ubiquinona	PRMT5
Degradación de Valina, Leucina e Isoleucina	ALDH1A1
Metabolismo de Glicina, Serina y Treonina	CHKA
Degradación de Lisina	ALDH1A1
Dolor/Sabor	TRPM5; TRPA1
Dolor	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2; Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a
Función Mitocondrial	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Neurología del desarrollo	BMP-4; Cordina (Chrd); Nogina (Nog); WNT (Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenina; Dkk-1; Proteínas relacionadas con Frizzled; Otx-2; Gbx2; FGF-8; Reelina; Dab1; unc-86 (Pou4f1 o Brn3a); Numb; Reln

5 Realizaciones de la invención también se refieren a métodos y composiciones relacionados con la inactivación de genes, la amplificación de genes y la reparación de mutaciones particulares asociadas con la inestabilidad en la repetición de ADN y trastornos neurológicos (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Segunda Edición, Academic Press 13 de oct. de 2011 - Medical). Se ha encontrado que aspectos específicos de secuencias repetidas en tándem son los responsables de más de una veintena de enfermedades humanas (New insights into repeat instability: role of RNA•DNA hybrids. Mclvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep.-Oct.;7(5):551-8). El sistema de CRISPR-Cas puede ser aprovechado para corregir estos defectos de la inestabilidad genómica.

Un aspecto adicional de la invención se refiere a la utilización del sistema de CRISPR-Cas para la corrección de defectos en los genes EMP2A y EMP2B que han sido identificados que están asociados con la enfermedad de Lafora. La enfermedad de Lafora es una afección autosómica recesiva que se caracteriza por la epilepsia mioclónica progresiva que puede comenzar como convulsiones epilépticas en la adolescencia. Unos pocos casos de la enfermedad pueden ser provocados por mutaciones en genes aún por identificar. La enfermedad provoca convulsiones, espasmos musculares, dificultad para caminar, demencia y, finalmente, la muerte. Actualmente no existe una terapia que haya demostrado ser eficaz contra el progreso de la enfermedad. Otras anomalías genéticas asociadas con la epilepsia también pueden ser fijadas como objetivo por el sistema de CRISPR-Cas y la genética subyacente se describe con más detalle en Genetics of Epilepsy and Genetic Epilepsies, editado por Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009).

Todavía en otro aspecto de la invención, el sistema de CRISPR-Cas puede utilizarse para corregir defectos oculares que surgen de varias mutaciones genéticas descritas adicionalmente en Genetic Diseases of the Eye, Segunda Edición, editado por Elias I. Traboulsi, Oxford University Press, 2012.

Varios aspectos adicionales de la invención se refieren a la corrección de los defectos asociados con una amplia gama de enfermedades genéticas que se describe con más detalle en la página web de National Institutes of Health, bajo la subsección Genetic Disorders (sitio web en health.nih.gov/topic/GeneticDisorders). Las enfermedades genéticas del cerebro pueden incluir, pero no se limitan a adrenoleucodistrofia, agenesia del cuerpo calloso, síndrome de Aicardi, enfermedad de Alper, enfermedad de Alzheimer, síndrome de Barth, enfermedad de Batten, CADASIL, Degeneración cerebelosa, enfermedad de Fabry, enfermedad de Gerstmann-Straussier-Scheinker, enfermedad de Huntington y otros trastornos de la repetición de tripletes, enfermedad de Leigh, síndrome de Lesch-Nyhan, enfermedad de Menkes, miopatías mitocondriales y colpocefalia NINDS. Estas enfermedades se describen con más detalle en la página web de National Institutes of Health, bajo la subsección Genetic Brain Disorders.

En algunas realizaciones, la afección puede ser la neoplasia. En algunas realizaciones, en los casos en los que la afección es la neoplasia, los genes a ser fijados como objetivo son cualquiera de los enumerados en la Tabla A (en este caso PTEN, etcétera). En algunas realizaciones, la afección puede ser la degeneración macular relacionada con la edad. En algunas realizaciones, la afección puede ser un trastorno esquizofrénico. En algunas realizaciones, la afección puede ser un trastorno de repeticiones de trinucleótidos. En algunas realizaciones, la afección puede ser el síndrome X frágil. En algunas realizaciones, la afección puede ser un trastorno relacionado con secretasa. En algunas realizaciones, la afección puede ser un trastorno relacionado con Priones. En algunas realizaciones, la afección puede ser ALS. En algunas realizaciones, la afección puede ser una adicción a los fármacos. En algunas realizaciones, la afección puede ser el autismo. En algunas realizaciones, la afección puede ser la enfermedad de Alzheimer. En algunas realizaciones, la afección puede ser la inflamación. En algunas realizaciones, la afección puede ser la enfermedad de Parkinson.

Ejemplos de proteínas asociadas con la enfermedad de Parkinson incluyen, pero no se limitan a α -sinucleína, DJ-1, LRRK2, Pink1, Parkin, UCHL1, Sinfilina-1 y NURR1.

Ejemplos de proteínas relacionadas con la adicción pueden incluir ABAT, por ejemplo.

Ejemplos de proteínas relacionadas con la inflamación pueden incluir la proteína quimioatrayente de monocitos 1 (MCP1) codificada por el gen Ccr2, el receptor de quimioquina C-C tipo 5 (CCR5) codificado por el gen Ccr5, el receptor IIB de IgG (FCGR2b, también denominado CD32) codificado por el gen Fcgr2b o la proteína Fc epsilon R1g (FCER1g) codificada por el gen Fcer1g, por ejemplo.

Ejemplos de proteínas asociadas a enfermedades cardiovasculares pueden incluir IL1B (interleuquina 1, beta), XDH (xantina deshidrogenasa), TP53 (proteína tumoral p53), PTGIS (prostaglandina 12 (prostaciclina) sintasa), MB (mioglobina), IL4 (interleuquina 4), ANGPT1 (angiopoyetina 1), ABCG8 (casete de unión a ATP, subfamilia G (WHITE) miembro, 8), o CTSK (catepsina K), por ejemplo.

Ejemplos de proteínas asociadas a la enfermedad de Alzheimer pueden incluir la proteína receptor de lipoproteínas de muy baja densidad (VLDLR) codificada por el gen VLDLR, la enzima activante del modificador de tipo ubiquitina 1 (UBA1) codificada por el gen UBA1 o la proteína de la subunidad catalítica E1 de la enzima activante de NEDD8 (UBE1C) codificada por el gen UBA3, por ejemplo.

- 5 Ejemplos de proteínas asociadas al trastorno del espectro del autismo pueden incluir la proteína 1 asociada al receptor de benzodiazepina (periférico) (BZRAP1) codificada por el gen BZRAP1, la proteína del miembro 2 de la familia AF4/FMR2 (AFF2) codificada por el gen AFF2 (también denominada MFR2), la proteína homóloga 1 autosómica del retardo mental X frágil (FXR1) codificada por el gen FXR1 o la proteína homóloga 2 autosómica del retardo mental X frágil (FXR2) codificada por el gen FXR2, por ejemplo.
- Ejemplos de proteínas asociadas con la degeneración macular puede incluir la proteína de la casete de unión a ATP, sub-familia A (ABC1) miembro 4 (ABCA4) codificada por el gen ABCR, la proteína apolipoproteína E (APOE) codificada por el gen APOE o la proteína quimioquina (motivo C-C) Ligando 2 (CCL2) codificada por el gen CCL2, por ejemplo.
- 10 Ejemplos de proteínas asociadas con la esquizofrenia pueden incluir NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, y combinaciones de las mismas.
- Ejemplos de proteínas implicadas en la supresión de tumores pueden incluir ATM (ataxia telangiectasia mutada), ATR (ataxia telangiectasia y relacionada con Rad3), EGFR (receptor del factor de crecimiento epidérmico), ERBB2 (homólogo 2 del oncogén de la leucemia eritroblástica v-erb-b2), ERBB3 (homólogo 3 del oncogén de la leucemia eritroblástica v-erb-b2), ERBB4 (homólogo 4 del oncogén de la leucemia eritroblástica v-erb-b2), Notch 1, Notch2, Notch 3 o Notch 4, por ejemplo.
- 15 Ejemplos de proteínas asociadas con un trastorno secretasa pueden incluir PSENEN (homólogo de potenciador de presenilina 2 (C. elegans)), CTSB (catepsina B), PSEN1 (presenilina 1), APP (precursor de la proteína amiloide beta (A4)), APH1B (homólogo B defectuoso de la faringe anterior 1 (C. elegans)), PSEN2 (presenilina 2 (enfermedad de Alzheimer 4)) o BACE1 (enzima de escisión de APP del sitio beta 1), por ejemplo.
- 20 Ejemplos de proteínas asociadas con la esclerosis lateral amiotrófica pueden incluir SOD1 (superóxido dismutasa 1), ALS2 (esclerosis lateral amiotrófica 2), FUS (fusionado en sarcoma), TARDBP (proteína de unión a ADN TAR), VAGFA (factor de crecimiento vascular endotelial A), VAGFB (factor de crecimiento vascular endotelial B) y VAGFC (factor de crecimiento vascular endotelial C), y cualquier combinación de las mismas.
- 25 Ejemplos de proteínas asociadas con las enfermedades priónicas pueden incluir SOD1 (superóxido dismutasa 1), ALS2 (esclerosis lateral amiotrófica 2), FUS (fusionado en sarcoma), TARDBP (proteína de unión a ADN TAR), VAGFA (factor de crecimiento vascular endotelial A), VAGFB (factor de crecimiento vascular endotelial B) y VAGFC (factor de crecimiento vascular endotelial C), y cualquier combinación de las mismas.
- 30 Ejemplos de proteínas relacionadas con enfermedades neurodegenerativas en los trastornos por priones pueden incluir A2M (alfa-2-macroglobulina), AATF (factor de transcripción antagonizante de la apoptosis), ACPP (fosfatasa ácida prostática), ACTA2 (actina alfa 2 aorta músculo liso), ADAM22 (dominio de metalopeptidasa ADAM), ADORA3 (receptor de adenosina A3) o ADRA1D (receptor adrenérgico alfa-1D para el adrenorreceptor alfa-1D), por ejemplo.
- Ejemplos de proteínas asociadas con inmunodeficiencia puede incluir A2M [alfa-2- macroglobulina]; AANAT [arilalquilamina N-acetiltransferasa]; ABCA1 [casete de unión a ATP, subfamilia A (ABC1), miembro 1]; ABCA2 [casete de unión a ATP, subfamilia A (ABC1), miembro 2]; o ABCA3 [casete de unión a ATP, subfamilia A (ABC1), miembro 3]; por ejemplo.
- 35 Ejemplos de proteínas asociadas con Trastornos por Repetición de Trinucleótidos incluyen AR (receptor de andrógenos), FMR1 (retardo mental X frágil 1), HTT (huntingtina) o DMPK (distrofia miotónica-proteína quinasa), FXN (frataxina), ATXN2 (ataxina 2), por ejemplo.
- 40 Ejemplos de proteínas asociadas con los Trastornos de Neurotransmisión incluyen SST (somatostatina), NOS1 (óxido nítrico sintasa 1 (neuronal)), ADRA2A (receptor adrenérgico alfa -2A), ADRA2C (receptor adrenérgico alfa-2C), TACR1 (receptor de taquiquinina 1) o HTR2C (receptor de 5-hidroxitriptamina (serotonina) 2C), por ejemplo.
- Ejemplos de secuencias asociadas al neurodesarrollo incluyen A2BP1 [proteína de unión 1 ataxina 2], AADAT [aminoadipato aminotransferasa], AANAT [arilalquilamina N-acetiltransferasa], ABAT [4-aminobutirato aminotransferasa], ABCA1 [casete de unión a ATP, subfamilia A (ABC1), miembro 1] o ABCA13 [casete de unión a ATP, subfamilia A (ABC1), miembro 13], por ejemplo.
- 45

Ejemplos adicionales de afecciones preferidas tratables con el presente sistema se pueden seleccionar de: Síndrome de Aicardi-Goutieres; Enfermedad de Alexander; Síndrome de Allan-Herndon-Dudley; Trastornos Relacionados con POLG; Alfa-Manosidosis (Tipos I y III); Síndrome de Alström; Síndrome de Angelman; Ataxia-Telangiectasia; Lipofuscinosis Ceroide Neuronal; Beta-Talasemia; Atrofia Óptica Bilateral y Atrofia Óptica (Infantil) Tipo 1; Retinoblastoma (bilateral); Enfermedad de Canavan; Síndrome Cerebro-oculofacioesquelético 1 [COFS1]; Xantomatosis Cerebrotendinosa; Síndrome de Cornelia de Lange; Trastornos Relacionados con MAPT; Enfermedades Priónicas Genéticas; Síndrome de Dravet; Enfermedad de Alzheimer Familiar de Aparición Temprana; Ataxia de Friedreich [FRDA]; Síndrome de Frys; Fucosidosis; Distrofia Muscular Congénita de Fukuyama; Galactosialidosis; Enfermedad de Gaucher; Acidemias Orgánicas; Linfocitosis Hemofagocítica; Síndrome de Progeria de Hutchinson-Gilford; Mucopolisacaridosis II; Enfermedad de Almacenamiento de Ácido Siálico Libre Infantil; Neurodegeneración Asociada a PLA2G6; Síndrome de Jervell y Lange-Nielsen; Epidermolisis Bullosa de la Unión; Enfermedad de Huntington; Enfermedad de Krabbe (Infantil); Síndrome de Leigh y NARP asociado a ADN mitocondrial; Síndrome de Lesch-Nyhan; Lisencefalia Asociada a LIS1; Síndrome de Lowe; Enfermedad de la Orina de Jarabe de Arce; Síndrome de Duplicación de MECP2; Trastornos de Transporte de Cobre Relacionados con ATP7A; Distrofia Muscular Relacionada con LAMA2; Deficiencia de Arilsulfatasa A; Mucopolisacaridosis Tipos I, II o III; Trastornos de la Biogénesis de Peroxisomas, Espectro del Síndrome de Zellweger; Neurodegeneración con Trastornos de Acumulación de Hierro en el Cerebro; Deficiencia de Esfingomielinasa Ácida; Enfermedad de Niemann-Pick Tipo C; Encefalopatía de Glicina; Trastornos Relacionados con ARX; Trastornos del Ciclo de la Urea; Osteogénesis Imperfecta Relacionada con COL1A1/2; Síndromes Mitocondriales de Deleción de ADN; Trastornos relacionados con PLP1; Síndrome de Perry; Síndrome de Phelan-McDermid; Enfermedad del Almacenamiento de Glucógeno Tipo II (Enfermedad de Pompe) (Infantil); Trastornos Relacionados con MAPT; Trastornos Relacionados con MECP2; Condrodisplasia Punctata Rizomélica Tipo 1; Síndrome de Roberts; Enfermedad de Sandhoff; Enfermedad de Schindler - Tipo 1; Deficiencia de Adenosina Desaminasa; Síndrome de Smith-Lemli-Opitz; Atrofia Muscular Espinal; Ataxia Espinocerebelosa de Brote Infantil; Deficiencia de Hexosaminidasa A; Displasia Tanatófórica Tipo 1; Trastornos Relacionados con Colágeno Tipo VI; Síndrome de Usher Tipo I; Distrofia Muscular Congénita; Síndrome de Wolf-Hirschhorn; Deficiencia Lisosomal de Lipasa Ácida; y Xenoderma Pigmentosum.

Como resultará evidente, se prevé que el presente sistema se pueda utilizar para fijar como objetivo a cualquier secuencia de polinucleótidos de interés. Algunos ejemplos de afecciones o enfermedades que podrían ser tratadas útilmente utilizando el presente sistema se incluyen en las Tablas anteriores y también se proporcionan allí ejemplos de genes asociados actualmente con esas afecciones. Sin embargo, los genes ejemplificados no son exhaustivos.

EJEMPLOS

Los siguientes ejemplos se dan con el propósito de ilustrar diversas realizaciones de la invención.

Ejemplo 1: Mejora del sistema Cas9 para la aplicación in vivo

Las solicitantes llevarán a cabo una búsqueda metagenómica para una Cas9 con pequeño peso molecular. La mayoría de los homólogos de Cas9 son bastante grandes. Muchos homólogos de Cas9 conocidos son grandes y contienen más de 1300 amino ácidos. Por ejemplo, la SpCas9 es de alrededor de 1368aa de larga, que es demasiado grande para ser empaquetada fácilmente en vectores virales para el suministro. En la Figura 6 se muestra un gráfico que representa la distribución de longitud de homólogos de Cas9: El gráfico se genera a partir de secuencias depositadas en GenBank. Algunas de las secuencias pueden haber sido anotadas erróneamente y, por lo tanto, la frecuencia exacta para cada uno de los tramos no tiene que ser necesariamente fiable. No obstante, proporciona un esbozo de la distribución de proteínas Cas9 y sugiere que existen homólogos de Cas9 más cortos.

A través de análisis computacional, la solicitante encontró que en la cepa bacteriana *Campylobacter* existen dos proteínas Cas9 con menos de 1000 aminoácidos. La secuencia para una Cas9 de *Campylobacter jejuni* se presenta a continuación. En esta longitud, CjCas9 puede empaquetarse fácilmente en AAV, lentivirus, adenovirus y otros vectores virales para el suministro sólido en células primarias y en modelos animales *in vivo*. En una realización preferida de la invención, se utiliza la proteína Cas9 de *S. aureus*.

Cas9 de *Campylobacter jejuni* (CjCas9)

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARR
 KARLNHLKHLIANEFKLNEDYQSFDESLAKAYKGLISPYELRFRALNELLKQDFAR
 VILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYSVGEYLYKEYFQKFKENSK
 EFTNVRNKKESYERCIASFLKDELKLFKKQREFGFSKFFEEVLSVAFYKRALKDFS
 HLVGNCFFTDKRAPKNSPLAFMFVALTRIINLLNKNTEGILYTKDDLNALLNEVLK
 NGTLTYKQTKKLLGLSDDYEFKGEKGTIFYEFKFKYKEFIKALGEHNLSQDDLNEIAKDIT
 LIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNLSFKALKLVTPMLEGKKYDEACNE
 LNLKVAINEDKKDFLPAFNETYYKDEVTPVVLRAIKEYRKVLNALLKKYGKVKINIE
 LAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGKINSKNILKRLRFKEQKEFCA
 YSGEKIKISDLQDEKMLEIDHIYPYSRFDSDSYMNVLVFTKQNQEKLNTQPFQAFGNS
 AKWQKIEVLAKNLPKQKRILDKNYKDKKEQKNFKDRNLNDTRYIARLVNLYTKDYL
 DFLPLSDDENTKLNQKGSKVHVEAKSGLTALSALRHTWGFSKDRNNHLHHAIDAVI
 IAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLKIDEIF
 VSKPERKKPSGALHEETFRKEEFYQSYGGKEGVKALELGKIRKVNKIVKNGDMFR
 VDIFKHKKTNKIFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCSLYK
 DSLILIQTKDMQEPEFVYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAS
 IGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK.

El elemento tracrRNA putativo para esta CjCas9 es:

TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTGCGGG
 ACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTTAAAATT

5 La secuencia de Repetición Directa es:

GTTTTAGTCCCTTTTAAATTTCTTTATGGTAAAAT

Un ejemplo de un ARN guía quimérico para CjCas9 es:

NNNNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAAAUAAAGAGUU
 UGCGGGACUCUGCGGGGUACAAUCCCCUAAAACCGCUUUU

Ejemplo 2: Diversidad de Cas9 y ARNs quiméricos

10 El sistema de CRISPR-Cas de tipo salvaje es un mecanismo inmune adaptativo contra la invasión de ADN exógeno
 empleado por diversas especies a través de bacterias y arqueas. El sistema de CRISPR-Cas de tipo II consiste en
 un conjunto de genes que codifican proteínas responsables de la "adquisición" de ADN extraño en el locus de
 CRISPR, así como un conjunto de genes que codifican la "ejecución" del mecanismo de escisión de ADN; éstos
 incluyen la ADN nucleasa (Cas9), un cr-RNA transactivador no codificador (tracrRNA) y una serie de espaciadores
 15 derivados de ADN extraños flanqueados por repeticiones directas (crRNAs). Tras la maduración por Cas9, el dúplex
 tracrRNA y crRNA guía la nucleasa Cas9 a una secuencia de ADN diana especificada por las secuencias guía del
 espaciador, y media en las roturas de la doble cadena en el ADN cerca de un corto motivo de secuencia en el ADN
 diana que se requiere para la escisión y que es específico para cada uno de los sistemas CRISPR-Cas. Los
 sistemas de CRISPR-Cas de tipo II se encuentran en todo el reino bacteriano y son muy diversos en la secuencia y
 20 el tamaño de proteínas Cas9, y la secuencia de repetición directa de tracrRNA y crRNA, la organización del genoma
 de estos elementos y el requisito del motivo para la escisión diana. Una especie puede tener múltiples sistemas de
 CRISPR-Cas distintos.

La solicitante evaluó 207 Cas9s putativos de especies bacterianas identificadas sobre la base de homología de
 secuencia con Cas9s y estructuras ortólogas conocidas a subdominios conocidos, incluyendo el dominio HNH de
 25 endonucleasa y los dominios RuvC endonucleasa [información de Eugene Koonin y Kira Makarova]. El análisis
 filogenético basado en la conservación de secuencias de proteínas de este conjunto reveló cinco familias de Cas9s,
 incluyendo tres grupos de Cas9s grandes (~ 1400 aminoácidos) y dos Cas9s pequeños (~ 1100 aminoácidos)
 (Figuras 4 y 5A-F).

En algunas realizaciones, las secuencias de emparejamiento tracr o las repeticiones directas se descargan ya sea desde la base de datos CRISPRs o se identifican in silico mediante el rastreo de motivos repetitivos que 1. se encuentran en una ventana de 2 kb de secuencia genómica flanqueante del locus de CRISPR de tipo II, 2. abarcan de 20 a 50 pb, y 3. están separados por 20 a 50 pb. En algunas realizaciones, se pueden utilizar 2 de estos criterios, por ejemplo, 1 y 2, 2 y 3, ó 1 y 3. En algunas realizaciones, se pueden utilizar los 3 criterios. En algunas realizaciones tracrRNA candidatas son predichas posteriormente predichas por 1. homología de la secuencia con repeticiones directas (rastreo de motivos en Geneious con hasta apareamientos erróneos de hasta 18 pb), 2. presencia de un terminador transcripcional independiente de Rho predicho en la dirección de la transcripción y 3. la estructura secundaria de horquilla estable entre tracrRNA y la repetición directa. En algunas realizaciones, se pueden utilizar 2 de estos criterios, por ejemplo 1 y 2, 2 y 3, ó 1 y 3. En algunas realizaciones, se pueden utilizar los 3 criterios. En algunas realizaciones, diseños de ARNs quiméricos guía sintéticos (sgRNAs) incorporan al menos 12 pb de la estructura dúplex entre la repetición directa y tracrRNA.

La lista de las secuencias de ortólogos Cas9 humanas, optimizadas en codones, a emparejarse con los ARNs quiméricos proporcionados en las Figuras 8 A-J se proporciona en las Figuras 9 A-O. La solicitante también ha demostrado que los ortólogos Cas9 pueden escindir sus dianas en ensayos de escisión in vitro (Figura 16). Los loci de CRISPR en algunas de estas familias se representa en la Figura 11. Las secuencias de ARN guía correspondientes se muestran en la Figura 12. La solicitante analizó sistemáticamente la secuencia de ADN genómico dentro de ~ 2 kb de las proteínas Cas9 utilizando el código de análisis computacional habitual e identificó repeticiones directas que oscilaban entre 35 pb y 50 pb, con espaciadores intermedios que oscilaban entre 29 pb y 35 pb. Sobre la base de la secuencia de repetición directa, la solicitante rastreó de forma computacional secuencias candidatas tracrRNA con los siguientes criterios: fuera de la matriz de crRNA, pero que contienen un alto grado de homología con repeticiones directas (tal como se requiere para el apareamiento de bases repetición directa:tracrRNA; análisis computacional habitual), fuera de las regiones codificadoras de los componentes de proteínas, que contienen señales de terminación de la transcripción Rho-independientes, ~ 60 pb-120 pb aguas abajo de la región de homología de repeticiones directas, y co-plegamiento con repeticiones directas para formar un dúplex, seguido por dos o más estructuras de horquilla en el extremo distal de la secuencia de tracrRNA. Sobre la base de estos criterios de predicción, la solicitante seleccionó un conjunto inicial de 18 proteínas Cas9 y sus repeticiones directas asociadas de forma única y tracrRNAs distribuidos en todas las cinco familias Cas9. La solicitante generó, además, un conjunto de 18 estructuras de ARN quimérico que conservaron la secuencia y estructuras secundarias del dúplex repetición directa nativa:tracrRNA al tiempo que acortaron la región de apareamiento de bases y fusionaron los dos elementos de ARN a través de un bucle artificial (Figuras 8A-J).

Ejemplo 3: Ortólogos Cas9

La solicitante ha generado ortólogos Cas9 optimizados en codones para progresar en la expresión en células eucariotas.

Un ejemplo de una secuencia humana optimizada en codones (es decir, optimizada para la expresión en seres humanos.) : SaCas9: se proporciona a continuación

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAA
 GGTTCGAAGCGTCCATGAAAAGGAACTACATTCTGGGGCTGGACATCGGGATTACAAGCGTG
 GGGTATGGGATTATTGACTATGAAACAAGGGACGTGATCGACGCAGGGCTCAGACTGTTCA
 AGGAGGCCAACGTGGAAAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGA
 AACGACGGAGAAGGCACAGAATCCAGAGGGTGAAGAACTGCTGTTTCGATTACAACCTGCT
 GACCGACCATTCTGAGCTGAGTGGAAATTAATCCTTATGAAGCCAGGGTGAAAGGCCCTGAGTC
 AGAAGCTGTCAGAGGAAGAGTTTCCGCAGCTCTGCTGCACCTGGCTAAGCGCCGAGGAGT
 GCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAACAGATC
 TCACGCAATAGCAAAGCTCTGGAAGAGAAGTATGTCCGAGAGCTGCAGCTGGAACGGCTGA
 AGAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTCAAGACAAGCGACTACGTCAAAGA
 AGCCAAGCAGCTGCTGAAAGTGCAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGAT
 ACTTATATCGACCTGCTGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCC
 CCTTCGGATGGAAAGACATCAAGGAATGGTACGAGATGCTGATGGGACATTGCACCTATTTT
 CCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTGTACAACGCCCTGAATG
 ACCTGAACAACCTGGTCAACACCAGGGATGAAAACGAGAACTGGAATACTATGAGAAGTT
 CCAGATCATCGAAAACGTGTTTAAGCAGAAGAAAAAGCCTACACTGAAACAGATTGCTAAG
 GAGATCCTGGTCAACGAAGAGGACATCAAGGGCTACCGGGTGACAAGCACTGGAAAACCAG
 AGTTCACCAATCTGAAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAAAGAAATCAT
 TGAGAACGCCGAACCTGCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGG
 ACATCCAGGAAGAGCTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGAT
 TAGTAATCTGAAGGGGTACACCGGAACACACAACCTGTCCCTGAAAGCTATCAATCTGATTC
 TGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACC GGCTGAAGCTGGTC
 CCAAAAAAGGTGGACCTGAGTCAGCAGAAAAGAGATCCCAACCACACTGGTGGACGATTTCA
 TTCTGTCACCCGTGGTCAAGCGGAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATC
 AAGAAGTACGGCCTGCCCAATGATATCATTATCGAGCTGGCTAGGGAGAAGAACAGCAAGG
 ACGCACAGAAGATGATCAATGAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTG
 AAGAGATTATCCGAACTACCGGGAAAGAGAACGCAAAGTACCTGATTGAAAAAATCAAGCT
 GCACGATATGCAGGAGGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCCTGGAGGACCTGC
 TGAACAATCCATTCAACTACGAGGTGATCATATTATCCCCAGAAGCGTGTCCTTCGACAAT
 TCCTTTAACAACAAGGTGCTGGTCAAGCAGGAAGAGA ACTCTAAAAAGGGCAATAGGACTC
 CTTTCCAGTACCTGTCTAGTTT CAGATTC CAAGATCTCTTACGAAACCTTTAAAAAGCACATT
 TGAATCTGGCCAAAGGAAAGGGCCGCATCAGCAAGACCAAAAAGGAGTACCTGCTGGAAG

AGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCGGAATCTGGTGGACACA
 AGATACGCTACTCGCGGCCTGATGAATCTGCTGCGATCCIAATTTCCGGGTGAACAATCTGGA
 TGTGAAAAGTCAAGTCCATCAACGGCGGGTTACATCTTTTCTGAGGGCGCAAATGGAAGTTA
 AAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCGAAGATGCTCTGATTATCGCAAATGC
 CGACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAAGCCAAGAAAGTGATGGAGAACCAG
 ATGTTTCAAGAGAAGCAGGCCGAATCTATGCCCGAAAATCGAGACAGAACAGGAGTACAAGG
 AGATTTTCATCACTUCTCACCAGATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCT
 CACCGGGTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACAAGAA
 AAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGATAA
 TGACAAAGCTGAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCATGAT
 CCTCAGACATATCAGAAACTGAAGCTGATTATGGAGCAGTACGGCGACGAGAAQAACCCAC
 TGTATAAGTACTATGAAGAGACTGGGAACCTGACCAAGTATAGCAAAAAGGATAATGG
 CCCCCTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGCCCATCTGGACATCACA
 GACGATTACCTAACAGTCGCAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGATTTCG
 ATGTCTATCTGGACAACGGCGTGTATAAAATTTGTGACTGTCAAGAATCTGGATGTCATCAAA
 AAGGAGAACTACTATGAAGTGAATAGCAAGTCTACGAAGAGGCTAAAAAGCTGAAAAAG
 ATTAGCAACCAGGCAGAGTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATGG
 CGAACTGTATAGGGTTCATCGGGGTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATG
 ATTGACATCACTTACCGAGAGTATCTGGAAAACATGAATGATAAGCGCCCCCTCGAATTAT
 CAAAACAATTGCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAAC
 CTGTATGAGGTGAAGAGCAAAAAGCACCCCTCAGATTATCAAAAAGGGCTAAGAATTC

5 La solicitante analizó ortólogos Cas9 para identificar las secuencias de PAM relevantes y el correspondiente ARN guía quimérico tal como se indica en la Figura 13A-II. Este conjunto expandido de PAMs proporciona una fijación como objetivo más amplia en todo el genoma y también aumenta significativamente el número de sitios diana únicos y proporciona un potencial para identificar nuevas Cas9s con niveles incrementados de especificidad en el genoma. La solicitante determinó que el PAM para subespecies de *Staphylococcus aureus* subespecie Aureus Cas9 era NNGRR (Figura 14). *Staphylococcus aureus* subespecie Aureus Cas9 también se conoce como SaCas9. Las Figuras 15 a-d proporcionan diseños únicos o múltiples de vectores para SaCas9.

10 La Figura 7 muestra logotipos de secuencia para PAMs putativos tal como se indica mediante complementos inversos. Ortólogos Cas9 y sus respectivos sgRNAs se utilizaron para escindir un banco de dianas que portan un PAM aleatorio (secuencia de 7 pb inmediatamente 3' de la secuencia diana). Los productos escindidos se aislaron y secuenciaron profundamente para proporcionar secuencias candidatas de 7 pb que eran permisivas a la escisión para cada uno de los ortólogos Cas9. Para cada uno de los ortólogos Cas9 se determinaron PAMs consenso alineando todas las secuencias candidatas de 7 pb (Figuras 7 y 21).

15 El trabajo adicional de examinar la termodinámica y la estabilidad *in vivo* de dúplex de sgRNA y ADN proporcionará probablemente un poder predictivo adicional para actividades fuera de objetivo, mientras que la exploración de mutantes y ortólogos SpCas9 puede proporcionar nuevas variantes con una especificidad mejorada. La especificidad de ortólogos Cas9 puede evaluarse adicionalmente, examinando la capacidad de cada una de las Cas9 de tolerar apareamiento erróneos entre el ARN guía y su diana de ADN.

20 *Ejemplo 4: Mutaciones Cas9*

En este ejemplo, la solicitante demuestra que las siguientes mutaciones pueden convertir SpCas9 en una enzima que produce mella: D10A, E762A, H840A, N854A, N863A, D986A.

La solicitante proporciona secuencias que muestran donde se encuentran los puntos de mutación en el gen SpCas9 (Figuras 10A-M). La solicitante también muestra que los nickases todavía son capaces de mediar en la

recombinación homóloga. Además de ello, SpCas9 con estas mutaciones reduce (individualmente) el nivel de ruptura de la doble cadena. Ortólogos Cas9 comparten todos la organización general de los dominios 3-4 RuvC y un dominio HNH (Figura 19). El dominio RuvC más próximo a 5' escinde la cadena no complementaria y el dominio HNH escinde la cadena complementaria. Todas las notaciones son en referencia a la secuencia guía.

- 5 El residuo catalítico en el dominio 5' RuvC se identifica a través de la comparación de homología de la Cas9 de interés con otros ortólogos Cas9 (de locus de CRISPR tipo 11 de *S. pyogenes*, locus 1 de CRISPR de *S. thermophilus*, locus 3 de CRISPR de *S. thermophilus* y el locus de CRISPR tipo II de *Franciscilla novicida*), y el residuo Asp conservado está mutado a alanina para convertir Cas9 en una enzima que produce mella de la cadena complementaria. DE manera similar, los residuos His y Asn conservados en los dominios HNH están mutados a alanina para convertir Cas9 en una enzima que produce mella de cadena no complementaria.
- 10

Ejemplo 5: Optimización funcional de Cas9

Para la función mejorada o para desarrollar nuevas funciones, la solicitante genera proteínas quiméricas Cas9 mediante la combinación de fragmentos de diferentes ortólogos Cas9.

- 15 Por ejemplo, la solicitante fusiona el extremo N de St1Cas9 (el fragmento de esta proteína está en negrita) con extremo C de SpCas9 (el fragmento de esta proteína está subrayado).

> Sti(N)Sp(C)Cas9

MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRRTNRQGR
RLARRKKHRRVRLNRLFEEESGLITDFTKISINLNPYQLRVKGLTDELSNEELF
IALKNMVVKHRGISYLDASDDGNSSVGDYAQIVKENSQKQLETKTPGQIQLER
YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDE
FINRYLEILTGKRKYHGPNGEKSRTDYGRYRTSGETLDNIFGILIGKCTFYP
DEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVNEKAM
GPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQ
MDRETLDKLAYVLTLNTEREGIQEALHEFADGSFSQKQVDELVQFRKANS
SIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYID
EKLLTEEIYNPVVAKSVRQAIVNAAIKEYGDFDNIVIEMARENQTTOKGOK
NSRERMKRIIEGKELGSOILKEHPVENTOLONEKLYLYYLONGRDMYVDOELDI
NRLSDYDVDHIVPOSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW
RQLNAKLITQRKFDNLTKAERGGELSEIDKAGEIKRQLVETROITKHVAQILDSR
MNTKYDENDKLIREVKVITLKSCLVSDFRKDFOFYKVREINNYHHAHDAYLNAV
VGTALIKKYPKLESEFVYGDKVYDVRKMIKSEOEIGKATAKYFFYSNIMNFF
KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVO
TGGFSKESILPKRNSDKLIARKKDWDPKKGFFDSTVAVSVLVVAKVEKGKSK
KLKSVKELLGITIMERSSEFEKNPIDFLEAKGYKEVKKDLIILPKYSLFELENGRKR
MLASAGELQKGNELALPSKYVNFYLYASHYEKLGKSPEDNEQKOLFVEQHKHY
LDEIIEQISEFSKRVLADANLKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPA
AFKYEDTTIDRKRYTSTKEVLDATLIHQSIITGLYETRIDLSQLGGD

> Sp(N)St1(C)Cas9

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDS
 GETAEATRLKRTARRRYTRRNKRICYLQEIFSNEMAKVDDSEFFHRLEESFLVEED
 KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
 GHFLIEGDLNPDNSVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
 LENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
 NLLAQIGDOYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHODI
 TLLKALVRQOLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
 ELLVKLNREDLLRKORTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKIL
 TFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
 KNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
 KTNRKVTVKOLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
 NEENEDILEDIVLTLTFEDREMIEERLKYAHLFDDKVMKQKRRRYTGWGRL
 SRKLINGIRDKQSGKTIIDFLKSDGFANRNFMOIHHDDSLTFKEDIQKAQVSGQG
 DSLHEHIANLAGSPAIKKGILOTVKVDELVKVMGRHKPENIVIEMARETNEDD
 EKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRL
 WHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATA
 NQEKQRTPYQALDSMDDAWSFRELKAFVRESKTLNKKKEYLLTEEDISK
 FDVRKKFIERNLVDTRYASRVVNLALQEIIIFRAHKIDTKVSVVRGQFTSCLR
 RHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIET
 GELISDDEYKESVFKAPYQHFDVTLKSKEFEDSILFSYQVDSKFNKISDAIY
 ATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHD
 PQTFEKVIPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGP
 EIKSLKYDSDLGNHIDITPKDSNNKVVLLQSVSPWRADVFNKTTGKYEILG
 LKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSSEFKFTLYKNDLLLKVD
 TETKEQQLFRFLSRTMPKQKHVELKPYDKQKFEGGEALIKVLGNVANS
 QCKKGLGKSNISYKVRTDVLGNQHIIKNEGDKPKLDF

La solicitantes también ha generado proteínas quiméricas Sp_St3 y ha demostrado la escisión *in vitro* por SpCas9, St3Cas9, Sp_St3 quimera y St3_Sp quimera (Figura 17).

- 5 El beneficio de hacer Cas9 quimérica incluye:
 - a. reducir la toxicidad
 - b. mejorar la expresión en células eucariotas
 - c. mejorar la especificidad
 - d. reducir el peso molecular de la proteína, hacer la proteína más pequeña que mediante la combinación de los dominios más pequeños de diferentes homólogos Cas9.
 - e. Alterar el requisito de la secuencia de PAM

Ejemplo 6: Suministro de Cas9 in vivo utilizando partículas o vectores AAV

Suministro *in vivo* – método AAV

AAV es ventajoso frente a otros vectores virales por un par de razones:

- 15
 - Baja toxicidad (esto puede ser debido al método de purificación que no requiere ultracentrifugación de partículas celulares que pueden activar la respuesta inmune)
 - Baja probabilidad de provocar una mutagénesis insercional porque no se integra en el genoma del huésped.

- 5 Mientras que determinados vectores AAV actuales pueden alojar hasta 4300 bases de ADN insertado, como un límite superior o un límite de empaquetamiento, el AAV puede tener ADN insertado de 4,5 o 4,75 KB. Esto significa que el ADN que codifica una enzima Cas9 así como un promotor y terminador de la transcripción tiene que encajar todos en el mismo vector viral. Construcciones mayores que 4,5 o 4,75 KB conducirá a una producción de virus significativamente reducida. SpCas9 es bastante grande, el gen por sí es de más de 4,1 kb, lo que hace que sea difícil de empaquetar en un AAV. Por lo tanto, realizaciones de la invención incluyen utilizar ortólogos de Cas9 que sean más cortos. Por ejemplo:

<u>Especie</u>	<u>Tamaño de Cas9</u>
Corynebacter diphtheria	3252
Fubacterium ventriosum	3321
Streptococcus pasteurianus	3390
Lactobacillus farciminis	3378
Sphaerochaeta globus	3537
Azospirillum B510	3504
Gluconacetobacter diazotrophicus	3150
Neisseria cinerea	3246
Roseburia intestinalis	3420
Parvibaculum lavamentivorans	3111
Staphylococcus aureus	3159
Nitratifactor salsuginis DSM 16511	3396
Campylobacter lari CF89-12	3009
Streptococcus thermophilus LMD-9	3396

La Figura 3 proporciona representaciones esquemáticas de vectores de AAV que se pueden utilizar en los métodos y composiciones de la invención. El empaquetamiento se comentó anteriormente.

10 *Ejemplo 7: Modificación por Ingeniería Genética de Microalgas utilizando Cas9*

Métodos de suministrar Cas9

Método 1: La solicitante suministra Cas9 y ARN guía utilizando un vector que expresa Cas9 bajo el control de un promotor constitutivo tal como Hsp70A-Rbc S2 o beta2-tubulina.

- 15 Método 2: La solicitante proporciona Cas9 y T7 polimerasa utilizando vectores que expresan Cas9 y T7 polimerasa bajo el control de un promotor constitutivo tal como Hsp70A-Rbc S2 o beta2-tubulina. ARN guía será suministrada utilizando un promotor T7 que contiene vector que impulsa el ARN guía.

Método 3: La solicitante suministra ARNm de Cas9 y ARN guía transcrito in vitro a las células de algas. ARN puede transcribirse in vitro. ARNm de Cas9 consistirá en la región codificadora de Cas9 así como 3' UTR de Cop1 para asegurar la estabilización del ARNm de Cas9.

5 Para la recombinación homóloga, la solicitante proporciona un molde de reparación dirigido a una homología adicional.

Secuencia para una casete que impulsa la expresión de Cas9 bajo el control del promotor de beta2-tubulina, seguido por la 3' UTR de Cop1.

```
TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGAGACGGCTTC
CCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCGAAGCTCCTTCGGGGCTG
CATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCG
ATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTAC
CACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTT
CCTCTTCGTTTCAGTCACAACCCGCAAACATGTACCCATACGATGTTCCAGATTACG
```

CTTCGCCGAAGAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACAGCATCGGC
 CTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGT
 GCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAG
 AACCTGATCGGAGCCCTGCTGTTTCGACAGCGGGCGAAACAGCCGAGGCCACCCGGCT
 GAAGAGAACCGCCAGAAGAAGATACACCAGACGGGAAGAACCGGATCTGCTATCTG
 CAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACT
 GGAAGAGTCTTCTGTTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCG
 GCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCACCATCTACCACCTG
 AGAAAAGAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGC
 CCTGGCCACATGATCAAGTTCCGGGGCCACTTCTGATCGAGGGCGACCTGAACCC
 CGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGC
 TGTTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCT
 GCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCAGCTGCCCGGCGA
 GAAGAAGAATGGCCTGTTTCGGCAACCTGATTGCCCTGAGCCTGGCCTGACCCCA
 ACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGAC
 ACCTACGACGACGACCTGGACAACCTGCTGGCCCGAGATCGGCGACCAGTACGCCGA
 CCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAG
 AGTGAACACCGAGATCACCAAGGCCCCCTGAGCGCCTCTATGATCAAGAGATACG
 ACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGGCGCAGCAGCTGCCT
 GAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACAT
 TGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAA
 AGATGGACGGCACCGAGGAACCTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG
 AAGCAGCGGACCTTCGACAACGGCAGCATCCCCACCAGATCCACCTGGGAGAGCT
 GCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGG
 AAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCA
 GGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCC
 CTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCAGAGCTTCATCGAGC
 GGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGC
 CTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACC
 GAGGGAATGAGAAAGCCCGCCTTCTGAGCGGCGAGCAGAAAAAGGCCATCGTGG
 ACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTAC

TTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC
 AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTT
 CCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACAC
 TGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCACCTGTTC
 GACGACAAAGTGATGAAGCAGCTGAAGCGGGCGGAGATACACCGGCTGGGGCAGGC
 TGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
 GATTCCTGAAGTCCGACGGCTTCGCCAACAGAACTTCATGCAGCTGATCCACGAC
 GACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGA
 TAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCA
 TCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAG
 CCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGAC
 AGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGG
 CAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAG
 CTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAAGTGA
 CATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTCTGAA
 GGACGACTCCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCAGGGGCAAG
 AGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCA
 GCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCG
 AGAGAGGCGCCTGAGCGAACTGGATAAAGCCGGCTTCATCAAGAGACAGCTGGTG
 GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACAC
 TAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCTGAAGT
 CCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCA
 ACAACTACCACCACGCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTG
 ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTA
 CGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCC
 AAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCC
 AACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGA
 TCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGGGAAAGTGCTGAGCATGCCC
 CAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTC
 TATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACC
 CTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG

CCAAAGTGGAAAAGGGCAAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGG
GATCACCATCATGGAAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAG
CCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCC
CTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCA
GAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCA
GCCACTATGAGAAGCTGAAGGGCTCCCCGAGGATAATGAGCAGAAACAGCTGTTT
GTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTC
CAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACA
AGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACC
CTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGG
AAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCAT
CACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACAGCCCCA
AGAAGAAGAGAAAAGGTGGAGGCCAGCTAAGGATCCGGCAAGACTGGCCCCGCTTG
GCAACGCAACAGTGAGCCCCTCCCTAGTGTGTTTGGGGATGTGACTATGTATTCGTG
TGTTGGCCAACGGGTCAACCCGAACAGATTGATACCCGCCTJGGCATTTCCTGTCAG
AATGTAACGTCAGTTGATGGTACT

accggaatggttagcttaccgccc aaaatgctggcgtagtaggtcaagactctgagactcgaactcgcacctgaatacgetgaggctat
 cgcaacccgctgcaggctgcctggctggcctcctccgatgttccaacctgtgctgaattcctcctaagccgctggactggcattactggtggggg
 ctatggggttaacggctgctgctcctctggcgtggtgctgactcaca gtaagaaagcactgatgcgctacgaagacgtttacatgccigagg
 gtacaaagcgatracatggcgcaaacaccgcatggaaaatcaacaagaaagtcctagcggctgccaacgtaatacceaagtggaagca
 ttgtccggtcgaggacatccctgcgattgagcgtgaagaactccgatgaaacgggaagacatgucacatgautcctgaggctctcaccgg
 tygaaacgtgctgcgctgctgtgtaccgcaaggacaaggctcgaagctcgcgctatcagccttgaagtcattgagcaagccaataa
 gtttgctaaccataaggccactctggttcccttacaacatggactggcgcggctcgtgtttacgctgtgtcaatgttcaacccgcaaggtaacgat
 atgaccaaaggactgcttacgctggcgaaaggtaaaccaalcggtaaggagggttactactggctgaaatccacggctgcaactgtgctg
 ggtgtcgacaagggtccgttcctgagcgcatcaagttcatgaggaaaaccacgagaacatcatggtctgctgtaagctccactggagaa
 cactgggtgggtgagcaagattctccgttctgctccttgccttctgctttgagtaagcgtgggtacagcaccacoggcctgagctataactgct
 ccttccgctggcgttggacgggtcttctcctggcatccagcacttctccgctgctccgagatgagggtaggtggtcgcgcggttaacttgc
 ttccctagtgaaacggttcaggacatctacgggattgttctaaagaaugtcaacgagattctacaagcagacgcaatcaatgggaccgataac
 gaagtagtaccgtgaccgatgagaacactggtaaatctctgagaaagcaagctgggcactaaggcactggctggtcaatggctggtt
 aoggtgttactcgcagttgactaagcgttcagtc atgacgctggctfacgggttccaaagagttcggctccgctcaacaagtgtggaagata
 ccattcagccagctattgattccgcaagggtctgatgttactcagccgaatcaggctgctggatacatggctaaagctgattgggaalctgt
 gagcgtgacgggtgtagctgcggtgaagcaatgaactggctaaagtctgctgtaagctgctgctgctgagggtcaagataagaagact
 ggagagattctcgcgaagcgttgcgtgctgattgggtactcctgatggttccctgtgtggcaaggatacaagaagccatctcagacgcgc
 ttgaacctgatgttccctggctcagttccgcttacagcctaccatcaacccaacaagala gcgagattgatcacacaacaggaatcgtggt
 atcgtcctaactttgtacacagccaagacggtagccacctcgtaaagactgtatgtgggcacacgagaaagtaoggaatcgaatctttgca
 ctgattcacgactcctcgggtacgattccggctgacgctgcaacctgttcaaaagcagtgccgaaacatagggtgacacafatgagcttggg
 atgtaactggctgatttctacgaccagttcgtgaccaagttgcacagagtcfaattggacuaaatgccagcacttccggctaaaggtaactttaa
 cctccgtgacatcttagagctggactcgcgttcgcgttaaGGATCCGGCAAGACTGGCCCCGCTTGGCAACG
 CAACAGTGAGCCCCCTCCCTAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTGTGG
 CCAACGGGTCAACCCGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTA
 ACGTCAGTTGATGGTACT

Secuencia de un casete que impulsa la expresión de T7 polimerasa bajo el control del promotor beta-2 tubulina, seguido por la 3' UTR de Cop1:

ES 2 576 126 T3

TCTTTCTTGGCCTATGACACTTCCAGCAAAAAGGTAGGGCGGGCTGCGAGACGGCTTC
CCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCGAAAGCTCCTTCGGGGCTG
CATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATAGCCAGGCCCCCG
ATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAACACCTAGATCACTAC
CACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCTAAGGGGGCGCCTCTT
CCTCTTCGTTTCAGTCACAACCCGCAAAATgctcctaagaagaagaggaaggtaaacacgattaacatcgctaag
aacgacttctctgacatcgaactggctgctatcccgtcaacactcggctgaccaffacgggagcgttagctcgcgaacagttggccctg
agcatgagcttacgagatgggtgaagcacgctccgcaagatggttagcgtcaactaaagctggtaggttgcggataacgctgccgcc
aagcctctcatcactaccctactccctaagatgaitgcacgcataacgactgggttaggaaagtgaaagctaacgcggcaagcggccga
cagcctccagttctgcaagaaatcaagccggagccgtagcgtacatcaccattaagaccactcgtgcttgcctaaccagtgctgacaat
acaaccgtcaggctgtagcaagcgaatcggcggccattgaggacgaggctcgttcggctgataccgtgacctgaagctaacgact
tcaagaaaaacgttaggaaacaactcaacaagcggtagggcacgtctacaagaagcattatgcaagttgtaggagctgacatgctctct
aagggtctactcggtagcaggcgtggtcttcgtggcataaggaagactctattcatgtaggagtagcgtgcatcgagatgctcattgagtc

Secuencia de ARN guía impulsado por el promotor T7 (promotor T7, Ns representa secuencia de fijación de objetivo):

gaaat**TAATACGACTCACTAT**ANNNNNNNNNNNNNNNNNNNNNNNgtttagagctaGAAAtagcaa
gttaaaataaggctagctccgtatcaactgaaaaagtgaccaggagtcggtgctttttt

5 Suministro del gen:

La cepa *Chlamydomonas reinhardtii* CC-124 y CC-125 del *Chlamydomonas Resource Center* se utilizará para la electroporación. El protocolo de electroporación sigue el protocolo recomendado estándar del kit GeneArt *Chlamydomonas Engineering* (información de sitio web en tools.invitrogen.com/content/sfs/manuals/geneart_chlamy_kits_man.pdf).

10 También, la solicitante generó una línea de *Chlamydomonas reinhardtii* que expresa Cas9 de forma constitutiva. Esto se puede hacer utilizando pChlamyl (linearizado utilizando PvuI) y seleccionando las colonias resistentes a higromicina. La secuencia para pChlamyl que contiene Cas9 figura debajo. De este modo, para lograr una inactivación de genes simplemente se tiene que suministrar ARN para el ARN guía. Para la recombinación homóloga, la solicitante suministra ARN guía, así como un molde de recombinación homóloga linearizado.

15 pChlamyl -Cas9:

TGCGGTATTTACACCCGCATCAGGTGGCACTTTTCGGGGAAAATGTGCGCGGAACCCC
 TATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGATTATCAAAA
 AGGATCTTACCTAGATCCTTTTAAATTA AAAATGAAGTTTTAAATCAATCTAAAGT
 ATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATC
 TCAGCGATCTGICTATTTTCGTTTCATCCATAGTTGCCCTGACTCCCCGTCGTGTAGATAA
 CTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAC
 CCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCCAGCCAGCCGGAAGGGCCGA
 GCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCG
 GGAAGCTAGAGTAAGTAGTTTCGCCAGTTAATAGTTTGGCGCAACGTTGTTGCCATTGC
 TACAGGCATCGTGGTGTACCGCTCGTTCGTTGGTATGGCTTCATTCAGCTCCGGTTCC
 CAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCC
 TTCGGTCCCTCCGATCGTTGTCAGAAGTAAAGTTGGCCGCAGTGTTATCACTCATGGTT
 ATGGCAGCACIGCATAATTCTCTTACTGTTCATGCCATCCGTAAGATGCTTTTCTGTGA
 CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCT
 CTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTG
 CTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTG
 AGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT
 TCACCAGCGTTTCTGGGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
 AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTG
 AAGCATTTATCAGGGTTATTGTCTCATGACCAAAAATCCCTTAACGTGAGTTTTCGTTC

CACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTT
 CTGCGCGTAAATCTGCTGCTTGCAAACAAAAAACCACCGCTACCAGCGGTGGTTTTGT
 TTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAGTGGCTTCAGCAGAGCG
 CAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAAC
 TCTGTAGCACCGCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGTTGCCA
 GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAG
 GCGCAGCGGTCCGGCTGAACGGGGGGTTCTGTGCACACAGCCCAGCTTGGAGCGAAC
 GACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTC
 CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGA
 GCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCTGTCGGGTT
 TCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGGCGGAGCCT
 ATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTT
 GCTCACATGTTCTTTCCTGCGTTATCCCCTGATTTCTGTGGATAACCGTATTACCGCT
 TTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTG
 AGCGAGGAAGCGGTGCTGAGGCTTGACATGATTGGTGCGTATGTTTGTATGAAGCT
 ACAGGACTGATTTGGCGGGCTATGAGGGCGGGGGAAGCTCTGGAAGGGCCCGCATG
 GGGCGCGGGCGTCCAGAAGGCGCCATACGGCCCCGCTGGCGGCACCCATCCGGTAT
 AAAAGCCCGCGACCCCGAACGGTGACCTCCACTTTCAGCGACAAACGAGCACTTAT
 ACATACGCGACTATTCTGCCGCTATACATAACCACTCAGCTAGCTTAAGATCCCATC
 AAGCTTGCATGCCGGGGCGCGCCAGAAGGAGCGCAGCCAAACCAGGATGATGTTTGA
 TGGGGTATTTGAGCACTTGCAACCCTTATCCGGAAGCCCCCTGGCCCACAAAGGCTA
 GGGCCCAATGCAAGCAGTTCGCATGCAGCCCCCTGGAGCGGTGCCCTCCTGATAAAC
 CGGCCAGGGGGCCTATGTTCTTTACTTTTTTACAAGAGAAGTCACTCAACATCTTAA
 AATGGCCAGGTGAGTCGACGAGCAAGCCCCGGCGGATCAGGCAGCGTGCTTGCAGAT
 TTGACTTGCAACGCCCGCATGTGTGTCGACGAAGGCTTTTGGCTCCTCTGTGCTGTCT
 CAAGCAGCATCTAACCTGCGTCGCCGTTTCCATTTGCAGGAGATTTCGAGGTACCAT
 GTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAG
 CGTCCGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGG
 GCCGTGATCACCGACGAGTACAAGGTGCCAGCAAGAAATCAAGGTGCTGGGCAA
 CACCGACCGGCACAGCATCAAGAAGAACCCTGATCGGAGCCCTGCTGTTTCGACAGCG
 GCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAG

ACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGG
 TGGACGACAGCTTCTTCCACAGACTGGAAGAGTCTTCTGCTGGTGGAAAGAGGATAAG
 AAGCACGAGCGGCACCCCATCTTCGGCAACAFCGTGGACGAGGTGGCCTACCACGA
 GAAGTACCCACCATCTACCACCTGAGAAAAGAACTGGTGGACAGCACCGACAAGG
 CCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCGGGGCCACT
 TCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTTCATC
 CAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGGAAAACCCCATCAACGCCAGCGG
 CGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAA
 ATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTTCGGCAACCTGATT
 GCCCTGAGCCTGGGCCTGACCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGA
 TGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGG
 CCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACG
 CCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCTG
 AGCGCCTCTATGATCAAGAGATACGACGAGCACACCAGGACCTGACCCTGCTGAA
 AGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGA
 GCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTAC
 AAGTTCATCAAGCCATCCTGGAAAAGATGGACGGCACCCGAGGAACTGCTCGTGAA
 GCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCC
 CCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTT
 ACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATC
 CCCTACTACGTGGGCCCTCTGGCCAGGGGAAAACAGCAGATTCGCTGGATGACCAG
 AAAGAGCGAGGAAACCATCACCCCTGGAACCTTCGAGGAAGTGGTGGACAAGGGC
 GCTTCCGCCAGAGCTTCATCGAGCGGATGACCAACITCGATAAGAACCTGCCAAC
 GAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGA
 GCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAAGCCCGCCTTCTGAGCG
 GCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
 GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGA
 AATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCT
 GAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGG
 AAGATATCGTGTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGG
 CTGAAAACCTATGCCACCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCG

GAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACA
 AGCAGTCCGGCAAGACAATCCTGGATTTCCCTGAAGTCCGACGGCTTCGCCAACAGA
 AACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAA
 AGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCG
 GCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTC
 GTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGA
 GAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATC
 GAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAA
 ACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGAT
 ATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCA
 TATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGACCAG
 AAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTGCTGAAG
 AAGATGAAGAATACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAA
 GTTCGACAATCTGACCAAGGCCGAGAGAGGGCGGCCTGAGCGAACTGGATAAGGCCG
 GCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAG
 ATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGA
 AGTGAAAGTGATCACCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCA
 GTTTTACAAAGTGCGCGAGATCAACAATACTACCACCACGCCACGACGCCTACCTGA
 ACGCCGTGCTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTC
 GTGTACGGCGACTACAAGGTGTACGACGTCCGGAAGATGATCGCCAAGAGCGAGCA
 GGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTT
 CAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGA
 CAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTG
 CGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGAC
 AGGCGGCTTCAGCAAAGAGTCTATCCTGCCAAGAGGAACAGCGATAAGCTGATCG
 CCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTG
 GCCTATICTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAA
 GAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGA
 ATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATC
 ATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAAACGGCCGGAAGAGAATGCT
 GGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATG

TGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCGAGGAT
 AATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCAT
 CGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACA
 AAGTGCIGTCCGCCTACAACAAGCACCCGGGATAAGCCCATCAGAGAGCAGGCCGAG
 AATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTAC
 TTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGC
 CACCCTGATCCACCAGAGCATCACCGGCCGTGACGAGACACGGATCGACCTGTCTCA
 GCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAA
 CATATGATTGGAATGTCTTTCTTTCGCTATGACACTTCCAGCAAAAGGTAGGGCGGG
 CTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCGAA
 GCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTAAA
 TAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAA
 CACCTAGATCACTACCCTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGCT
 AAGGGGGCGCCTTTCCTCTTCGTTTTAGTCAACAACCCGCAAACATGACACAAGAAT
 CCCTGTTACTTCTCGACCGTATTGATTGGGATGATTCCTACGCGAGCCTGCGGAACG
 ACCAGGAATTCTGGGAGGTGAGTCGACGAGCAAGCCCCGGCGGATCAGGCAGCGTGC
 TTGCAGATTGACTTGCAACGCCCGCATTGTGTGCGACGAAGGCTTTTGGCTCCTCTGT
 CGCTGTCTCAAGCAGCATCTAACCTGCGTCCCGTTTTCCATTTGCAGCCGCTGGCC
 CGCCGAGCCCTGGAGGAGCTCGGGCTGCCGGTGCCGCCGGTGTGCGGGTGCCCCGG
 CGAGAGCACCAACCCCGTACTGGTCCGCGAGCCCCGGCCCCGGTATCAAGCTGTTCCG
 GCGAGCACTGGTGCGGTCCGGAGAGCCTCGCGTCCGGAGTCCGGAGGCGTACGCGGTC
 CTGGCGGACGCCCGGTGCCGGTGCCCCGCCCTCCTCGGCCCGGGCGAGCTGCGGCC
 CGGCACCGGAGCCTGGCCGTGGCCCTACCTGGTGATGAGCCGGATGACCGGCACCA
 CCTGGCGGTCCCGGATGGACGGCACGACCGACCGGAACGCGCTGCTCGCCCTGGCC
 CGCGAATCGGCCGGGTGCTCGGCCGGCTGCACAGGGTGCCGCTGACCGGGAACAC
 CGTGCTCACCCCCATTCAGAGGTCTTCCCGAACTGCTGCGGGAACGCCGCGGGC
 GACCGTCGAGGACCACCGCGGGTGGGGCTACCTCTCGCCCCGGCTGCTGGACCGCC
 TGGAGGACTGGCTGCCGGACGTGGACACGCTGCTGGCCGGCCGCGAACCCCGGTTT
 GTCCACGGCGACCTGCACGGGACCAACATCTTCGTGGACCTGGCCGCGACCGAGGT
 CACCGGGATCGTCGACTTACCGACGTCTATGCGGGAGACTCCCGCTACAGCCTGGT
 GCAACTGCATCTCAACGCCTTCCGGGGCGACCGCGAGATCCTGGCCCGGCTGCTCGA

 CGGGGCGCAGTGGAAGCGGACCGAGGACTTCGCCCCGGAAGTCTCGCCTTACCT
 TCCTGCACGACTTCGAGGTGTTTCGAGGAGACCCCGCTGGATCTCTCCGGCTTACCG
 ATCCGGAGGAACTGGCGCAGTTCTCTGGGGGCGCCGGACACCGCCCCGGCGCC
 TGATAAGGATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCCTCC
 TAGTGTGTTTGGGGATGTGACTATGTATTCTGTGTTGGCCAACGGGTCAACCCGAA
 CAGATTGATAACCCGCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGATGGTACT

Para todas las células de *Chlamydomonas reinhardtii* modificados, la solicitante utiliza PCR, ensayo de nucleasa SURVEYOR y secuenciación de ADN para verificar la modificación con éxito.

Ejemplo 8: Reconocimiento de SaCas9 y PAM para aplicaciones in vivo

5 El proyecto comenzó ya que la solicitante quería explorar adicionalmente la diversidad del sistema de CRISPR/Cas de tipo II tras la identificación del sistema de CRISPR/Cas de *Streptococcus pyogenes* (Sp) y *Streptococcus thermophilus* (St) CRISPR/Cas como una herramienta de modificación por ingeniería genética del genoma funcional en células de mamífero.

Al definir nuevos sistemas funcionales de CRISPR/Cas de tipo II para su aplicación en células de mamífero, la solicitante será capaz, en potencia, de encontrar:

- 10 (1) Un sistema de CRISPR/Cas con una mayor eficiencia y/o especificidad
- (2) Un sistema de CRISPR/Cas con diferente Motivo Adyacente al Protoespaciador (PAM) que permite la fijación como objetivo de una gama más amplia de loci genómicos
- 15 (3) Un sistema de CRISPR/Cas con un tamaño más pequeño, de modo que la solicitante podría suministrarlos in vivo en un único vector con el sistema de suministro vírico en mamíferos tales como vectores de virus adeno-asociados (AAV) que tiene un límite de tamaño de empaquetamiento (el actual sistema Sp o St excede este límite de 4,7kb)

y otros rasgos deseables.

Identificación y Diseño del sistema de CRISPR/Cas en Sa para aplicación in vivo

20 La solicitante sometió a ensayo un nuevo sistema de CRISPR/Cas de tipo II en *Staphylococcus aureus* (Sa) que funciona in vitro en un ensayo de escisión de ADNds e identificó un PAM putativo de NNGRRT. Los componentes de este sistema son una proteína Cas9 de Sa, un ARN guía de CRISPR con repeticiones directas (DR) de Sa que formará un complejo de ARN guía funcional con tracrRNA de Sa. Este sistema de tres componentes es similar a todos los demás sistemas de CRISPR/Cas de tipo II. Por lo tanto, la solicitante diseñó un sistema de dos componentes, en donde la solicitante fusionó el tracrRNA de Sa al ARN de CRISPR guía de Sa a través de un corto bucle de tallo para formar un ARN guía quimérico, exactamente como lo hizo la solicitante con el sistema de CRISPR/Cas de *Streptococcus pyogenes* (Sp). Esta ARN guía quimérico fue capaz de apoyar la escisión de ADNds in vitro. Por lo tanto, la solicitante decidió clonar el sistema completo de dos componentes: cas9 y el ARN guía quimérico, en un vector de AAV para probar su funcionalidad en organismos vivos.

30 La solicitante eligió el sistema de AAV, ya que es un virus de mamífero no inmunogénico, basado en ADNss y no integrante, que tiene un amplio espectro de tropismo en diferentes tejidos/órganos dependiendo del serotipo, que se ha demostrado que es seguro para la aplicación in vivo y también apoya la expresión a largo plazo del transgén en los organismos vivos.

35 El diseño del vector de AAV inicial tiene (1) un promotor de CMV que impulsa la proteína SaCas9 con un solo NLS y un marcador de epítipo HA. (2), el promotor U6 humano que impulsa el ARN quimérico. Éstos se colocan entre dos Repeticiones Terminales Invertidas (ITRs) del serotipo 2 de AAV mejor estudiado que sirven como la señal de empaquetamiento viral.

40 El ensayo de la secuencia de PAM en genoma de mamíferos endógeno es el siguiente: Se seleccionaron espaciadores de la diana SaCas9 a través de múltiples genes para cubrir diferentes secuencias de PAM potenciales. Diferentes espaciadores se clonaron en la casete de ADNds de expresión de U6-sgRNA (ARN guía sencillo). Los espaciadores en la casete de ADNds de expresión fueron co-transfectados en líneas de células de mamíferos (293FT para dianas humanas, N2a y Hepa para dianas de ratón). 72 horas después de la transfección, se extrajo todo el ADN genómico y se sometió al ensayo de nucleasa surveyor. Se hizo pasar a través de Gel TBEPAGE para detectar la escisión genómica. Se cuantifica la eficacia de escisión de ADN genómico y se representa.

Sumarios de la Eficiencia de Escisión del Genoma y otras Estadísticas en Todas las Dianas Testadas

Secuencias SaCas9 PAM	Recuentos Dianas	Recuentos Dianas Cambiadas	Porcentaje Dianas Escindidas (%)	Eficiencia de Escisión Acumulativa (%)	Contenido Medio GC de Espaciador (%)
GAAA	1	1	100.0	5.4	65.0
GAAC	2	2	100.0	6.1	55.0
GAAG	8	8	100.0	47.1	65.0
GAAT	9	8	88.9	138.9	66.1
GAGA	3	3	100.0	17.5	69.2
GAGC	8	8	100.0	44.2	61.0
GAGG	12	12	100.0	63.3	58.8
GAGT	44	20	45.5	434.0	58.2
GGAA	2	2	100.0	4.7	60.0
GGAC	3	2	66.7	39.9	60.0
GGAG	12	8	75.0	38.9	58.8
GGAT	20	10	50.0	180.2	59.0
GGGA	7	5	71.4	39.1	63.8
GGGC	11	9	81.8	70.3	65.5
GGGG	8	5	62.5	53.3	70.0
GGGT	45	18	40.0	402.3	58.2
Total General	156	120	61.2	1618.6	59.4

Sumario de la Eficiencia de Escisión del Genoma y otras Estadísticas en Todas las Dianas Testadas

Secuencias de Sa Cas9 PAM	Recuento de Dianas	Recuento de Dianas Cambiadas	Porcentaje de Dianas Escindidas (%)	Eficiencia de Escisión Acumulativa (%)	Eficiencia de Escisión Media (%)	Contenido GC de Espaciador Medio (%)
GAAA	1	1	100.0	5.4	5.4	65.0
GAAC	2	2	100.0	6.1	3.0	50.0
GAAG	8	8	100.0	47.1	5.9	65.0
GAAT	4	4	100.0	66.4	17.1	65.0
GAGA	2	2	100.0	12.5	6.3	67.5
GAGC	8	8	100.0	39.2	7.8	61.0
GAGG	11	11	100.0	68.3	8.0	58.2
GAGT	13	10	70.0	199.0	15.3	58.2
GGAA	2	2	100.0	4.7	2.3	60.0
GGAC	3	2	66.7	39.9	13.3	60.0
GGAG	12	8	75.0	38.9	3.1	58.8
GGAT	13	9	69.2	181.2	12.4	58.8
GGGA	7	5	71.4	39.1	5.6	63.8
GGGC	11	9	81.8	70.3	6.4	65.5
GGGG	8	5	62.5	53.3	6.7	70.0
GGGT	14	8	57.1	182.8	13.0	54.0
Total General	116	92	79.3	1053.6	9.1	60.5

5 Los resultados del ensayo de PAM se muestran en las Figs. 22-27. Un ensayo completo de más de 100 dianas identificó que el PAM para SaCas9 podría describirse como NNGRR (pero no NNGRRT como se indicó anteriormente).

Sumario del Ensayo de PAM: (1) NNGRR para SaCas9 PAM general – de ayuda para el diseño de nuevas dianas, (2) Testado de doble nickase con nuevas dianas, (3) NNGRG puede ser un PAM más potente.

Referencias:

- 10 1. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823 (2013)
 2. Mali, P. et al. RNA-Guided Human Genome Engineering via Cas9. Science 339, 823-826 (2013).

3. Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* 2, e00471 (2013).
4. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31, 230-232 (2013).
- 5 5. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471,602-607 (2011).
6. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816-821 (2012).
7. Wang, H. et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* 153, 910-918 (2013).
- 10 8. Guschin, D.Y. et al. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* 649,247-256 (2010).
9. Bogenhagen, D.F. y Brown, D.D. Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. *Cell* 24, 261-270 (1981).
- 15 10. Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31, 227-229 (2013).
11. Bultmann, S. et al. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res* 40, 5368-5377 (2012).
12. Valton, J. et al. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J Biol Chem* 287, 38427-38432 (2012).
- 20 13. Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757-761 (2010).
14. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29, 143-148 (2011).
15. Mussolino, C. et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic acids research* 39,9283-9293 (2011).
- 25 16. Hsu, P.D. y Zhang, F. Dissecting neural function using targeted genome engineering technologies. *ACS chemical neuroscience* 3, 603-610 (2012).
17. Sanjana, N.E. et al. A transcription activator-like effector toolbox for genome engineering. *Nature protocols* 7, 171-192 (2012).
- 30 18. Porteus, M.H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* 300, 763 (2003).
19. Miller, J.C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* 25, 778-785 (2007).
20. Sander, J.D. et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* 8, 67-69 (2011).
- 35 21. Wood, A.J. et al. Targeted genome editing across species using ZFNs and TALENs. *Science* 333, 307 (2011).

22. Bobis-Wozowicz, S., Osiak, A., Rahman, S.H. y Cathomen, T. Targeted genome editing in pluripotent stem cells using zinc-finger nucleases. *Methods* 53, 339-346 (2011).
23. Jiang, W., Bikard, D., Cox, D., Zhang, F. y Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31, 233-239 (2013).
- 5 24. Qi, L.S. et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152, 1173-1183 (2013).
25. Michaelis, L.M., Maud "Die kinetik der invertinwirkung.". *Biochem. z* (1913).
26. Mahfouz, M.M. et al. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci U S A* 108, 2623-2628 (2011).
- 10 27. Wilson, E.B. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 22, 209-212 (1927).
28. Ding, Q. et al. A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* 12, 238-251 (2013).
- 15 29. Soldner, F. et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* 146, 318-331 (2011).
30. Carlson, D.F. et al. Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A* 109, 17382-17387 (2012).
31. Geurts, A.M. et al. Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. *Science* 325, 433-433 (2009).
- 20 32. Takasu, Y. et al. Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. *Insect Biochem Molec* 40, 759-765 (2010).
33. Watanabe, T. et al. Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nat Commun* 3 (2012).
- 25 34. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30, 460-465 (2012).
35. Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509-1512 (2009).
36. Moscou, M.J. y Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* 326,1501 (2009).
- 30 37. Deveau, H., Garneau, J.E. y Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* 64,475-493 (2010).
38. Horvath, P. y Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167-170 (2010).
- 35 39. Makarova, K.S. et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9, 467-477 (2011).
40. Bhaya, D., Davison, M. y Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45, 273-297 (2011).

41. Garneau, J.E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67-71 (2010).
42. Gasiunas, G., Barrangou, R., Horvath, P. y Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109, E2579-2586 (2012).
- 5 43. Umov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. y Gregory, P.D. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11, 636-646 (2010).
44. Perez, E.E. et al. Establishment of HIV-1 resistance in CD4(+) T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* 26, 808-816 (2008).
- 10 45. Chen, F.Q. et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* 8, 753-U796 (2011).
46. Bedell, V.M. et al. In vivo genome editing using a high-efficiency TALEN system. *Nature* 491, 114-U 133 (2012).
47. Saleh-Gohari, N. y Helleday, T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. *Nucleic Acids Res* 32, 3683-3688 (2004).
- 15 48. Saprunauskas, R. et al. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 39, 9275-9282 (2011).
49. Shen, B. et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res* 23, 720-723 (2013).
50. Tuschl, T. Expanding small RNA interference. *Nat Biotechnol* 20, 446-448 (2002).
- 20 51. Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A. y Kucherlapati, R.S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* 317, 230-234 (1985).
52. Thomas, K.R., Folger, K.R. y Capecchi, M.R. High frequency targeting of genes to specific sites in the mammalian genome. *Cell* 44, 419-428 (1986).
53. Hastly, P., Rivera-Perez, J. y Bradley, A. The length of homology required for gene targeting in embryonic stem cells. *Mol Cell Biol* 11, 5586-5591 (1991).
- 25 54. Wu, S., Ying, G.X., Wu, Q. y Capecchi, M.R. A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. *Nat Protoc* 3, 1056-1076 (2008).
55. Oliveira, T.Y. et al. Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. *J Immunol Methods* 375, 176-181 (2012).
- 30 56. Tremblay et al., Transcription Activator-Like Effector Proteins Induce the Expression of the Frataxin Gene; *Human Gene Therapy*. agosto de 2012, 23(8): 883-890.
57. Shalek et al. Nanowire-mediated delivery enables functional interrogation of primary immune cells: application to the analysis of chronic lymphocytic leukemia. *Nano Letters*, 12 de diciembre de 2012;12(12):6498-504.
58. Pardridge et al. Preparation of Trojan horse liposomes (THLs) for gene transfer across the blood-brain barrier; *Cold Spring Harb Protoc*; abril de 2010 (4)
- 35 59. Plosker GL et al. Fluvastatin: a review of its pharmacology and use in the management of hypercholesterolaemia; *Drugs* 1996, 51(3):433-459).

60. Trapani et al. Potential role of nonstatin cholesterol lowering agents; IUBMB Life, Volumen 63, Cuaderno 11, páginas 964-971, noviembre de 2011

61. Birch AM et al. DGAT1 inhibitors as anti-obesity and anti-diabetic agents; Current Opinion in Drug Discovery & Development, 2010, 13(4):489-496

5 62. Fuchs et al. Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi), Oncogene 2002,21(37):5716-5724.

63. McManaman JL et al. Perilipin-2 Null Mice are Protected Against Diet-Induced Obesity, Adipose Inflammation and Fatty Liver Disease; The Journal of Lipid Research, jlr.M035063. Publicado por primera vez el 12 de febrero de 2013.

10 64. Tang J et al. Inhibition of SREBP by a Small Molecule, Betulin, Improves Hyperlipidemia and Insulin Resistance and Reduces Atherosclerotic Plaques; Cell Metabolism, Volumen 13, Cuaderno 1, 44-56, 5 de enero de 2011.

65. Dumitrache et al. Trex2 enables spontaneous sister chromatid exchanges without facilitating DNA double-strand break repair; Genetics. agosto de 2011; 188(4): 787-797

15 Aun cuando realizaciones preferidas de la presente invención se han mostrado y descrito en esta memoria, resultará obvio para los expertos en la técnica que estas realizaciones se proporcionan a modo de ejemplo. Numerosas variaciones, cambios y sustituciones se les ocurrirán ahora a los expertos en la técnica sin apartarse de la invención. Ha de entenderse que en la práctica de la invención se pueden emplear diversas alternativas a las realizaciones de la invención descrita en esta memoria.

La invención se describe también con ello con referencia a las siguientes cláusulas.

20 Cláusula 1. (A) Una composición que se produce de forma no natural o mediante ingeniería genética, que comprende un sistema de vectores que comprende uno o más vectores que comprenden

I. un primer elemento regulador enlazado operativamente a una primera secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la secuencia de polinucleótidos comprende

25 (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota,

(b) una secuencia de emparejamiento tracr, y

(c) una secuencia de tracr, y

II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de enzimas que codifica una enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear,

en donde (a), (b) y (c) están dispuestas en una orientación 5' a 3',

30 en donde los componentes I y II están ubicados en el mismo o en diferentes vectores del sistema,

en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y

en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr,

o

(B) una composición que se produce de forma no natural o mediante ingeniería genética, que comprende un sistema de vectores que comprende uno o más vectores que comprenden

I. un primer elemento regulador enlazado operativamente a

40 (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota, y

(b) al menos una o más secuencia de emparejamiento tracr,

II. un segundo elemento regulador enlazado operativamente a una secuencia codificadora de enzimas que codifica una enzima de CRISPR, y

III. un tercer elemento regulador enlazado operativamente a una secuencia de tracr,

45 en donde los componentes I, II y III están situados en el mismo o en diferentes vectores del sistema.

en donde cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana,

y en donde el complejo de CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, y

50

en donde, en (A) o (B), la enzima CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Fiaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*.

5 Cláusula 2. Una enzima de CRISPR modificada, en donde la enzima es un ortólogo Cas9 de un género perteneciente al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Fiaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*, y en donde diferencias de la correspondiente enzima de CRISPR de tipo salvaje comprenden:

10 la enzima de CRISPR modificada está truncada en comparación con una enzima de CRISPR de tipo salvaje, o

la enzima de CRISPR tiene una longitud de al menos 500 aminoácidos, al menos 800-899 aminoácidos, al menos 900-999 aminoácidos, al menos 1000-1099 aminoácidos, al menos 1100-1199 aminoácidos, al menos 1200-1299 aminoácidos, al menos 1300-1399 aminoácidos, al menos 1400-1499 aminoácidos, al menos 1500-1599 aminoácidos, al menos 1600-1699 aminoácidos o al menos 2000 aminoácidos, y/o

15 la enzima de CRISPR es una nucleasa que dirige la escisión de las dos cadenas en la localización de la secuencia diana, o la enzima de CRISPR es una nickase que dirige la escisión de una cadena en la localización de la secuencia diana, y/o

20 la enzima de CRISPR está optimizada en codones o está optimizada en codones para la expresión en una célula eucariota, y/o

la enzima de CRISPR comprende una o más mutaciones, y/o

la enzima de CRISPR comprende una enzima CRISPR quimérica.

Cláusula 3. La composición de acuerdo con la cláusula 1, o la enzima de acuerdo con la cláusula 2, en donde la enzima de CRISPR tiene una o más mutaciones en cualquier dominio de la enzima.

25 Cláusula 4. La composición o enzima de acuerdo con cualquier cláusula precedente, en donde la enzima de CRISPR tiene una o más mutaciones en un dominio catalítico.

Cláusula 5. La composición o enzima de acuerdo con la cláusula 3 ó 4, en donde la una o más mutaciones es en un dominio RuvC I, RuvC II, RuvC III o HNH de la enzima de CRISPR.

30 Cláusula 6. La composición o enzima de acuerdo con la cláusula 4 ó 5, en donde la enzima de CRISPR comprende una o más mutaciones correspondientes a la numeración de posición de SpCas9 en las posiciones D10A, E762A, H840A, N854A, N863A o D986A.

Cláusula 7. La composición o enzima de acuerdo con cualquier cláusula precedente, en donde la enzima comprende, además, un dominio funcional.

35 Cláusula 8. La composición o enzima de acuerdo con la cláusula 7, en donde el dominio funcional es un dominio activador transcripcional.

Cláusula 9. La composición o enzima de acuerdo con la cláusula 8, en donde el dominio activador transcripcional es VP64.

Cláusula 10. La composición o enzima de acuerdo con las cláusulas 7 a 9, en donde el dominio funcional es un dominio represor transcripcional.

40 Cláusula 11. La composición o enzima de acuerdo con la cláusula 10, en donde el dominio represor transcripcional es KRAB, SID o SID4X.

Cláusula 12. Un método para modificar un organismo o un organismo no humano mediante la manipulación de una secuencia diana en un locus genómico de interés, que comprende

45 suministrar una composición que se produce de forma no natural o modificada por ingeniería genética que comprende un sistema de vectores que codifica operativamente una composición de cualquier cláusula precedente para la expresión de la misma.

Cláusula 13. El método de acuerdo con la reivindicación 12, en el que el suministro es a través de uno o más vectores virales.

Cláusula 14. Una composición que se produce de forma no natural o que se produce mediante ingeniería genética, que comprende un sistema de vectores que comprende uno o más vectores que comprenden

I. un primer elemento regulador enlazado operativamente a una primera secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la primera secuencia de polinucleótidos comprende

(i) una primera secuencia guía capaz de hibridarse con una primera secuencia diana en un primer locus genómico en una célula del organismo,

(ii) una primera secuencia de emparejamiento tracr, y

(iii) una primera secuencia de tracr, y

II. un segundo elemento regulador enlazado operativamente a una segunda secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la segunda secuencia de polinucleótidos comprende

(i) una segunda secuencia guía capaz de hibridarse con una segunda secuencia diana en un segundo locus genómico en la célula del organismo,

(ii) una segunda secuencia de emparejamiento tracr, y

(iii) una segunda secuencia de tracr, y

III. un tercer elemento regulador enlazado operativamente a una secuencia codificadora de enzimas que codifica una primera enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear y está enlazada operativamente a un primer dominio funcional,

IV. un cuarto elemento regulador enlazado operativamente a una secuencia codificadora de enzima que codifica una segunda enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear y está enlazada operativamente a un segundo dominio funcional,

en donde (i), (ii) y (iii) en I y II están dispuestos en una orientación 5' a 3',

en donde los componentes I, II, III y IV se encuentran en el mismo o en diferentes vectores del sistema,

en donde, cuando se transcriben, cada una de las secuencias de emparejamiento tracr se hibrida con su secuencia tracr correspondiente y la primera y segunda secuencias guía dirigen la unión específica para la secuencia del primer y segundo complejos de CRISPR a la primera y segunda secuencia diana,

en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, y en donde la expresión de la enzima de CRISPR proporciona la manipulación de la secuencia diana,

en donde la primera y segunda enzimas de CRISPR comprenden cada una dos o más mutaciones, y

en donde la primera y segunda enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*.

Cláusula 15. Un método de modular la expresión de dos o más loci genómicos de interés en un organismo, que comprende

suministrar una composición que se produce de forma no natural o que se produce mediante ingeniería genética que comprende un sistema de vectores que comprende uno o más vectores que comprenden

I. un primer elemento regulador enlazado operativamente a una primera secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la primera secuencia de polinucleótidos comprende

(i) una primera secuencia guía capaz de hibridarse con una primera secuencia diana en un primer locus genómico en una célula del organismo,

(ii) una primera secuencia de emparejamiento tracr, y

(iii) una primera secuencia de tracr, y

II. un segundo elemento regulador enlazado operativamente a una segunda secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la segunda secuencia de polinucleótidos comprende

(i) una segunda secuencia guía capaz de hibridarse con una segunda secuencia diana en un segundo locus genómico en la célula del organismo,

(ii) una segunda secuencia de emparejamiento tracr, y

(iii) una segunda secuencia de tracr, y

III. un tercer elemento regulador enlazado operativamente a una secuencia codificadora de enzima que codifica una primera enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear y está enlazada operativamente a un primer dominio funcional,

IV. un cuarto elemento regulador enlazado operativamente a una secuencia codificadora de enzima que codifica una segunda enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear y está enlazada operativamente a un segundo dominio funcional,

en donde (i), (ii) y (iii) en I y II están dispuestos en una orientación 5' a 3',

5 en donde los componentes I, II, III y IV se encuentran en el mismo o en diferentes vectores del sistema,

en donde, cuando se transcriben, cada una de las secuencias de emparejamiento tracr se hibrida con su secuencia tracr correspondiente y la primera y segunda secuencias guía dirigen la unión específica para la secuencia del primer y segundo complejos de CRISPR a la primera y segunda secuencia diana,

10 en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, y en donde la expresión de la enzima de CRISPR proporciona la manipulación de la secuencia diana,

en donde la primera y segunda enzimas de CRISPR comprenden cada una dos o más mutaciones,

15 en donde la primera y segunda enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*, y

20 en donde el primer locus genómico es modulado por la actividad del primer dominio funcional y el segundo locus genómico es modulado por la actividad del segundo dominio funcional.

Cláusula 16. La composición de acuerdo con la cláusula 14, o el método de acuerdo con la cláusula 15, en donde el primer dominio funcional se selecciona del grupo que consiste en un activador transcripcional, un represor transcripcional, una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo y un dominio inducible/controlable por la luz o un dominio inducible y/o controlable químicamente.

25 Cláusula 17. La composición de acuerdo con la cláusula 14 ó 16, o el método de acuerdo con la cláusula 15 ó 16, en donde el segundo dominio funcional se selecciona del grupo que consiste en un activador transcripcional, un represor transcripcional, una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo y un dominio inducible/controlable por la luz o un dominio inducible y/o controlable químicamente.

30 Cláusula 18. La composición o el método de acuerdo con la cláusula 16 ó 17, en donde el primer o segundo dominio funcional es un dominio activador transcripcional.

Cláusula 19. La composición o el método de acuerdo con la cláusula 47, en donde el dominio activador transcripcional es VP64.

Cláusula 20. La composición o el método de acuerdo con la cláusula 16 ó 17, en donde el primer o segundo dominio funcional es un dominio represor transcripcional.

35 Cláusula 21. La composición o el método de acuerdo con la cláusula 20, en donde el dominio represor transcripcional es KRAB, SID o SID4X.

Cláusula 22. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Sutterella wadsworthensis*.

40 Cláusula 23. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Filifactor aloicis*.

Cláusula 24. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Lactobacillus johnsonii*.

Cláusula 25. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Campylobacter lari*.

45 Cláusula 26. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Corynebacter diphtheriae*.

Cláusula 27. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Parvibaculum lavamentivorans*.

- Cláusula 28. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Mycoplasma gallisepticum*.
- Cláusula 29. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Staphylococcus aureus* y subespecies *Aureus*.
- 5 Cláusula 30. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Legionella pneumophila* Paris.
- Cláusula 31. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Treponema denticola*.
- 10 Cláusula 32. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Staphylococcus pseudintermedius*.
- Cláusula 33. La composición o el método de acuerdo con cualquiera de las cláusulas 14 a 21, en donde la primera o segunda enzima de CRISPR es una Cas9 de *Neisseria cinerea*.
- Cláusula 34. Un método de tratar o inhibir una afección provocada por un defecto en una secuencia diana en un locus genómico de interés en un sujeto o un sujeto no humano que necesita del mismo, que comprende modificar al sujeto o a un sujeto no humano mediante la manipulación de la secuencia diana y, en el que la afección es susceptible de tratamiento o inhibición mediante la manipulación de la secuencia diana, que comprende proporcionar tratamiento que comprende:
- 15 suministrar una composición que se produce de forma no natural o modificada por ingeniería genética que comprende un sistema de vectores AAV que comprende uno o más vectores AAV, que comprende codificar operativamente una composición de cualquiera de las cláusulas 1 ó 3 a 11 para la expresión de la misma, en donde la secuencia diana es manipulada por la composición cuando se expresa.
- 20 Cláusula 35. Un método de la cláusula 34, que incluye inducir la expresión.
- Cláusula 36. El método de la cláusula 34 ó 35, en el que el organismo o sujeto es un eucariota o un eucariota no humano.
- 25 Cláusula 37. El método de cualquiera de las cláusulas 34 a 36, en el que el organismo o sujeto es una planta.
- Cláusula 38. El método de cualquiera de las cláusulas 34 a 36, en el que el organismo o sujeto es un mamífero o un mamífero no humano.
- Cláusula 39. El método de cualquiera de las cláusulas 34 a 37, en el que el organismo o sujeto son algas.
- Cláusula 40. El método de cualquiera de las cláusulas 34 a 39, en el que el vector viral es un AAV.
- 30 Cláusula 41. Una composición o enzima de acuerdo con cualquiera de las cláusulas 1-11, 14, o 16-33 para uso en medicina o para uso en terapia.
- Cláusula 42. Una composición o enzima de acuerdo con cualquiera de las cláusulas 1-11, 14, o 16-33 para uso en un método de acuerdo con cualquiera de las cláusulas 12, 13 ó 15-40.
- 35 Cláusula 43. Uso de una composición o enzima de acuerdo con cualquiera de las cláusulas 1-11, 14, o 16-33 en la edición de un gen o genoma *ex vivo*.
- Cláusula 44. Uso de una composición o enzima de acuerdo con cualquiera de las cláusulas 1-11, 14, o 16-33 en la fabricación de un medicamento para la edición de un gen o genoma *ex vivo* o para uso en un método de acuerdo con cualquiera de las cláusulas 12, 13 ó 15-40.
- 40 Cláusula 45. Una composición o enzima de acuerdo con cualquiera de las cláusulas 1-11, 14, o 16-33, en donde la secuencia diana está flanqueada en su extremo 3' por 5'-NRG (en que N es cualquier nucleótido), especialmente en que la Cas9 es (o se deriva de) *S. pyogenes* o *S. aureus* Cas9.

Cláusula 46. Un método de preparar un AAV para suministro de la composición o enzima de acuerdo con cualquiera de las cláusulas 1-11, 14 ó 16-33, que comprende transfectar un plásmido o plásmidos que contienen o consisten esencialmente en una molécula o moléculas de ácido nucleico que codifican el AAV en células infectadas con AAV, y suministrar AAV rep y/o cap obligatorios para la replicación y el empaquetamiento del AAV.

- 5 Cláusula 47. El método de la cláusula 46, en el que el AAV rev y/o cap obligatorio para la replicación y el empaquetamiento del AAV se suministran transfectando las células con un plásmido o plásmidos helper o un virus o más virus helper.

Cláusula 48. El método de la cláusula 47, en el que el virus helper es un virus de la viruela, adenovirus, virus herpes o baculovirus.

- 10 Cláusula 49. El método de la cláusula 48, en el que el virus de la viruela es un virus vacuna.

Cláusula 50. El método de cualquiera de las cláusulas 46 a 49, en el que:
- las células son células de mamífero; o
- las células son células de insectos y el virus helper es baculovirus.

- 15 Cláusula 51. Un método de suministrar la enzima CRISPR de cualquiera de las cláusulas 2-11, que comprende suministrar a una célula ARNm que codifica la enzima de CRISPR.

Cláusula 52. Una enzima Cas quimérica, en donde la enzima comprende uno o más fragmentos de un primer ortólogo Cas y uno o más fragmentos de un segundo ortólogo Cas.

Cláusula 53. La enzima de acuerdo con la cláusula 52, en donde el uno o más fragmentos del primer o segundo ortólogo Cas son del extremo C o N del primer o segundo ortólogo Cas.

- 20 Cláusula 54. La enzima de acuerdo con la cláusula 52 ó 53, en donde el primer o segundo ortólogo Cas se selecciona de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*.

- 25 Cláusula 55. Uso de la composición o enzima de acuerdo con cualquiera de las cláusulas 1-11, 14, 16-33 ó 52-54 en la preparación de un medicamento para la modificación de una secuencia diana.

Cláusula 56. La composición de acuerdo con cualquiera de las cláusulas 1 a 11, en donde la enzima de CRISPR está truncada en comparación con una enzima de CRISPR de tipo salvaje o la enzima de CRISPR comprende de 500-2000 aminoácidos.

- 30 Cláusula 57. La composición de acuerdo con cualquiera de las cláusulas 1 a 11 ó 56, en donde la enzima de CRISPR es una nucleasa que dirige la escisión de las dos cadenas en la ubicación de la secuencia diana, o la enzima CRISPR es una nickase que dirige la escisión de una cadena en la ubicación de la secuencia diana.

Cláusula 58. La composición de acuerdo con cualquiera de las cláusulas 1 a 11 ó 56-57, en donde la secuencia guía comprende al menos quince nucleótidos.

- 35 Cláusula 59. La composición de acuerdo con cualquiera de las cláusulas 1 a 11 ó 56-58, en donde CRISPR está optimizada en codones o está optimizada en codones para la expresión en una célula eucariota.

Cláusula 60. La composición de acuerdo con cualquiera de las cláusulas 1 a 11 ó 56-59, en donde la enzima de CRISPR comprende una o más mutaciones.

- 40 Cláusula 61. La composición de acuerdo con cualquiera de las cláusulas 1 a 11 ó 56-60, en donde la enzima de CRISPR comprende una enzima de CRISPR quimérica.

Cláusula 62. Una composición que se produce de forma no natural o modificada por ingeniería genética, que comprende:

A) - I. una secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la secuencia de polinucleótidos comprende:

- (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota,
- (b) una secuencia de emparejamiento de tracr, y
- (c) una secuencia de tracr, y

II. una secuencia de polinucleótidos que codifica una enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear,

en donde (a), (b) y (c) están dispuestas en una orientación 5' a 3',

en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y

en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, y la secuencia de polinucleótidos que codifica una enzima de CRISPR es ADN o ARN,

o

(B) I. polinucleótidos que comprenden:

- (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula procarionota, y
- (b) al menos una o más secuencias de emparejamiento tracr,

II. una secuencia de polinucleótido que codifica una enzima de CRISPR, y

III. una secuencia de polinucleótidos que comprende una secuencia tracr,

en donde cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y

en donde el complejo de CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, y la secuencia de polinucleótido que codifica una enzima de CRISPR es ADN o ARN, y

en donde la enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Fiaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*.

Cláusula 63. Un método de modificar un organismo o un organismo no humano mediante manipulación de una secuencia diana en un locus genómico de interés, que comprende suministrar una composición que se produce de forma no natural o modificada por ingeniería genética, que comprende:

A) - I. una secuencia de polinucleótidos de ARN quimérico (chiRNA) de un sistema de CRISPR-Cas, en donde la secuencia de polinucleótidos comprende:

- (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula eucariota,
- (b) una secuencia de emparejamiento de tracr, y
- (c) una secuencia de tracr, y

II. una secuencia de polinucleótidos que codifica una enzima de CRISPR que comprende al menos una o más secuencias de localización nuclear,

en donde (a), (b) y (c) están dispuestas en una orientación 5' a 3',

en donde, cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y

en donde el complejo de CRISPR comprende la enzima de CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia de tracr, y la secuencia de polinucleótidos que codifica una enzima de CRISPR es ADN o ARN,

o

(B) I. polinucleótidos que comprenden:

- (a) una secuencia guía capaz de hibridarse con una secuencia diana en una célula procarionota, y
- (b) al menos una o más secuencias de emparejamiento tracr,

II. una secuencia de polinucleótido que codifica una enzima de CRISPR, y

III. una secuencia de polinucleótidos que comprende una secuencia tracr,

en donde cuando se transcribe, la secuencia de emparejamiento tracr se hibrida con la secuencia tracr y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a la secuencia diana, y

en donde el complejo de CRISPR comprende la enzima CRISPR formando complejo con (1) la secuencia guía que se hibrida con la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida con la secuencia tracr, y la secuencia de polinucleótido que codifica una enzima de CRISPR es ADN o ARN, y

en donde la enzima de CRISPR es un ortólogo Cas9 de un género que pertenece al grupo que consiste en *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Fiaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* y *Campylobacter*.

Cláusula 64. La composición de acuerdo con la cláusula 62, o el método de acuerdo con la cláusula 63, en donde la enzima de CRISPR está truncada en comparación con una enzima de CRISPR de tipo salvaje o la enzima de CRISPR está constituida por 500-2000 aminoácidos.

5 Cláusula 65. La composición o el método de acuerdo con las cláusulas 62-64, en donde la enzima de CRISPR es una nucleasa que dirige la escisión de las dos cadenas en la localización de la secuencia diana, o la enzima de CRISPR es una nickase que dirige la escisión de una cadena en la localización de la secuencia diana.

Cláusula 66. La composición o el método de acuerdo con las cláusulas 62-65, en donde la secuencia guía comprende al menos quince nucleótidos.

10 Cláusula 67. La composición o el método de acuerdo con las cláusulas 62-66, en donde la enzima de CRISPR está optimizada en codones o está optimizada en codones para la expresión en una célula eucariota.

Cláusula 68. La composición o el método de acuerdo con las cláusulas 62-67, en donde la enzima de CRISPR comprende una o más mutaciones.

15 Cláusula 69. La composición o el método de acuerdo con las cláusulas 62-68, en donde la enzima de CRISPR comprende una enzima de CRISPR quimérica.

REIVINDICACIONES

1. Una composición que comprende:
componentes del complejo de CRISPR, que comprenden:
 - I. una o más secuencias de polinucleótidos del sistema de CRISPR-Cas, que comprende(n):
 - (a) una secuencia guía modificada por ingeniería genética constituida por ARN y capaz de hibridarse a una secuencia diana en un locus del polinucleótido,
 - (b) una secuencia de emparejamiento tracr constituida por ARN, y
 - (c) una secuencia de tracrRNA constituida por ARN, y
 en donde (a), (b) y (c) están dispuestas en una orientación 5' a 3',
 - II. una proteína Cas9 de Tipo II,
en donde la secuencia de emparejamiento tracr se hibrida a la secuencia tracrRNA y la secuencia guía dirige la unión específica para la secuencia de un complejo de CRISPR a una secuencia diana,
en donde el complejo de CRISPR comprende la proteína Cas9 de Tipo II formando complejo con (1) la secuencia guía que se hibrida a la secuencia diana, y (2) la secuencia de emparejamiento tracr que se hibrida la secuencia de tracrRNA, y
en donde la proteína Cas9 de Tipo II es o comprende una Cas9 de *Staphylococcus aureus* (SaCas9).
2. Una composición de acuerdo con la reivindicación 1, en donde la secuencia de polinucleótidos del sistema de CRISPR-Cas es un ARNm quimérico (chiRNA).
3. Una composición de acuerdo con la reivindicación 2, que comprende un par de chiRNAs.
4. Una composición de acuerdo con la reivindicación 3, que comprende dos pares de chiRNAs.
5. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, que comprende un sistema de vectores que comprende uno o más vectores, y en donde la secuencia de polinucleótidos que comprende o codifica dichos componentes I y II están situadas en el mismo o en diferentes vectores del sistema.
6. Una composición de acuerdo con la reivindicación 5, en donde el uno o más vectores comprenden uno o más vectores virales.
7. Una composición de acuerdo con la reivindicación 6, en donde el uno o más vectores virales comprenden uno o más vectores de retrovirus, lentivirus, adenovirus, virus adeno-asociados o virus herpes simplex.
8. Una composición de acuerdo con una cualquiera de las reivindicaciones 1-4, que comprende una nanopartícula, liposoma, exosoma, sistema de levadura o microvesícula.
9. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, que incluye uno o más dominios funcionales.
10. Una composición de acuerdo con la reivindicación 9, en donde el uno o más dominios funcionales comprenden un dominio activador transcripcional.
11. Una composición de acuerdo con la reivindicación 9, en donde el dominio funcional comprende VP64 o KRAB, SID o SID4X, o una recombinasa, una transposasa, un remodelador de histona, una ADN metiltransferasa, un criptocromo, un dominio inducible/controlable por la luz o un dominio inducible/controlable químicamente.
12. Una composición de acuerdo con una cualquiera de las reivindicaciones 5-11, en donde la composición de vector comprende un vector sencillo.
13. Una composición de acuerdo con una cualquiera de las reivindicaciones 5-12, en donde la célula es una célula eucariota.
14. Una composición de acuerdo con una cualquiera de las reivindicaciones 5-13, en donde la secuencia de polinucleótidos que codifica la proteína Cas9 de Tipo II está optimizada en codones para la expresión en una célula eucariota.

15. Una composición de acuerdo con una cualquiera de las reivindicaciones 5-14, en donde las secuencias de polinucleótidos que comprenden o codifican dichos componentes del complejo de CRISPR comprenden un promotor específico para tejidos.
- 5 16. Una composición de acuerdo con la reivindicación 15, en donde el promotor específico para tejidos dirige la expresión de dichos componentes del complejo de CRISPR en el músculo, neurona, hueso, piel, sangre, hígado, páncreas o linfocitos.
17. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la secuencia guía es capaz de hibridarse a una secuencia diana en una célula eucariota.
- 10 18. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la secuencia de tracrRNA es de 30 o más nucleótidos de longitud.
19. Una composición de acuerdo con una la reivindicación 18, en donde el tracrRNA es de 50 o más nucleótidos de longitud.
20. Una composición de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde la proteína Cas9 de Tipo II comprende, además, una o más secuencias de localización nuclear (NLSs).
- 15 21. Una célula huésped o una línea celular *ex vivo* o *in vitro* que comprende la composición de acuerdo con una cualquiera de las reivindicaciones precedentes o progenie de las mismas, en donde la célula huésped o la línea celular no es una célula de la línea germinal humana.
22. Una célula huésped, línea celular o progenie de la misma *ex vivo* o *in vitro* de acuerdo con la reivindicación 21, que es una célula madre o una línea celular madre.
- 20 23. Un método de modificar un organismo mediante manipulación de una o más secuencias dianas en loci genómicos de interés, que comprende suministrar al organismo la composición de acuerdo con una cualquiera de las reivindicaciones 1-20, en donde el organismo es un organismo no animal.
- 25 24. Un método *ex vivo* de modificar una célula de un organismo mediante manipulación de una o más secuencias dianas en loci genómicos de interés, que comprende suministrar a la célula una composición de acuerdo con una cualquiera de las reivindicaciones 1-20, en el que el método no comprende un proceso para modificar la identidad genética de la línea germinal de un ser humano.
25. Un método de acuerdo con la reivindicación 23 ó 24, en el que el organismo es una planta o alga.
- 30 26. Uso de la composición de acuerdo con una cualquiera de las reivindicaciones 1-20 en la edición de un gen o genoma *ex vivo*, en donde el uso no comprende un proceso para modificar la identidad genética de la línea germinal de un ser humano.

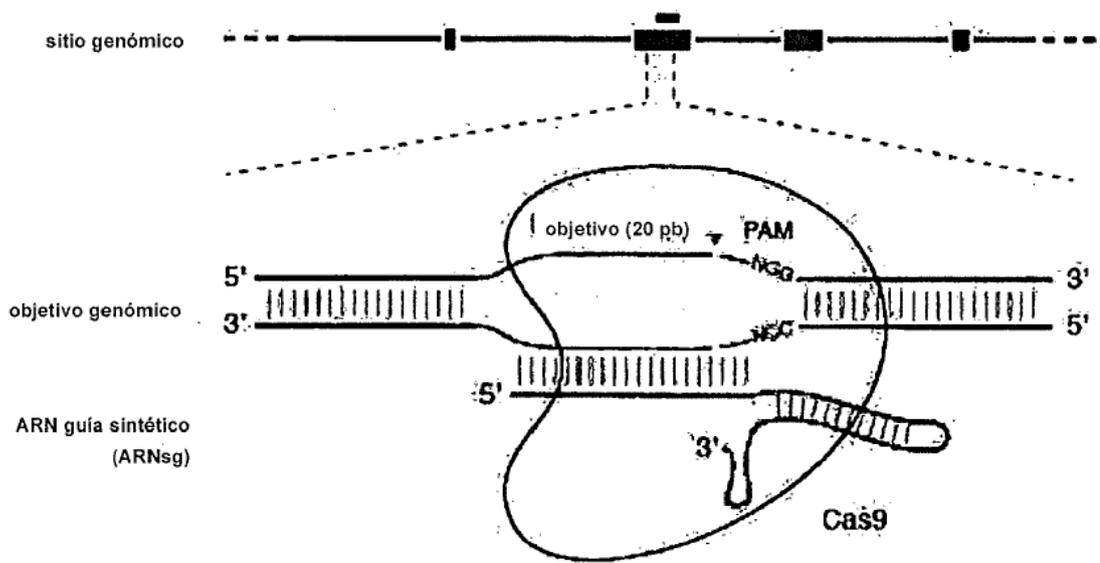


FIG. 1

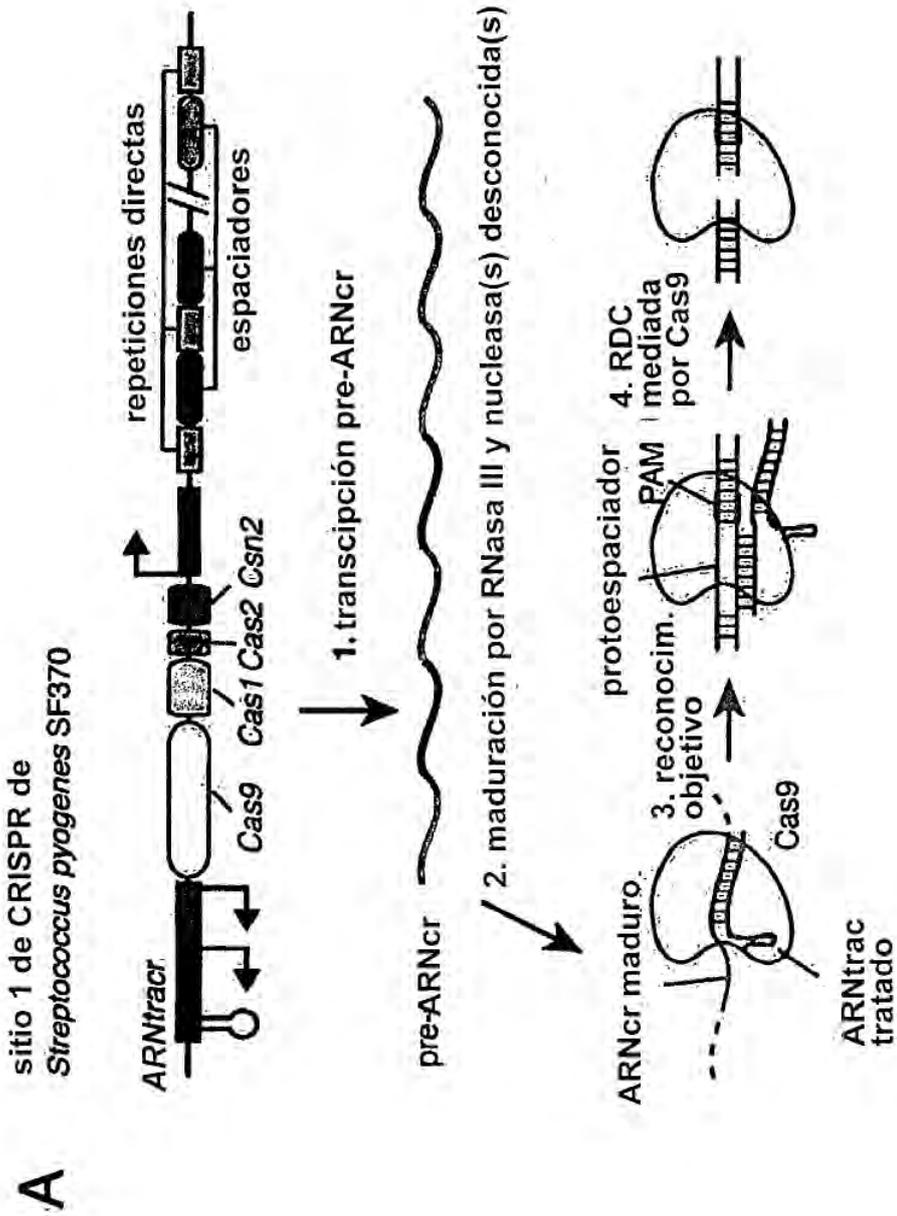


FIG. 2A

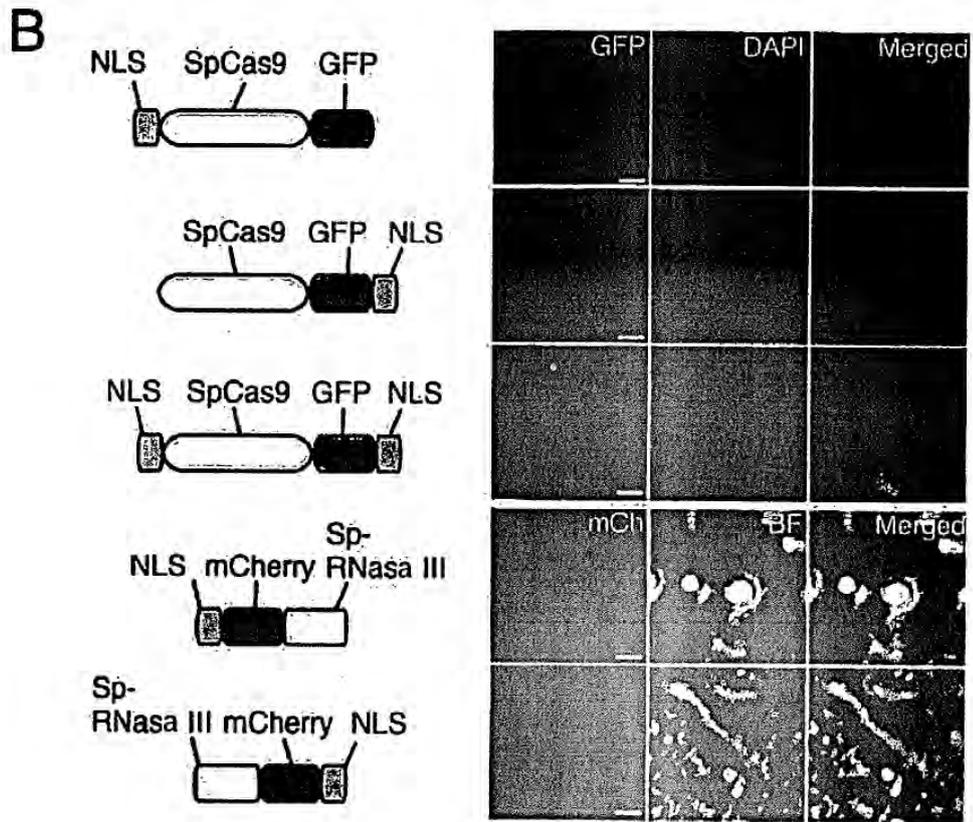


FIG. 2B

C

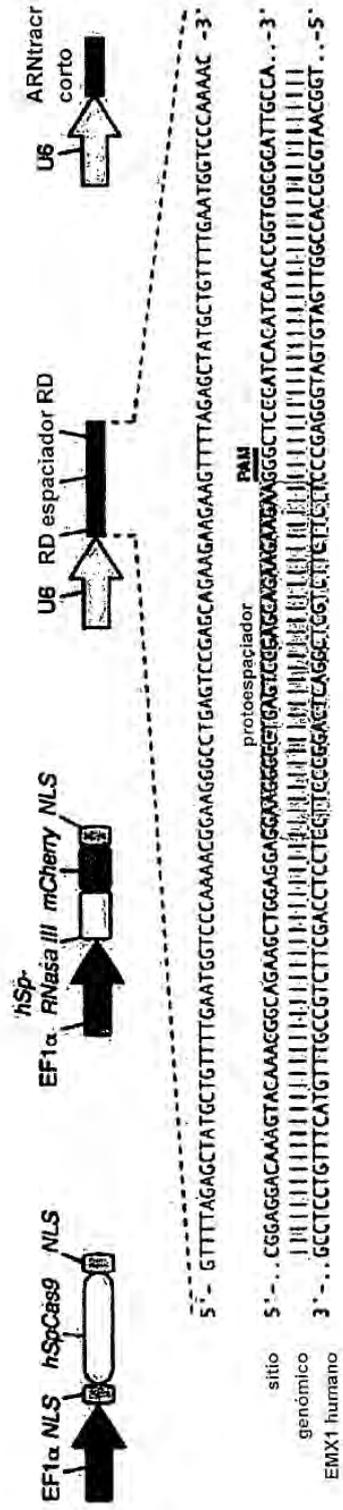
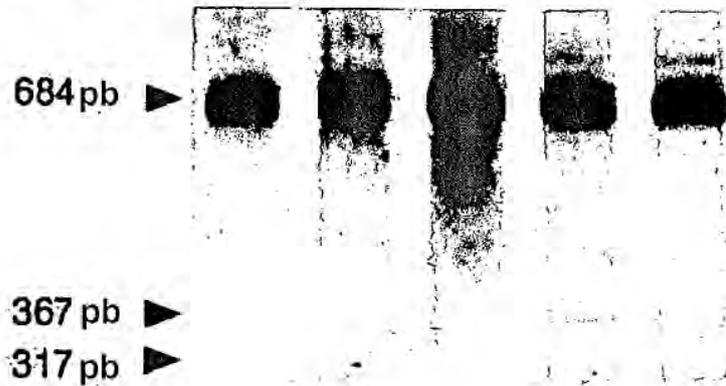


FIG. 2C

D

2xNLS-SpCas9	+	+	+	+	+
SpRNAsa III	-	+	+	-	+
ARNtracr corto	-	+	-	+	+
RD-EMX1-RD	+	-	+	+	+



inserción o supresión (%): 4,7 5,0

FIG. 2D

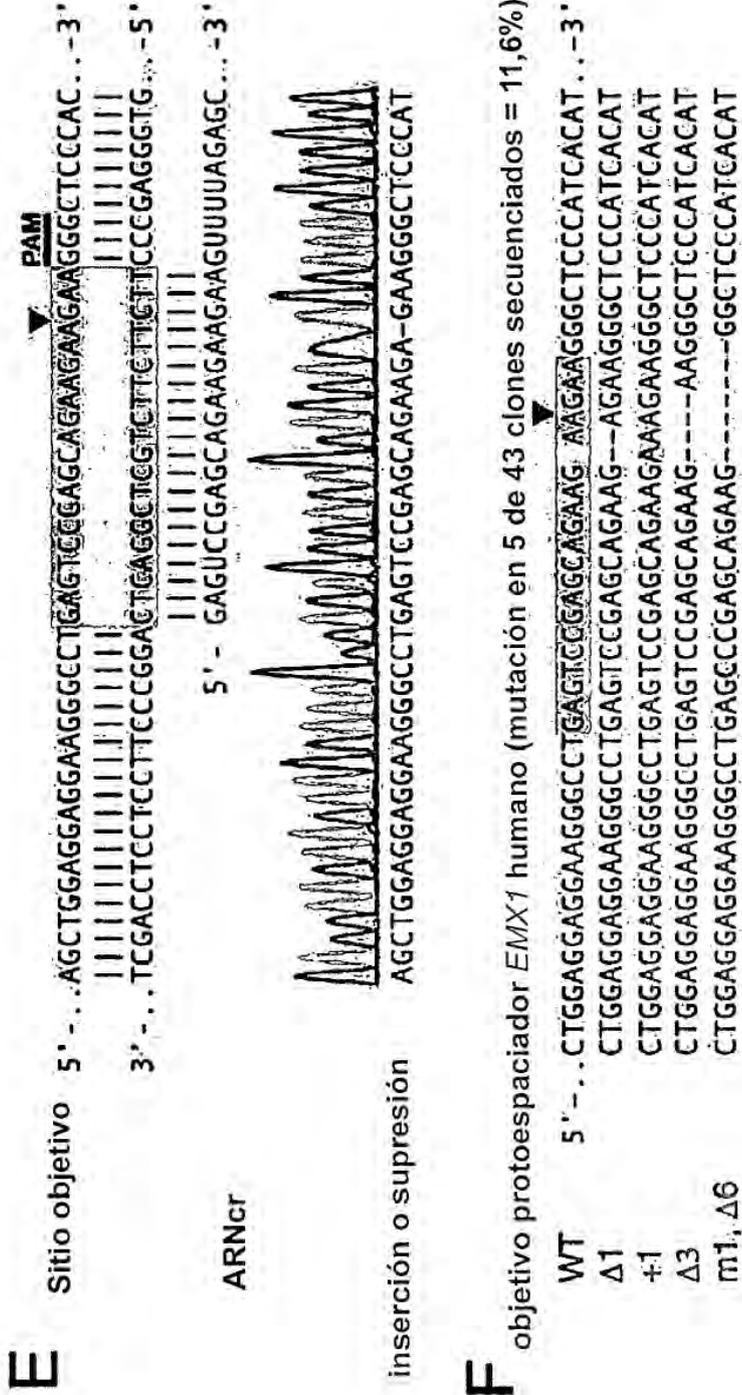
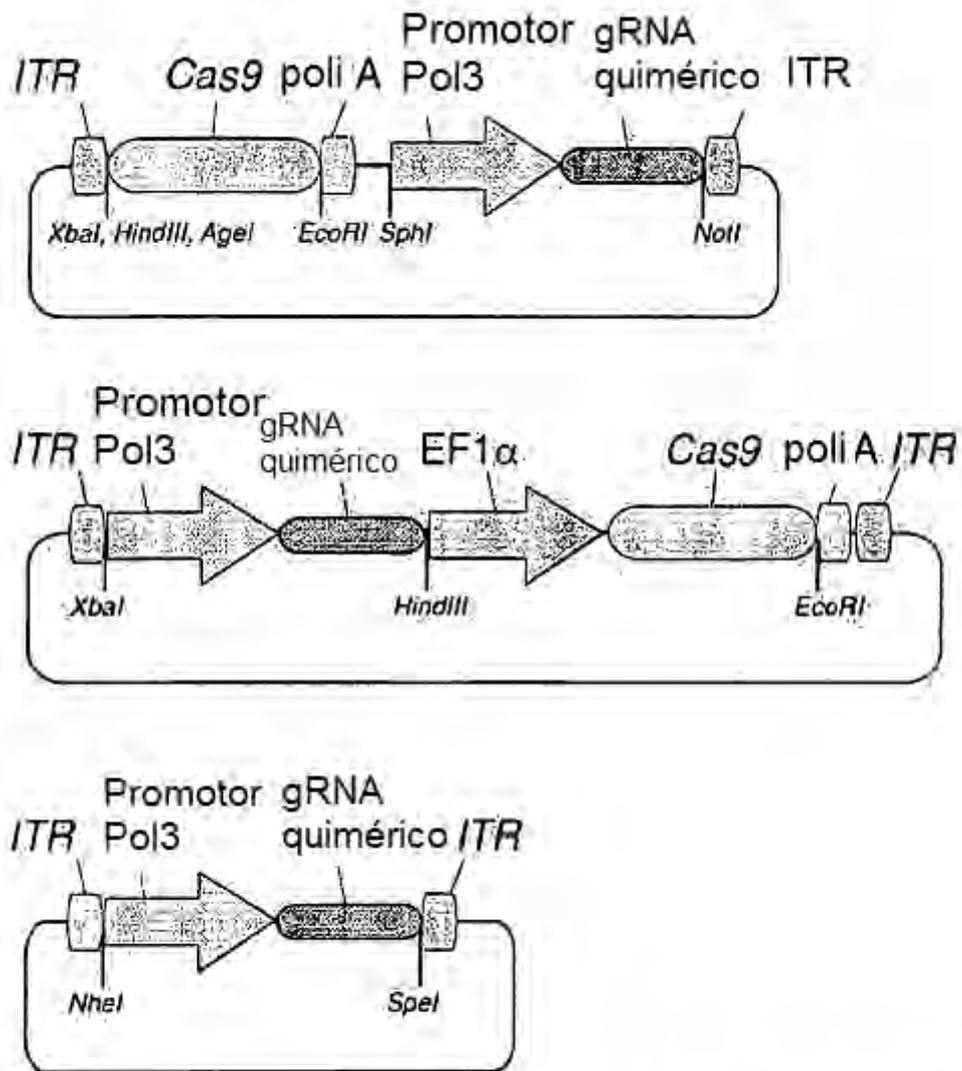


FIG. 2E-F



- dependiendo del tamaño de la construcción completa, puede contener una o más casetes de ARN guía (gRNA quiméricas Pol3)

FIG.3

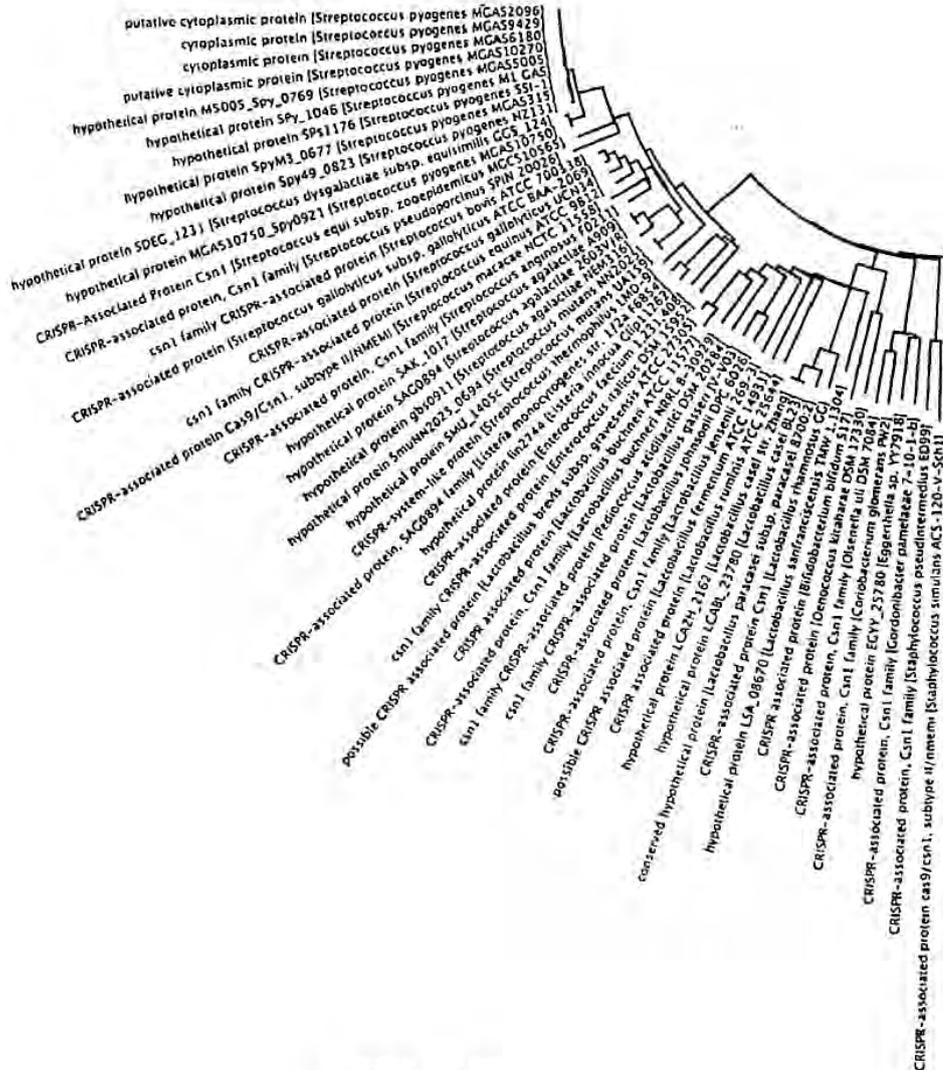


FIG. 4C

Putative cytoplasmic protein = proteína citoplasmática putativa; cytoplasmic protein = proteína citoplasmática; hypothetical protein = proteína hipotética; CRISPR-associated Protein Csn1 = Csn1 Proteína asociada a CRISPR; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated Protein Cas9/Csn1, subtype II = Proteína Cas9/Csn1 asociada a CRISPR, subtipo II; possible CRISPR-associated protein = Proteína asociada a CRISPR posible; conserved hypothetical protein = proteína hipotética conservada;

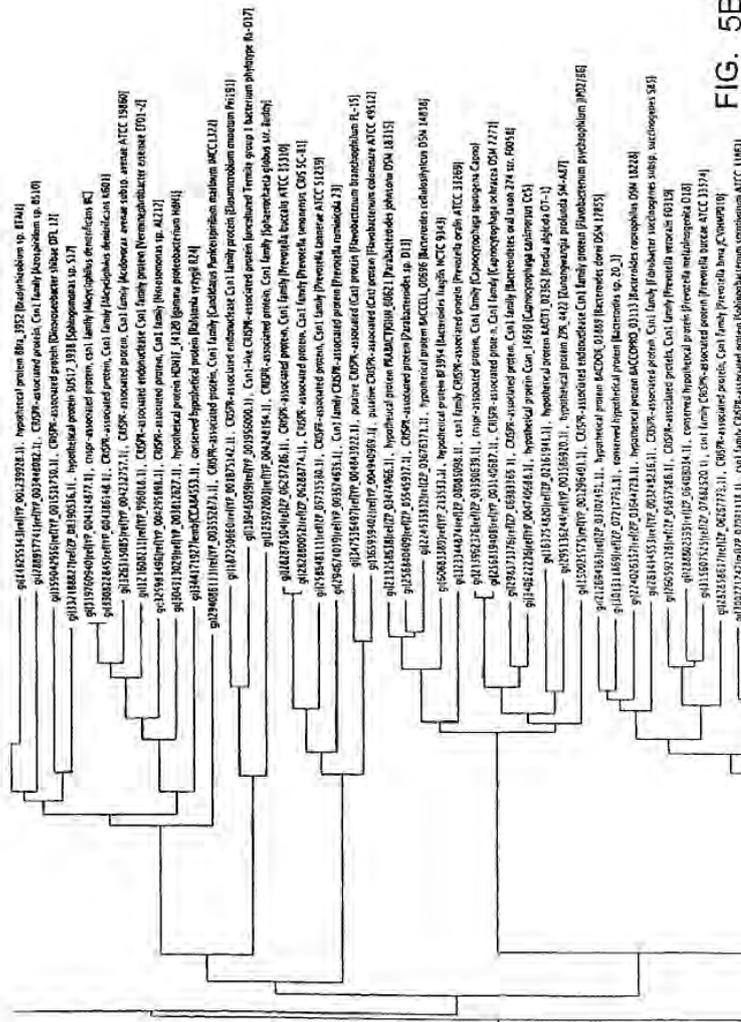


FIG. 5B

Putative CRISPR-associated (Cas) Protein = proteína (Cas) asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1;

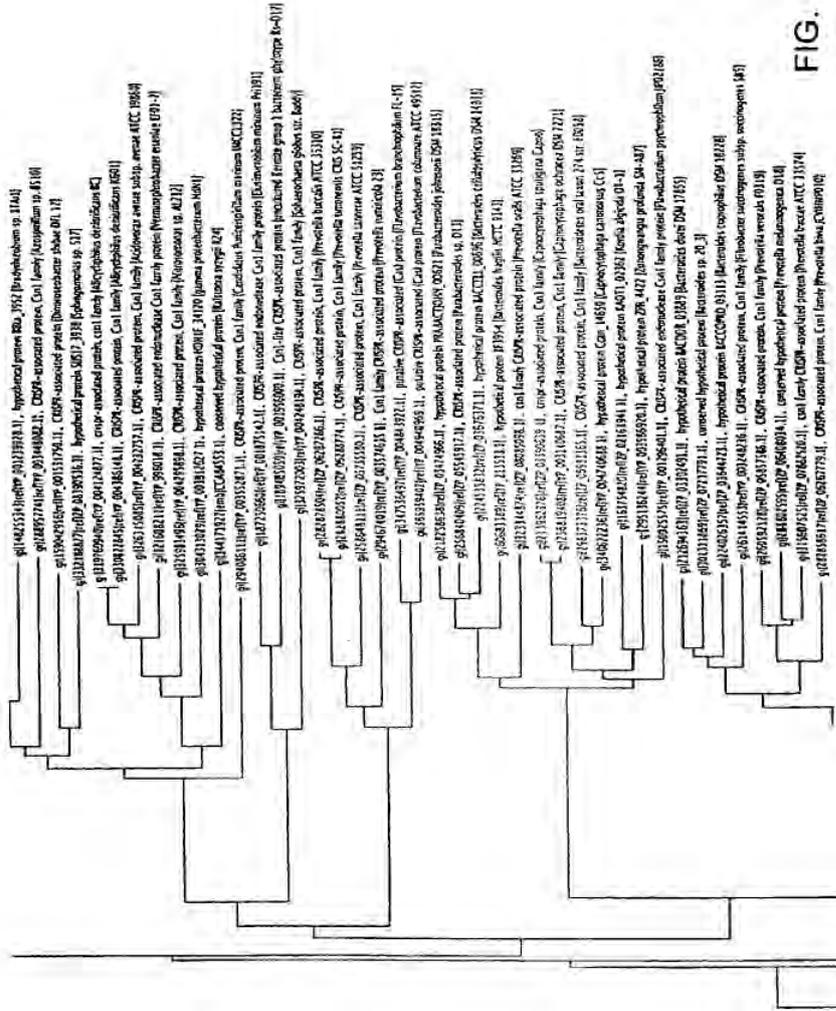


FIG. 5C

Putative CRISPR-associated (Cas) Protein = proteína (Cas) asociada a CRISPR putativa; hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1;

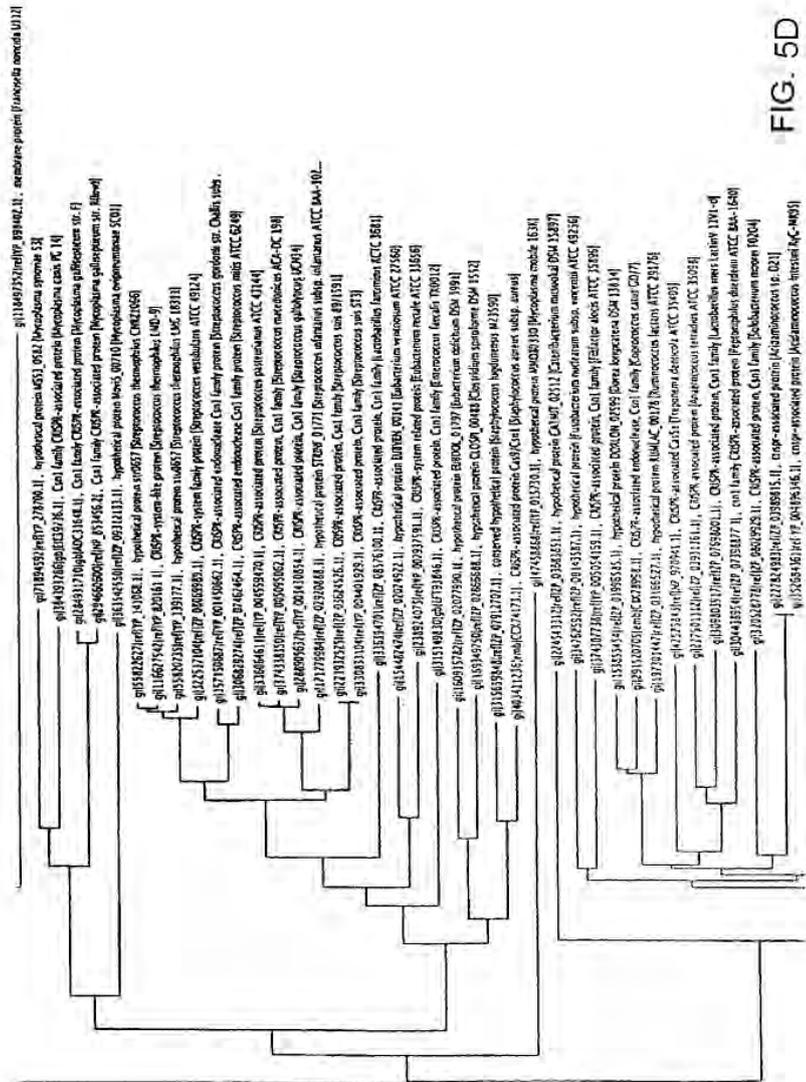


FIG. 5D

Hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated endonuclease Csn1 family protein = proteína de la familia Csn1 de endonucleasa asociada a CRISPR; conserved hypothetical protein = proteína hipotética conservada; Csn1-like CRISPR-associated protein = proteína asociada a CRISPR de tipo Csn1; CRISPR-system-like protein = Proteína de tipo sistema CRISPR; CRISPR-associated protein Cas9/Csn1= proteína Cas9/Csn1 asociada a CRISPR

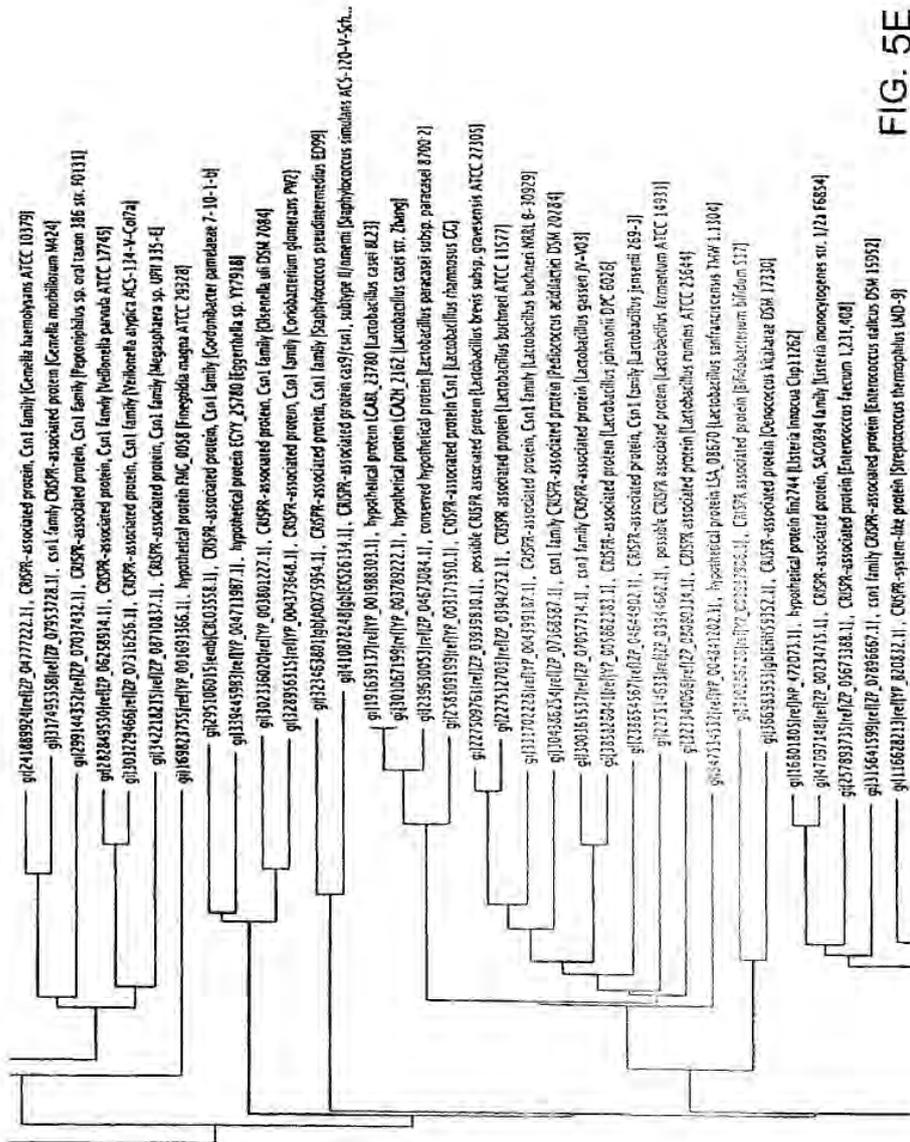


FIG. 5E

Hypothetical protein = proteína hipotética; CRISPR-associated Protein Csn1 family = familia Csn1 Proteína asociada a CRISPR; possible CRISPR-associated protein = Proteína asociada a CRISPR posible; conserved hypothetical protein = proteína hipotética conservada.

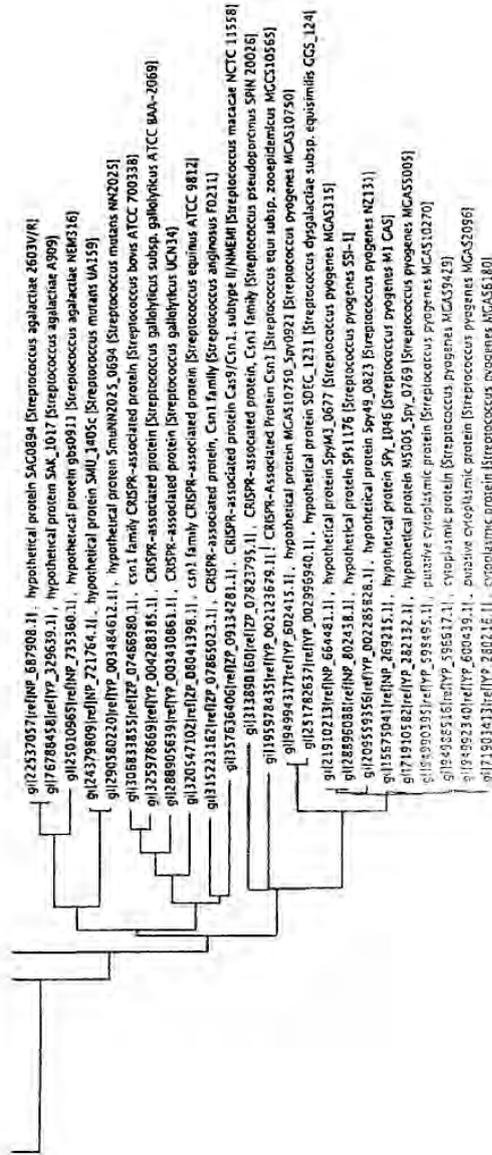


FIG. 5F

Hypothetical protein = proteína hipotética; CRISPR-associated Protein, Csn1 family = Proteína asociada a CRISPR, familia Csn1; CRISPR-associated protein Cas9/Csn1, subtype II = proteína Cas9/Csn1 asociada a CRISPR, subtipo II; Putative cytoplasmic protein = proteína citoplasmática putativa; cytoplasmic protein = proteína citoplasmática

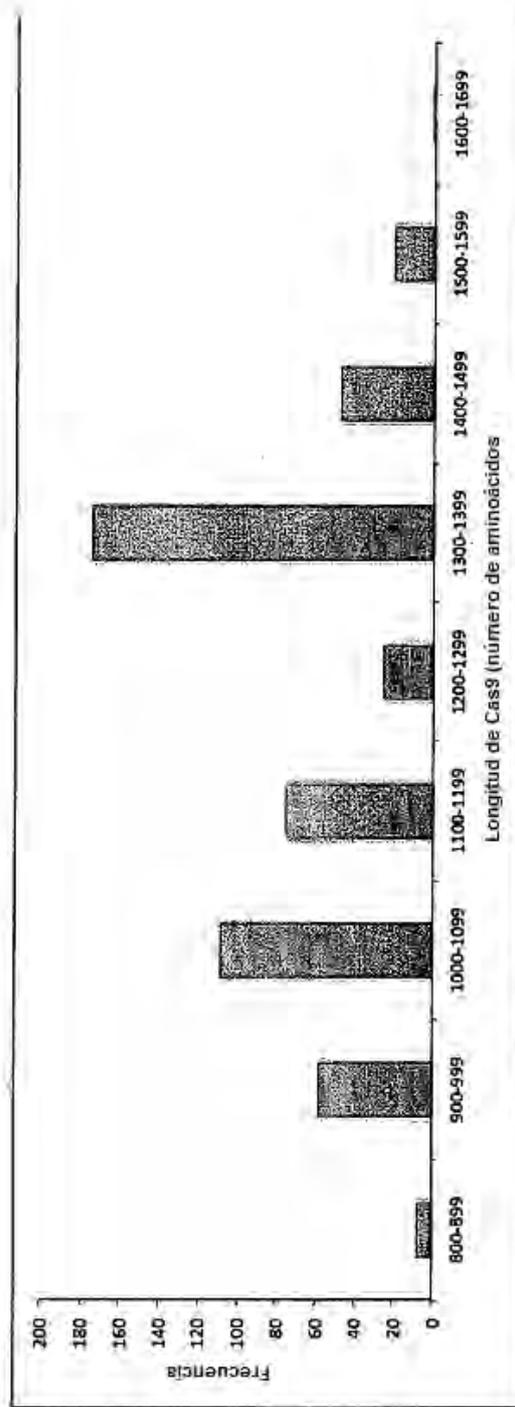


FIG. 6

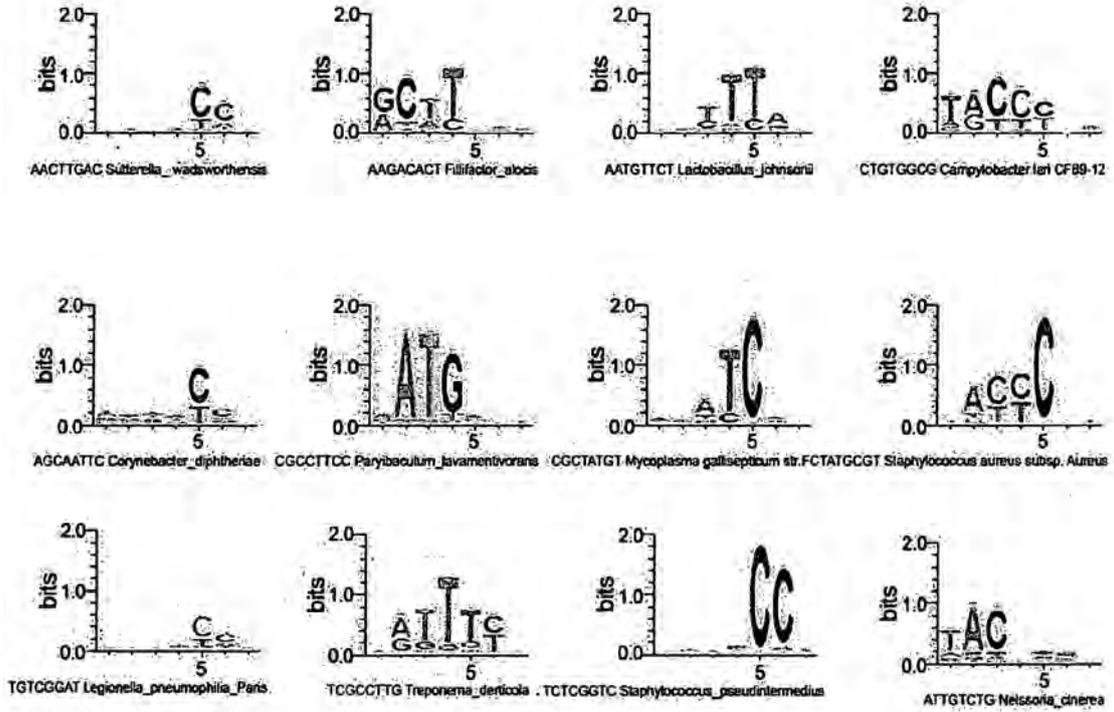
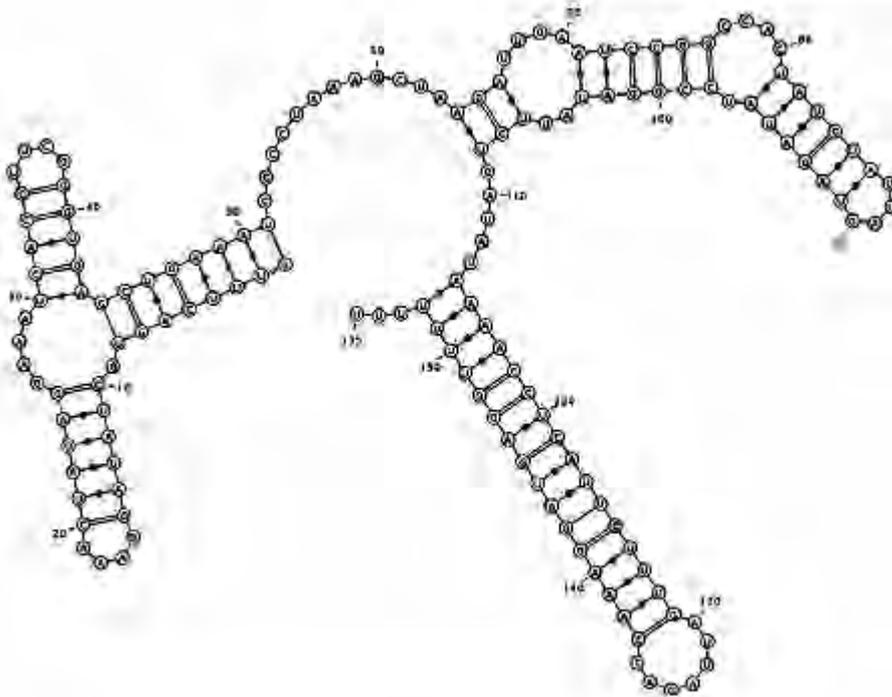
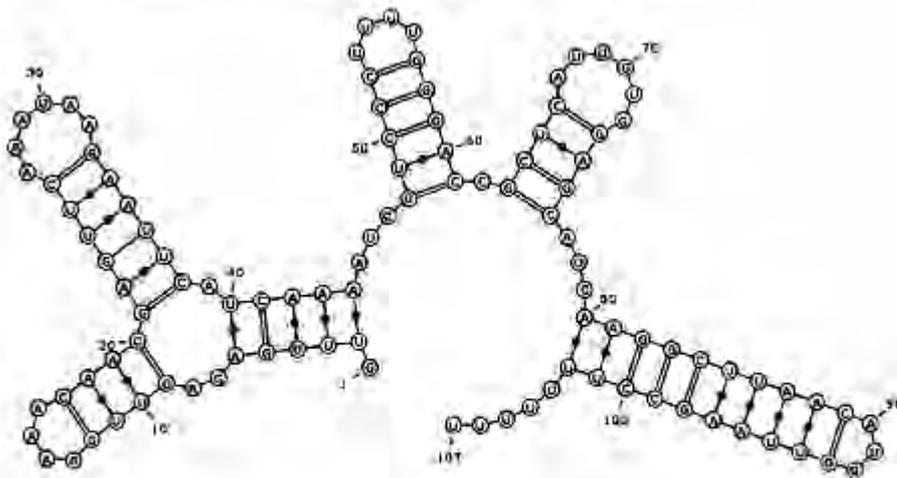


FIG. 7

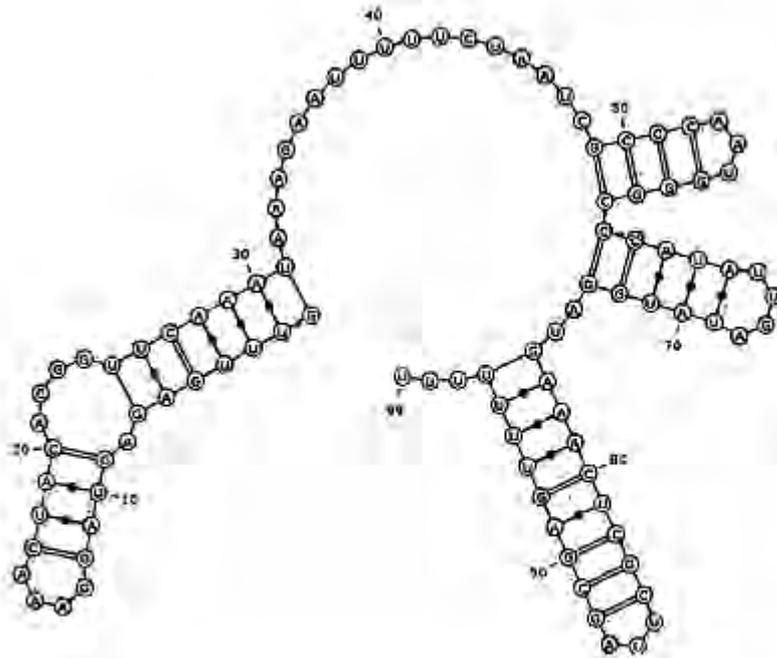


ARN guía quimérico para homólogo 2 de Cas9

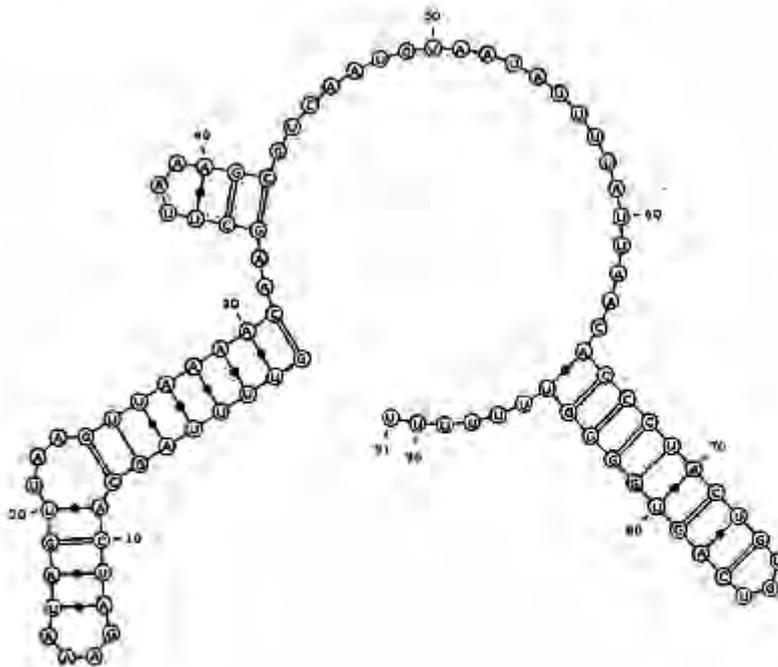


ARN guía quimérico para homólogo 4 de Cas9

FIG. 8A

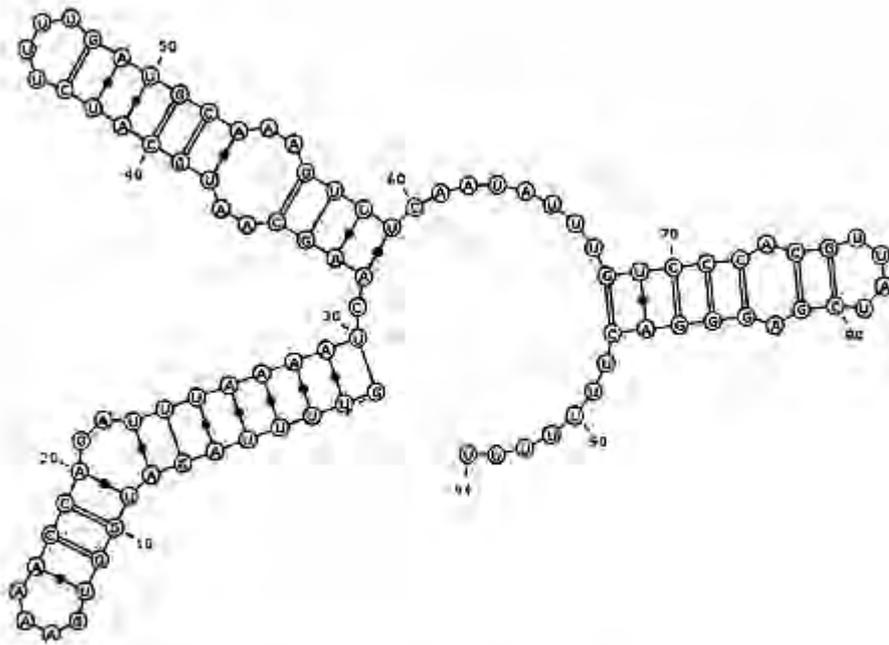


ARN guía quimérico para homólogo 5 de Cas9

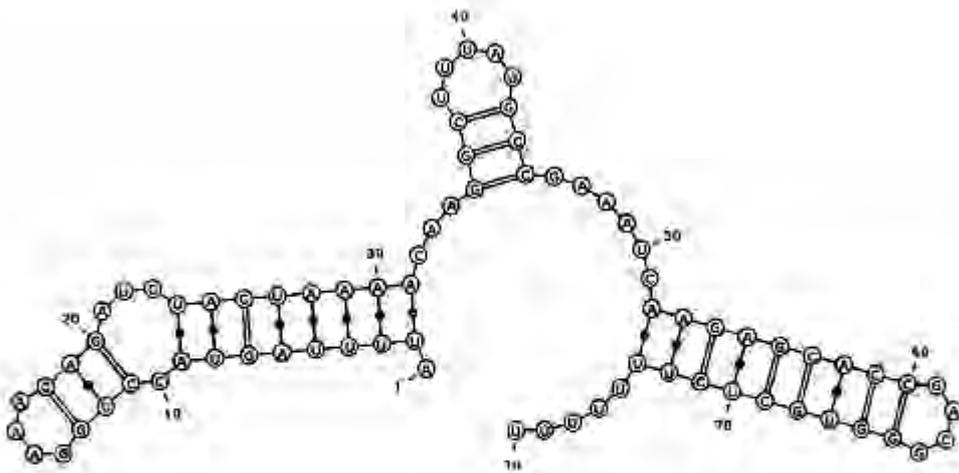


ARN guía quimérico para homólogo 6 de Cas9

FIG. 8B

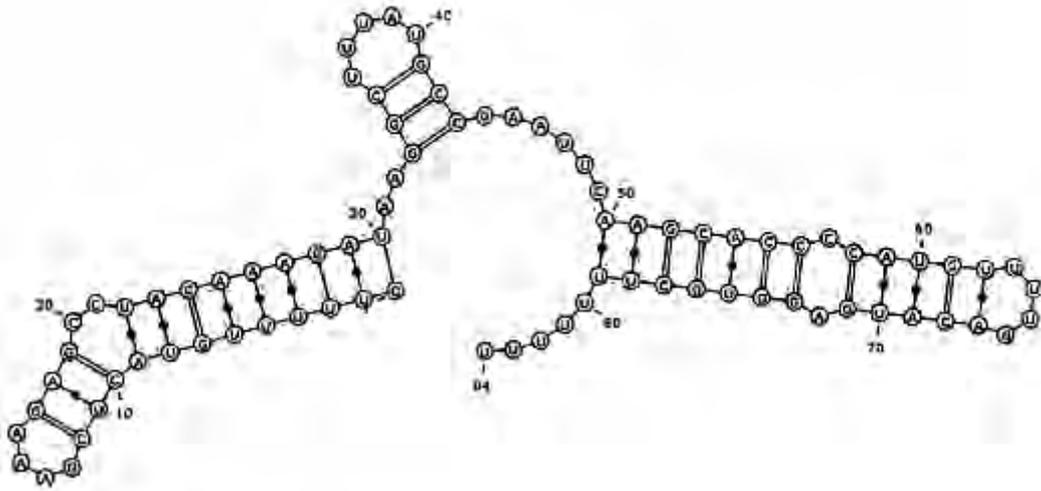


ARN guía quimérica para homólogo 7 de Cas9

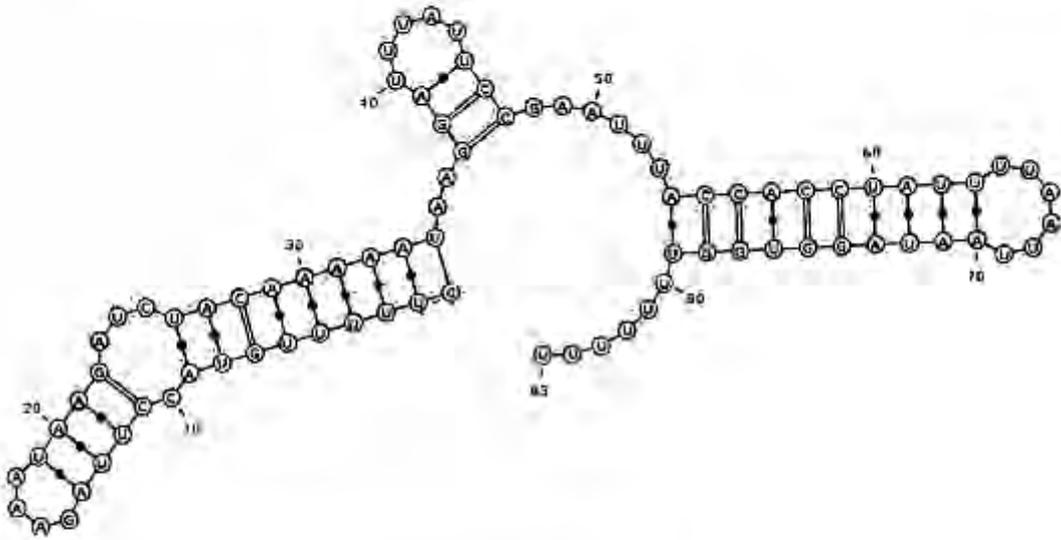


ARN guía quimérica para homólogo 8 de Cas9

FIG. 8C

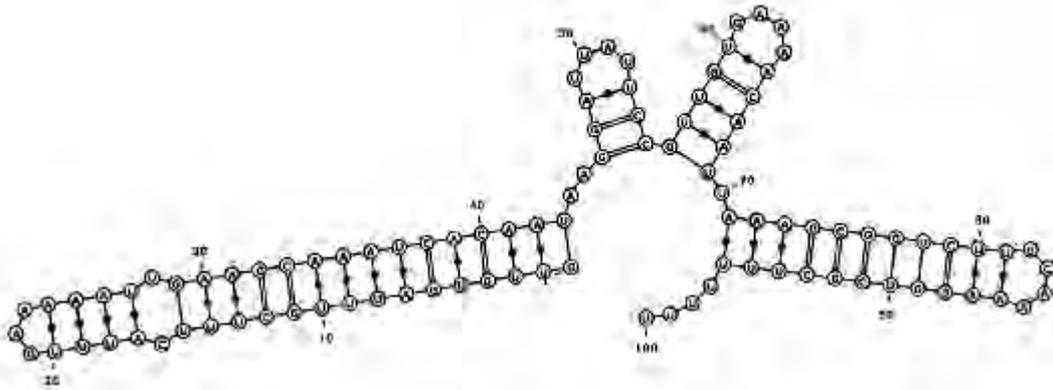


ARN guía quimérico para homólogo 9 de Cas9

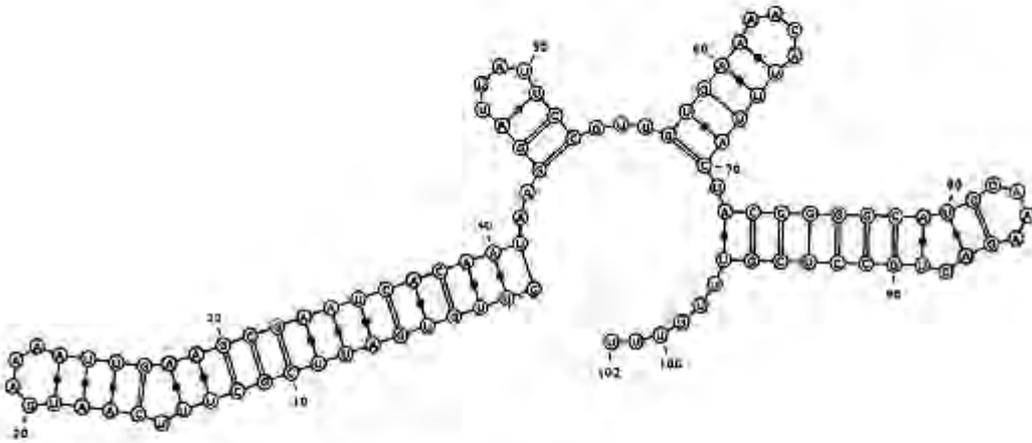


ARN guía quimérico para homólogo 10 de Cas9

FIG. 8D

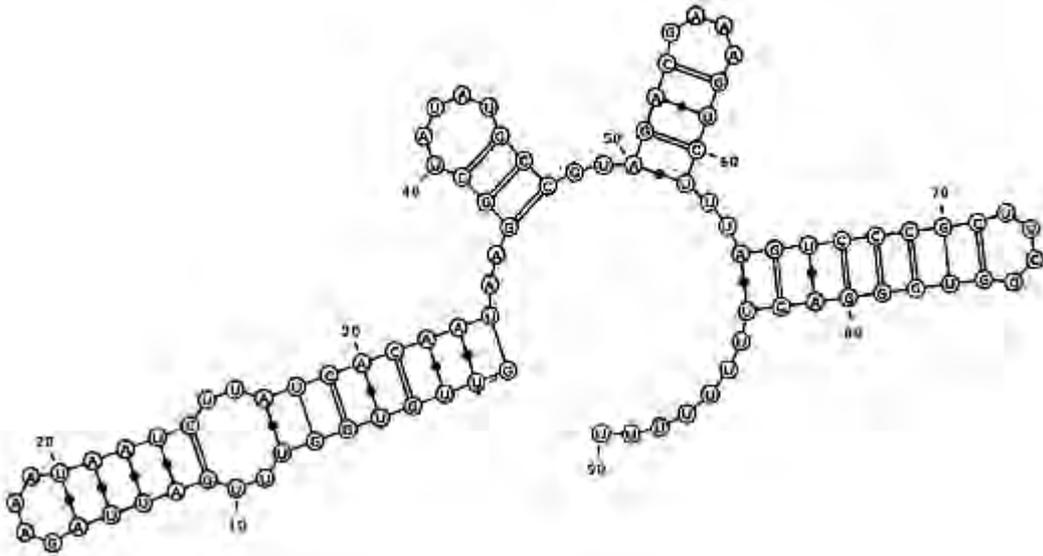


ARN guía quimérico para homólogo 12 de Cas9



ARN guía quimérico para homólogo 13 de Cas9

FIG. 8E



ARN guia quimérico para homólogo 14 de Cas9

FIG. 8F

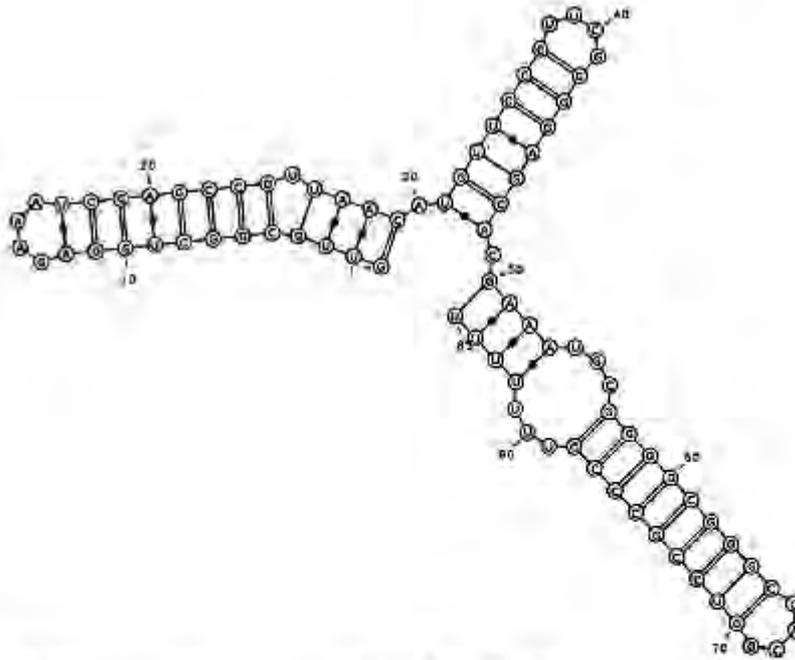
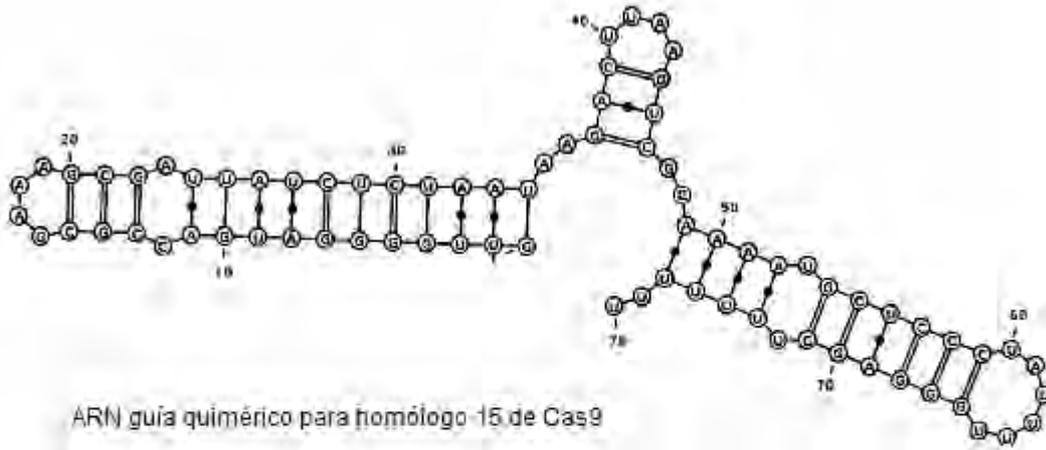
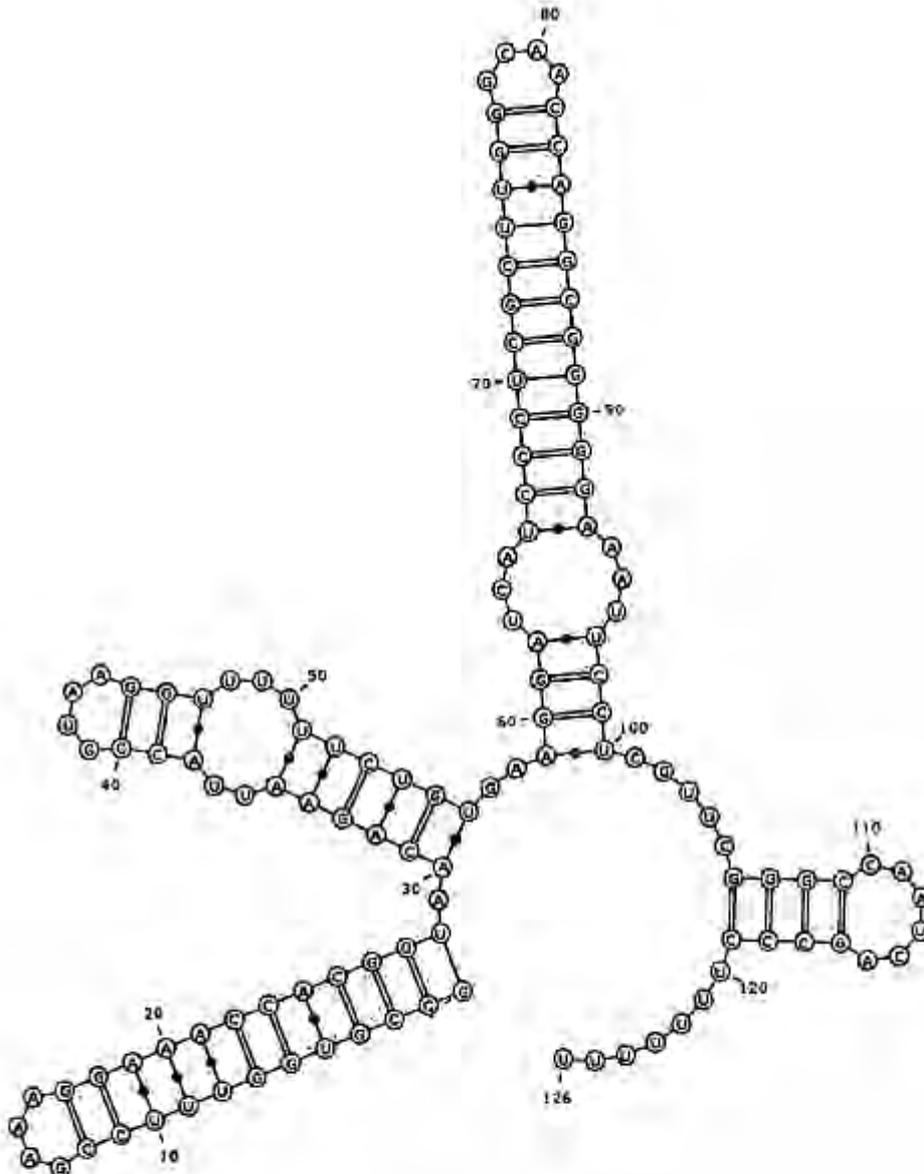
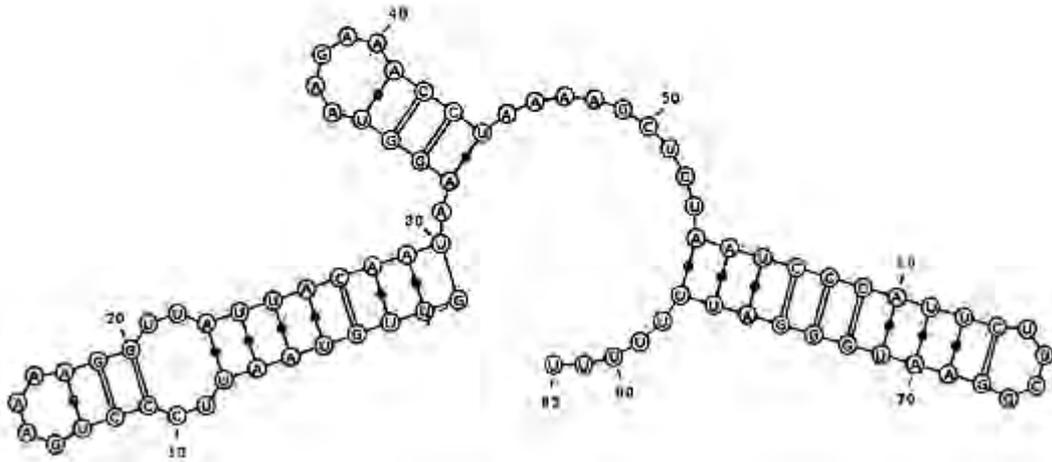


FIG. 8G



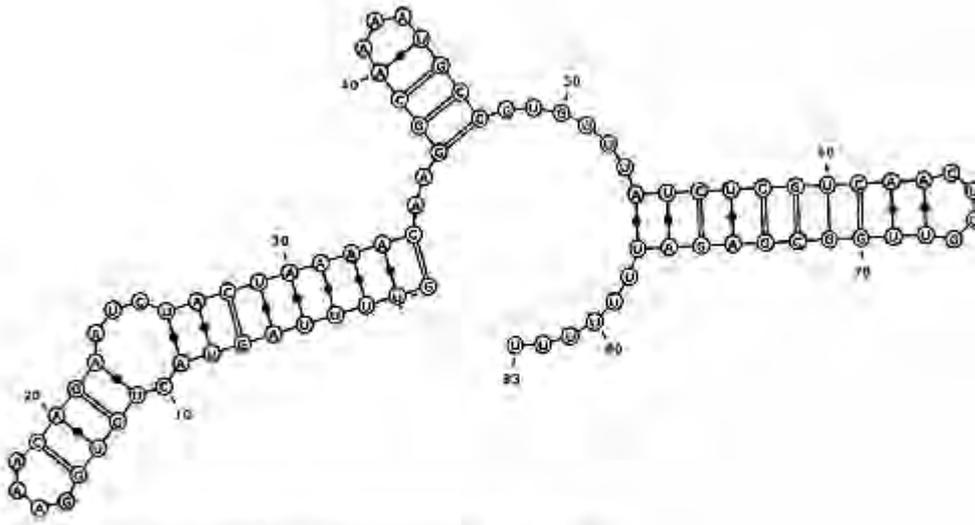
ARN guía quimerico para homólogo 17 de Cas9

FIG. 8H

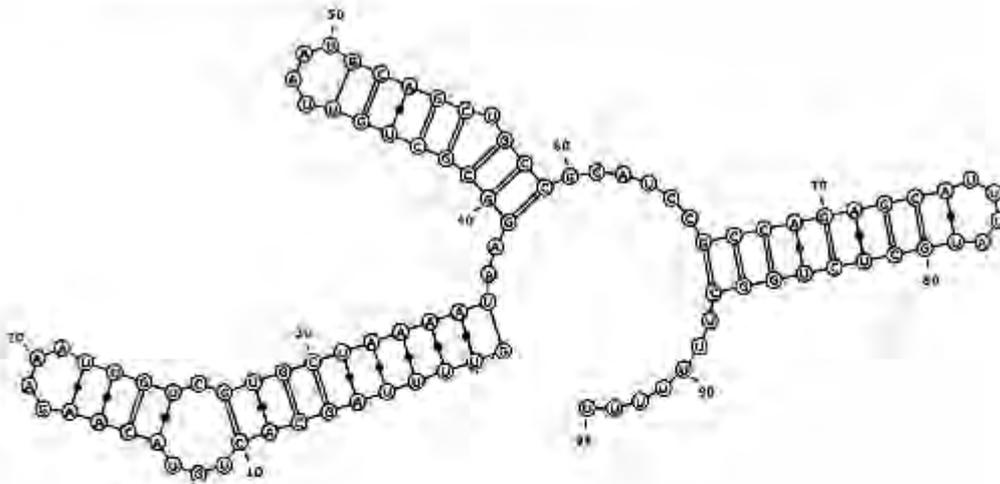


ARN guía quimérico para homólogo 19 de Cas9

FIG. 8I



ARN guía quimérico para homólogo 21 de Cas9



ARN guía quimérico para homólogo 23 de Cas9

FIG. 8J

FIG. 9A

secuencias Cas9 optimizadas en el codon humano

#2

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGACTCAG
 AGCGAGCGACGATTTTCTGACGATTTGGCATTTGACATGGGGGCTAAGTACACTGGGGTGTCTACGCACTGTTCCAGCGGGAG
 GAACCGCCACAAACCTGAACAGCAAGGCCATGACCTGGTCACTGECTGAGACAGGGCCAGATACGTCAGGCGACAGAGAAGT
 GCCGTACAGACAGGGCTGCGCGACAGAGAGATATACCTGGCTAGGAAACTGGCATTCTGGTGGTCGACGATATGATCAAG
 AAAAGGAAAAAGGGCTGACTGATGAGGAATGGAACAGGAGCGGGAGGGCCCTGTCCGGCTGCTGAAGCGGGAGAGGGTACTCT
 CGGCCAACGCTGACGGCGAAGATCTGACCCCTCTGGAGAATGTGAGAGCAGACGTTGTCGGCGCTCATCTCTGCCCTTCAGCA
 YATTTTTCCGAAGTGGCTCTCTGGCTGAGCAGTGGGAGGAGTTACCGCAAAACATCAGCAATGTCGAGAAGTTTCTGGGGCAG
 CCAACATCCCCGCGATAAAGAGTTCATTGAATTTGCCGTGGCTGAAGGGTGTATTGACAAGCCGAGAAGAAAGCCATCCAG
 TCAGCTCTGAGCACCTGAGGGCAACGCCAATGTGCTGACAGGACTGCGGCGAGATGGGCCACAAGCCTAGAFCAGAATTTTT
 AAAGCAATCGAGGCGGACCTGAAGAAAGATAGCCGCTGGCCAGATTAACGAAGCATTGGAGGAGCAGAGCGCTGGCTCGA
 CTGCTGGGAAACCTGTCCAATCTGCAAGTGCAGGCGCAAGAGATGGTACTTCAATGCCCCGACATCATGAAGGATAGGGGCTGG
 GAGCTGATCGCTCAAGAAAACACTGGTGGGGCTTTTAAAGTCTTTTACCCAGCAAAGGACCAGAAACAAAGCAGATCTGGAA
 CTGATCAAAACAGATTGAGAACAGCGAAGATATCAITGAGACTCTGTGCACTGGACCCAAACAGAACCATCCCCCTTACGAG
 GATCAGAAACAATAGGGCGCCACCCCTGGACAGACTCTGCTGCTGAGTCCCGAAAAGCTGACCCGGCAGTATGGCGAGATCTGG
 AAAACATGGAGCGCCAGACTGACCTCCGCTGAACCCACACTGGCACCTGCAAGCGAGATTCTGGAAAGATCTACCGACAGGAA
 AGTCCGGTGGCAGTCAACGGACAGAGCCACTGCTTACACTGGCTTACCAGCTGAGTTATGCACTGCAGAGAGCTTCCGACAGG
 TCAAAAGCCCTGGATCATATGCTCTGAGGGCACTGGCTGACGGCTCAAAAAGCAATAAGCTGACATCCGCCCGCAGTCTCTG
 GAGAACTGCATCGGAGGCCAGAAATGTGAAACCTTCTGACTGTGCCCCGACGGTACTATCGGAAAGCAGACATGGCCAAAGTC
 GGGCTGTGGTTCGACAAAGCCGATGGACTGCTGGAGAGTCTGACTGCACTCCCAATGAAGAAAGAAAGTCTGCCCTGAGACTG
 GTGGCCAAATATCTGCAGACAGATGAAACCAAGGCGAGAAAGTTCTGGACGAGATCTGGCGAAAACAGATTAAGGGGCGGGAA
 ACTGTGGCTAGCCGATGTGACAGGATCGAGACAGTGGGAAATCTTCCGGGGAGGGCTTAAACATGGCTACAAATCCGGCTCAG
 TATAGGAGGTTGAACAAGCTGCCCGCAATGCCAGGATAAAGAACTGCTGACAAATCAGAGATAGGGTGGCTGAGACTGCAGAC
 TTCATTGCCGTAACCTGGGGCTGTCTGACGAGCAGAAAAGAAAGTTCCGCAATCTTTTAGTCTGGCTCAGTTCTACACCCCTG
 ATCGAGACAGAAAGTGTCCGGATTTCTGCAACTACCTGGCCGTCCACCTGGAGAACCGCTGGAGGATGACAATCAAGGATGCT
 GTGATTAATGGGGAACCTGTGAGGAGCAGTGCAGAGGCTGCTGCTGACAGACAGCTCCGCCATTCGATGGACTGGAGAG
 AGGCTGGTGCAGACAGGCTTGGGAGATCGCAAAGAGGGTGTCAACTGACATTCAGAGCAAGTTCGATTTCTCAAACGGCATC
 GTGGAGCTCAGCATTTTGTGGAGGAAAAAAGTTCCGAGTCTGTGGCCGATCTGAAAAGAAACAAACGGGTCAAA
 GACAAGATGCTGTCCAGACCGGAAAGCTGGAAACAGATGGCTGATCAAAAATGAGCGGATCAAGAAAGCCAGCCGGGAACT
 TGTCCCTACACCGGCGATAGGCTGGCTGAGGGGGGAGAAATCGACCACATTCTGCCCGAAGCTGATCAAGGATGCCCGGGGA
 ATTTGTGTTTAAAGCTGAGCCTAATCTGATCTATGCAAGCTCCCGCGCAACAGCTGAAAAAGAAATCAGCGATACAGTCTGTC
 GATCTGAGGCCAATCTCGAATGAGATCTTCAAACTAGCAACATGCTGCAATACCGCGGAGATTTGAGGACGTTGGTCACT
 AAGCTGCAGCAGACCCATAGACTGAAAATCTTGAATCTGCTGAATGAGCAGCAACAGGACTGCGTGGGCGACGCCCTGTCTG
 GACGATGGCAGCGAAGCTGCGGACGCACTGCTGGAGCTGCTGGCAACACAGCGCCGAACTCGCGTCAACGGGACACAGATCTGG
 ATGATTAAGAACTGGCCAAACAGATCCGAGAGGAACTGCAGAAATGGTGAAGCAACTAACAAATGACTGCACTTTCCAGCC
 GTGCAACTAACGCTGTCGATGCAAGAAATCTGAGGCTGAAACTGGCCAGAAACAGCCGACTTCGAGAAGCCAGATATCCAG
 CCCATTTGCCAGCCATTCCATCGACGCCCTGTGCTCTTTCGCTGTGGGGAGTGTGACGACAGAACCGGATCAGAAATGGATTTGAC
 TACCTGATGGCAAGACCTGTCTGGGACTGTATCCACAGAGCTGTGAGTCACTCACCTGACGGCCAGGCCCAGGAGGAAAA
 AGTCAITTCGATTCAGTGGCTATCTTTAAGGAAGGCATCTACGAGAGCAGTTCTGCTATCTTTACCTGAACGAAAAGATC
 TGGATTTGGATATGAGACACTGAATGCCAAAAGGCGAAAGATGGGGGGCTATTGAGGTGAGTGGCAACAGCCAAAGGAGTGTCTG
 GAAATGCTGGCCCTTCTTTAAACAGCTGTGGGGCAGCTGTGAGCCACGCTACTTACCGGATCTGAAAAAGCCTGCATAT
 GAGTTCTGGCAAAAGGACGCTCTGACGCCACTGAGCGCAGAGGAAAAAGACTGGCAGCCCTGCTGGATGCTCTGGCTACTGT
 ACCAGTCGAAAGTCACTGATGAGCCTGTTTCAATGGCTGCAAAAGGGAAATCCCTGAAAAAGCGGGAGGAGCTGTGAAACCCAAG
 CTGTTCCAGCTGAAGGTGAGCTGAAAGGGGAAAGAGCTTCAAGCTGAACGGGAGCCTGACCTGCTGTGAAAGCAGGACTGG
 CTGAGAATCTGCGATAGCCAGAACTGGCAGACGCTTTTGGCAACCTGTTCCGCGGATGAGCTGACATCTAAGCTGGCTGCG
 ATTTGGAAACGACTGTGATGCGGGATCTGGCTCATGCAACAGTCCGGAGAGATTCAGCCTGCCCGCAATCGACAACCCAAAGT
 GGAGGGTTCAAGATTAGGGCGACCAACCTGTTTGGCAATGAGCTGTACAGGTGACGCCATCAACGCTAAAAAGTATCGCGCT
 TTCGCTCCGCTGGTCTAATGTGACTGGTCCAAGGGGATCTGTTTAAACGAGCTGACGATGAAAAATCTGACCGAGTGGGA
 GGCAGGTTCAATCAAGCGCCGATGTGACTCTATGTCCGAATGGCGCAAGGTGGTGCAGAGGACAACCTGTCTATCTGGATT
 GCTCCAGGGACAGAAAGGACGAGCTGAGGGTCTGAGACAACATTCATCCAGGCCAGTCACTGGTTTGAACAGTCACTGGAG
 AATTGGGCCATTAAGTCTCTGTCACTGCCAGCTTCTTCAAGGTGGACAAACAGCTGAGTTTCAGAAGGCACTCGGAACC
 GAGCTGTGAGAACTGCTGGCCAGCCAGGAGCGAAATCTTCAITGAGAACGTGGGCAATGCCAAGCATATCCGCTTTTGGTAC
 ATTTGGTGGTGGCAGCAACAAAAAGATGAACGAGTCTTACAACAATGTGTCTAAGAGTTAAGAAATTC

#4

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAAAAA
 GAAATCAAAGACTACTTCTGGGGCTGGATGTGGGACTGGGAGCGTGGGGTGGCTGTGACCGTACTACTACAACTGCTG
 AAGGCTAACCGGAAAGACTGTGGGCAATGAGATGCTTCGAGACAGCCGAAACTGCTGAGGTGCGGAGACTGCAAGGGGAGCC
 AGGCGCCGAATCGAGCGGAGAAAAGAAACGCAATTAAGCTGCTGCAAGGAGCTGTTCTCTCAGGAAATCGCCAAACCGATGAGGGC
 TTCTTTAGAGAAATGAAGGAAAGCCCTTTTACGCTGAGGACAAAAAATCTTGCAGGAAAAACTCTGTTCAATGACAAGGAT
 TTTGCTGATAAGACTTACCACAAAGCATATCTACCAATTAATCATCTGATCAAGGCTTGGATTGAGAAACAGGTGAAACCGAGC
 CCCCGACTGCTGTACTGGCATGTCAACAATCATTAAGAAAAGGGGACATTTCTGTTTGAAGCGGACTTCGATTCAGAGAAAT
 CAGTTTGATACCAGCATCCAGGCACTGTTTCGAGTATCTGCGGAGGACATGAGGTTGGACATCGATGCCGACAGCCAGAGGTC

FIG. 9B

AAAGAGATTCTGAAGGATAGCTCCCTGAAGAAGCTGAAAAACAGAGTCGGCTGAATAAGATCCTGGGGCTGAAGCCTTCCGAC
 AACAGAGAAAGCCATCAAAAACCTGATTTCTGGAAAACAGATCAATTTCCGCCGATCTGTACGCAATCCAGATCTGAAGGAC
 GCTGAGAAAAACTCAATCAGCTTCTCCAAGGACGATTTTGTAGCAGTGAAGTCAATCGGGGACGAGCAGTACCTGAGCTTCCGCC
 GAACTGCTGCTGAAGGCCAAAGCTGTCTATAACTGCTCTGTGCTGAGTAAGGTCAATCGGGGACGAGCAGTACCTGAGCTTCCGCC
 AAGGTGAAAATCTACGAAAAGCACAAAACCGATCTGACAAAAGCTGAAAAACGTGATCAAGAAAACATTTCCCAAGGACTACAAG
 AAGGTCTTTGGATACAACAAGAACGAGAAAAACAACAATTAATCTCCGGCTATGTGGGAGTCTGTAAGCAAGAGTAAGAAA
 CTGATCATTAAACAACCTAGTCAACCCAGGAAGATTTCTACAAGTTTCTGAAAACTATCCTGTGAGCCAAAGAGCGAGATCAAGGAA
 GTGAATGACATCCTGACCCAGATTGAAAACCTGGCACCTTTCTGCCAAAGCAGATCTCTAAAAGTAACGAGAGATCCCTATCAG
 CTGAGGAAAAATGGAGCTGGAAAAGATCCTGTCCAATGCCGAAAAGCACCTTCTTTTCTGAAAGCAGAAAAGCAGAAAAAGACTG
 TCACATAGCGAGAAGATCATTATGCTGCTGACCTTCAAGATCCCTTACTATATTGGCCCAATCAACGATAATCAACAAGAAATTC
 TTTCCCGACAGATGCTGGGTGGTCAAGAAAGAGAAAATCCCTTCTGGCAAGACCACCCATGGAATCTTTTGTATCATATCGAC
 AAGGAAAAAACAGCAGAGGCTTCACTTCTTAGGACCAATTTTTCGACTTACCTGGTGGGAGAGAGCGTCCCTGCCTAAGTCT
 AGTCTGCTGACTCCGAATATACCGTGTGAACGAGATCAACAATCTGCAGATCATTATCGATGGCAAGAATATTTGTGACATC
 AAGCTGAAACAGAAAGTCTACGAGGACCTGTTCAAGAAAGTACAAGAAAATTAACCCAGAAAGCAGATCAGCACCTTCATCAAGCAC
 GAAAGCATCTGCAACAAAACCGATGAAGTGTATCCTCGGGATTGACAAGGAATGTACATCAAGCCTGAAAAGCTACATCGAG
 CTGAAAAACATTTTCCGCAAGCAGGTGGATGAGATCTCCACTAAGAATATGCTGGAGGAAATTAATCAGATGGGCTACCATCTAC
 ACGAGGGGGAAAGGAAAGACCATCTGAAAAACAAGATCAAGGCAAGAAATCGGAAAGTATTGCTCCGACGACGAGATTAAGAAA
 ATCCTGAACTGAAAGTTCTCCGGCTGGGGGCGACTGCTCGGAAATTTCTGGAGACAGTACTAGTAAAATGCCCCTTCTCA
 GAACCTGTCAATATTATCACCGCCATGAGGGAGACACAGAAATCTGATGGAGCTGCTGCTCCTGAGTTCACTCCCGGACCGGAG
 AACATTAAGAAAAATCAATTTCTGGATTGAAAGATGCCGAGAAGCAGTTAGTTACGACGGCTGGTGAACCACTGTTTCTGAGT
 CCCTCAGTCAAGAAAATGCTGTGGCAGACCTGAAGCTGGTGAAGAGATTAGCCATATCACACAGGCCCCCCCTAAGAAAATTT
 TTCATCGAAAATGGCAAAGGGGGCCGAGCTGGAACCTGCTCGGACTAAGACCAGACTGAAAAATCTGACAGGATCTGTATAACAAT
 TGTAAAGAACGATGCTGACGCTTCAAGTCAAGAGTCAAAAGCCTGAGCGGAAAGATTGAGAACGAAGATAATCTGAGGCTGCGC
 TCCGACAAGCTGACTGTACTATACTCAGTGGGGAAATGCATGTATTGTGAAAGCCAAATGAGATCGGCCACGCTGTTCGAT
 ACCTCAAACTACGATTTGACCATATCTATCCCAAGGCAAGATCAAAAGCAGATAGCATTTCGAATCGGGTCTGGTCTGCAGC
 TCCTGTAACAAGAACAAGGAGGACAAGTACCCACTGAAATCAGAGATCCAGAGCAAGCAGCGGGCTTCTGGAATTTCTGCAG
 CGAAAAATTTCAATTTCTCGGAGAAGCTGAATAGACTGACAAGGGCCACTCAATCAGTACGATGAGACAGCCAAAGTTTAT
 GCTAGGACGCTGGTGAAGAACTCGCCAGGTACCAAGGTGGCCGCTAAAGTCTGGAAGAAAGTGTTCGGAACAAGAAATCTGTG
 TACAGCAAGGCCGAGACTGTCTCCATGTTCCGGAACAAGTTTGTATATCGTGAAGTGCAGAGAAAATTAACGATTTTCAACATGCT
 CACGACGCATACCTGAATATCGTGGTCCGCAACGTGTATAATACCAAGTTCACAAAACAATCCTTGGAACTTTATCAAGGAGAAA
 AGAGATAATCCAAAGATTGCTGACACCTACAAC TACTATAAGGTGTTGATTATGACGCTCAAAGGAAACAATATCACAGCATGG
 GAGAAAGGGAAAAATTTATCACCGTGAAGACATGCTGAAGAGAAAACACACCAATCTACACTAGGCAGGCAAGCTGTAAAGAAA
 GGGGAGCTGTTCAATCAGACCTTATGAAGAAAGGACTGGGCCAGCACCCCTGAAGAAAGAAAGGACCTTTTCAATATCTCT
 AAATACGGCGGTATAACAAGGTGAGCGCTGCA TACTATACTGATTGAGTATGAGGAAAAGGGCAACAAAATCCGCTCCCTG
 GAAACTATTCCCTCTCAGCGAGAGAAGATCCGGAAGAACTAGCCACTGAGGACGCTGAAAGTCTTATCTGACAGACTGTGGGGAAG
 AAAGAGTTCAAGATCCTGGTGCCCAAGATCAAGATCAACAGCCTGCTGAAGATCAATGGGTTTCTTGCCATATTACAGGAAAA
 ACTAACGATAGTTTCTGCTGCGCCCTGCGGTGAGTTTGTGTTCAAACAATGAGGTCCTGACTTCAAGAAAATTTATCCGG
 TTTTCCGAAATCCGCTCTCAGCGAGAGAAGATCCGGAAGAACTAGCCATAACGAGGACCTGAGCTTCCGGTCATATATCAAG
 GAGAACCTGTGGAAGAAAACTAAGAACGATGAAATCGGAGAGAAGGAAATTTTACGACCTGCTGCAGAAGAAAAACCTGGAGATC
 TATGATATGCTGCTGACTAAGCAACAAGACACCATCTACAAGAAACGCCCTAATTTCTGCCACCTTGTATCTCTGGTGAAGGGG
 AAAAGGAAAGTTCAAAGCCTGATTAATCGAAAACAGTTTGAAGTGTACTGAGATCCTGAGATCCTGAAGCTGTTTCTGCAACACGGAAT
 GTCAGTGACCTGCAGCATATCGGAGGACGCAAGTACTCCGGCTGGCCAAAATCGGGAACAAGATCTCTAGTCTGGATAACTGT
 ATCCTGATCTATCAGTCCATCACCGGCATCTTCGAGAAAACGGATCGACCTGCTGAAGGTGAAGAAATTC

#5

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCCGCCAAGAAAAAGCGCAAGGTGCAAGCGTCCATGACCAAG
 GAGTATTACCTGGGGCTGGATGTGGGGACCAATCCCGTGGGATGGGCAAGTCTCAGTACAACCTGTGCAAGTTTAAAG
 AAAAAAGGATATGTGGGCATCCGCTGTTTCGAAAAGCCCAACAGCAAGGACCGGAGACTGCAGAGAGGGAATAGGCGCCGA
 CTGGAGCGGAAAAAGCAGAGAAATTGATCTGCTGCAGGAAATCTTCTCCCAAGATCTGCAAGATTGACCCCACTTTCTTTATC
 CGACTGAACGAATCCCGGCTGCACCTGGAGGACAAGTCTAACGATTTCAAATACCCACTGTTTATTGAGAAAGGACTATTCTGAT
 ATCGAGTACTATAAAGAGTTCCCACTTTTTCACCTGAGGAAGCATGATCGAGAGTGAGGAAAAACAGGATATCCGGCTG
 ATCTACCTGGCCCTGCACAACATCATTAAAGACCCGAGGACATTTTCTGATTGACGGCGATCTGCAGAGCGCCAAGCAGCTGAGG
 CCCATCCTGGATACATTTCTGCTGTCCCTGCAGGAGGAAACAGAACTGTGAGTGAAGCTGTCGAAAAATCAGAAGGACGAGTAT
 GAGGAAATTTGAAAAACCGCAGCATCGCAAGTCCGAAAAGTGAAGAAAGCTGAAGAAATCTGTTTGTAGATCTCAGACGAGCTG
 GAAAAAGAGGAGAAGGAGCCAGAGCGCGTGTGAGAACTTCTGCAAGTTTATCGTGGGAAAATAAGGCGAGTCTGTGAAA
 TTCTGCGGGTGTCTAAGGAGGAACTGGAGATTGACTCTTTCAGTTTTTAGAGGGCAAGTACGAGGACGACATCGTGAAGAAC
 CTGGAGGAAAAAGTGCCTGAAAAGGTCTACCTGTTTGTAGCAGATGAAGGCAATGTATGATTGGAATATTCTGGTGCACATCCTG
 GAAACCGAGGAATACATCAGCTTTCGCAAAAGTGAAGCAGTATGAGAAAACAAGACTAACCTGCGGCTGCTGAGAGACATCATT
 CTGAAATACTGCACCAAGGATGAGTATAATCGGATGTTTAAACGACGAGAAGGAAAGCTGGCAGCTACCCGCATATGTGGGGAAA
 CTGAAAAAGAACCAAGAAGTACTGGATCGAGAAAAAGAGAAATCCCGAGGAGTTCTACAATCCCTGGGCAAGCTGCTGGAT
 AAAATTGAGCCTGAAAGGAGGACCTGGAAGTGTGACTATGATCGAGGAGTGAAGAACCAACCTGCTGCCAATTCAG
 AAAAAAAGGACCAAGCGCTGATCCCCACAGGTGATGAGGTCGAACTGAAAAAGATCCTGGAAAATGCAAAAAGTACTAT
 TCCTTCTGACCGAGACAGACAAGGATGGGACTCAGTGGTCCAGAAAATCGAGAGCATTTTCAGGTTTGCATCCCTACTAT

FIG. 9C

GTGGGGCCTCTGAGTACCCGGCACCAGGAAAAGGGATCAAACGTGTGGATGGTCAGAAAACCTGGCAGGGAGGATCGCATCTAC
 CCATGGAATATGGAGGAAATCATTGACCTTGGAGAAGAGCAACGAAAATTTTATTACACGGATGACTAACAAATGTACATATCTG
 ATCGGGGAAGATGTCCTGCCCAAGCATTCTCTGCTGTACAGTAAATATATGGTGTGAATGAGCTGAACAATGTGAAGGTCAGA
 GGAAAAAGCTGCCTACATCTCTGAAACAGAAGGTGTTGAGGACCTGTTTGAAAACAAATCCAAAGTGACTGGAAAGAAATCTG
 CTGGAGTACCTGCAGATCCAGGACAAAGATATCCAGATTGACGATCTGTCTGGCTTCGACAAGGACTTCAAGACCAGCTGAAG
 AGCTATCTGGACTTCAAAAAGCAGATTTTTGGGGAGGAAATCGAGAAGGAAAGCATTGAGAACATGATCGAAGATATCATTAAAG
 TGGATCACCATCTACGGCAATGACAAGGAGATGCTGAAACGAGTGATTGCGGCTAATTATAGCAACCAGCTGACAGAGGAACAG
 ATGAAAAAGATCACTGGATTTTCAGTACAGTGGCTGGGGAACTTCTCAAAGATGTTTCTGAAAGGGATCAGCGGATCCGACGCTG
 AGCACCGGCGAAACATTGACATCATTACCGCAATGTGGGAGACAGACAACAATCTGATGCAGATCCTGTCAAAAAAGTTCACC
 TTTATGGACAACGTCGAGGACTTCAACAGCGGCAAGGTGCGGAAAATCGACAAGATTACTTACGATAGCACCTGAAGGAAATG
 TTCTGTCCCTGAGAACAAAAGGGCCGTCTGGCAGACCATTCAAGTGGCTGAGGAGATCAAGAAAGTGATGGGCTGCGAGCCA
 AAAAAAGATCTTTATTGAAATGGCACGGGGCGGGGAGAAAGTTGAAAGAAAGGACAAAATCTCGCAAGGCCAGCTGCTGGAGCTG
 TACGCCGCTTGCAGGAAAGATTGTAGAGAATGATCAAGGAGATTGAGGACCGGGACGAGAGGGACTTCAATAGCATGAAGCTG
 TTTCTGACTATACCCAGTTCGGGAAATGATGTATTCCGGCGACGACATCGATTTAACGAGCTGATTCGGGCAATTTCTAAG
 TGGACCCGAGATCACATCTACCCCGAGACAAAATTAAGGACGATTCCATCGATAACCTGGTCTGGTCAATAAGACATATAAT
 GCCAAAAAGTCCAATGAGCTGCTGTCTGAGGACATCCAGAAAAAGATGCATTATTCTGGCTGAGCCTGCTGAACAAAAAGCTG
 ATCACTAAAAGCAAGTACGACCGCCTGACTCGAAAGGGCGACTTTACCGATGAGGAACTGAGTGGGTTTCTGCTAGACAGCTG
 GTGAAAACAAGGCAGTCAACTAAGGCAATCGCCGATATCTTCAAGCAGATCTACAGCTCCGAGTGGTCTATGTGAAGAGCAGC
 CTGGTGAGGCTCTCAGGAAAAAGCCACTGAACTACTTCAAGCAGTCTGGAGAGTCAATGATTAACCACTGAAAAAGACGCTTAT
 CTGAACATTTGGTGGGAAAGCTGTACAACAAAAAGTTTACCAGTAATCCATCCAGTGGATGAAAAGAATCGCGATACAAC
 TATAGCCTGAACAAGGTGTTGAAACACGACGCTGGTCTTAAACGGAGAAGTGTCTGGGAAAAGTGCACATACCATGAGGACACT
 AATACCTATGATGGAGGCACTCTGGACCGAATCCGGAAGATTGTGGAACGCGATAACATTCTGTACACCGAGTACGCTTATTGT
 GAGAAGGGCGAAGCTGTTAATGCAACCATCCAGAACAAAATGGAACTCCACAGTCTCTGAAAAAGGGCTGGACGTGAAA
 AAGTACGGGGGACTTCTCAGCGCAACACAAGTTACTTCTCAGTGTAGTGGTGGAGCAAGAAGGGGGATAGAGCAAGGAC
 ATCATTTGGAGTGCCTATCTATATTGCAAACTGCTGGAGCATTCTCAAGTGCCTTCTGGAGTACTGCGAAACAGAAGGGGTAT
 CAGAATGTGCGGATTTCTGGTGCAGAAAATCAAAAAGAACAGCCTGCTGATCATTAAATGGATACCTCTGCGCATTGAGGGCGAG
 AACGAAGTGGATACTTCTTTAAGAGGGCCATCCAGCTGAAGCTGGACCAGAAAAACTATGAGCTGGTCCGCAATATCGAGAAG
 TTCTGGAAAAATACGTGGAGAAAAAGGGAACTATCCAATTGACGAGAATAGAGATCACATCACATGAAAAGATGAACCAAG
 CTGACGAGGTGCTGCTGCTCAAAAATGAAAAAGTTCAACAAGAAAGGGCATGGCCGACCCCTCTGATAGGATCGAAAAGAGTAAG
 CCTAAATTCATCAAGCTGGAGGACCTGATCGATAAGATTAATGTGATCAACAAAATGCTGAACCTGCTGCGCTGTGACAATGAT
 ACTAAGGCCGACCTGTCTCTGATTGAGCTGCCAAAAACGCTGGGAGTTTCTGGTCAAAAAGAAATACCATCGGAAAGTCAAAA
 ATCATCTGGTGAATCAGAGCGTACTGGACTGTACGAGAATAGACGGGAACGTGAAGAATTC

#6

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCGGAAGAAAAGCGCAAGGTGCAAGCGTCCATGGGGAGG
 AAACCTTACATTCGTCTCTGGATATTGGAACCTGGGTCGCTCGGCTACGCTTGCATGGAATAAGGATTCAACGTGCTGAAGTAC
 CAGGACAAAGATGCCCTGGGAGTGTATCTGTTTCGACCGCGCTGACTGCAACAGGAGCGGAGACAGTTTAGGACCTCCAGGCGC
 CGAAAGAACCAGGAATCAACCGCTGGGCTGCTGCAGAACTGCTGGCACCCCTGGTGCAGAACCTAATTTTACCAGTTT
 CAGCGGAGTTTCGCTGGAAAGAACGACAATATGGATTTAAGAAACAAGAGCCTGCTGAGGTGCTGAGCTTCTGGGATAGAA
 TCCAAGAAATACCTACCATCTACCACCTGCAGGAGGCTCTGCTGCTGAAAGACGAGAAGTTTGTCCAGAATGATCTACATG
 GCCTGTATCATCTGGTGAATAACAGAGGCACTTTCTGTTTCGATCATCTGAAGATCGAAGAACCTGACTAACAAATGACAATATG
 CACGATTTCTGGAGTATTGAAACCTATGAGAACCTGAAGCTGAATCTGGACTGAGAAAACAGGAAAGTATC
 TATGAGATTTCTGAAAGCAACGAAATGACTAAGAATGATAGAGCCAAAAGGGTCAAGAACATGGAGAAGAACTGGAACAGTTC
 TCTATCATGCTGCTGGGGCTGAAGTTCAATGAGGGAAAACCTGTTAACCACGCCGATAATGCTGAGGAACTGAAGGGGGCTAAC
 CAGAGCCATACATTTGCAGACAACCTACGAGGAAAATCTGACTCCCTTCTGACCGTGGAAACAGTCAAGTTTATTGAAAGGGCC
 AACAAAAATCTATCTGAGCCTGACTCTGCAGGATATCCTGAAGGGCAAGAAAATCAATGGCTATGAGCAAAAGTGGCCGCTTACGAC
 AAGTTCAGAAATGAGCTGAAACAGGTGAAGGACATTGTCTAATAGGCTGATTCTACCAGGACACAGTTCAAGAAAATCTTTGTG
 AGCTCCAAGAAAAGTCTGAAGCAGTACGACGCAACTCCCAACGATCAGACCTTCTCTAGTCTGTGCTGTTTGACCAAGTACCTG
 ATTCGCCCAAGAAAACAGTATAGCCTGCTGATCAAGGAGCTGAAGAAAATCATTCCCAAGGACTCCGAACCTGACTTTGAGGCA
 GAAAATGATACCCTGCTGAAGGTGCTGAACACCACAGACAATGCTAGCATCCCTATGCAGATTAACCTGTACGAGGCGAAGAAC
 ATCCTGCGAAATCAGCAGAAATATCACGCCGAGATCACAGATGAGATGATTGAAAAGGTGCTGTCTGATCCAGTTCGCAATT
 CCATACTATGTGGGGCCCTGGTCAACGACCATAACGCAAGTAAAGTTGGATGGATGGAGCGCAAAAAGTAAACGAATCAATCAAG
 CCTTGGAAATTTGACGAGGTGGTGCATCGAAGTAAATCAGCCACTCAGTTTATTAGGCGCATGACCAACAAGTGTTCCTACCTG
 ATCAATGAGGATGTGCTGCCAAAAAATCTCTGCTGTATCAGGAGATGGAAGTCTGAAACGAATGAATGCCACACAGATCAGG
 CTGCAGACTGACCCAAAAAACCAGTACCGAATGATGCCCGAGTTAAGCTGTTGCTGTGGAGCACATCTTTAAGAAATAT
 AAAACCGTCAGCCATTTCAAAGTTCCTGGAAAATATGCTGACAGCAATCAACAGGGAGAACTTTATGAATCATGGAGAAAAGCTG
 AGTATCTTCGGCACACAGGACGATAAGAAATTTGCATCAAAGCTGTCAAGCTACCAGGACATGACTAAAATCTTCGGGGATATT
 GAGGGAAAGCGCGCCAGATTGAGGAAATCATTCAAGTGGATCACCATTTTTGAGGACAAAGAAATCCTGGTGCAGAAAGCTGAAA
 GAGTGTACTCTGAACTGACATCAAGCAGATCAACCAGCTGAAGAACTGAATTACTTGGCTGGGGAGGCTGAGTGAGAAG
 CTGCTGACTCACGCCATCAGGGCCATAGCATCATTGAACTGTGCGCCACTCCGATGAGAAATTTATGAAATTTCTGACCAAC
 GACGTGACGGGTTCCAGAAATTTATCAAAGAGGAAAACAGGTTCCAGAGCAATAAGATCCAGCATCAGGATATTGCCAACCTG
 ACTACCTCTCCGCTCTGAAGAAAAGGCACTCTGGAGTACAATTAAGCTGGTGGGGAGCTGACTTCCATTTTCGGGGAGCCTGAA
 AAGATCATTATGGAGTTTGTACCGAGGACCAGCAGAAAGGCAAGAAAACAGAAATCAAGAAAAGCAGCTGTGGGACGATAACATC

FIG. 9D

AAGAAAAATAAGCTGAAAAGCGTGGACGAGTACAAAATATCATTTGATGTCGCCAATAAGCTGAACAATGAGCAGCTGCAGCAG
 GAAAAACTGTGGCTGTACCTGAGCCAGAACGGCAAGTGTATGTATAGCGGGCAGTCCATCGACCTGGATGCCCTGCTGTCCCC
 AATGCTACCAAGCACTACGAGGTGGATCATATTTTCCCTCGGAGCTTCATCAAGGACGATAGCATTGACAACAAGGTGCTGGTC
 ATCAAGAAAAATGAATCAGACAAAAGGGCGATCAGGTGCCCTGCAGTTCATTCAGCAGCCTTACGAGAGAATCGCATATTTGGAAG
 AGCCTGAAACAAAGCCGGGCTGATCTCTGATAGTAAACTGCACAAGCTGATGAAACCAGAGTTCACAGCTATGGACAAGGAAGG
 TTCATCCAGCGGCAGCTGGTGGAGACTAGACAGATCAGCGTGCATGTCGGGATTTTCTGAAAGAGGAATACCCTAATACCAAA
 GTGATCCCAATGAAGGCCAAAATGGTGAAGCGAGTTCGGAAAGAAATTTGACATCCCAAAGATTAGACAGATGAACGACGCACAC
 CATGCCATCGATGCTTACCTGAATGGCGTGGTCTATCACGGGGCACAGCTGGCCTACCCCAACGTGGACCTGTTTGTATTTCAAT
 TTTAAGTGGGAGAAAAGTCCGAGAAAAGTGGAAAAGCCCTGGGAGATTTCAACACAAAGCAGAAAATCTCGGAACTGTTCTTTTTC
 AAGAACTGGAGAAGATGGAAGTGTCCAGGGCGAGCGGTGATCTCAAGATCAAGCTGGACATGAACCACTTCAAGATCAAC
 TACTCCAGAAAAGTGGCCAACATCCCTCAGCAGTTTTAATCAGACCGCAGTGTCTCAAAGACAGCCGAGCTGAAATACGAA
 TCTAACAAAGAGTAATGAGGTGGTCTATAAGGGACTGACCCATACCAGACTTATGTGGTCCCATCAAGAGCGTGAACAAGAAA
 GGCAAGGAGAAAATGGAATACCAGATGATCGACCCTACGTGTTGATTTTTATAAATCCAGAAGCGCAATGAGAAGGAACTG
 GCTCTGACCTGGCAGAGGGGAGAACAAGGACGAAGTCTGGATGCTCAGATTTGCTATAGTCTGAATAAGGGGGATCTGCTG
 TACATCAACAATCATCCCTGCTATTTCTGTGTACGCAAGAGGTTCATCAACGCAAGCAGTTTGAAGTGCACCGTGGAAACAGCAG
 CTGTCTCTGACCAAGTGTGATGAACAACAAGGAGACAAAATGTCGAAAAGCTGCTGATCGAGTATGACTTCATTTGCCGAGAAAAGT
 ATCAACGAATACCACCATTTATCTGAATAGCAAGCTGAAAAGAAAAGCGAGTCCGGACCTTTTTCTCAGAGAGCAACAGACACAC
 GAGGACTTCATCAAGGCCCTGGACGAGCTGTTAAGGTGGTCAACGCATCCGCCACAAGGTCTGATAAAAATCGGGAGTGCGAAG
 AACAGCATGACTCATCGAGCCTTCTGGGAAAAGGCAAGGACGTGAAGATTGCTTACACCTCCATCTCTGACTGAAAACAACT
 AAACCTAAGAGTCTGTTAAGCTGGCCGAGTCAAGAAACGAACGTGAAGAATTC

#7

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAGCGCAAGGTGGAAGCGTCCATGACAAAA
 ATCAAAGCAGACTACATCGTGGGACTGGACATCGGCACAGACTCTGCGGGTGGTGGCTATGAACAGCAATAATGACATTCG
 AAATGCAGGGCAAGACCCTGATCGGGTCAAGCTGTTGAGGGAGGGAAAGCGCAGCTGAACGGAGACTGTTTCGACCACA
 CACAGGCGCATCAAACGACGAGATGGCGACTGAAGCTGCTGGAGGAGTTCCTCGACCCCTACATGGCAGAGGTGGATCCTTAT
 TTCTTTGCCGGCTGAAGGAATCTGGCTGAGTCCACTGGACAAAAGAAAGACCCTGAGCTCCATTTGTGTTCCCCACATCCGCC
 GAGGATAAAGAGTTCTACGACGATTACCCTACAATCTACCATCTGAGGTATAAACTGATGACTGAGGACGAAAAGTTGATCTG
 CGCGAAGTGTACCTGGCTATCCACCATATCATTAAAGTACCAGGAAAATTCCTGTATAAATACAGTGTGAAAGACTTCAAGGCA
 TCAAAGATCGATGTCAAATCTAGTATCGAGAAGCTGAACGAGCTGTATGAAAATCTGGGCTGGACCTGAACGTGGAGTTCAAC
 ATTAGCAATACTGCCGAGATCGAAAAGGTGCTGAAAAGCAAGCAGATCTTCAAGCGGGATAAAGTCAAGAAAATTCGCCGAGCTG
 TTTGCTATCAAACCGACAAACAAGGAACAGAGCAAGAAATCAAAGATAATTTCAAACAGGTGGCCAACTGCTGCTGGGGTAC
 AAGACCAGGTTTCGACACAATCGCTCTGAAAAGAGATTTCAAGGACGAACCTGTCTGATTGGAACTTCAAATGTGACAGATCGAT
 GCAGACAGCAAGTTTGGAGCCCTGATGGGAAACCTGGAATGAGAATGAACAGGCCATCCTGCTGACTATTAAAGGAGCTGTTAAC
 GAAGTGACCTGAATGGAATTTGTCGAGGACGGCAACACCTGAGCGAATCCATGATCAACAAGTACAATGATCACCGGGACGAT
 CTGAAAGCTGCTGAAAAGAGTGTGCGAAAATCATATTGACAGAAAAGGCAAGGAGCTGGCACTGGCCTACGATCTGTATGTC
 AACAATAGGCACGGACAGCTGCTGCAGGCTAAGAAAAGCTGGCAAAAATCAAGCCCCCTCTAAGGAGGACTTCTCAAAGTG
 GTCAAACAAGAACTGGACGATTCACGGGCAAGCAAGGAGATCAAAAAGAAAATGAACTGGACAGCTTTATGCTAAGCAGAGA
 ACAACGCCAATGGCGTGTATCCCATACCAGCTGCAGCAGCTGGAGCTGGATAAGATCATCGAAAACCAAGTCAAGTACTATCCA
 TTCTGAAAGGAGATTAATCCCGTGTCAAGCCACTGAAAAGAGGCCCTATAAGCTGGACGAACCTGATCCGATTTCCGGTGCCT
 TACTATGTCCGCCCTGATTTCTCCTAACGAGAGTACCAAGGATATCCAGACAAAAGAAAACCAAGAAATTCGCCTGGATGATT
 CGCAAAGAGGAAGGGCAATCACACTTGGAACTTTGACCAGAAGGTGGATCGAATGAGAGCGCCAATAGTTTCACTCAAACGG
 ATGACTACCAAGGACACTTACCTGTTTGGGGAGGATGTGCTGCCAGCTAACAGCCTGCTGTATCAGAAGTTCCAGCTCCTGAAC
 GAACTGAACAACATCCGGATTAATGGAAAAGAAATCTCCGTGGACCTGAAGCAGGAGATCTACGAAAACCTGTTAAGAAAACAC
 ACAACTGTGACCGTCAAGAACTGGAGAAATTAATCGAAGAAAACCAATCTGGTGAAGTGCAGATCAAGGGGCTGGCCGAT
 GAAAAGAAAATTCACAGCGGACTGACCACATACAATAGATTCAAGAACCCTGAACATCTTTGACAACCAGATTGACGATCTGAAG
 TACAGGAACGATTTGAGAAGATCATCGAATGGTCTACAATTTTTGAGGACAAGAGTATCTCAAAGAAAAGCTGAGGAGCATC
 GATTGGCTGAACGAGAAGCAGATTAACGCTCTGTCTAATATCAGACTGCAGGGGTGGGGAAAGGCTGAGTAAGAAAATGCTGGCA
 CAGCTGCACGACCAATAATGGCCAGACCATCATTGAGCAGCTGTGGGATTTCCAGAAACAATTTATGACAGATTGTGACACAGGCC
 GACTTTAAAGATGCTATCGAAAAGGCCAACCAAGAACTGCTGTTGGTACCTCAGTCCAGGACAATCTGAACAATGCATACACA
 AGCCCCGAAAACAAGAAAGCCATCAGACAGGTCATCAAGGTGCTGACGATATCGTGAAGGCAGCCTCCGGAAAAGTCCCAAAA
 CAGATCGCCATTGAGTTCCTAGGGATGCTGACGAAAATCCCAAGAGAAGTGAACAGGGGCTCAAAGTGCAGAAAAGTGTAC
 AAGGACCTGAGCACTGAGCTGGCCTCCAAGACCATTGCTGAGGAACCTGAACGAAGCAATCAAAGACAAGAAAATGGTGCAGGAT
 AAGTACTATCTGTACTTTATGACGCTGGGGCGGACGCTATACAGGAGAGCCTATCAATATCGATGAAATCCAGAAGTACGAT
 ATCGACCACATTTGCCACAGTCTTTTCAAGGACGATGCCCTGGACAACAGGGTGTGGTGAAGCGGGCTGTGAACAATGGC
 AAATCTGATAATGTGCTGTCAAGCTGTTTGGCAACGAGATGGCTGCAAATCTGGGGATGACTACAGGAAAATGTGGGAGGAA
 TGGAAGAACATCGGCTGATTAGCAAAAACAAGTACAACAATCTGCTGACTGATCCCGACCACATTAACAAGTATAAGAGTGCC
 GGGTTCATCAGGCGCCAGCTGGTGGAGACATCACAGATCATCAAGCTGGTGAAGCACTATCCTGCAGAGTGCCTCAACCTAACACT
 GAAATCATTACCGTGAAGGCTAAGTACAATCATTCTGCGGGAGAAATTTGACCTGTATAAGAGCAGAGAAGTCAACGACTAC
 CACCATGCTATTGATGCATATCTGTCGCCATCTGCGGAAATCTGTGATACCAGAACTATCCAAATCTCGGGCCCTCTTTGTG
 TACGGCCAGTATAAGAAAATCTCCTCTGATCCTGACAAAAGAGAAGGCCATTTTAAACAAAACCCGCAAGTTCTCCTTTATCTCT
 CAGCTGCTGAAAACAAGAGTGAAGACAGCAAGGAAATCGCTAAGAACTGAAACGGGCATACCAGTTCAAGTATATGCTGGT
 TCTCGAGAGACTGAAACCCGGGACCAAGGAGATGTTCAAATGACCGTGTACCCCGGTTCAAGCCAGTACAGTCAAGGCTCT

FIG. 9E

AGGAACCTGATTCAAAGAAAATGGGCATGTCCCCTGACATCTACGGAGGCTATACAAAATTTCTGACGCATACATGGTCATC
GTCCGCATTGATAAGAAAAGGGAACCTGAGTATAAGATCCTGGGCATTCACCCCGGGAACCTGGTGAATCTGAAAAAGGCCGAG
AAGGAGGACCATTAACAAAAGCTATCTGAAGGAGTCTGACACCAAGGATTCTGTACAAACAAAATGGGAAGCGCGATAAAAAG
ATCACTTCCCTCGAAAATTGTGAAATCTAAGATCCCCTATAAGCAGGTCATCCAGGATGGGGACAAAAGTTTATGCTGGGAAGT
TCAACATACGTGTATAACGCAAAGCAGCTGACACTGAGCTGAGTCCATGAAAGCCATCACTAACAAATTTGATAAAGGACAGC
GATGAGAACGACGCTCTGATTAAGGCATACGATGAAATCCTGGACAAAAGTGGATAAGTATCTGCCACTGTTGACATCAACAAG
TTCCGGGAGAAGCTGCACAGTGGGCGAGAAAAGTTCACTCAAGCTGAGCCTGGAGGACAAAAGGATACCATCCTGAAAGTGTCTG
GAAGGACTGCATGATAACGCTGTATGACAAAAGTCCCCTACTATTGGCCTGTCCACACCCTGGGGTTTCATGCAGTTTCCCAAC
GGCGTGATTTCTGAGCGAGAATGCCAAAATGATCTACCAGTCCCCACCGGGCTGTTCAAAAAGTCAGTGAAGATCAGCGACCTG
TAAGAATTC

#8

ACCGGTGCCACCATGTACCATACGATGTTCCAGATTAACGCTTCGCCGAAGAAAAGCGCAAGGTGCAAGCGTCCATGGGCTAC
ACTGTGGGACTGGATATTTGGGGTGGCTTCCGTCGGGGTGGCTGTGCTGGATGAGAATGACAAACATCGTGGAGGCTGTGTCAAAC
ATCTTTGATGAAGCCGACACAAGCAACAATAAGGTGCGGAGAACTCTGAGGGAGGGCAGGGCGCAAAAGCGGGCGAGAAAACC
CGCATTGAGGACTTCAAGCAGCTGTGGGAGACTTCAGGCTACATCATTCTCACAGCTGCATCTGAATATCATTGAGCTGCCG
AACAAAAGGGCTGACCGAATGCTGAGCTGGATGAGCTGATTGCGTGTCTGCTGCTGATGCTGAAAGCACCGGGGATCTCTAC
CTGGAGGACCGCGACGATGGCGAGAAGGGGAATGCCTATAAGAAAAGGACTGGCTTTTAAACGAAAACAGCTGAAGGAGAAAATG
CCATGTGAGATCCAGCTGGAAACGATGAAGAAAATACGGGAAGTACCATGGAGAGTTCATCATCGAAAATTAATGATGAGAAGGAA
TACCAGAGCAACGTGTTCCACCAAAAGCTTATAAGAAGGAGCTGAAAAGATCTTCGAGACACAGCGGTGCAACGGCAACAG
ATCAACACAAAAGTTTATTAAAGAAAATACATGGAGATCTACGAAACGAAAGCGGGAATACATATCGGACAGGCAATGAGAAAAGC
AGAACAGACTACCGCATCTATACTACCAGGACTGATGAGGAAGGGAAATTCATCGACGAGAAGAACAATTTTGGCAAACCTGATC
GGGAAGTGTAGTGTGTACCCCGAGGAATATAGAGCAAGCTCCGCTCATACACCGCCAGGAGTTCATCTGCTGAACGATCTG
AACAACTGAAAAATCAACAATGAGAAGCTGACAGAAATTCAGAAGAAAAGAGATTGCGAAAATCATTAAAGGACGCTTCTAGTGTG
AACATGAGGAAAATCATTAAAGAAAGTCATCGATGAGGACATTAACAGTACAGCGGAGCACGAATCGATAAGAAAAGGCAAGGAA
ATCTACCACACCTTCGAGATCTATCGGAAGCTGAAGAAAAGAGCTGAAAACAATCAATGTGGATATCGACTTTTTACTAGAGAG
GAACTGGATAAAGACCATGGACATCCTGACCTGAACACAGAGAGGGAAAAGTATTGTGAAGGCCTTGACGAAACAGAAAATTTGTC
TACGAGGAAAATCTGATCAAGAACTGATTGAGTTTCCGGAAGAACAATCAGAGACTGTTGACGGCTGGCATAGTTTTTTCATAC
AAGGCTATGCTGCAGCTGATCCAGTGTGTACAAGGAGCCCAAAGAACAGATGCAGCTGCTGACCGAAAATGAACGTTTCAAAC
AGTAAGAAAGAGAAGTACGTCAACTACAAGTACATCCCAGAGAACGAAGTGGTCAAGGAGATCTATAACCCCGTGGTGTGTAAG
AGCATTAGAACAACTGTGAAAATTTCTGAATGCATGATCAAGAAAATACGGGTATCCTGAATCCGTGATCGAGATGCAAGG
GATAAGAACTCTGACGATGAGAAGGAAAAGATCGACATGAACCAGAAAAGAAAACAGGAGGAATACGAGAAAATCCTGAACAAG
ATCTACGATGAGAAGGGAACTGAAATACCAACAAGGACTACAAGAAAACAGAAAAGTGGTGTGAAGCTGAAACTGTGGAAC
GAGCAGGAAAGGACTGTGCCTGATTTCCGGCAAGAAAATCGCTATTGAGGATCTGCTGAATCACCCGAGTTCTTTGAAAATGAC
CATATCATTTCTAAGAGCATCTCCCTGGACGATTTCTCGCAGTAAACAGGCTCCTGGTGTACAAAACAGAAAATTTCTATCAAGGAG
AACGATACCCCTACCCTATCTGACACGGATTAACCGAAAAGTGGGGCTTTGACGAAATAAAAGCTAATGTGCTGGAGCTGAGA
AGGCGCGGCAAGATCGACGATAAGAAAAGTGAACAATCTGCTGTGATGGAGGATACCTAAGATTGACGCTGTAAGAGGTTT
ATTAACCGCAATCTGAACGACACCAGATACGCAATCCAGGGTGGTGTGACGAAATGCAGTCTTCTTTGAGTCTCGAAAAGTAC
TGTAATACTAAGGTCAAAGTGTCCGAGGCTCTCTGACCTATCAGATCGCGCAGGATCTGCACCTGAAGAAAACAGAGAGGAA
TCATAAGCCACCATGCTGTGGACGCAATGCTGATCGCAATCTCCAGAAGGGGTACGAGGCTATAGGAAAATCCAGAAAAGT
TGCTACGACTTTGAGACAGGGCGAAAATTTCTGGACAAGGAAAATGGAATAAGTACATTTGACGATGACGAGTTTGATGACATCCTG
TATAAAGAGAGGATGAACGAAAATCCGCAAGAAAATCATTGAGGCGAGGAAAAGGTGAAGTACAACATAAGATCGATAAGAAAG
TGCAATCGCGGGCTGTGTAACCAGACTATCTACGGGACCCGAGAAAAGGACGAAAATCCACAAGATTTCAAGCTACAACATC
TATGATGACAAGGAGTGAATCCCTGAAGAAAATGATTAACAGTGGGAAAGGATCAGATCTGATGTACAACAATGATCTCT
AAGACATATCGGACATGCTGAAAATCCTGGAACTTACTCTCTGAGAAGAATCCATTCTGGCATATAACAAGAGACAGGA
GACTACTTTTCGAAAATTTCTAAGAAATCAACAACGACCCAAAGGTCGAGAAGGTGAAATACTATAGCGGCCAGATCAACTCCTG
ATCGATATTTCTACAAGTACGGCCATGCCAAAATAGTAAGAAAAGTCTGCTGCTGGTGTCACTGAACCTTATAGAACCAGCCTC
TACTATGATAATGACACAGGCAAGTACTATCTGGTGGGGTGAAGTACAATCATAACAAATGTGTGCGAAAACAAAGTACGATG
GATAGCGAGACATATAACGAACTGCTGAGGAAGGAGGGCGTGTGAAACAGCGATGAGAACCTGGAGGACCTGAAACAGCAAAAAC
ATCACTTACAAGTTCAGTCTGTACAAGAACGACATCATCCAGTACGAGAAGGGCGGGGAATACTATAACAGAGCGCTTTCTGAGC
CGAATCAAAGAACAGAAAGAACTGATTGAGACTAAACCCATCAATAAGCCTAACCTTCCAGCGGAAGAAATAAGAAAAGGCGAGTGG
GAAAATACAGAAACAGATCGCCCTGGCTAAGACTAAATACGTGGGAAGCTGTTCAACCGATGTGCTGGGAAAATGTTACATC
GTGAACATGGAGAAGTTCTCCCTGGTCTGGACAATAAGAAATTC

#9

ACCGGTGCCACCATGTACCATACGATGTTCCAGATTAACGCTTCGCCGAAGAAAAGCGCAAGGTGCAAGCGTCCATGACTAAC
GGCAAGATTTCTGGGGCTGGACATTTGGCATCGCAAGCGTGGGGTGGGATTTATTGAGGCAAAAATGGAAAAGGTGGTGCATGCC
AATTCCTGGCTGTTCTCTGCGGCTAACGCTGAGAACAATGCAGAACGGAGAGGGTTTGGGGATCTAGGCGCCTGAATCGACGG
AAGAAAACCCGCTGAAGCGAGTCCGGGATCTGTTTCGAGAAAATACGGAATCGTCAACCGACTTTTCGCAACCTGAATCTGAACCT
TATGAGCTGCGAGTGAAGGGCTGACCGAACAGCTGAAAACAGGAACTGTTTCGAGCCTGAGAACAATCTCTAAGAGAAGG
GGGATTAGTTACCTGGACGATGCCGAGGACGATAGTACCGGATCAACAGACTATGCTAAGAGATCGATGAGAATCCGCGACTG
CTGAAAACAAAGACACCAGGCCAGATTGAGCTGGAGAGGCTGAAAAGTACGGCCAGCTGCGCGGGAAATTTCAACCGTCTATGAC
GAGAACGGGGAAGCCATCGCCTGATCAATGTGTTTAGTACATCAGATTAACGAGAAGAAAGCACGGGAAGTCTGGAGACACAG

FIG. 9F

GCCGACTACAACAAGAAAATCACAGCTGAGTTTCATTGACGATTATGTGGAAATCCTGACCCAGAAAACGAAAGTACTATCACGGC
 CCCGGGAACGAAAAGAGCCGGACTGACTACGGACGGTCCGGACCGATGGGACCACACTGGAGAATATTTTCGGAACTCTGATT
 GGCAGTGCACCTTTTACCCTGATGAATATCGAGCAAGCAAGGCCAGCTACACCCGACAGGAGTATAATTTCTGAAACGACTG
 AACAACTGGAAGTGAGCACCGAAAAGGGAAGCTGTCAACAGAGCAGAAAAGAAAGCCTGGTGGAGTTTCCAAAGAATACTGT
 ACCCTGGGACCCGTAACCTGCTGAAGGAGATCGCAAAAATTTCTGGACTGTAAGGTGGATGAGATCAAAGGATACAGAGAGGAC
 GATAAAGGCAAGCCAGATCTGCATACCTTCGAGCCCTATAGGAACTGAAGTTTAACTGGAAAGCATCAACATTGACGATCTG
 TCCCGCAAGTGTGCAAGCTGGCTGATATTCTGACTCTGAACCCGAGAGAGAAGGAATCGAGGACGCAATTAAGAGGAAT
 CTGCCAAAACAGTTTACAGAGGAACAGATCAGCGAGATCATCAAGGTGCGGAAGAGCCAGTCCACTGCTTTCAATAAGGGCTGG
 CACTCTTTTAGTGCAAAAAGTATGAAACGAGCTGATCCCCGAACTGTACGCCACCTCCGACGAGCAGATGACAAATTCTGACTCGG
 CTGGAAAAATTAAGGTCAATAAGAAAAGCTCCAAAAACAAAAGACTATCGACGAGAAGGAAGTCACTGATGAGATCTACAAT
 CCTGTGGTCCCAAGAGCGTGAGACAGACCATAAAAATCATTAAACGCTGCAGTCAAGAAATATGGCGACTTCGATAAGATCGTG
 ATTGAAATGCCCGGATAAAAATGCTGACGATGAGAAGAAATTCGACTCGACAAGAAAATAGGAGAACAAAGGAAAAGGAC
 GATGCCCTGAAAAGGGCCGTTACCTGTATAATTCTAGTGACAAGCTGCCCGATGAGGTGTTCCACGGCAACAAGCAGCTGGAA
 ACCAAAATCCGACTGTGGTATCAGCAGGGGGAGCGGTGCTGTATAGTGGAAAAGCCCATCAATTCAGGAGCTGGTGCAATG
 TCTAACAAATTTGAAAATTCGATCACTTCTGCTCTGTCACTGAGCTTTGACGATAGTCTGGCCAAATAGGTGCTGGTCTACGCT
 TGGACAAACCAGGAGAAAGGCCAGAAAACCCCTTATCAGGTCACTGACTCCATGGATGCAGCCTGGTCTTTCAGGGAGATGAAG
 GACTACGTGCTGAAACAGAAAGGACTGGGCAAGAAAAGCGCGACTATCTGCTGACTACCGAGAACATCGATAAGATTGAAGTG
 AAGAAGAAGTTCATCGAGAGGAATCTGGTGGATACTCGTACGCATCTCGAGTGGTCTGAACTCTCTGCAGAGTGCCTGAGA
 GAGCTGGGAAAAGACATAAGGTGTCTGTGGTCAAGGACAGTTCACCCAGTCAAGTCAAGGAGAAAATGGAAGATCGATAAGAGC
 CGCGAGACATACCACCATCAGCAGTGGACGCCCTGATCATTGTGCTCATCAAGCCAGCTGAACTGTGGGAGAAGCAGGACAAT
 CCCATGTTTGTGGATTATGGCAAGAACCAGGTGGTGCAGAAAACAGACTGGGGAGATCCTGTCCGTGCTGACGATGATGATAAG
 GAACTGGTGTCCAGCCCCCTTATCAGGGCTTTGTGAATACCATCTCCTCTAAAGGGTTCGAGGACGAAAATTCGTTTAGCTAC
 CAGGTGGATTCCAAATATAACCGGAAGGTGAGTACGCAACCATCTACTCAACAAGAAAAGCCAAGATTGGCAAGGATAAGAAA
 GAGGAAACCTACGTGCTGGGAAAAATCAAGGACATCTACTCCAGAATGGCTTCGATACCTTCATCAAGAAGTACAACAAAGAT
 AAGACTCAGTCTGTATGATCAGAAGGACTCTGACATGGGAGAAGTGTGCAAGTCAATCTGAGGGACTACCCAAACACT
 AAGAAAAGCGAGGACCGCAAAAATGATGTGAAGTGCACCCCTTTGAGGAATACAGGCGGAGAAATGGGCTGATGTAAGTAT
 TCCAAGAAAAGGAAAAGGAACTCCCATCAAGAGCTGAAAGTACTATGACAAGAACTGGGGAAGTGCATCGATATTACCCAGAG
 GAATCAGCAATAAGGTCACTCTGCAGAGCATTAAACCTTGGCGAGCCGACGTGTACTTCAATCCAGAGACACTGAAGTACGAA
 CTGATGGGCTGAAAATTCAGATCTGAGCTTTGAAAAGGCACTGGGAACTACCAATATCAGCCAGGAGAAAATGACGCTATC
 AAAGAGAAGGAAGGAATTGGCAAGAAAATCCGAGTTCAAGTTTACTGTACCAGCAACGACCTGATCCTGATCAAGGATATCGCC
 AGTGGCGAGCAGGAAAATCTACAGATTCCTGTCAAGAACTATGCCAATGTGAACCACTACGCTCGAGCTGAAGCCTTACGACAAG
 GAAAAGTTCGATAACGTGACAGGAGCTGGTCAAGCACTGGGAGAGGCAGATAAAGTGGGACGATGATCAAGGACTGAATAAG
 CCAAAACATCAGCATCTACAAGGTGAGAACCCGACTCCTGGGAAAACAAATATTTCTGTAAGAAAAGGGCGCAAAACCAAGCTG
 GATTTTAAAGAACAAAGAAGTAAGAAATTC

#10

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAGCGCAAGGTGCAAGCGTCCATGGCCGAC
 CGAATCTCTTGGGGCTGGACATTTGGGTTGGCAAGCGTGGGTTCTCAGTGTGGACATTTGATAAGGGAAAAGTCAATTGAGCTG
 GCGCCAGGCTGTTCTCTGCTACTGTGGCCGCTGGCAACAGGATCGAAGAGACATGCGAGGAGCCAGGCGCTGCTGAACCCG
 AACAAAGCAGCGACGGCAGGATACCGAAAAGCTGTTCAAGAAAATTTGGCCTGATCGACGATTTTGTAAAGGGCAGCTTCTACGAC
 AACTTTAATCAGAACCTGAATCCTTATGAGCTGAGAGTGAAGGGCTGACAGAACAGCTGACTAAGGAGGAACTGGCCGAGTCT
 CTGTACCAGATCGTGAACATAGGGGGAATTAGTTATGCACTGAAGGACGCCGATGTGGACGAAGGGCGGACAGACTACTCAGT
 AGCCTGAAAAATCAACAGCCAGGAGCTGGCAGAAAAGACTCCAGCCGGAATTCACTGCAAGAGACTGAATGATTATGGAAAAGGTG
 AGGGGCCAGGTGGTCACTCGGCGACGATCCAGACAACCAGAAGGTGCTGCTGAATGTGTTCCCCACATCAGCTTACGAGAAAGAA
 GCAAAGCAGATCATTGCCACTCAGCAGCAGTCTATCCTGAGAGCCTGACCCGACAAGTTACCCGAGGAATACTGCCAGATCCTG
 ACTCGCAAGCGAGATTATTTGTGGGGCCAGGAAAACGAGAAAAGCCGACTACGGGATCTACAAGACTGATGGAAAGAAC
 CTGGACAATCTGTTGAGGAACTGATCGGCCACGATAAAGTCAACCCGAGGAACTGCGGGCATCTGACGCCAGTTTATACCGCC
 CAGCTGTTTAACTGTGCTGAATGACCTGAACAATCTGAGAATCCTGAACACAGGATGGGAAACTGACAAAAGGAGGACAAGGAA
 AAGATCATCGCTGAAATTAAGAACAAACCCACAACATCAACATGCTGAATGTGATTAAGAAAAGTCCGCGGGTGTTCGAAGGAC
 GATATCAAAGGATTCAGAGTGAATGAGAAGGATAAAACCCGAAATCAGTCCATGCCCTGTGATCCGCAAAAATTCATAAGGACCTG
 CTGAAGGCCGCGTGGATATCTCAGACTGGCCCGTCAAGTTCATCGACGAACTGAGCTTTATTCTGACTGAACTGAGAAAT
 GGGGAAATTCGCAAAACAGCTGAACAATCGACTGGCCCTAAGTTCGATTTTCTGAACGCTGACCTGATCCAGCTGATCATTGAT
 AATAAGGACTCCTTTGAGATTAAGACTAACAAAGTGGCACAGATTGAGCTGAAAACCATGAACAACTGATCCAGAGATG
 ATGAAAAGACCCGTGGAGCAGATGACCTGCTGAATGAAAATGGGACTGGTCAAGAAAAGATAAGAAAACGCTTTGAGAACAAATAG
 TACCTGCCTTACAAGGAAATCGCAAGGACATTTTCAACCCAGTGGCCTCAAATCTGTCCGCGAGGCCCTGAAGATCGTGAAT
 GCTGTCTGAAGAAAATACGGCCACATTTGATTATCTGGTGGTGAATGCCTCGGGATAAAAACTGAAGGAGGAACAGGACAAT
 ATCAAGGAGTTCCAGAACAAAAATAAGAAAAGCTAAGGACGCTCAATTCGAAGCATTTGTGAAATCAGTCCGGAGCCGACAGAGA
 GTGAAGGAAGCCCTGTCTAAAAACCGGAAGCTGCAGATGAAGATGAGACTGTGGTATCAGCAGCAGGAGATCGATCCATATAAT
 GGAAAGACAATCGATGCCACTGACCTGATTAACAATCCTGATAAGTTGAGATTGACCATATCATTTCCACAGAGTATCTCATT
 TACGACAGATTAACAATAAGACCCTGTCTTCCGCTCAATGAACCAAGTGAAGGACGAAAACCCCTACGAGTTTATGCTG
 GAAGGCCACGGGCACTTATGACAAGTTCAAAGCTACAGTGTGGCAAAACAAAGAAATTTGGCAAGGCTAAAAGGGCAAACTAC
 CTGTTCCAGGAAAATGTGAGCGATATCGAGACTCGGAAGAGATTCCTGTCCCGCAACCTGGTGGACACCCGATATTTAGTCCG
 GTGGTGTGAAACAGCCTGCAGGATTTCTTTCGGGAGAAAATCTGCCGACCAAGGTGACAGTCAATTCGCGGCAAGTTTACCTCC

FIG. 9G

AACATGCGAAAAATTGGCACATCGATAAGACTAGGGAGACATTCACCATCACGCCATTGACGCTTCTATCATTGCCGCTACA
 CCATTTCTGCGCATGTGGAAAGAAAGGAGGCACTATCTTCCCGTGAAGGTGGGAGAAGAAAGTATCGATATTGAGACAGGCGAA
 ATTCTGGACGATAAGAAATTTGACAAAGCAATGTACGAGGAACCTATAGTGGCTTCGTGTCAGAGATCATGAACGCCGACGAT
 CGGATCAAGTTTCCAGCCAGGTGGATAAGAAAATGAATAGGAAGGTGAGCGACGCCACCATCTACAGTACTCGCAACGGGAAA
 CTGGCTAAGGATAAGAAAGACGCTGAGTACATCGTGGCAAAGGTCAAAGATATCTACAGCGTGGACGGATTCAAGAAGTTCAAG
 AAAGTCTACGATAAGGACAAAACCAAGTTTCTGCTGTACAAATATGATCCTAGGACATTTCTCAAAGCTGGAGCGCATATTAGC
 GATTGCCAGACAAGTGGAAAAGGTCCAGACAACCGCAAAGTGAAGGCTGTCGATATCAGTCCATTGAGATGTACAGAAGG
 GACCATGGGATGATCAAGAAAATACTCAAAGAAAGGAAAACGGCCCGCCATCAAGCAGCTGAAGTACCTGGATAAGAAAAGTGGG
 AGCCACATCGACATTACCCCGCAAACGCCAATGGAAAACACGCTGATCCTGCAGAGTCTGAAGCCTTGGAGAACCGACGCTCTAT
 CTGAACCAGAGACAGGCGAGTACGAAATCATGGGGAATTAAGTATAGCGATCTGAAGTTCAACAAGAATGAGGGGTACGGAATC
 AAGAAAAGACAAGTATCTGGAAAATTAAGAAAAGTGGAGGAAGTCTCCGAGAAGTCTGAGTTTATGCTTTAGCCTGTACAGGAAGGAT
 CGCGTGAAGTCCAGGACATGAAAACCGCGAGTCCGTGGAAGTCTGTTCTGGAGCAGGAACTTTCCAATAAGAAAATACGCT
 GAGCTGAAGCCATCTCCAGGCGAGAAAACGACAAGAACTGCCTGTGTATGGCAAAGGGAGACTGATCAAGAGGCTGATTCCC
 AAAAATGTAAGTCTGGAAAAGTGAATACCAAAATCTGGGCGATCCCTACTATCTGGAGAAAAGGCACTCCCTAAGGAT
 ATCTGGACTAAGAATC

#12

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCCGCGAAGAAAAGCGCAAGGTGCAAGCGTCCATGAAGAAA
 ACATGGGACTGGACCTGGGGACTAACAGCATTGGGTGGGCGCTCATCAACTCAAACATCGACTCAGAAGGAAAAGAAAAGCTG
 TGGGGATCAGTCTCGGGAAGCCGGATCATTCTTACGACCTACCACACTGGGAGATTTGGAAAGGGCAACACTAAAAGT
 CCAGTGGCAGAGCGAACCCGACTGCGGGGCAATCCGAGACTGCTGGAAAGGTCACTGCTGAGGCGGAGCGACTGCACCGAGTG
 CTGTCACTCATGGGGTTCTGCGCAGAGCATTACGCTAGCCAGCTGGACCCTATGGAAGTTTCTGCCAGAAAACCGAGCCCAAG
 CTGGCATGGTACAAAGACGATAGCGGCGAGTATCAGTTCCTGTTTCAAGAAGTCTTCCAGGATGCTGGAGGACTTTTCGACAG
 CATCAGCCAGAACTGGTGGCAGGAGAGAAAGAAAATCCCTTACGATGGCAATCTACTATCTGCGGAAAGAAAGCACTGTCTCAG
 GAGTACTAAGGAGGAACCTGGCTGGATTCTGCTGAATTTCAACCAGAAAACCGGCTACTATCAGCTCGAGGGGAGGAAGAG
 CAGGAAGAGAACAATAAGAGCGTGGAGTATCACGCCCTGAAAGTGGTGAAGCTCGAGGACTTGGAGAAAAGAAAGGGCAAAGAT
 ATCTGGTACAATGTGACTCTGGAGAACCGGATGGTCTATCGACGGGCTAGCAACATTTCCCTGGACTGGACCGGCAAGGTGAAA
 GAGTTCGTGGTCACTACCGAGCTGGACGATGCCGGGAACCCAAAGAAAGATAAGGAAGGAAATGTGAAAAGATCCTTTAGGGCA
 CCAAAGGAGGACGATTTGGGGGCTGCTGAAAACFAGGACCCAGGCTCAGATCGACGAATCCGGCAAGACCGTGGGAACTTACATC
 TACGAGTCTGTGTGTATGCCAACCCAGAAGATCCGGGAAAACCTGGTGAAGACCAATTGAACGCAAGTACTATAAAGACGAG
 CTGAGACAGATCCTGGTGAAGCAGAGCGAGTTCACGCCGCTCTGCAGGATCATAATCTGCTGTCTGTATCGAAGAGCTG
 TACCCTAAACATGAGGGCCACAGAAGGCTGTGAGCGCCAGCAGCTCATCTACTTCTGTATCGAGGACATTTGTTTTATCAG
 CGCCACTGAAGTCCCAGAAAAGGGCTGATCGATAACTGCCCTACGACTCAGATCTACATCTACAAGGATAAGAAAAGTGAAGTCTG
 CACCATGTGCCCTCTGAAAGTGTGTGAGCAAAATCCATCCACTGTTCCAGGAAATTTCCCTGTGAGCAGTTCTGTCTCAACTCGGA
 ATCTACAGAGAGAGAGGATGATGGACGGCAGTCTGAAAACCTGGACGCTGGATGTCAACCAGGAGTTTCTGCCCTCAGAAGAGGAC
 TACGTGAAGCTGTTTGAATGGCTGAATGAGAGAAAGGAAATCTCAGAAAATTTCTGCTGGCTTATAAACTTTTGGGCTGAAG
 AAAAAAGCAACAGGCAAAATACAGATGGAACTATGTGGAGGACAAAGGACTACCCCTGCAACGAGACAGGGGCAAGAAATCAAGAGC
 AGACTGTCCAAAGCCGAGTGCCTGAAGAGTTTCTGACTGAAGAGAAGGAAGAGGCCCTGTGGCACATCTGTATTCTATTAGT
 GATAAGAAAAGAGCTGACTAAGGCTCTGGGCACCTTCCGAGCCAAAACCTGTCTGAATGAGTCTTTCTGGAAGTCTTTGCCAAG
 ATCCCCCTTTTGGTCAAACTACGCTGCATATAGCCTGAAGGCTATTAGGAAAAGTGGCACTGATGCGCATGGGGAAGTAC
 TGGAATGAACAGGCCATCGACAGGCGACTCGCGATCGAATCGAGAAAATTTGACCCGAGAGTATGACGAAAACATCCGGAGC
 AGAGTGAAGGAGAAGGCAATGCTGCTGACCGATATTAGCAGTTCGCGGGGCTGCTCTGTGGCTGGCTGTACATCGTGTAT
 GACCGCCAATCAGAGAGCAAGGAACTGGTCAAATGGGAGACACCAGCCGACATCGATCAATTTCTGTCCAAGTTTAAACAGAAC
 AGCCTGCGGAACCCCATCGTGGAACAGGTCAATTACAGAGTCCCTGCGCACTGTGCGAGACATTTGGAAGCAGGAGGGAAAAATC
 GATGAGATTCAGTGGAACTGGGCGGGAGATGAAGAACCTGCAAAAAGAGCGCGCCGAATACAGCTCAGGTGCAGGAAAAT
 GAGAACAATACTGAGAATCAAGGCTCTGCTGGCAGAGTTCATGAACCCGAAATTTGAGATTGAAAATGTGCATCCATACTCA
 CCCGGCCAGCAGGAAATCTGCGGATCTACGAGGACGGGCTGTGAGCGGGATCGCTGAGAAGGATCTGCTGAGGACATCCAGATAGCTTC
 GCAATTTCTGAAGAAATTCAGAGAAAACGACGCTGAAGAAAACGGCAACAACTAGCGAAGTCTGCGGTAACAACTGTGGCTGGAG
 CAGCGGTACAGATCCCATATACCGGAAGAGTGTACCCCTGGGCAAGCTGTTACACCTGCTTACGAGATCGAACCGTGTATT
 CCCAGACCGGATTTTTGACGATCCATCTCTAACAAAGTATCTGCGAAAAGTGGCGTCAATAAGCTGAAAAGATAACTGTCTG
 GGCTATGAGTTCATCAAGAAAATTTCCGGGAGATGGTGGAACTGGGGAATGGAGAGACAGTCCCCTGTTACGCGTGGAGAG
 TACGAACGGTTCGTAAGGAGTCTTACTTCCGCAACAGTAAGAAAATGAAGAAAAGTGTGCTGGAGGACATCCAGATAGCTTC
 ATTTGAGAGACAGCTGAATGACAGTTCGATACATCTCACGGTGGTGCACATCTGCTGAGTAACTGGTGTGCGAAGAGGGAGAG
 CAGGATGGCCTGTCCAAGAATGTGATCGTGTACCCGGCGGATTACAGACAGGCTGAAGAAAAGATTGGGAGTGCAGGAAGT
 TGGAAACCGCATATTCTGCCTCGGTTCTGAGACTGAATGAGATCACCGGACGGACAGACTTTACAAGTACTTCAGTGAACCGC
 CACTGCTGCCCTGCCCTGCACTGTACTGCAGAAGGCTTTAATAAGAAAAGAAATGACCATAGGCACCATGCCATGGATGCT
 ATCGTATTGCCCTGCGTAACCGGAATATCGTCAACTACTGAAACAATTTCTCTGTAGAAAAGAACAGCGAAATAGCCGATAT
 GACCTGCAGCGGCTGCTGTGTGAGAAGGTGAAAACCGATGCCAACGGCAATTAACAAGTGGATCCTGAGGAAAACCATGGGAGACA
 TTCACCCAGGATGTGTATGCCGCTCTGACAAAACATCGTGGTCACTTTAAGCAGAATCTGCGCGTGAATTAATCGAACCAAAAC
 TACTATCAGCACTACAACGAGCGAGGGGAAAAGCGCATGATCCCTCAGACCAAAAGGGCAGACATGGGCAATTAAGAAAAGCAATG
 CATAAAGATACTGTGTATGGCGAAGTCAATCTGAGAAAAGGAGAAAACCTGCCACTGAGGGACGTGGTCAAGAACCCAGTATC
 GTGGTCGATAAATCACTGAAGAAACAAGCTGTACGAGCTGCTGAAGAGCCAGTATGACCTGAAGGCAATCGCCAAATACTTCGAG
 ACACACAGGACGTGGGGCAGATGTCAACCTGAAGAAAATTAAGGTGACTACTTCAAGGAGACAAAACGAAAAGATCTTT

FIG. 9H

GCCACTAGGAAGAGCCTGGACCCATCCTTTGATCAGAAGAAAATCGAAGAGGAAGTGACTGATACCGGGCTCCAGAAGATTCTG
 CTGCACCATCTGCAGCAGAAACAATAACGACCCTGATATGGCCTTTTCCCAGACGGCATCGATAGGATGAACCCAGAATATGACC
 ATTCTGAATGACGGGAAGTGGCCACGACCCTACTCAAAGTGCACACATATGAAAAGGCAGATAAATTCGCCGTGGCGAGTCT
 GGGAAACAAGGCCAAGAAAATTTGTGGAAGCAGCCAAAGGCACCAATCTGTTCTTTGCCGTGTACGAGAGTGTCCAGGAGGACGAA
 GCTATCGGGAAGCAGGTCTGCAACCGGACATTCGCCACTATCCCCCTGAACGAAGTGATCAAGAGAAAAGAAACAGGGCCTGCC
 GCTGCACCTGAGGACCTGAACGGGAATCTGCCAAGTTTGTGCTGAGCCTAACGATCTGGTCTACCTGCCACCGAGGAAGAG
 AGGAATAGTTCACGCATCATTACGCTCTGGACAGGGAGCGCATCTAAGATGGTGAGCTCCTCTGGGAGTCAGTGCTCTTT
 ATCAAAGTGTTCGTCGCAATTCAAATTTGGGATAGAAGCAATACAGCAGCCTGAACAAGATGGAGAGGGCTATTACAAACGAA
 ATGATCAAGGAGATTTGTGTCCTATCAAATTTGACCGCTGGCAATGTCAGCTGATCCAGATTTAAGAATTC

#13

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTCGCTTCGCCGAAGAAAAGCGCAAGGTGCAAGCGTCCATGACAAA
 ACTATTCTGGGACTGGATCTGGGACTAACAGCATTGGATGGGCACTGATTAACCAGGACTTCGATAACAAAAGGGGGAGATC
 CTGGGACTGGGAGCCGGATCATTCTATGACACAGGATTAAGGACGAGTTCGAAAAGGCAACTCTATCAGTCAGACAGCT
 GAGCGAACTCGACTCGGGGAGTCCGGAGACTGATCCACAGAACTCTGCTGAGGCGCGAGCGGCTGCATAGAGTCTGAATATC
 ATTTGGCTTCTCCGAACTATGCCAACCAGATTGACTTTGAGAAGAGGTTCCGCAAGTTCAAGCCTGAATCTGAGCAGAAA
 ATCAGTTTCGATGGAATGACGTGTTCTGTTGAGAAGAGCTACCAGGAAATGCTGGTCAAGTTTAAATCCACCAGCCAG
 CTGGTGAGCAACGGCAAGAAAATCCCCATGACTGGACAATCTACTATCTGAGAAAAGAAAGCTCTGTCTAAGAAAATTGAGAAG
 GAGGAACTGGCATGGATCTGTGTAACCTCAATCAGAAAAGGGGATACTATCAGCTGAGAGGGGAGGAAAGGAAAGAGACCA
 AATAAGCTGGTGGAGTTTCATAGCCTGAAAATTTGTGGAGTCAACGGCGATGAACCCAGAAAAGGAAAGCCTGAGATCTGGTAC
 TCACTGGTGTGGAAAAGCGGCTGGGTCTATCGACGGGTAGCAAGACTCCCTGTTTCGATTGGAAGACAAGATTAGGGAGTTT
 ATCGTGACCACAGAAATCAACAATGATGGAATGTCAAAACCGCAAGGAGGGCACCAGAAAAGAGGAGCTTCGCGCACAAAA
 CCCGAGGACTGGACTCTGCAGAAGAAAAGACCAGTTCGAGCTGTCAAAGAGCGGGAAATGAAGTGGGAGCCTTCATCTACGAA
 AGTATCTGCAGAAACCAACCAAGATTAGAGGCAACTGATTTCAACCATCGAGAGGAAATTTTATAAGGAAGAGCTGAAA
 ACCATCTGAAAACAGCTGTTCTTTCACAAAAGAACTGAAGGATGAGAAAAGTACAAATGCCTGCATCGAAGAGCTGTATAAG
 AACAACTGAAGCTCACCGGAGCCTGCTGTCCAACAAGGGTTCGAGCATCTGTTTATTAACGACATCCTGTTCTACCCAGCCT
 CTGCGGTCTAAAAGAGTCAAGTCTCAAACTGCCACTGGAGAAGCGCACATATAAAAAGGAGGGGATTGAAATCACTGAGGGC
 ATCAAAGTGTCTCAAATCTAATCCAATCTACAGGAGTTCGGGTGTGGCAGTGGATTAGCAACCTGTCCCTGATTTGATC
 GAACCCACCGAGCAAAATGTGACTTCAACCTTTCTGAACAGCATTAAGATTACGAGAACTCTGTTTCGAAATTTCTGAACAATCG
 AAGGAAATCGAGCAGAAGCACCTGCTGAAATATCTGCTGGAGAACCAGGGGTTTAAAGGAAAGCTGTGACAAAAGCACTGGAG
 AAGTTCCGCTGGAATTTTGTGCTGACAAAAGTACCCCTGTAATGAGACAGGACGCTGCTGCATCTCGGCTGAGCAAAAGT
 AAGAATTTCCCTGATTTCTGACCAAGGAAATCGAGCACCAGCTGTGGCATACTCATCTACAGCTGACCGCAAGATTGAA
 TATGAGCAGGCCCTGAAAACCTTTGCTCGGAAAACAACTCTGGATGTGACTCCTTTTGGACACTTCAAAAAGATCCCCCCT
 TTTGAGTCTACCTACGGAGCATATAGTGAAAAGGCCATCAAGAAAGCTGCTGCCACTGATCAGACTGGGCAAAATCTGGAAGTGG
 GAGGCCATTGATAGTATCTCAAAGGACAGGATTAAGTAAAATCCTGTCAAGGGAAATACGATGAGAAATTAAGAACAGAGTGAGG
 GAGAAAAGCTCACCTCCGAAACCAATTTCCAGGGACTGCAGGAGTGGCTGGCCAAGTACATCTGCTATGATCGCCAT
 TCTGAGGGCAATGACCTGGGGAAGTGGACTAGCGTGTCCGACTGGAGACATACCTGAAGGAGTTCAAGCAGCATAGCTGCGG
 AACCTATTGTGGAGCAGGTCACTACAGAAACTCTGAGAGTGGTCAAGGATTTGGATCAAGCAGCGGAAAGGAAACCGAAAAT
 TTTCTTTGACGAAATCCATGTGGAGCTGGCCGGGAAATGAAGAAACAATTCGAGGATCGCAACAGCATGACCAACAATTA
 GAAAACGAGAATACAAACCTGAGAATCAAGGCCCTGCTGATGGAATGATGAATGATAACGACGTGGAGAACGTGAGCCCTTAC
 TCTCAAAGTCAAGCAGGAGATTCTGAAGATCTATGAGGACGGAGCTGGAATAGCAACATCGAGCTGGACGATGAAATTTGTAAG
 ATCTCCAAAAGGCAGAGCCACCAAAATCTGAACCTGCAAGCTACAAGCTGTGGCTGGAGCAGAAATCCGATCCCTTATACT
 GGCCAGGTCACTCCACTGAACAAGCTGTTCACTCTGAATATGAGATCGAACACGTGGTCCCTCAGAGTCTGCTCTTTGACGAT
 AGCTTCAGCAACAAAGTGTCTGCGAGTCAAGCCTCAACAAGCGGAAGGATAACCAAGCTGGGCTGCAGTTCATCAAGAACCAT
 AGCGGAGAAAAGTGGAGCTGGGCTTCGGGAAGTGGTCCAGGTCTTTACAGAAGAGCAGTACCTGGATTTTGTGAAGGACAC
 TATAGCAAAAATCGCTCAAGCATAACAAACTGTGCTGGAAGAGATTCCCAGAAAGATGATCGAAAAGGACGCTGAATGACACT
 CGCTACATCAGTAAGTTCGTGAGCTCCATTCTGTCCAACATCTGTCAGATCTGAGAAGGACGATGACGGCCTGAATAGCAAAAAC
 ATTCTGCCTGGAATGGCAAGATCACTACCGAACTGAAAAGGACTGGGGCTGAATGATGTGTGGAACGACCTGATTTCTGCCA
 AGATTCGAGAGGATGAATACCATCACAAACAGCGATCTGTTTACAACCTTACAACGACAAGTATCAGAAAACCTGCCACCGTG
 CCTTTGAGTACTCCAAGGGCTTTGAGAAAAGCGCATTGATCACCGACACCATGCTATGGACGCACTGGTTCATCGCATGTGCC
 ACACGGGATCATCTGAATCTGATGAACAATCAGTCTGCCAAGAGTGAACGAAACGATACGACTGCAGAACAAAGCTGCGGAAA
 AAGGAGCCTTACTTCAACAGAAAGGAGAACAAAAGAAAGGAAAGCCTTCAAGGATTTTATCAAACCTGGGGCACTTTCAACCGAG
 GACAGCAAGAAATGCTCTGAAAAAATCATTATCTCCTTTAAGCAGAACTGAGAGTGTCAACAAGGCAACAAACTCATACGAG
 AGCTATAAGGATGAGGACGGGAATCTGAGGATCGGCAAGGATGGGAAACAGAGAAGGGCCTGATCAAGCAGAAGGGGCTGAAC
 TGGGCAATCAGAAAAGCCCTGCACAAAGACACCGTGTGAGGAGGATTAACCTGAGCAGGATCAAGCTGCCAATGGGAAAATC
 CTGACAGCCACTCGCAAGAATCTGGATACCAATTTGACCTGAAACCAATTTGAGAATCAATCACCGACACAGGCATTCAGAAA
 ATCCTGAAGAAATACCTGCTGTCCAAGGAATCTCCAGAGCTGGCCTTCTCTCCGAGGGCCTGGAAGAGATGAACAATGAGATC
 GAAAAGTACAACAATGGGAAATTTCAACATCTATTAACAAGGCTAGGATCTATGAACTGGGAGCAAAATCAATGTGCGACAG
 ACAGGCAACAGAAAGGATAAGTTTGTGGAGGCGCTAAAGGAACTAATCTGTTCTTTGCTATCTACCAAGGACGAGAATAAGAAAC
 CGCTCTTATGAAACTATTCCCTGAACGAAGTGTGAGCACCAGAAAGTGGGAGCAAGTCTGCCTAAGGAAGAGCAGGAGAAA
 ATTCCTGCTGGTCCCGTCAACAAGCTGAAGGGGACCTTCACTTTTCCCTGTCTCCCAACGATCTGGTGTACGTCCCTTCAAT
 GACGAGCTGGAACGAAGTCTCAATTTGACTTCTAAGCTGAAAAGGAAACAGATCAACCGGCTGTATAAAATGGTGTCTAGT
 TCAGGATCCAGTCTTCTTTGTGAAGAGTGGTTCGCAACCTCAGTGGTCAACAAAATGGAATACAGCAGCCTGAACAAGATG

FIG. 9I

GAGAAGTCTATCGATAACCTGATGGTGAAGGAGATTTGTATCAAACCTGAAGATTGACAGGCTGGGCAACATCAGCAAGGCCTAA
GAATTC

#14

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTGCAAGCGTCCATGACCAAA
ATCCTGGGGCTGGACCTGGGGACTAAATAGCATCGGGTGGGCAATCCGCGACACAGAAAAAGAGGGCATCAATCAGATCCTGGAC
AAGGGCGCCGCATTTTCAGCGAGGGAGTGAAGTCCGAGAATGGCAAAGAAATCAGTAGGGCAGCTGAACGAACCCGTTACCGG
AGCGCAAGAAAGATCAAATATCGGAGAAAACCTCGGAAGTACGAGACACTGAAGGTGCTGTCAATTAACGGAATGTGCCCCCTG
AGCGTGGAGGAAGTGAACAGTGGAAAGAAATCTGGCTTCAAGGAATATCCCTGAATCCTGAGTTTCTGGATTGGCTGCGGACC
AACGAGGACAAGAACATCAATCCTTACCCTGTTACAGAGATAAAGCTAGCAAGCAGAAAAGTGAACACTGTTTGAACCTGGGAAGAGCT
CTGTATCAGATCGCACAGAGGGCGGGCTTCTGAGTAAACAGGCTGGACCAGTCAAGCCGAGGGCGTCTTTGAGGAACATAATCCT
CAGATCCAGAACCTGATTGAGGACTGGACAGCTCCAACAATTTCTGAATGAGCTGAAGGAATACTATATCAATCTGGGGATC
ATTGACGAGACTGAAAAGTCCGGCTTCAAGAAAGACCTGGATGAGGGGAAAAGAACTGAAGTCACTGTACAACAGCCTGGT
GCCATCACAAGAAAAACGCTAACGATATCGAAAACCTGCAAGCAGGAGCTGATCGTAGACTGAATAAGAAAGAGGACCTGGGC
AAGGTCAGAGGGAAGTCAAAGATATTTCTCAGGCAATGCTGGACAGGAACCTTAAAGACTCTGGGACACTTTTCTTTCACTG
TACAACAGGAGCGAATCGGAAACAGTATACCAGCCGAGGAACACTACCTGGAGGAGTTCATCATTATGTCAGACAGCAG
GGGATCGAGGGAATTAACCAAAATGAGAAGCTGCTGAGAAGAAAGTTCACCGGCCTGGCAAAAAGATCTGTATCGAGCATTTC
TTTCAAGCGGCACTGAAATCCCAAGAGGGCCTGATCGGGAAGTGTCTTTGAGAAAGAACAGTCTCGTGTGCCGTGAGTCA
CCAGACTATGAGGAATTCGAATGTGGAGCTACCTGAACACAAATGGAACCTGACTCCGAGAAAACCTCGCTGCTGCTGCTG
ACACTGGAGGAAACAGCTGAAAACCTGGTGCCTAAGTCTACCGAAAGAGTGAATTTCACTTGGAGGTGCTGCTAAGGAGCTGGT
GAAAAGGAGCATTCTCGGCTACTACAAGTCTAGTAAAGAAAATGAGTCTTTTACTGGTTCAACTATAAGCCACCCGATTC
GTGAGCGCTGCGTGGTGAAGCTTCTGAGAAACGCAATCGGCAAGGACTGGAAGTCAAGACTTTCAAGATCAGACTGAC
AACACCGGAAAGAAATGAAGTGAACCAAAATCCGTGATTAACAAGGACTGTGGCACCTGCTGTCCGTGGCAACACTGTGATCT
CTGTATGACTTTGCCATCGAAAAGCTGAAAACCTGGAGCCTAAAACCGCAAGGCTTTCAGCAAGACAAAACCTGAAGAAAGACT
GCCAGTCTGCTACTGGCAGCCATCGCTAAAATCTGCTCATATCTGAAGCAGGGCCTGCTGTACTCCACGCGCTGTTCTGAT
AACATCGAGAAATATGTCGACGCGATATCTGGAAGGATGAGGAACAGCAGAAAGTTCATCCAGTCCAAGATTGTGGAGCTGGT
GACAATTATATCGTGGAAAAGTCTAAAACCTGGAGCTGATCAATGGGGCTGCTGAAGATCTACAACACCGAGGATAAGGAAGGACGG
AAAGTGTACTATTCAAAGGAGGCTGAAAGCCTGTTGAGGCGAGCCTGAGAAAGAAACTGGTCCCCTTCAACAAGGCTAACATC
ATCATCGAGGAAACAGCAGCAGGAAATCATTTTCCAGGATCTGTTTTCTATCTCTGATGGACCAAGTGAAGAAACAGGAGTTTATC
AAAATTTGGCAGACTGGATAAGCAGATTGAAGCCTTCTGGAGGGGAAAAATGAGGAAGGACAGATCTTTTGAACACACAGAT
AAGCTGAAGAAACTGTACCATCAAAGCAGATCGAGGTGTTTAAAGAAAAGACTATCAAAGATGAGTGGGGAAATGAAAAGGTG
GTCTGGGATCCCACTGACCACATCTATCAAGAACCCATGGCAATGAGAGCCCTGCACCAGCTGAGGAAGGTGCTGAATACA
CTGATTGCCAACGACAGATCGAGCAGGATACTCGGATCCATATTGAGATGGCAGAGAACTGAACGATGCTAATAAGCGAAAA
GGCATCCAGGACTTCCAGAACGAGAACAAAGAAAGTTCGGGAGGAAGCCATCAAGGAGATCAAGAAAGTGTACCTGGAGGATGC
CACAAAGACGTGAACCCACCGAGGACGATATCTGAGGTATCAGCTGTGGCTGGAAACAGGGAAGTGGGAGATCTACGAGGAA
GGCAACAATATCAGCATTGTGATATCAATGGCAGCAATCCCTCCTATGACATCGAGCATACCATCTCTCGGAGCATCTCCAG
GATAACAGCCAGATGAACAAGACACTGTGTAGCCTGAAGTTCAACAGATCCATCAAAAAGCAGAAAGTGCAGTGGAGCTGTCC
AATCACTCAATTTCTGCCCAGGATCGCACACTGGA AAAAGGAGGCGGAGGAACTGACTAGGCAGATCGAAAACCTTTCTCGC
AGCATCAAGAGCGCTGCAACCAAGGTGGCCAAAGGATAAGAACATCAGAAAGAGGCAATTACTGACACTGAAGCGAGACTATATT
CTGGGAAATACGAGCGGTTCACTTGGGAGGAACTAAAGTGGGCTTTAAAACTCCAGATCCAGACACTGGGATCATTACC
AAGTACGCTCAGGCATATCTGAAATCTTACTTCAAAAAGGTTGAGAGTGTCAAGGGAGGAGCAAGTGGCTGAGTTCCGGAAAAC
TGGGGCATCCAGGAATCTTTATCGATGAGAAGTGGTGAAGCACTATAAGGACAAGATAGAGACAAAACATACCACCATACA
ATCGACGCAATCACTATTGCTGTATGCCCCAAGGATAAAACAGACCTGCTGGCACACGCTGGAGGCTGGAGGATGAACAGGAC
AAAAAGGCGCTAAGGTGCTGATCGAGCAGGCCAAAACCTGGA AAAACCTTCAAGGAGGATATCGAAAAGATTGAGACTGAAATC
CTGGTGAGCCATTTTACCCCGACAACGTCAAAAAGCAGTCAAAGAGCATCATCAAGAACTCGCGCAAAAAGGTTGACGTCCTG
AAGAACGAGCTGCCTGTGAACCTCAAGAACAAAGATCGAAGGGAAGGATTATTTCAAGCTGAAATTTGACAGCAAGATTCTGTAC
AAAATCCCAAAAAGAAAAGAGAAGCAGACCGATACATTTCTATGAGGAACTGCCTAAAAACTACCTGAATGGAGTGGAAAGGCAAG
GACTACTTCAAATCAATACTACCGGCAAGACTTCTACAAAATCCCAATTTTAAACAGGGCGACACAATCCGGGGGAGCCTG
CACCAGGAGACAATACCGGAGCAATTAAGCTGCCAGATATCGACATTGAAAACAAAGAAACCCCTGCACTACTGATAAGGGAGGC
TTCAATCTGAAGAAAAGACATCAAAGGCAATGAGATTGTGTTCTTTGTTGGTCCGCAAGGAAATCTCTAAAATTAAGTGAAGCAG
GTGCAGAAATCGTGCACAACGTGGTCCGGAAGAAAATGAGAATGCTATCGAAAACCTCCTGATCACTTTTAAAGATCGTGAAG
AAAAAGAAAGTGGCCGTCAATCAAGAGGGGGTCACTATTTGGATGAGGGAGCCCAACTTGA AAAAGGGGATCGAGGGAATCCCT
ATTAAGAAAGTGCATCAATTAACAATTTCTGTCAAAAACCTATCGAGATTAAGGTGCAAGTCCAATGTCCAAGTCTCGCCAC
AAGCATAAACAGAAAGTCTACGGCCAGAACGATGAGAATTAAGCCATGGCTGTGACGAACTGGACGGGAAGCGGGAGTTGAA
CTGATCAACAATTTAATCTGGCCAAGCTGCTGAAAACAGAGTCAAGTCACTATCCACTGCATAAAGAGAAGGAAATCAAGGGGA
AAGAAAATTTCTGGTGCCCATCGAGAAGAGAAAACAACAGGATGCTATCTGAAAGAGGGGCCAGCAGGTGGTGTCTACGACAAA
ACCGTGGAGAACCCCAAGGATATCTCTGAAATCAATTGACTTTGCGGAGCGAATCTACATCAATTGAAGGGCTGACCATCCAGAGA
CAGAAAGATAAGAAAACATCCAAGTGAATGAGTACGGAATCAATTCAGCTGCGCCACTTCAAGGAAAGCTGAAAAGTGAAGGAA
ATCTCAAAGGATAACTTCAAACCTGACGGCGAGTTCAAGATCAACAGGAAATAAGCCAACCTAGGAAAATGAACCATATCAGTT
ACCGCTTTGTGGAGGGGATCGACTTCAAGGTCTGCTAGCGGAAAGTTTCAGAAAATTAAGAAATTC

#15

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTGCAAGCGTCCATGAGCAAG

FIG. 9J

AAAGTCAGTAGACGCTATGAAGAACAGGCCACAGGAGATTTGTCAGAGGCTGGGGAGTAGACCTTATTCATCGGGCTGGACCTG
GGAGTGGGATCTATCGGAGTGGCAGTCGCCGCTTACGACCAATTAAGAAAACAGCCCTCCGATCTGGTGTTCTGTCAGCTCCAGG
ATTTTTATCCCTTCTACCGGCGCAGCCGAGCGGAGACAGAAGAGAGGACAGAGGAACAGCCTGCGCCACCAGCAAATCGCCTG
AAATTCCTGTGGAAGCTGTGGCTGAACGAAACCTGATGCTGTCAATATAGCGAGCAGGACGTGCCAGATCTGCACGGCTGAGA
TTTGAGGACGCTGTGGTCCGAGCAAACCATACGAGCTGCGGCTGAAGGGCTGAATGAACAGCTGACCCTGAGCGAGCTGGG
TATGCACTGTACCACATCGCCAATCATAGGGGATCTAGTTCAGTGCACACATTCTGGATGAGGAAAAGAGCTCCGACGATAAG
AAACTGGAGGAACAGCAGGCTATGACAGAACAGCTGGCAAAGAGAAGGGAATTTCCACTTTCATCGAAGTGTGACAGCCTT
AACACTAATGGCCTGATCGGGTACAGGAACCTCGAGTCTGTGAAGAGTAAGGGCTGCCAGTCCCCACTCGCGACATCATTCA
AATGAGATTGATGTGCTGTGACAGCCAGAAGCAGTCTATCAGGAAATCTGTGACAGCAGTACTGCGATCGGATTGTGACG
GCAATCTGTTTGAACAGAGAAGATCGTGCCAGAAGCCGGCTGCTGTCCCTATTTCCCTGACGAGAAGAACTGCCAGATGT
CACTTCTGAATGAGGAAAGGCGCCTGTGGGAAGCCATTAACAATGCTAGGATCAAGATGCCCATCGAGGAGGGCGCTGCAAAA
CGCTACCAGAGTGTTCATTACGCGACGAGCAGAGACATTTGTTTCATATCGCAAGGAGCGGACTGATATCACCCCTAAA
CTGGTGACAGAAGGAGTCCCAGCCCTGAAAACCTCCATCATTGTGCTGACAGGAAAAGAGAAGGCTATTGAGAAGTCCGAGGC
TTCCGATTTCCGACGGCTGGAGGAAAAATCTTTTTGGAAGAGACTGAGTGAGGAAACAGAAGGACGATTTCTTTAGCGCCTGGACA
AACACTCTGACGATAAAAGACTGTCCAAGTACTGTGAAACACTGCTGCTGACAGAAAATGAGGTGGTCCGACGCCCTGAAA
ACCGTGAGCCTGATGGAGATTATGGCCCAATCGGCAAGACCCGCAACACAGCTGCTGATGAAACATCTGGAGGATGGCCTGACT
TACACCGAAGCCTGGAGCGGGGAATGAAAACCGGCGAGTTCAGGAACGTGCTGAGTGGGAGCAGCAGAGCCTGCTGCCCTAC
TATGGGCAGATCTGACAGGATCTACTCAGGCCCTGATGGGGAAGTATTGGCACTCTGCTTTAAAGAAAAGAGAGACAGTGA
GGATTTTAAAGCCTAACACAAAATAGCGATGAGGAAAAATACGGCAGGATTTGCCAACCCAGTGGTCCATCAGACTGAAACGAA
CTGGCAAGCTGATGAATGAGCTGATTACCCTCTGGGAGCTAAACCTCAGGAGATCACAGTGGAACTGGCAGGAGAGCTGAAG
GTCCGGAGCTGAGAAAAGAGGACATCATTAAAGCAGCAGACCAACAGGAAAAGGAGGCTGTGGCATAAGCAAGTATGAC
GAGCCCAACAATCTGGACAAAAGGTATATTGAAAGTTTCCGCTGCTGAGGATCAGGCCCTTTGTGTGCCCTTACTGTCTGGAG
CACATTAGCGTGCAGATACTGCAGCTGGAAGGGCAGACGTGGATCATATCTTCCACGCGACGATACAGCTGACAACCTCTAC
GGGAAATAGGTGGTCCACACCGCAGCTGTAACGATATCAAGGGAAGCGGACCCCTATGACGCTTTCAGTAATACATCAGCC
TGGGGCCCATCATGCATTACTGGACGAAACCCCTGGGATGTGGCGAAAAGAAGGAAGTTTGAAGCAACAGGAGGAGAT
GCTAAGTACCTGAGTCAAAAAGGCTTTCGTGAGCAGGTTTGAAGCGATAACTCTATATCGCAAAAGCTGCAAAAGGATCCTG
CGCTGCTGTTCAATCAAAACAATGTGACTGCCGTGGGTCCCTGAAGGGAATGGAGACATCTATCTCGGGAAGGCTGGAA
CTGCAGGGAATTGACGATCTGCTGGGACGCCGACTGGAGTAAGGACGCCGATACCAGCCCAATGGCAAAAACCGGGAC
GACAACTGGGACCATGGCTGGACGCCATCGTGGCTCTGTATTGTTCCAGATCTCTGGTCCAGATGATTAACACCATGTCCGAG
CAGGGCAAGCGAGCAGTGGAAATCGAGGCTATGATTCTATCCAGGGTACGATCCGAACCAATCTGTCTTTGAAAGCCAG
CGGGAGCTGTTTGAAGAAAATCCTGGAGTTCATGGACTGCACGCCCTTTGTGAGTATGAAAACCGCAACAGTACAAATGGC
GCCCTGCTGAAAGATACTGTATTCAATCTGGGAGCAGACACCCAGGGAGAGGATCTGGTGTTCGTTGTTCAAGAAAAGATT
AAGGACATCGGCGTGAAGTTCGGGGATTATGAAGAGTGCATCTGCCATTGAGGCGCGATCACCAGCAAAACAGCAAAAGTGG
TATCCCATGGAATGAAAGATAAGATCGAGCAGCTGCAGTCTAAGAACGAAGCCGCTCTGCAGAAATCAAGGAGAGTCTGGT
CAGGACGCCCTGCTCTGGAAGAGATTAAGGAAGCTGATTGAGTCAAGGCAAAAGCCCATCCAGCTGAGTGAAAAAACAATT
TCAAAAAAGGCCCTGAGCTGGTGGGCGGGTACTATTACTGATTAGCAACAACAAGCGCACAAAGACTTTCTGGTCAAGGAA
CCTTCAAACGAGGTGAAGGGTTCGCATTTGACACTGGAAGCAATCTGTGCTGGACTTTTATCAGATGCCAGGGAAAAGCTG
TGTGGCGAGATCATTAGAAAATCCAGGCTATGAACCTTCTATAAGCCAGCATAATGAAACAGGGGTATTCTCTGTACGTG
AGACTGTACCAGGGCGAGCTGTGCGAGCTGAGGGCAAGCGACTGACTGAAGCAGAGTCCAACCTGGCCAAAGCACACATGTC
CGCTGCCAATGCTAAACCTGGGCGAACCTTCTGTGATCATTACCTTTACAGAGATGGGGTCTGGATACAGATCTACTTC
AGCAACTGGCAAGTCCAAAAGGGACAGGACACTAGTTTTACCTGACTACCATCAAGAAATGATGTGCGGAAAAGTCCAG
CTGCTAGTCCGGGCTGGTGGAGATACGTGAGCCCTCTGCTGGTGGACAAAATCGAGAAGGATGAAGTCCGCTCTGTGTGGAGAG
TAAGAATTC

#16
ACCGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCCGCGAAGAAAAGCGCAAGGTCGAAGCGTCCATGGCAAGA
CCTGCATTTCCGGCCACTCGGAGAGAACACGTCACCGCTGGACCCCTGACCCACATCGGATTAGCAAACCTTTTTATCCTG
GTGAGCTGGCACCTGCTGTCCCGGGTGGTCAATTGACAGCTCCTCTGGATGCTTCCAGGCACAGCCGGGACCACACCGACAAG
TTTGCCGAGTGGGAATGTGCTGTGACAGCCCTACAGGCTGAGTTTTGACCTGGGGACCAACTCAATCGGATGGGGCTGCTGAAT
CTGGATGCCAGGAAAACCAAGGAGATCCGAGCACTGGGGTCCCGCATCTTACGCGACGGACGGGATCCCAGGACAAGGCT
TCTCTGGCTGTGGCAGGAGACTGGCCAGACAGATGAGGCGCCGACGGGACAGATATCTGACTGAAGGACCAGGCTGATGGGA
GCTCTGGTGGCTTCCGCTGATGCCAGCAGCCCTGAGCTAGGAAGCGCCTGGAAGTGGCCGTGCATCCATACCTGGCACGA
GAGCGGGCCACAAGAGAAAAGGCTGGAGCCTTCAAAATCGGGAGGGCCCTGTTTACCTGAAACAGCGCCAGGATATAAAACCC
GTGCGCACCGCCACAAGCCTGATGAGGAAGCCGGCAAGGTGAAGAGGCTGTGAAAGGCTGGAGGCAGCAATCGCTGCAGCC
GGAGCACCTACTCTGGAGCTTGGTTCGATGGCGAAAAACAGGAGAGAACTCTGCGAGCAGACTGGCTGGGAAGGGAAAA
GAGGCTGCATACCCATTTATCCCGCACGGGAATGCTGGAGGCCGAATTTGACACTGTGGGCGAGAGAGGCAAGGCACAT
CCAGATCTGCTGACCGCCGAAGCTCGCGAGATCCTGCGGCACAGAATTTTTATCAGCGGCCCTGAAGCCACTCCAGTGGGA
AGATGCACCTGTACCTGACGATGGGAGAGCTCTAGGGCACTGCCAAGCGCTCAGAGGCTGCGCCTGTTCCAGGAGCTGGCC
AGCCTGAGAGTGTCCACTGGACCTGTCCGAACGCCCTGACCCAGCTGAGCGAGATCGGATTTGGGCAATTTGTCAGGGC
AGACCCCTAAAGCCGGAAGGAAGCCTGGCAAGTGCAGAAGAGCGTCCATTCGAAAAGCTGAGGGGGCTGCTGGAGCTGCCA
CCAGGCACTGGGTTTTCTCTGGAGAGTGACAAAACGGCCTGAACTGCTGGGCGACGAGACAGGCCAGAAATCGCACAGCATT
GGACCTGGATGGACAGCTGCTCTGAGGAAACAGGACGCCCTGGTGAAGTCTGCTGACAGAGGCGAAGCCAGAGGGCA
ATTGACGCTGACTGACGATGGGCTTGGACGAGGCAACTGACGCAAAAGCTGGTGGAGCAACCTGCCAGATTTTACGGA

FIG. 9K

CGATATGGCAGGCGCGCAGTGGCTGAACTGCTGCCTGCTGGAACGCGAGACACGAGGCGACCCAGATGGGAGAGTGAGGCC
 ATCCGGCTGGACGAGGCGAGTAACTGCTGAGAGGCGGGAAGGATCACTCAGACTTCAGCCGGGAAGGAGCTCTGCTGGACGCA
 CTGCCCTACTATGGAGCCGTGCTGGAGAGACATGTCGCTTTTGGGACAGGAAACCCCGCAGATCTGAGGAAAAGCGGGTGGGA
 AGAGTCGCCAATCCCACTGTGCACATCGCTCTGAACCAGCTCGGCATCTGGTGAATGCAATTCTGGCCAGGCACGGCCGCCCT
 GAGGAAATCGTATTGAGCTGGCACGGGACCTGAAAAGGTCTGCCAAGATCGACGGAGAGAGGACAAGCGGCAGGCCGATAAC
 CAGAAAAGAAATGAGGAAACGCAAGCGACTGATCCTGAGTCTGGGAGAGCGCCAAACCCACGAAAACCTGCTGAAGCTGCGGCTG
 TGGGAGGAAACAGGGCCAGTGGAAAATAGGCGCTGCCCTACTCTGGGAGACAATTAGTATGAGAATGCTGCTGAGCGAGCAG
 GTGGACATCGATCACAATTCTGCCATTGAGCGTGTCCCTGGACGATTCCGCTGCAAAACAAGGTGGTCTGCTGCGGGAGGCCAAC
 AGAATCAAGCGGAATAGATCTCCCTGGGAGGCCCTCGGCCATGACAGTGGAGAGATGGGAGGGATTCTGGCACGAGCAGAACT
 CTGCCCAAGAAACAAAAGGTGGCGCTTTGCTCCTGACGCATCGGAGAAAACCTGGAAAGGAGGGAGGCCCTGCCAGCACGACACTG
 AATGATACAAGGCATCTGAGTGCCTGGCCGTGGAGTATCTGCGGTGCGTCTGTCTAAGGTGCGGGTGGCCAGGCCGACTG
 ACTGCACTGCTGCGAGCGAGATGGGCATCGACGCCAATTCTGGCAGAAGCAGATGGACCTCCACCAGAAAGTGCCTCCGAGAGACA
 CTGACCTGCTGCGAGGAGTGGGCACTGACGCGCAATTCTGGCAGAAGCAGATGGACCTCCACCAGAAAGTGCCTCCGAGAGACA
 TCAATGGTGCAGCGAGTCCAGTGGCCGCTGCAAGCGCAGAAAGAGGCGCTGCAAGGGAGGACAATACAGGCGCGTGTG
 GAGGGATTCAAAGAGGAACTTGGGATGGCTTTAGAGCTGAACTGGAGCGACGGGCACGGACCATCGTGGTGGAGCCACAGACCA
 GAAATGGGATTGGGGGAGCCCTGCATAAGGAGACAGCTTACGGGCTGTGGACCTCCAGAGGAAGGATTCAACTGGTGGT
 AGGAAACCAATCGACGGCCTGTCAAAGGATGAGATTAATAGCGTGGGGACCCCGGCTGAGAAGGGCAATACAGGCGCGTGTG
 GAGGGATTCAAAGAGGAACTTGGGATGGCTTTAGAGCTGAACTGGAGCGACGGGCACGGACCATCGTGGTGGAGCCACAGACCA
 GAAATGGGATTGGGGGAGCCCTGCATAAGGAGACAGCTTACGGGCTGTGGACCTCCAGAGGAAGGATTCAACTGGTGGT
 AGGAAACCAATCGACGGCCTGTCAAAGGATGAGATTAATAGCGTGGGGACCCCGGCTGAGAAGGGCAATACAGGCGCGTGTG
 GCCATTCGCGCAGCGGATGCTAACGACCCTGCTACCGCACTGGCCAAAGCAGCTGAGGATCTGGCAGCACAGCCAGCCTCCCG
 GGCATCAGAAGGTTGCGGGTCTGAAGAAAGAATCAACCCATTAGGGTGGAGCACGGCGGGAATCCAAGTGGACCCCGCTCA
 AGGAAACCAATCGACGGCCTGTCAAAGGATGAGATTAATAGCGTGGGGACCCCGGCTGAGAAGGGCAATACAGGCGCGTGTG
 GGACACTGGGTCACTGTTTCGAGGCACATGGGGGACGGGAGCAGAGCGGAGCTGCAGCCCACTAGACTGGGCGATGGGGAA
 AGATTTCTGATGAGGCTGCACAAGGAGACTGCTGAACTGGAGCATAAGGGCAGAGTGGGGTCACTGAGGTGGTCAAAGT
 GAACTAGTTCAAATAGCGTGGT
 CAGCTGCGCGCTCGAGGAGCACGGAGAGTGACCGTCTGATCCACTGGGACGGGTGAGAGTCCACGCCAGGAGCTAGGGTGGGA
 ATCGGAGGGGACGCCGACGAACCGCTATGGAACCCGAGAGGATATTAGCTAAGAATT

#17

ACCGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCAAGAAAAAGCGCAAGGTGGAAGCGTCCATGGGAGAG
 AATATGATTGACGAGAGTCTGACCTTCGGCATTGACCTGGGATTTGGGAGTTTGGGTGGGAGTCTGAGGCGGCCAAGCGCC
 TTCGGACGGAAGGGCGTGATCGAGGGAATGGGCTCCTGGTGTCTTTCAGCTCCCGGAGACATCAAAGAACGGACTCCTACCAAC
 CAGATCCGGAGATCCAATAGACTGCTGAGGCGCGTGTCCGACGGAGAAGGAAACGGATGGCCGCTATTCGCCGACTGCTGCAC
 GCAGCAGCCCTGCTGCCATCAACCGACAGCGATGCCCTGAAGCGGCCCGGACATGATCCTTGGGAACTGAGGGCACGAGGCCCTG
 GACAAGCCACTGAAACCCGTTGGAGTTTCGCTGTGGTCTGGGGCACATCGCAAAGCGGAGAGGGTTAAATCCGCTGCAAAGAGA
 AAAGCCACAAACATTAGCTCCGACGATAAGAAAATGCTGACAGCCCTGGAGGCTACTCGAGAACGGCTGGGGAGATACAGGACC
 GTGGGAGAAATGTTTCGCCAGGGACCCCTGATTTTTGCTTTCAGGCGCCGAAATCGCGAGGGCAAGTATGATAGGACCAAGCTCC
 GACGATCTGGAGCACGAAGTGCACGCCCTGTTTCGAGCTCAGCGGAGACTGGGACAGGGATTGGCCAGTCCAGAACTGGAGGAG
 GCCTTCACCGCTTCAGCATTTTCATCAGAGGCCATGCAGGACAGCGAGCGCTGGTGGGATTCTGCCCTTTTGGAGCAACCGAA
 AAGCGGGCAGCCAAACTGACACCCTCTTTCGAGCGCTTTCAGCTGCTGGCCCGGCTGCTGAACTGAGAATCACTACCCAGAG
 GGAGAGCGGCCCTGACAGTGGATGAAATTTGCTCTGGTCAACCGGGACCTGGGCAAGACCGCAAACTGAGTATCAAGCGGGTG
 AGAAGCTGATTTGGACTGGAGGCAATCAGAGGTTCAAACTATCCGCCCGGAGGACGAAGATCGAGACATTTGGGCTCGGACA
 GCGGGGCAATGCACAGGACTGCCACCCTGAGGAAGGCACCTGGGAGGGCCCTGTGGACTGATATGCAGAGCGCCCTGAACAG
 CTGGACGCTATCGTGCAGGTCTGAGCTCTTTGAGGCACACGAAACAATCACTGAGAAGTGGAGGAAATGGCCGACTCTG
 GCCGTGCTGGACGCTCTGCTGACCGCACTGGATGCCGGAGTGTTCGCCAAGTTAAAGGGCTGCACACATCAGCACCAAGGCC
 GCTAGGAATCTGCTGACCATCTGGAGCAGGGCAGGCCCTACGATGAGGCTGCACAATGGCAGGGTATGACCAAGCGCTCC
 CGCCTGTCTACCATGGCCAGATCGTGGCAAAGACAGTTCAACGCCCTGGTCACTGAGATCGGCGAATCCATTGCCAATCCA
 ATCGCTCGGAAGGCACTGATCGAGGGGCTGAAACAGATTTGGGCCATGAGAAAACCTGGGGGCTGCCCGGAAATATCCATGTG
 GAGCTGGCCCGGATGTCCGCAACTCAATTTGAAAAGCGACGGGAGATTGAAAAGCACATCGAGAAAAAATACTGCCCTGAGGGCT
 CGCGAGAGAAGGGAGGTGCATGATCTGCTGGACTGGAAGATGTCAATGGCGACCCCTGCTGGATACCGGCTGTGGAAGGAG
 CAGGGAGGCAAAATGCTGTATACAGGGAAGGCCATCCACATTCGGCAGATCGCTGCAACTGACAACCTCCGTGCAAGTGCATCAT
 ATTCGCTTGGAGCCGGTTCCGGCAGCATAGTTTTAAACAACAAGACCCTGTGTCTGGCCTGCTAATCAGCAGAAGAAAAGG
 TCAACACCATAACGAGTGGCTGAGCGGCCAGACTGGGATGCATGGAACGCCCTTCGTGAGCGCATCGAGACAAATAAGGAACTG
 AGAGGGTTTAAAGAAAAGGAACTATCTGCTGAAGAAATGCTAAAGAGGCGAGGAAAAATTCAGAAGCAGGAACCTGAATGACACC
 AGATACGCGCTAGGCTGTTCCGAGAGGCCCTGGGGCGGCTCTGCGGAGAGTGGAGCCACTTGGGAGAGCTTCCGGGACAGCTCGAG
 GTGTTTACTCGGCTGGAGCACTGACCGCAGCACTGAGACAGGCTTGGGAGTGGAGAGCTGAAGAAACAGGATGGGAAGCGC
 ATCAATGACGATCGACACCATGCCCTGGATGCTCTGACCGTGGCTGCACTGCGAGAGGCGGAAATTCAGAGGCTGACAAAATCA
 TTCCACGAGTGGGAAACAGCAGGGCCCTGGGGCGGCTCTGCGGAGAGTGGAGCCACTTGGGAGAGCTTCCGGGACAGCTCGAG
 GCTACCTACCTGAAGTGTGTCGACGGCCAGAGAGGGCCGAGCAAGAGGAGAAGGCCATGCCCTACCATCCGGCAGGTG
 AAGGAGAGAGAATGCACACCAATTTGTGTTTGAAGAAAAGGCTGTCTAGTCTGAAAGAGGCGAGCTGGAAACGAAATCAAAGAT
 GCGAGGCAACGAAAGCAATTTGGAGGCCATCAGGAGCTGGATTGCCACTGGACGCCAGCTGATGCACCACACGCTCCCGC
 CGAGGCGACATCAATACCAAGATCAGGCTGGCCACCACCATCAAGGCGAGCCGTGCTGTCCGCGGAGGGACCGCAGGAAAGGGGA
 GAAATGGTGCAGCAGATGTGTTTCAGCAAGCCAAACCGGAGAGGGAAGACGAGTGGTATCTGGTGCCTGTATCCACACCAG
 ATCATGAACAGGAAGCTTGGCCAAACCTCCAATGCGCTCAATTTGGCCAATAAGGATGAGGACGAATGGACCGAAGTGGGA
 CCTGAACACAGTTCGGTTTAGCTGATCCCTAGATCCAAATCGAGATCATTAGGCCATCTGGAGGATGATCGAAGGAT

FIG. 9L

TTCGTCGGCCTGCATCGCAACTGGCGCTCTGACCATCAGTGCACACAATGATCCCAAGAGTATCCATTAGGCACTTGGGACC
 AAGACTGCTGGCCATTTCCAAATACCAGGTGGACAGATTCGGCAGAAAGTCTCCAGTGCAGCAAGAGGTCCGAACTTGGCAC
 GGGGAAGCCTGTATCTCTCCACCCCCCTGGATAAAGAATTC

#19

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAGCGCAAGGTTCGAAGCGTCCATGAGAGAG
 AACGGGAGTGACGAAAGCGGGAGGAATATGGACGAAAAATGGATTACAGGATTGGGCTGGACATCGGCATCGCATCAGTGGGG
 TGGCCGCTCCTGCAGAACAATTCAGACGATGAGCCTGTGAGGATTGTGACCTGGGAGTGCGCATTTTCGATACCCGAGAGATC
 CCAAAGACAGGCGAAAGCCTGGCAGGACCCCGGAGAGCAGCTCGAACCAAGGCGCCGACTGCGGAGAGGAAACACCGCCTG
 GACCGAATCAAGTGGCTGTTGAGAACAGGGGCTGATCAATATTGACGATTTTCTGAAGAGATACAACATGGCCGGACTGCCA
 GATGTGTACCAGCTGCGGTATGAGGCTCTGGACAGAAAACTGACCGATGAGGAACTGGCCAGGTGCTGTCACATCGCTAAG
 CATCGGGCTTCAAGACCCAGGAAAGCCGAGACAGCAGCAAGGAAACCGGGCAGTGTGAAGGCCACAGACGAGAAATCAG
 AAACGAATGCAGGAAAGGGATACAGGACAGTGGGCGAGATGATCTACCTGGACGAGGCCCTCCGGACTGGTCTCTTGGAGT
 GAGAAAGGGTACATCTGACCCCGCAACAAGGCTGAAATTTATCAGCACACAATGCTGCGGCAATCTGCTGGGAGGAAATC
 AAGGAGATTTTCAAGTCCAGCAGCCGACTGGGCAACGAAAAAGCCACTGAGGAACTGGAGGAAAGTACCTGGAGATCATGACC
 TCCAGCGCTCTTTTACCTGGGGCTGGAATGCAGCCAGATGGGAAAGCCCTCCCTTATGCAATGGAGGGCTCTCTGACAGA
 TGGGGAAATGTACTTTTCTGGGGATCAGGGAGAGCTGAGGGGCGCTAAGGGGACCTACACAGCCGAATTTCTGGGCTCTG
 CAGAAAATCAACCACACAAGCTGGTCAATCAGGACGCGGAGACAAGGAATTTCACTGAGGAAGAGCGGAGAGCCCTGACTCTG
 CTGCTGTTTACCCAGAAAGAGGTGAAGTACGCTGCACTCCGCAAGAACTGGGCCCTGCCTGAGGACATCTGTTCTACAACCTG
 AACTACAAGAAGGCCGCTACTAAAGAAGAGCAGCAGAAAGGAGAACAGAAATACCAGAAAGCCAAAGTTTATCGGGATGCCATAC
 TATCACGATTACAAGAAATGCCTGGAAGAGAGAGTGAAGTATCTGACCGAGAACGAACTCAGGGACCTGTTTGTAGATCGGA
 ATGATTTCTGACTTGTACAAAAATGACGATTTCCCGCACCGAACGACTGGCCAAGCTGGGACTGGTGCCTACGAGATGGAAAGC
 CTGCTGGCTTACTCTACCAAAATCCAGCATCTGTCTATGAAGGCAATGCGGAACATCATTCCCTTTCTGGAGAAAGGGATG
 AACTACGACAAGGCTTGGCAAGAGGCGAGGATGACTTCAAAGCCGATAGCAAGGGGACAAAAGAAAGCTGCTGACCGGAGAG
 AACGTGAATCAGACAATCAACGAAATTAATAATCTGTGTTCAAACGCTCAGTGAAGCAGACAGTGAAGGTCAATTAACGCCATC
 ATTCGGACTTACGGCAGTCCACAGGCTATCAATATTGAGCTGGCAAGAGAAATGTCAAAGCCTTTGAAGAGAGGCGCAAAATC
 AAGGGGGACATGGAGAAAACCGCAGAAAGAAACAATGAAGATGTGAAGAAAACAGATTACGGAGCTGGGAAACTGTCTCTACAGGC
 CAGGACATCTGAAGTACAGACTGTGGCAGGAGCAGCAGGGGATTTGTATGTATAGTGGAAAAACCATCCACTGGAAGAGAGCTG
 TTCAAGCCCGCTACGACATCGATCACATTTCTGCCATTTCAATTACATTCGACGATAGCTTTAGGAACAAAGTGTGGTCCACA
 TCCAGGAGAACAGACAGAAAGGGCAATAGGACTCTTACGATATATGGGAAACGACGAAACAGCGTGGAAATGAGTTTGAACCC
 AGGGTAAAACTACCATCCCGGATTACAAGAAAACAGCAGAAAGCTGCTGAAGAAACATTTCTCTGAAGAGGAAAGGAGTGAATTT
 AAAGAACGGAACCTGACAGACACTAAGTACATCAACCCGTGATCTACAACATGATCAGACAGAAATCTGGAGATGGCCCCCTG
 AACCCCTTGAAGAAAGAAAGCAGGTGCGGGCTGTCAATGGCGCAATTAACCGCTACCTGCGAAACCGTGGGGCTGCCACAG
 AAGAATCGGGAGACAGACACACCATGCTATGGATGCAAGTGGTCACTCGCTGCTGTACCGACGGCATGATCCAGAAAATGAT
 AGATACACAAAGGTGAGAGAGAGGTGCTATTTCAAAGGGAACAGAGTTCTGTCGATGACAGAGACTGGCGAAATCTTTAGACCCGAG
 GACTACAGCAGGGCCGAGTGGGATGAAATTTTTCGGCGTGACATCCCAAAGCCCTGGGAGACATTTCCGCGCGAACTGGACGTC
 CGAATGGGGACGATCCAAAGGATTCCTGGACACTCATAGCGATGTGGCTCTGGAGCTGGATTATCCCGAGTACATCTACGAA
 AACCTCGGGCTATCTCTGAGCAGAAATGCCAAATCACAAGGTCACCGGAGCAGCCATGCTGACACAAATTCGGTCCCAAGA
 CACTTTAAAGATGAGGGCATCTGCTGACTAAGACCGCACTGACCGACTGAAACTGGACAAGGATGGGGAGATCGACGGATAC
 TATAACCCCGCTCCGATCTGCTGTACGAAGCACTGAAAAAGCAGCTGCTGCTGATGGCAATGATGCCAAAAAGGCCTTC
 GCTCAGGACTTTTATAAACCCCAAGGCGGATGGAACCTGAGGGCCCTGTGGTCAAGGAGGTGAAGATCCAGAAAAAGCAGACCATG
 GGAGTGTTCGTCGACTCTGGCAACGGGATTCGCCAGAATGGCGGATGGTGCATCGATGTGTTCCGAGTCAACCGCAAGTAC
 TATTTTGTGCCGCTACACCGCTGACGTGGTCAAAAAGGTGCTGCCTAATAGGGCCAGTACAGCTCAACAGCCATACGGCGAG
 TGGAAAGTGTGGAGGACAAGGATTTCTGTTTGTGCTGATTACCGCAGCTGATCCATATCAAGTCAAAAAGGATATCCCT
 ATTAAGATGTTGAACGGAGGATGGAGGGATCAAGGAAACCTACGCATACTATATTGGAGCCGACATCAGCGCTGCAAAATATC
 CAGGGCATGCCACAGTTCAGGTATAAATTCGCGGACTGGGCAATTCAGTCTGGACGTGCTGGAGAAAGTGTGAGATCGAT
 GTGCTGGGACATGTCAGCGTGGTCCGATCCGAAAAGCGGATGGGCTTTAGCTAAGAATTC

#21

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAGCGCAAGGTTCGAAGCGTCCATGAAAAGG
 AACTACATTTGGGGCTGGACATCGGGATTACAAGCGTGGGGATGGGATTTGACTATGAAACAAGGGACGTGATCGACGCA
 GCGTCAAGACTGTTCAAGGAGGCCAACGTGGAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGAAACGACGG
 AGAAGGCACAGAATCCAGAGGGTGAAGAACTGCTGTTGATTACAACCTGCTGACCGACCATTTCTGAGCTGAGTGGAAATTAAT
 CCTTATGAAGCCAGGGTGAAGGGCTGAGTCAGAAAGTGTGAGAGGAAAGGTTTTCCGAGCTCTGCTGCACCTGGTAAAGCC
 CGAGGAGTGCATAACGTCATGAGGTGGAAGAGGACACCGCAACGAGCTGTCTACAAGGAAACAGATCTCACGCAATAGCAAA
 GCTCTGGAAGAGAAGTATGTCGAGAGCTGACGCTGGAACGGCTGAAGAAAGTGGCGAGGTGAGAGGGTCAATTAATAGGTTT
 AAGACAAGCGACTACGTCAAAGAAGCAGCAGCTGCTGAAAGTGCAGAAAGGCTTACCAACAGCTGGATCAGAGCTTATCGAT
 ACTTATATCGACTGCTGGAGACTCGGAGAACCTACTATGAGGGACAGGAGAAGGGAGCCCTTCGGATGGAAGACATCAAG
 GAATGGTACGAGATGCTGATGGGACATTGCACCTATTTTCCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTG
 TACAACGCCCTGAATGACCTGAACAACCTGGTCATCAACAGGGATGAAAACGAGAAACTGGAATACTATGAGAAGTTCAGATC
 ATCGAAAACTGTTTAAAGCAGAAGAAAAAGCCTACACTGAAACAGATTTGTAAGGAGATCCTGGTCAACGAGAGGACATCAAG
 GGCTACCGGTTGCAAGACTGGAAAAACAGAGTTCACTCAATGCAAGGTGATACAGATTTAAGGACATCAACAGCAGGAAA
 GAAATCATTGAGAAGCCGAACTGCTGGATCAGATTGCTAAGATCCTGACTATCTACAGAGCTCCGAGGACATCCAGGAAGAG

FIG. 9M

CTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATCTGAAGGGGTACACCGGAACACACAACCTG
 TCCCTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGGCATACAAACGACAAATCAGATTGCAATCTTTAACCGGCTGAAGCTG
 GTCCAAAAAAGGTGGACCTGAGTCAGCAGAAAAGAGATCCCAACCACACTGGTGGACGATTTCAATCTGTACCCCGTGGTCAAG
 CGGAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAAGTACGGCCTGCCAATGATATCATTATCGAGCTGGCT
 AGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTGAAGAG
 ATTATCCGAACTACCGGAAAGAGAACGCAAAAGTACCTGATTGAAAAAATCAAGCTGCACGATATGCAGGAGGGAAAGTGTCTG
 TATTCTGGAGGCCATCCCCTGGAGGACCTGCTGAACAATCCATTTCAACTACGAGGTCGATCATATTATCCCAGAAGCGTG
 TCCTTCGACAAATCCTTTAAACAACAGGCTGGTCAAGCAGGAAGAGAAGCTTAAAAAGGGCAATAGGACTCCTTTCCAGTAC
 CTGTCTAGTTCAGATTCCAAGATCTTTACGAAACCTTTAAAAAGCACATTTCTGAATCTGGCCAAAGGAAAGGGCCGCATCAGC
 AAGACAAAAAGGAGTACCTGTGGAAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCGGAATCTGGTG
 GACACAAGATACGCTACTCGCGCCTGATGAATCTGCTGCCATCTATTTCCGGGTGAACAACTGGATGTGAAAAGTCAAGTCC
 ATCAACGGCGGGTTCACATCTTTCTGAGGCGCAATGGAAGTTTAAAAAGGAGCGCAACAAGGGTACAAGCACCATGCCGAA
 GATGCTCTGATTATCGAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGGCAAAAGCAAGAAAGTATGGAGAACCAG
 ATGTTCCGAAGAGAAGCGAGCCGAATCTATGCCGAAATCGAGACAGAAACAGGAGTACAAGGAGATTTTCATCACTCCTCACCAG
 ATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCTCACCGGGTGGATAAAAAAGCCCAACAGAGAGCTGATCAATGACACC
 CTGTATAGTACAAGAAAAGACGATAAGGGGAATACCTGATTGTGAACAATCTGAACGGACTGTACGACAAAAGATAATGACAAG
 CTGAAAAAGCTGATCAACAAAAGTCCCAGAAAGCTGCTGATGTACCACCATGATCTCAGACATATCAGAAAAGTGAAGCTGATT
 ATGGAGCAGTACGGCGACGAGAAGAACCCACTGTATAAGTACTATGAAGAGACTGGGAACTACCTGACCAAGTATAGCAAAAAG
 GATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGTGAATGCCCATCTGGACATCACAGACGATTACCCTAAC
 AGTCGAAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGATTGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACT
 GTCAAGATCTGGATGTCAAAAAGGAGAAGTACTATGAAGTGAATAGCAAGTGTACGAAAGGCTAAAAGCTGAAAAAG
 ATTAGCAACCAGGCAGAGTTTCATCGCCTCCTTTTACAACAACGACCTGATTAAAGTCAATGGCGAAGTGTATAGGGTTCATCGGG
 GTGAACAATGATCTGCTGAACCGCATGAAAGTGAATGATTGACATCACTTACCAGAGTATCTGGAAAAACATGAATGATAAG
 CGCCCCCTCGAATTTACAAAACAATTCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGAAAACTGTAT
 GAGGTGAAGAGCAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAGAAATC

#22

ACCGGTGCCACCATGTACCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTGCAAGCGTCCATGAAAAGG
 AACTACATTTCTGGGCTGGACATCGGGAATTACAAGCGTGGGGTATGGGATTAATGACTATGAAACAAGGGACGTGATCGACGCA
 GGCTGACGATGTTCAAGGAGGCCAACGTGGAAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGGCCCTGAACGACCG
 AGAAGGCACAGAAATCCAGAGGGTGAAGAACTGCTGTTGATTACAACCTGCTGACCGACCATTTCTGAGCTGAGTGGAAATTAAT
 CCTTATGAAGCCAGGGTGAAGGGCCTGAGTCAGAAGCTGTGAGGAAAGAGTTTTCCGCGACTCTGCTGCACCTGGCTAAGCGC
 CGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAACAGATCTCAGCAATAGCAAA
 GCTCGAAAGAGAAAGTGTGCGAGAGCTGCAGCTGGAACGGGTGAAGAAAGATGGCGAGGTGAGAGGTTCAATTAATAGGTTT
 AAGACAAGCGACTACGTCAAAGAAGCCAGCAGCTGCTGAAAGTGCAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGAT
 ACTTATATCGACCTGCTGGAGACTCGGAGAACCTACTATGAGGGACAGGAGAAGGGAGCCCTTCGGATGGAAAGACATCAAG
 GAATGGTACGAGATGCTGATGGGACATTCACCTATTTCCAGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACCGCAGATCTG
 TACAACGCCCTGAATGACCTGAACAACCTGGTTCATCACCAGGGATGAAAACGAGAACTGGAATACTATGAGAAGTCCAGATC
 ATCGAAAACGTGTTAAGCAGAAGAAAAAGCCTACACTGAAACAGATGCTAAGGAGATCTGGTCAACGAAAGGACATCAAG
 GGCACCGGGTGAACAAGCACTGGAAAACAGAGTTCACCAATCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAA
 GAAATCATTGAGAACGCCGAACTGCTGGATCAGATTGCTAAGATCTGACTATCTACCAGAGCTCCGAGGACATCCAGGAAGAG
 CTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAATCTGAAGGGGTACACCGGAAACACAACCTG
 TCCCTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGCTG
 GTCCAAAAAAGGTGGACCTGAGTCAGCAGAAAAGAGATCCCAACCACACTGGTGGACGATTTCAATCTGTCAACCGTGGTCAAG
 CGGAGCTTCATCCAGAGCATCAAAGTGTCAACGCCATCATCAAGAAGTACGGCCTGCCAATGATATCATTATCGAGCTGGCT
 AGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTGAAGAG
 ATTATCCGAACTACCGGAAAGAGAAAGCAAGTACTGATTGAAAAAATCAAGTGCACGATATGCAAGGAGGAAAGTGTCTG
 TATTCTCTGGAGGCCATCCCCTGGAGGACCTGCTGAACAATCCATTTCAACTACGAGGTCGATCATATTATCCCAGAAGCGTG
 TCCTTCGACAAATCCTTTAAACAACAGGCTGGTCAAGCAGGAAGAGAAGCTTAAAAAGGGCAATAGGACTCCTTTCCAGTAC
 CTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCACATTTCTGAATCTGGCCAAAGGAAAGGGCCGATCAGC
 AAGACAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCGGAATCTGGTG
 GACACAAGATACGCTACTCGCGCCTGATGAATCTGCTGCCATCTATTTCCGGGTGAACAACTGGATGTGAAAGTCAAGTCC
 ATCAACGGCGGGTTCACATCTTTCTGAGGCGCAATGGAAGTTTAAAAAGGAGCGCAACAAGGGTACAAGCACCATGCCGAA
 GATGCTCTGATTATCGCAAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAAGCAAGAAAGTATGGAGAACCAG
 ATGTTCCGAAGAGAAGCAGGCCGAATCTATGCCGAAATCGAGACAGAAACAGGAGTACAAGGAGATTTTCATCACTCCTCACCAG
 ATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCTCACCGGGTGGATAAAAAAGCCCAACAGAGAGCTGATCAATGACACC
 CTGTATAGTACAAGAAAAGACGATAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAAGATAATGACAAG
 CTGAAAAAGCTGATCAACAAAAGTCCCAGAAAGCTGCTGATGTACCACCATGATCTCAGACATATCAGAAAAGTGAAGCTGATT
 ATGGAGCAGTACGGCGACGAGAAGAACCCACTGTATAAGTACTATGAAGAGACTGGGAACTACCTGACCAAGTATAGCAAAAAG
 GATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGCCCATCTGGACATCACAGACGATTACCCTAAC
 AGTCGAAACAAGGTGGTCAAGCTGTCACTGAAGCCATACAGATTGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACT
 GTCAAGAAATCTGGATGTCAAAAAGGAGAAGTACTATGAAGTGAATAGCAAGTGTACGAAAGGCTAAAAGCTGAAAAAG
 ATTAGCAACCAGGCAGAGTTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAAGTGTATAGGGTTCATCGGG

FIG. 9N

GTGAACAATGATCTGCTGAACCGCATTTGAAGTGAATATGATTGACATCACTTACCGAGAGTATCTGGAAAACATGAATGATAAG
CGCCCCCTCGAATTATCAAAAACAAATTCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAACCTGTAT
GAGGTGAAGAGCAAAAAGCACCTCAGATTATCAAAAAGGGCTAAGAAATTC

#23

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCTGAAGCGTCCATGAACAAT
AGCATCAAATCTAAACCTGAAGTGACCATCGGGCTGGACTGGGAGTGGGAAGCGTGGGGTGGGCAATCTGGGTAACGAAAACA
AACATCATTACCATCTGGGCTCCAGGCTGTTTTCTCAGGCCAAGACTGCTGAGGATCGGAGATCTTTCCGCGGGGTGAGGCGC
CTGATCCGACGGAGAAAATACAAGCTGAAACGATTGCTCAATCTGATTTGGAGTACAACAGCTATTTCCGGCTCAAGAACAAA
GAGGACATCCTGAACAATATCAGGAGCAGCAGAAGCTGCACAATACCGTCTGAACTGAAATCAGAGGCACTGAATGCCAAG
ATCGATCCTAAAGCACTGAGCTGGATTCTGCACGACTACCTGGAAGAACAGAGGCCATTTTTATGAGGACAATAGGGATTTCAAC
GTGTACCCAAACAAAGGAGCTGGCCAAAGTACTTCGATAAGTACGGGTAACAAGGGAATCATTTGACAGCAAGGAGGACAATGAT
AACAACTGGAGGAAGAGCTGACAAAGTACAATTTCTCAAATAAGCACTGGCTGGAAGAGGTGAAGAAAAGTCTGTCTAACAG
ACTGGCTGCCAGAAAAGTTTTAAAGAAGAGTATGAGTCACTGTTCAAGTACGCTAGGAAATTTATCAGAGGCTCAAGAAAAGTCACT
AACTCTGTGCTCCTACGGGATCTACCATCTGGACGAAAAAGAGGGAAAAGGTGGTCCAGAAGTACAACAACATCTGGGATAAG
ACAATCGGAAAGTGAACATCTTCCCTGACGAGTATAGAGCTCCCAAGAACAGTCTATCGCAATGATTTTTCAATGAAATCAAC
GAGCTGTCCACAATCAGGTATA CAGCATCTACCTGACTGGCTGGTTCAATTAATCAGGAGTTCAAGAAAAGTCAAGAAAAGTCA
CTGCTGGATCTGCTGATCAAAAACCAAGGAGAGAAGCCAAATGACGCAAGGCAAGTCAAGAAAAGTCAAGAAAAGTCAAGAAAAGTCAAG
GAAAGCATTGGCAAGAGACACTGAAGGATGTGGAGAAATGAAGAGAAAAGTGGAAAAGGAGGACCACAAGTGGAAAAGTGAAGGGA
CTGAAGCTGAATACCAACGGCAAAAATCCAGTACAACGATCTGAGCTCCCTGGCTAAGTTTGTGCACAAAAGTGAAGCAGCATCTG
AACTGATTTTTCTGCTGGAGGACAGTATGCAACACTGGACAAGATCAATTTTTCTGCACTCCCTGTTTGTGACCTGGGCAAG
CACCTGAGATTCCAATAGGGTGGATTCTGCCAACCTGAAGGAAATTTCCGACTCTAACAAAAGTTCGAGCGCATCCTGCGAG
AAACAGAGGATGGCTGTTCAAGCTGTTTGAACAGACTGACAAAGAGCATGAGAAGATCCTGGCCAGACACATAGTCTGTCA
ACTAAGGCCATGCTGCTGGCTATTACCCGGATGACAAATCTGGACAACGATGAGGACAACAGAAAACAATGACAAGGGCTGG
AATTTTGAAGCCATCAAAAACCTCGATCAGAAGTTTATCGACATCACCAAGAAAACAACAACCTGAGCCTGAAAACAGAAATAG
CGCTACCTGGACGATCGATTCAACAGATGCTATTTCTGCTCCCTGGGGTGAAGCGAAATCCTGCGGGAGGCAACCAAGGTCTTT
AATGCCAATTTCTGAAAACAGTTCTCTGAAGAGTACGACGTGACAAAAGTGGTCACTCGAACTGGCTCGCGAGCTGAGCGAAGAGAAG
GAACTGGAGAACAACAAGAACTACAAGAACTGATCAAGAAAACGGGCAAGATAAGTGAAGGGCTGAAAGCACTGGGGATC
TCAGAAAGTGAAGTCAAAAGACATTTGAAAGAGTCCCACTAAATCATAAAGTTTCTGCTGTGGCTGCAGCAGGACCAATCGAT
CCTTATAGCCTGAAGGAGATCGCCTTCGACGATTTTTACCAAAAACAGAAAAGTTGAGATCGACCATATCATTCCCTACAGC
ATTTCTTTCGACGATTCTAGTTCAAACAAGCTGCTGGTGTGGCTGAAAGTAAATCAGGCAAGTCAAACAGACTCCTTATGAG
TTCACTAGCTCCGGAACCGCAGGCATTAAGTGGGAAGTACGAGGCTATTTGCCCAAGTTCAAGGATGGGACTCTAGTCTG
CTGGACAGCACCCAGCGGTCCAAGAAATTCGCCAAAATGAGAAAACCGATACCTCAAGCAAGTACGACATCGGATTTCTGGCT
CGAAATCTGAACGATCTCGGTACGCAACCAATTGTTTCCGGGACGCCCTGGAGGACTATGCTAATAAACCACTGGTGCAGGAC
AAACCCATGTTTAAAGGTGGTCTGATCAATGGGTCCGTGACCTCTTTCTGCGGAAGAATTTGACGATTCTCTTACGCCAAG
AAAGATAGAGACAAGAATAACCCATGCTGTGGATGCAAGTATCATCTCAATTTTCAAGCAACGAGACAAGAACTGTTTCAAC
TACTTGGCCAGAGTTGCTGACTATAAACTGTTCAAGAACACCGATGGCAGCTGGAAGAAAATCGACCTAAGACAGGGGTGGTC
ACTGAAGTGAACGAGCAGAAATGGAAGCAGATTAGGGTGCACAACAGGTGAGCGAAATCGCCAAAGTCAATGAGAAGTACATC
CAGGATAGCAACATCGAAAAGAAAGGCTAGGTATTTCCCGAAAATCGAGAATAAGACTAACATTTCCCTGTTTAAATGACACCGT
TACTTGGCCAAAGAAAGTCCGCTATGAGGATCAGATCAAAAAGAAAGAACTGAAAACCCCTGGACATTCACGAATCTGCTAAAAG
AATAAAGACAGTAAAGTGAAGCGGCAAGTTTGTCTACAGAAAGCTGGTGAATGTGACCTGCTGAATAACGATAAGCTGGCAGAC
CTGTTCCGCGAAAAGAGGATATCCTGATGTATAGGGCCAAATCCATGGGTCACTCAACCTGGCTGAGCAGATTTTCAATGAATA
ACTGAGAACAAGAAAATCAAGTCCAGAACGTTGTTTTGAAAATATATGCTGGACCTGACCAAGAGTTCCCGGAGAAGTTCAAGC
GAGTTTCTGGTGAAGTCCATGCTGAGAAAACAAGACCGCCATCATCTACGACGATAAGAAAACAATGTTCCATCGAATCAAACGG
CTGAAGATGCTGATTTCAAGACTGAAAGAGAAATAAGCTGTCTAACGTTGATCAATAGGTCTAAGAAATCAGAGTGGGACCAACTG
TCATACCAAGGATACAATCAACAGCTGGCCCTGATGATTATGCGCAGCATCGACCTACTGCTAAGAAAACAGTATAATTCGAGTG
CCACTGAATACCCTGAACCTGCACTGGGAGATCATGACTTTGATCTGCACAATATGGAATGCTTACCTGAAAGAAAACAAAATTC
GTGAAGTATCTGAAAGCAAAACGAAATCGGCGACGAGTACAAGCCCTGGAGGGTCTGACATCTGGCACTCTGCTGATCCATAAG
AAGGATAAGAAAAGTGTACATCAGCTCCTCCAGAATCTGAACGACGTTGATCGAAATTAAGAATCTGATCGAAACCGAGTAT
AAAGAGAACGACGATTTCTGATAGTAAAGAAAAGAAAAGGCAACCCGCTTCTGATGACCTGAGCAACAATCTGAAATGACTAC
ATTTCTGCTGGACGCAAGGATAACTTCGACATCCTGGGGCTGTCTAAAATCGGATCGATGAGATTCTGAACAGTAAAGCTGGGA
CTGGACAAGATTGTGAAAATAAGAAATTC

#24

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCTGAAGCGTCCATGAGGATT
CTGGGGTTTGACATTTGGCATTAAACAGCATCGGGTGGGCTTTTGTGGAGAACGACGAATGAAGGACTGCGGAGTGGGATCTTC
ACAAAGGCCGAGAACCCAAAAATAAGGAAAGCCTGGCACTGCCCGGAGAAAATGACGCGAGCTCCAGGCGCCGACTGAAACGG
AGAAAGGCCCGGCTGATCGCTATTAAGAGAAATCCTGGCCAAAGAGCTGAAGCTGAACTACAAGGACTATGTGCGAGCTGATGGA
GAGCTGCCAAAGGCTACGAAGGATCCCTGGCATCTGTGACGAGCTGGGATAAGGCCCTGACACAGAACCTGGAACTAAA
GATCTGGCCAGAGTGTCTGCTGCAATTTGTAAGCATAGGGGGTACATGAACAAGAACGAGAAAGAAATCAAACGACGGCTAAGAAA
GGAAAGATCCTGAGCGCTCTGAAAACAATGCACTGAAGCTGGAGAATACCAGAGCTGGGCGAATCTTCTACAAGGAGTTT
TTTCAGAAAATCAAGAAAACCAAAAGAACTTCAATCAAGATCCGCAACACTAAGGATAATTACAACAATTCGCTGCTGTCTAGT
GACCTGGAAAAGAGCTGAAGCTGATCCTGGAAAACAGAAAGGATTTCCGCTACAACACTCTGAAGATTTTCATCAACGAGATT

FIG. 90

CTGAAGGTCGCCTTCTTTGAGGAAGAGAAAAGG
 GCCTGTAAGAACAGCTACTCTGCCTGGGAGTTTGTGGCTCTGACCAAGATCATTAAACGAGATCAAGAGCCTGGAGAAGATCAGC
 GGCGAAATTGTGCCAACCCAGACAATCAACGAGGTCCTGAATCTGATCCTGGACAAGGGGTCTATCACCTACAAGAAATTCAGA
 AGTTGTATCAATCTGCATGAGAGTATCAGCTTCAAGAGCCTGAAGTATGATAAAGAAAACGCCGAGAATGCTAAACTGATCGAC
 TTCCGCAAGCTGGTGGAGTTTAAGAAAAGCCTGGGAGTCCACAGCCTGTCCCGGCAGGAACTGGATCAGATCTCCACTCATATC
 ACCCTGATTAAGGACAACGTGAAGCTGAAAACCGTCTGGAGAAATACAACCTGAGTAATGAACAGATCAACAATCTGCTGGAA
 ATTGAGTTC AACGATTATCAACCTGAGCTTCAAGGCCCTGGGAATGATTCTGCCACTGATGCGCGAGGGCAAACGATACGAC
 GAGGCTGCGAGATCGCCAATCTGAAACCTAAGACCGTGGACGAGAAGAAAGATTTCTGCCAGCATTTTGTGATTCCATTTTC
 GCCCAGAGCTGTCTAACCCCGTGGTCAATAGGGCTATCAGCGAATACCGCAAGGTGCTGAACGCACTGCTGAAGAAATATGGA
 AAGGTCCACAAAATTCATCTGGAGCTGGCTCGCGACGTGGCCTGTCCAAGAAAACGACGAGAGAAGATCGAAAAAGAGCAGAAG
 GAAAACAGGCCGTGAATGCATGGGCCCTGAAGGAATGCGAGAATATTGGCCTGAAGGCCAGCGCAAAGAACATCCTGAAACTG
 AAGCTGTGGAAAGAACAGAAGGAGATCTGTATCTACTCCGGAAAATAGATCTCTATTGAGCACCTGAAAGATGAAAAGGCCCTG
 GAGGTGGACCATATCTACCCCTATTCTAGGAGTTTTCGACGATTCTTTTATCAACAAAGTGTGGTGTTCACCAAGGAAAATCAG
 GAGAACTGAACAAGACACCTTTTCGAGGCCTTTGGCAAGAATATTGAAAAATGGAGCAAGATCCAGACCCTGGCTCAGAACCTG
 CCATACAAGAAAAAGAATAAGATTCTGGACGAGAACCTCAAAGATAAGCAGCAGGAGGACTTTATCTCTCGAAATCTGAACGAC
 ACCCGGTATATCGCTACACTGATTGCAAAAATACACAAGGAGTATCTGAACTTCTGCTGCTGAGCGAAAATGAGAACGCCAAT
 CTGAAGAGTGGCGAAAAAGGGTCAAAGATCCACGTGCAAGACTATTAGCGGGATGCTGACCTCCGTCTGAGGCACACATGGGGG
 TTTGACAAAAAGGATCGCAACAATCATCTGCACCATGCACTGGATGCCATCATTGTGGCCTACAGTACAAATTCATCATTAA
 GCTTTTCAGCGATTTCCGGAAAAACAGGAGCTGCTGAAGGCCAGATTCTACGCTAAAGAACTGACTTCCGATAACTATAAACAT
 CAGGTCAAGTTCCTTTGAGCCTTTCAAGAGTTTTAGAGAAAAAATCCTGTCAAAGATCGACGAGATTTTCGTGTCCAAACCACT
 CGAAAGCGAGCTAGGCGCGCACTGCACAAGGATACCTTTTATTCTGAGAACAAGATCATTGACAAGTGCAGCTACAACCTCAAAG
 GAAGGCCTGCAGATTGCCCTGAGCTGTGGAAGAGTGAGGAAAAATCGGCACTAAGTATGTGAGAAATGATACCATCGTGAGGGTC
 GACATTTTCAAAAAGCAGAACAAGTTTTACGCTATCCCAATCTACGCAATGGATTTTGGCCTGGGGATCCTGCCAATAAGATC
 GTGATTACTGGAAAAGATAAGAACAATAACCCCAAACAGTGGCAGACCATGACGAATCATACGAGTTCTGCTTTAGCCTGTAT
 AAGAATGACCTGATCCTGCTGCAGAAAAAGAATGCAAGAACCTGAGTTCCCTACTATAACGATTTTCAATCAGCACATCA
 AGCATTTGTGTGGAGAAACACGACAACAAGTTCCGAAAATCTGACTAGCAACCAGAAGCTGCTGTTTTCAATGCAAAAAGAGGGC
 TCTGTGAAGTTCGAAAGTCTGGGGATCCAGAACCTGAAAGTGTTCGAGAAGTACATCATTACCCCCCTGGGAGATAAAATTAAG
 GCTGACTTTCAGCCTCGAGAAAACATCAGCCTGAAAACAGTAAAAAGTATGGCCTGAGGTAAGAATTC

Posiciones de mutación de SpCas9

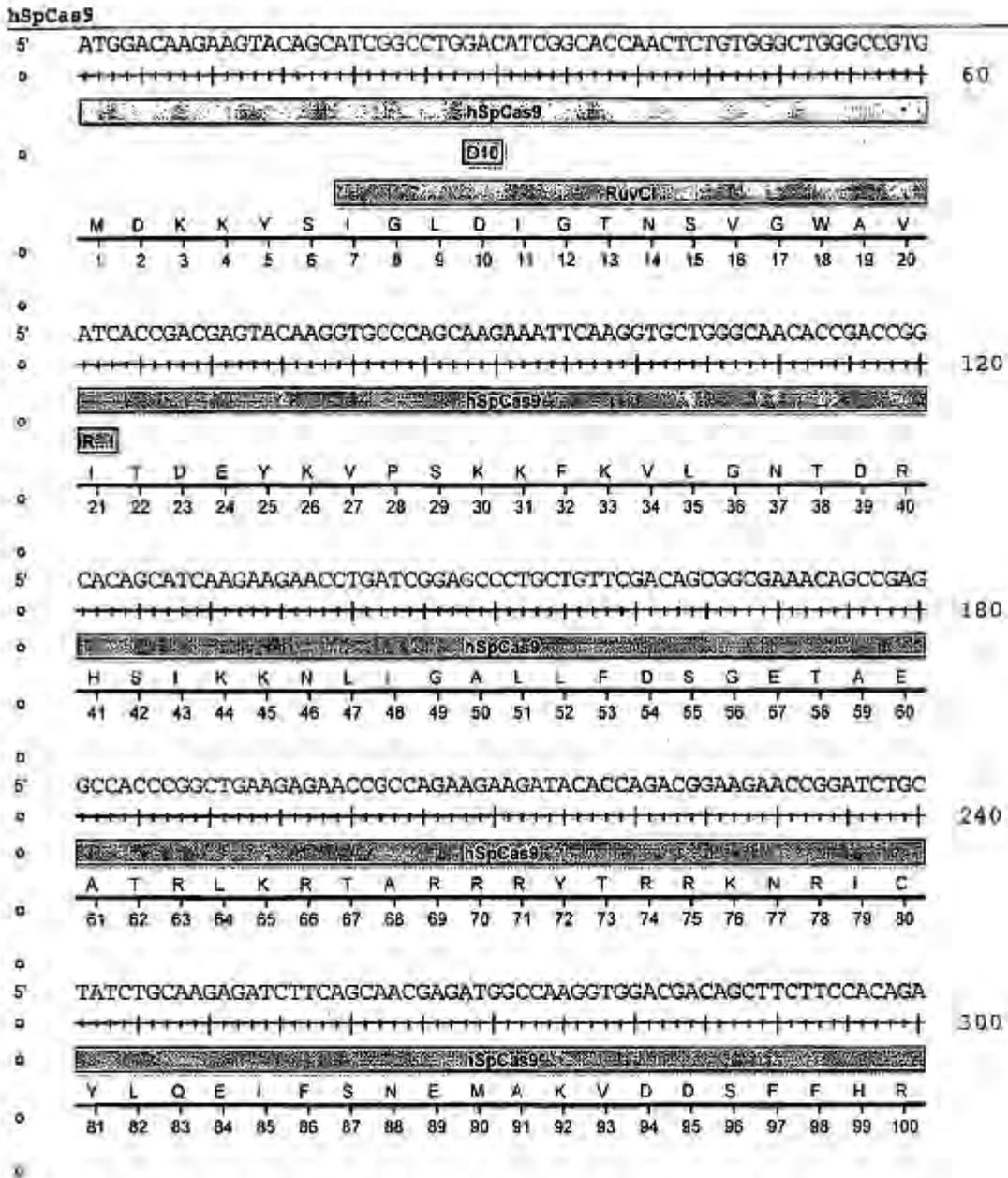


FIG. 10A

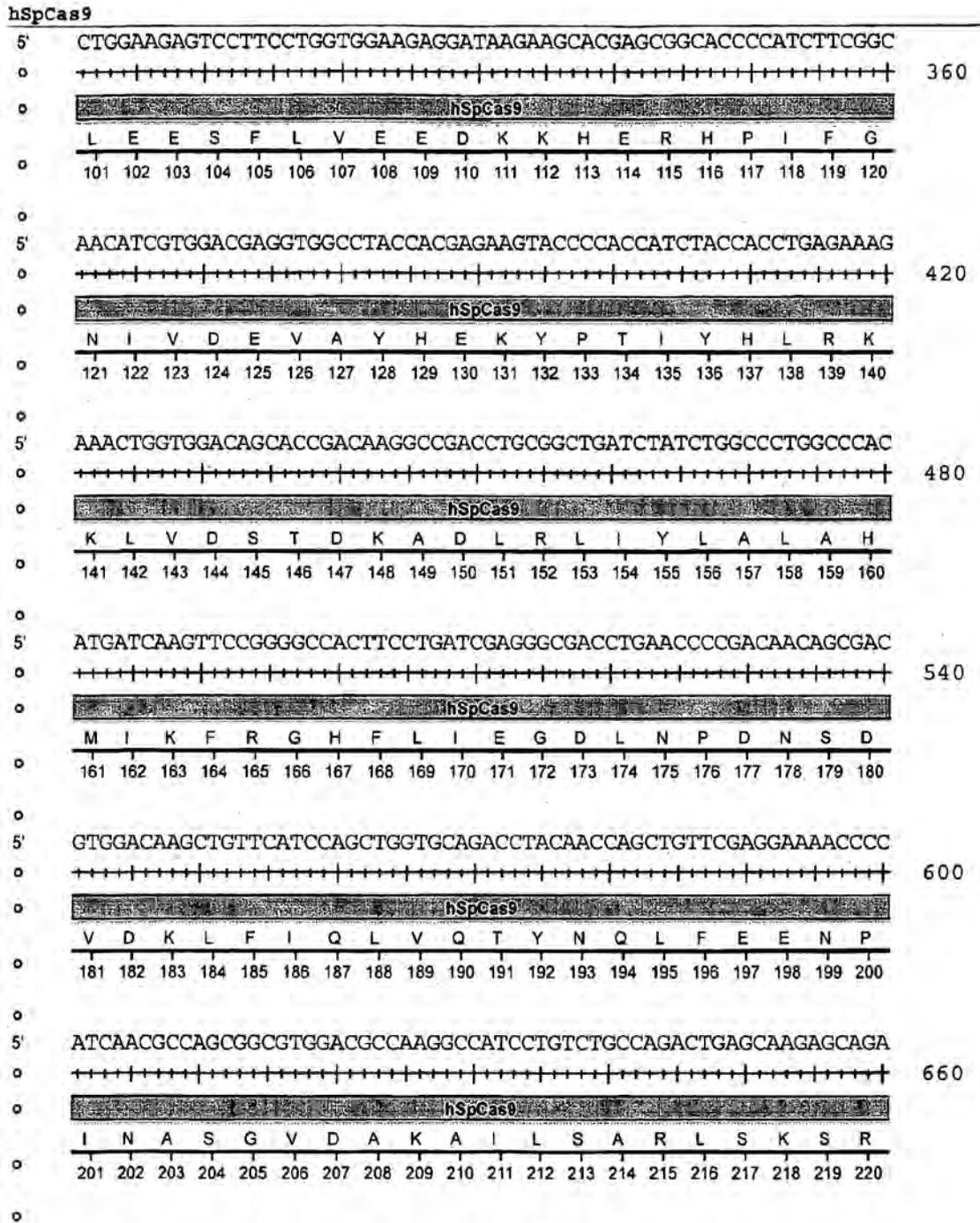


FIG. 10B

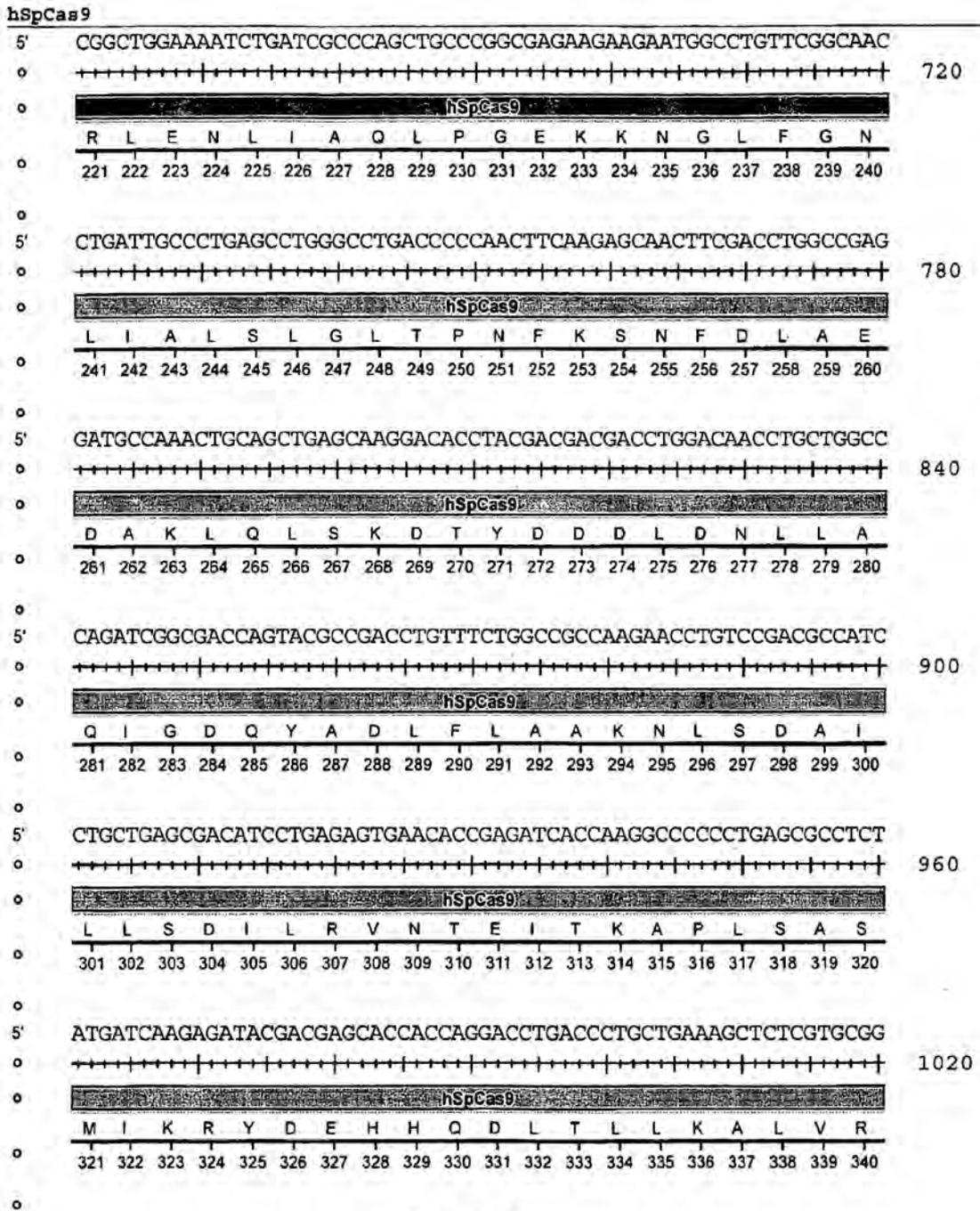


FIG. 10C

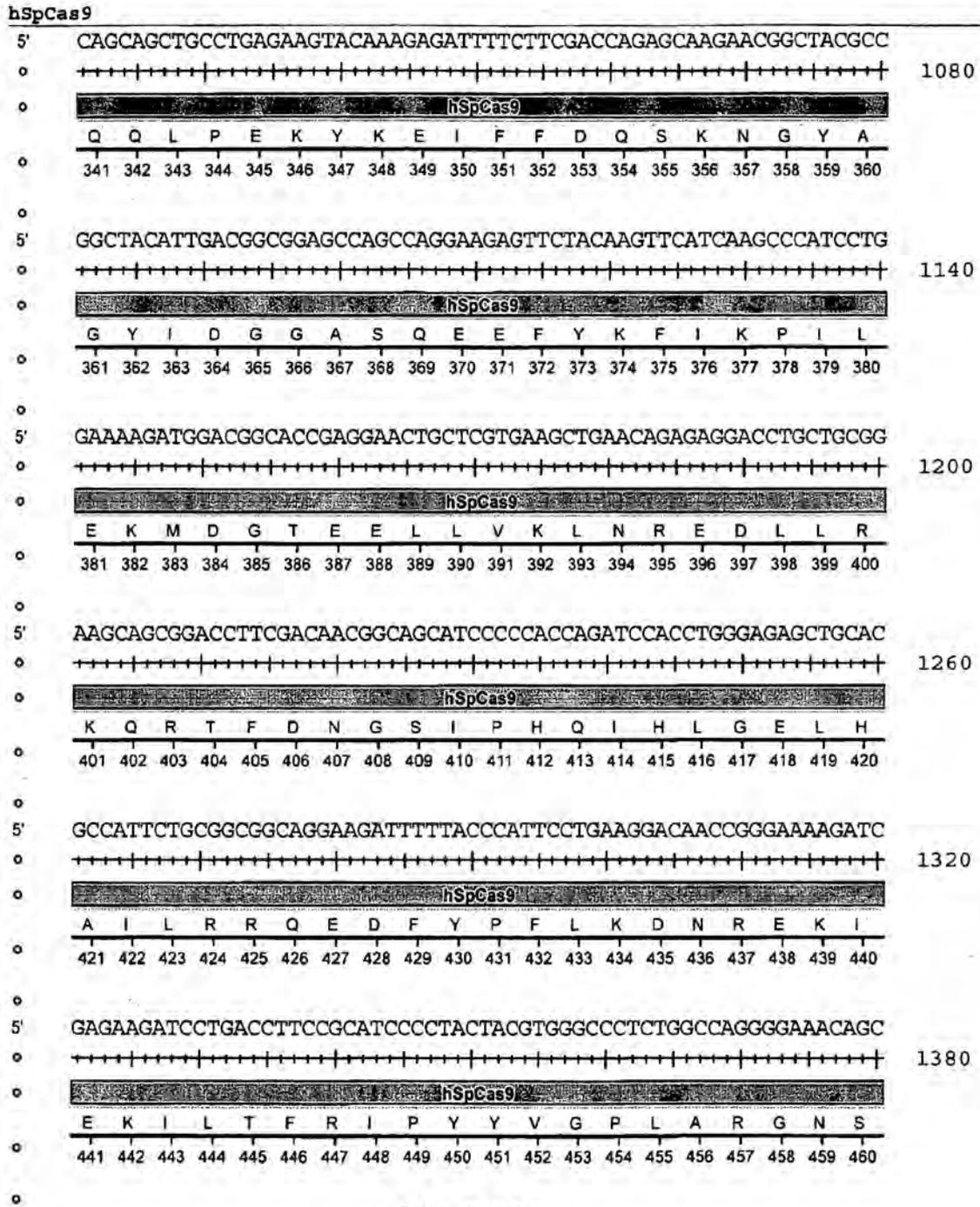


FIG. 10D

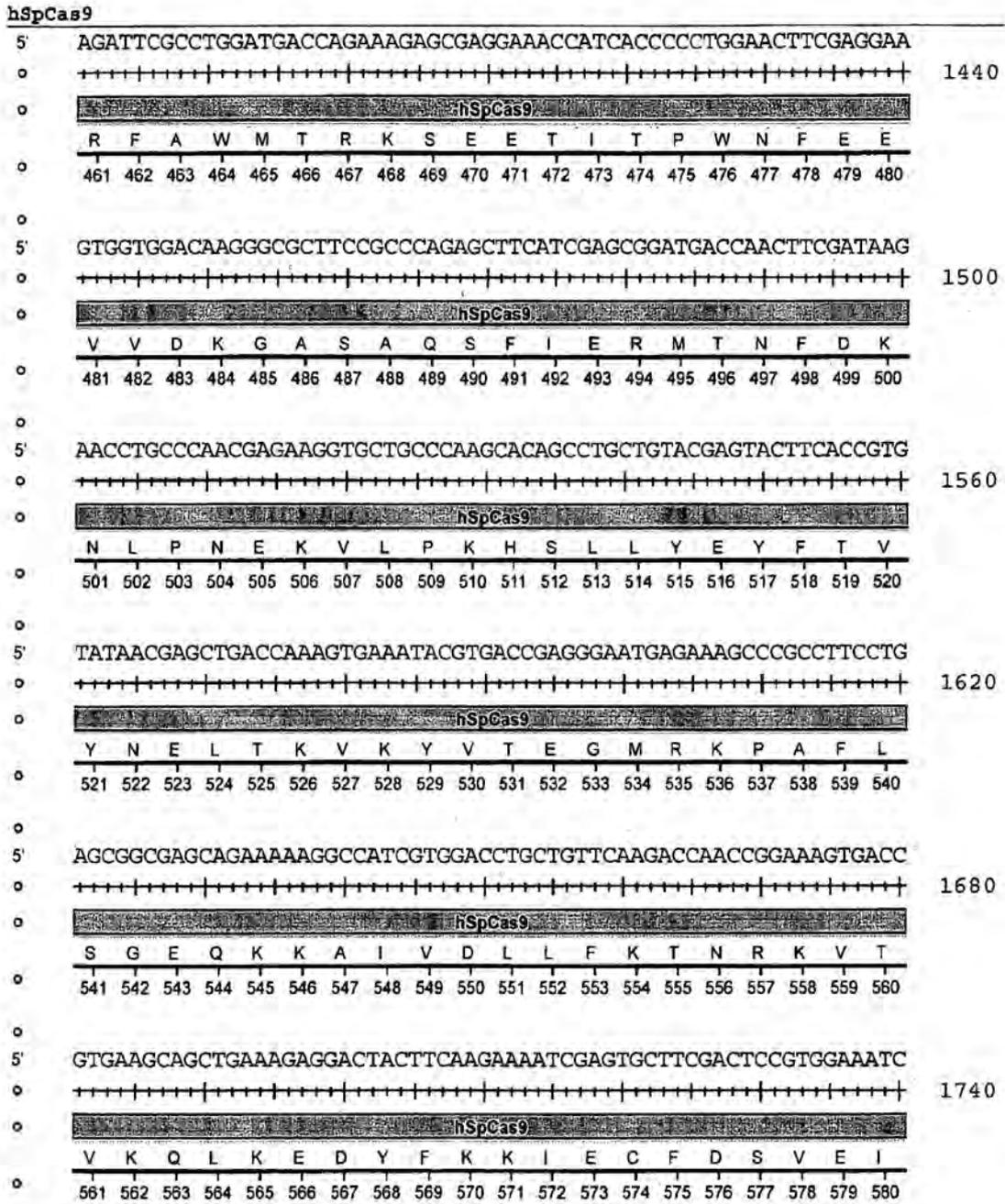


FIG. 10E

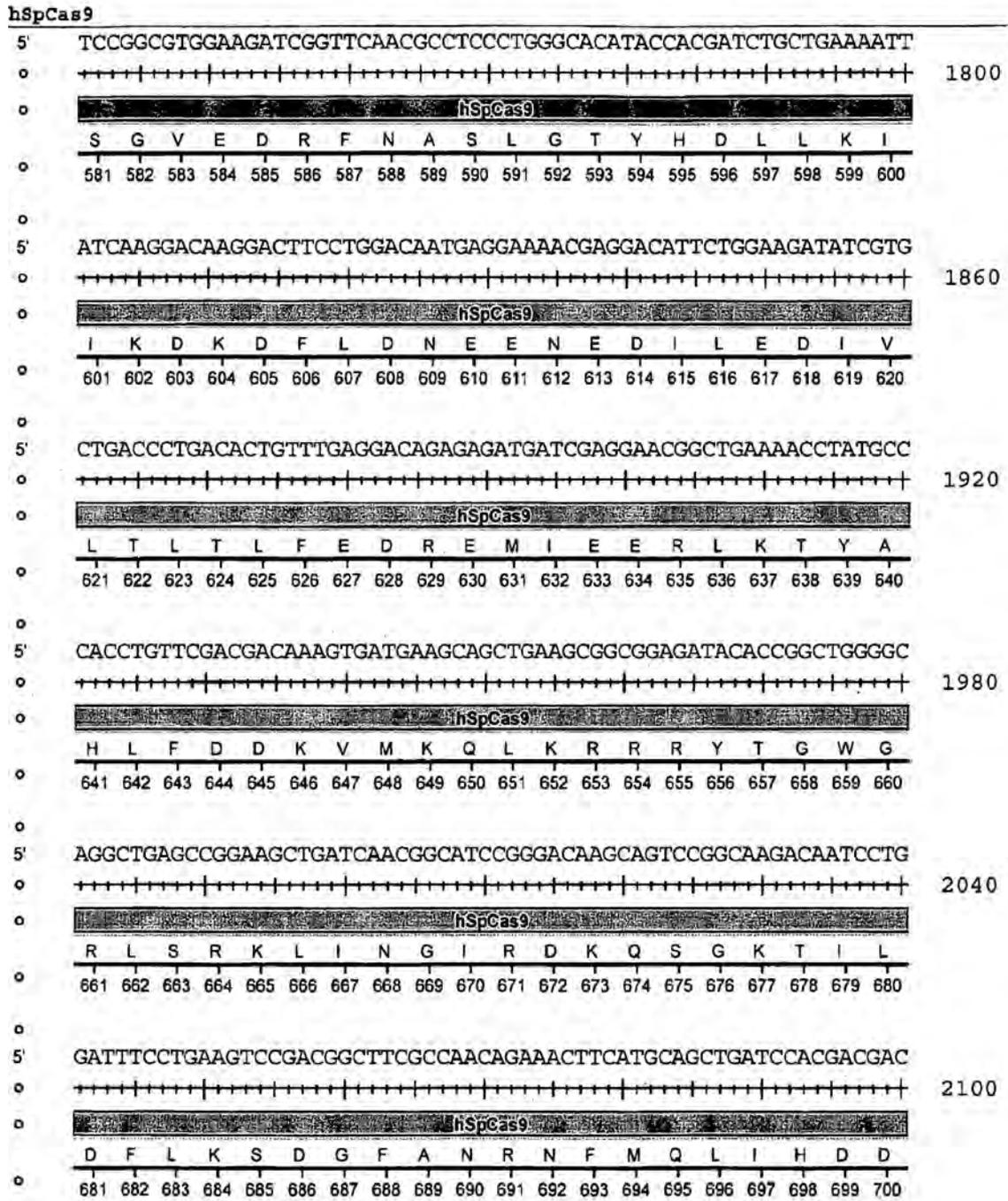


FIG. 10F

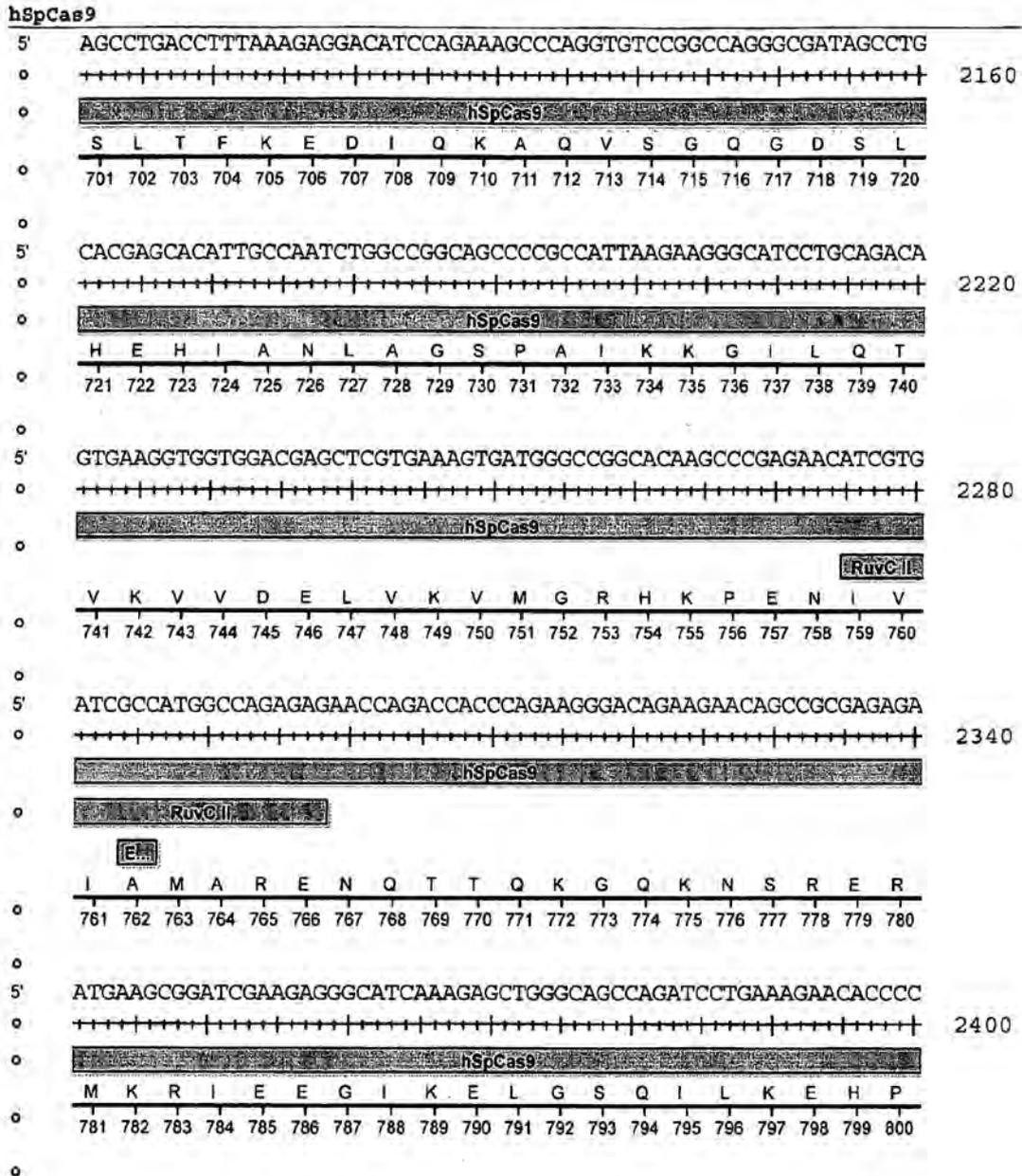


FIG. 10G

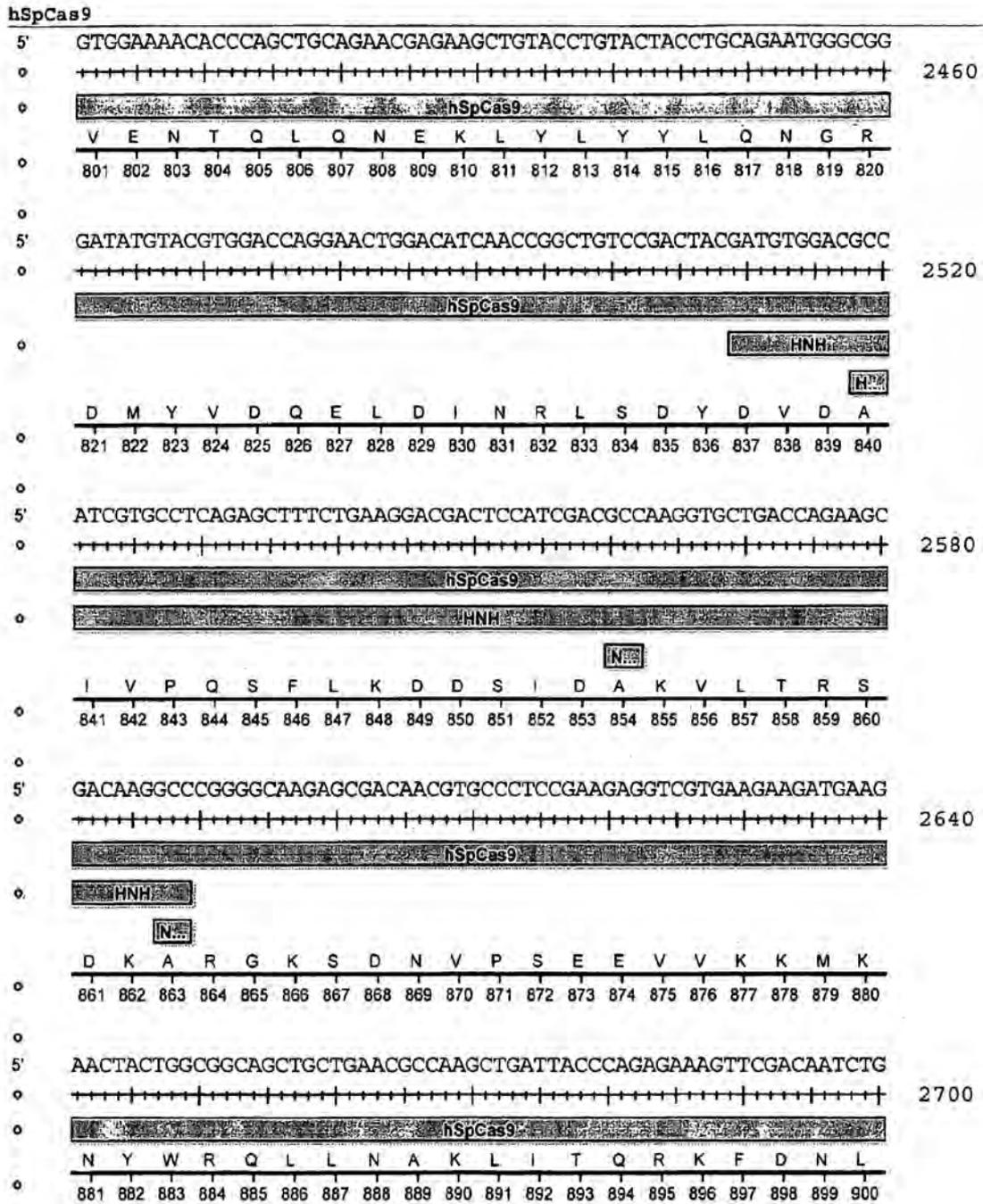


FIG. 10H

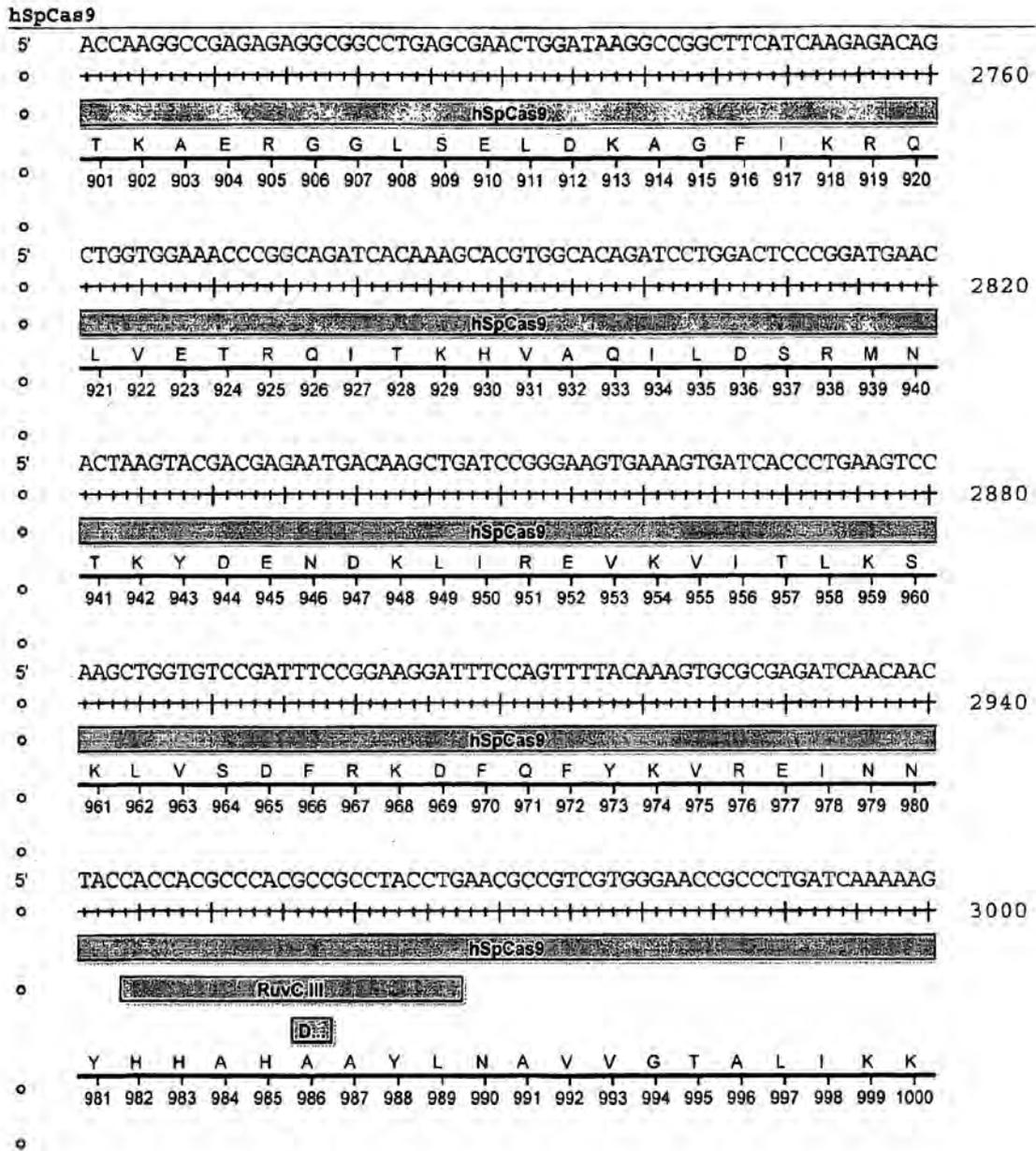


FIG. 10I

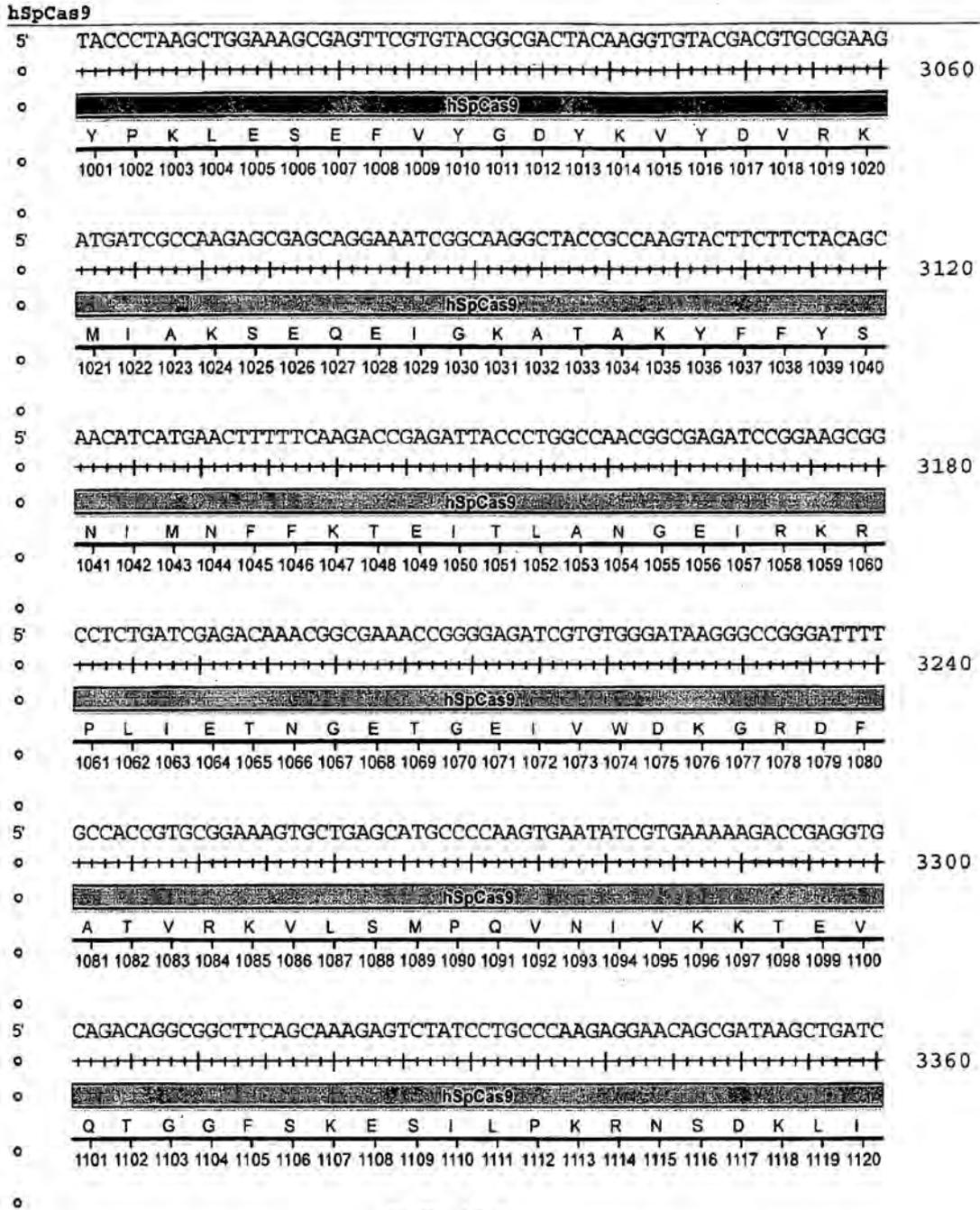


FIG. 10J

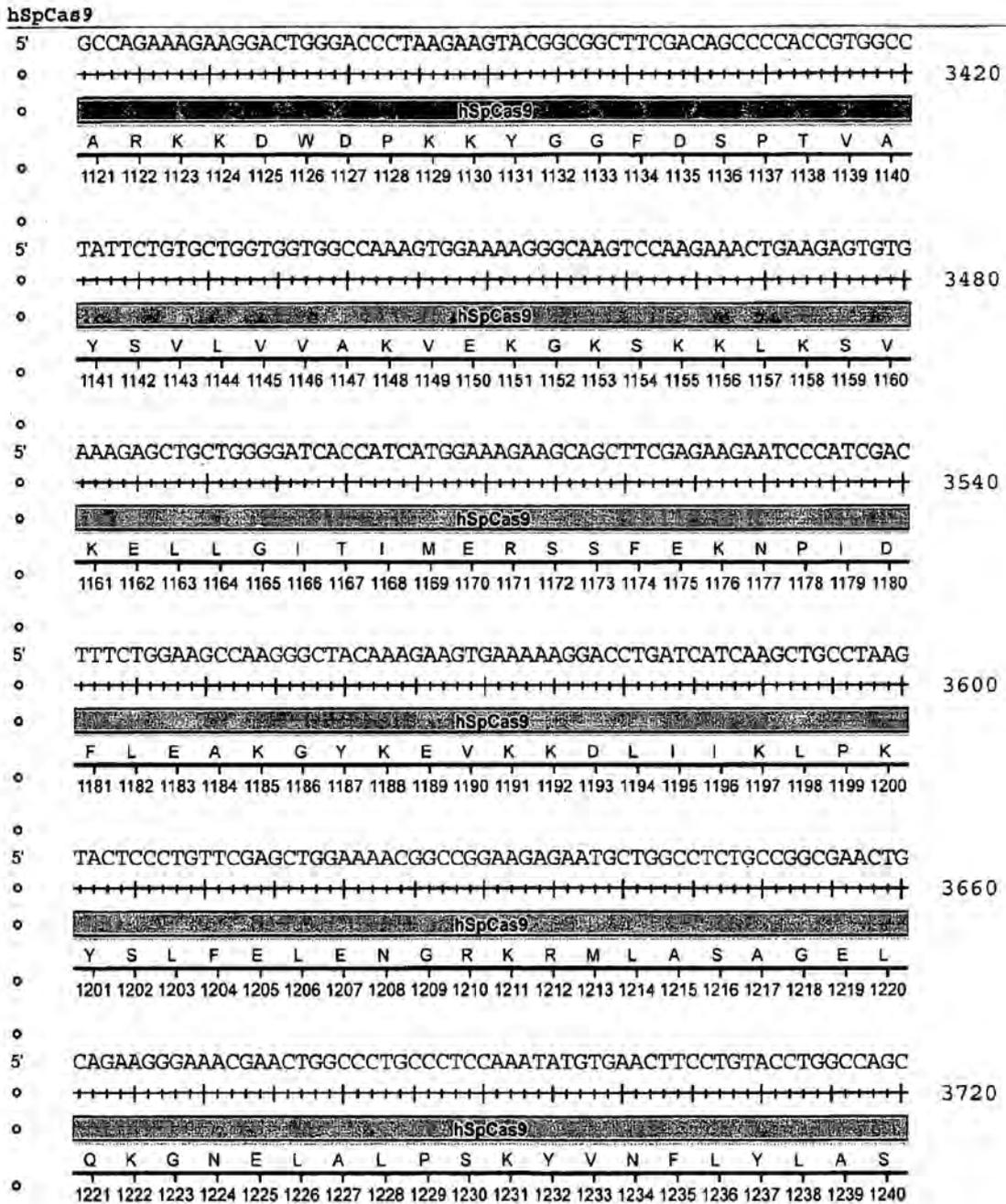


FIG. 10K

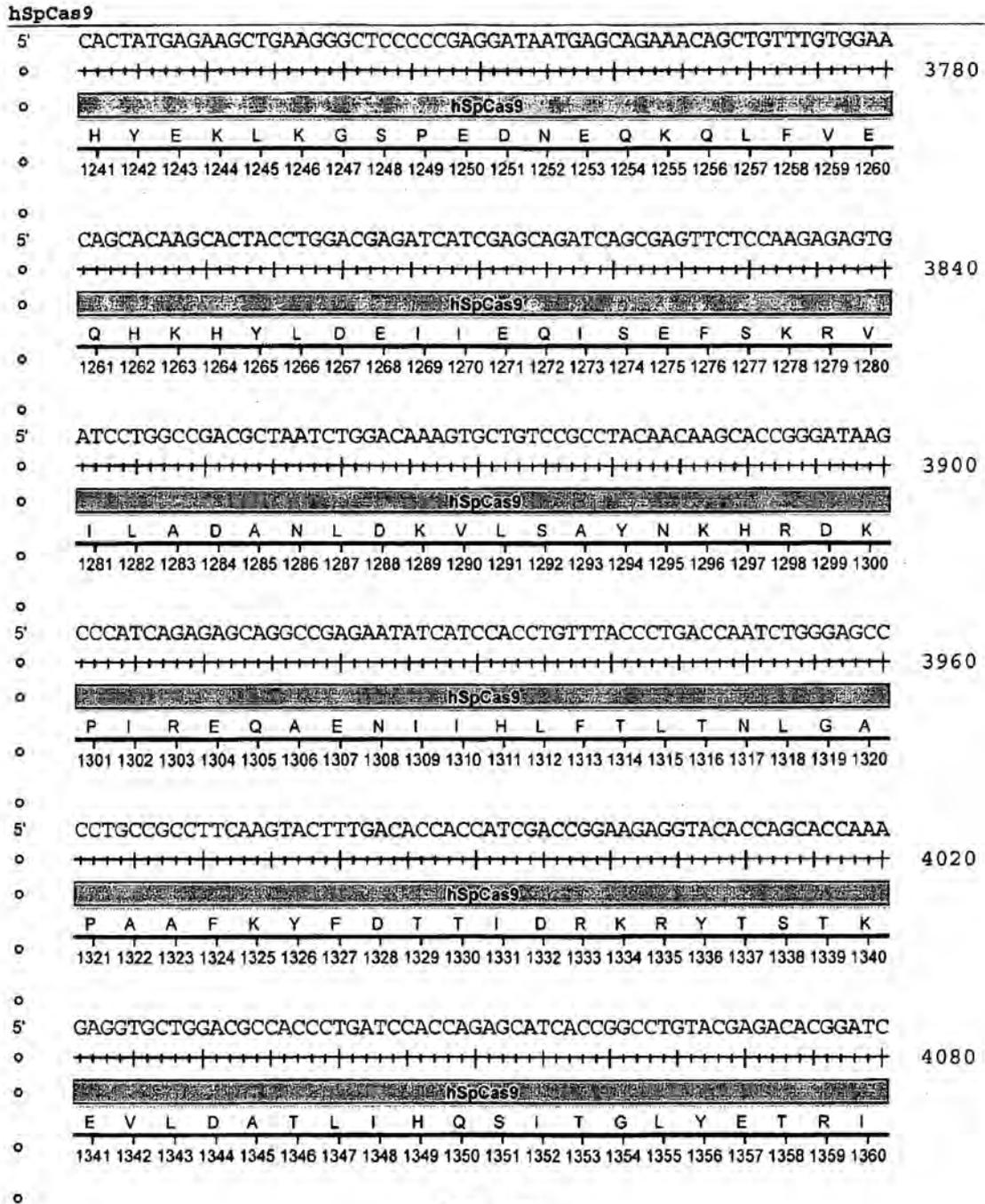


FIG. 10L

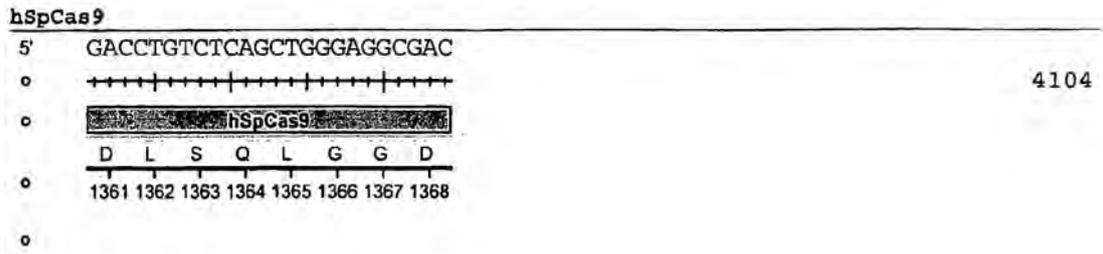


FIG. 10M

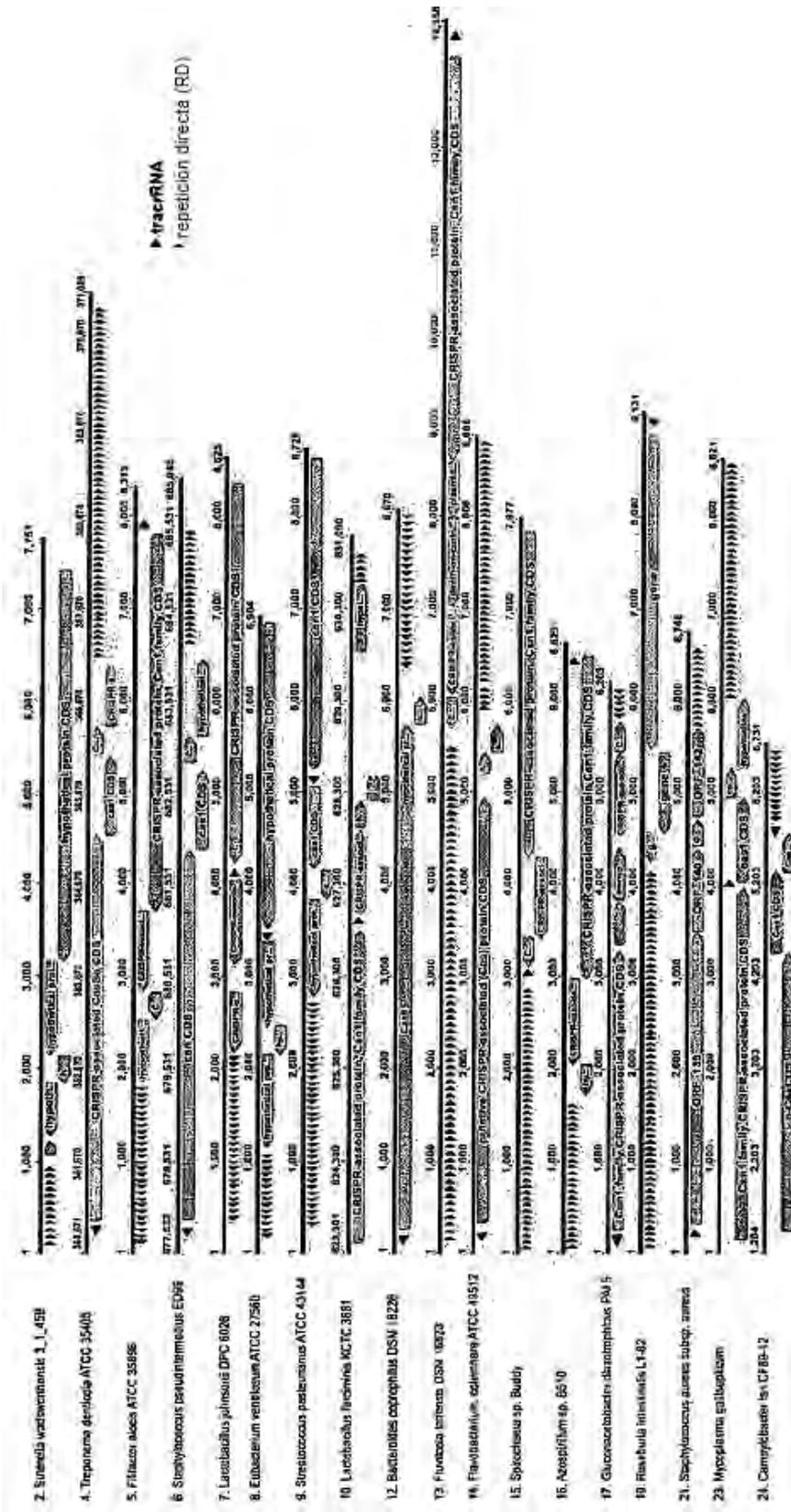


FIG. 11

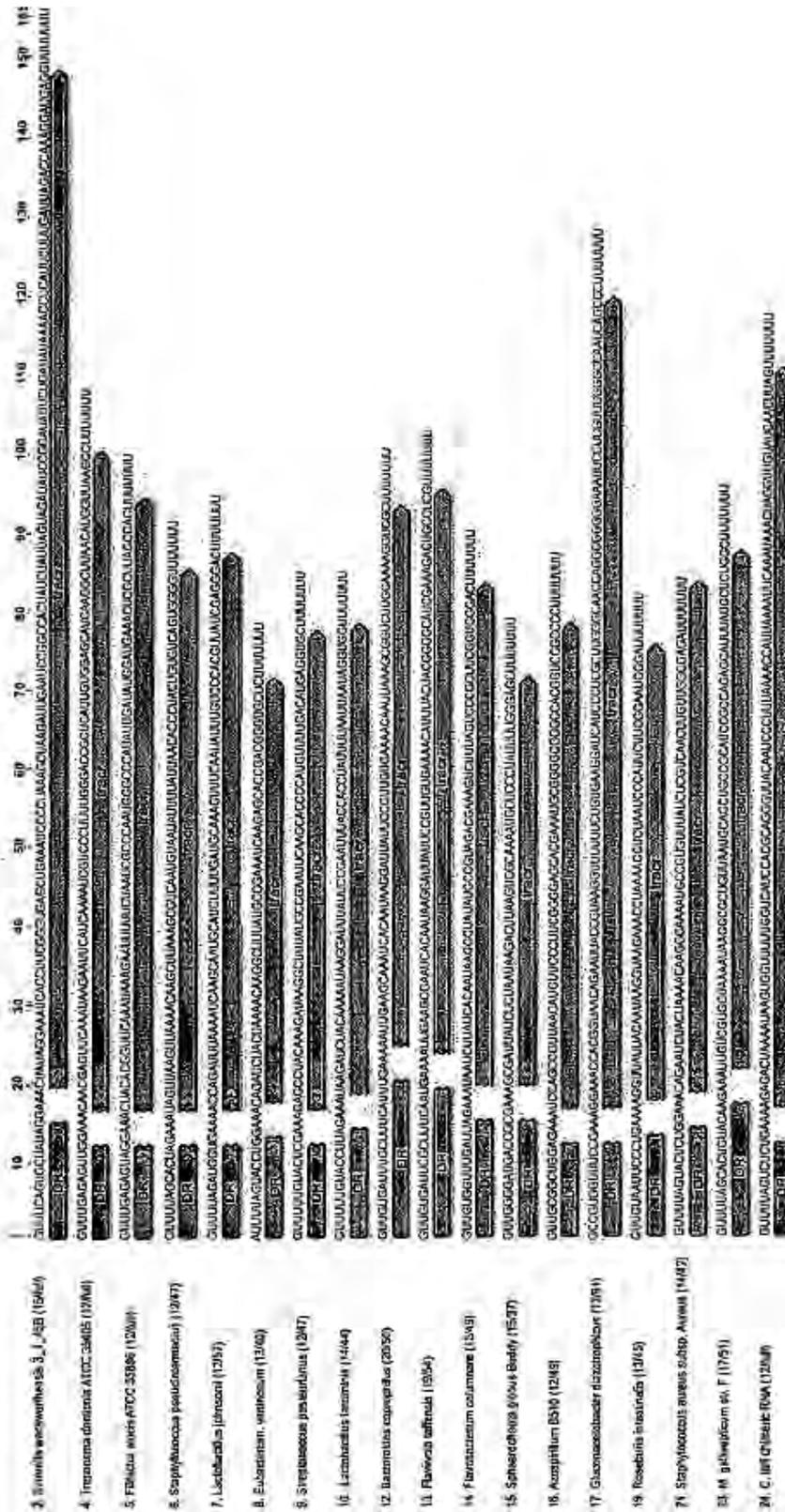


FIG. 12

FIG. 13A

<p>Especie: <i>Corynebacterium diphteriae</i></p>	
<p>Secuenciación código de barras</p>	<p>AGCAATTTC</p>
<p>ID</p>	<p>I</p>
<p>ARN guía quimérico</p>	<p>NNNNNNNNNNNNNNNAUCGGGGUUCAGGAAAACUGAACCCUCAGUAAGCAUUGGGUCGUUUUCCAAUG UUGAUUGCCCGGUGCUCCUUUUUUUUUAAAGGGCGCGCGUUUUUUUU</p>
<p>Secuencias de proteína Cas3 (de tipo salvaje en NLS eteolós)</p>	<p>MKYHVGDVGTFSVGLAAIEVDDAGMPIKTLVSHIHDSGLDPDEIKSAVTRLASSGIARRRLYRRKRRRLQQLDK FIORQGWPIVELEDYSDPLYPWKVRAELAASVIADEKERGEKLSVALRHARHRGWNPYAKVSSLYLPDGPDAFKA IREEIKRASGQPVPETAIVGQMVTLCELGTLKRCGGVLSARLQOSDYAREIQEICRMQEIGQELYRKIIDVVFAAES KGSASSRVGKDPQGNRAKASDAFORYIAALIGNLRVVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVVTIAEI LGIDRGLIGTATMTDDGERAGARPTHTDNRISVNSRIAPLVDWWKTASALEQHAMVKALSNAEVDVDFDSPEGAKV QAFFADLDDVHAKLDSLHLVGRAAVSEDTLVRLTRRMLSDGVLDLYTARLQEFIEPSWTPPTPRIGEVPVGNPAVDR VLKTVSRWLESAKTVGAPERVIEHVREGFVTEKRAREMDDGMRRAARNAKLFQEMQEKLNQQKPSRAIDLWR YQSVORQCOCAYCGSPITFSNEMDHVPRAGQSTNTRENLVAVCHRCNQSKGNTPFALWAKNTSIEGVSVEAVE RTRHWVTDTGMRSTDFFKFTKAVVERFQRAVMEDEIDARSMESVWMANELRSRVAQHFASHGTTVRYVYRGLTA EARASGISGKLFDFGVGKSRLLDRRHHADAIVAFVTSYVAETLAVRSNLKQSQHRQEAQWREFTGKDAEHRA AWRVWCOQKMEKLSALLTEDLRDRVVVMSNVYRLLRGLNGSAHKETIGKLSKVKLSSQLSVSDIDKASSEALWICALTR EPGFDIPEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAEELGSSFFHARVYKITSGKKPAPAFAMLRVY LLPYRNQDLPSVELKPKQIMSMRQAEKKLRDALATGNAEYLGWLVDDELVVDTSKIATDQVKAVEAELGTRRWKY DGIFFSPSKLRLRPLQMSKEGIKKESAPELSKIIDRPGWLPAVNLKFSDDGNVTVVRRDSLGRVRLRESTAHLPVTWKVQ</p>
<p>Secuencias optimizadas en codones de mariferos (Agcl-HA-NLS-Cas9-NLS-taa-EcoRI)</p>	<p>ACGGGTGCCACCATGTA CCGATACGATGTTCAGATTACGGCTTCGCCGGAAGA AAGCGCAAGGTTCGAAAGCGTC CATGAAGTACCATGTCCGAA TCGATGTCCGAA CTTTTCTGTGGGCTGGCTGCTATTGAA GTGGATGACCGCTGG AATGCC TATTAAGACCCCTGATGTGTGTGCACACATTCATGACTCAGGACTGGATCCTGACGAGATCAAGAGCGC TGTGACCAGGCTGGCA AGCTCCGGA TCGCCGGAGAACAAAGGCGCCTGTACCGA CCGAAGAGAAAGCGCCTGC AGCAGCTGGATA GTTTCATCCAGAGGCA GGGCTGCCAGTGTGATCGAGCTGGAAGATTACAGCGGACCCCTGTAT CCTTGAAGGTGGCGCCGAACTGGCCCTTCTTATATTGCTGACGAGAAAGAAACGGGGGGGAGAAACTGATGTG GGCTTGAGACACATCGCAAGGCA TCGGGATGGAAGAAACCCCTTACGCCAAGGTGTCTAGTCTGTATCTGCCAGA TGGCCCTCAGACGCTTCAAGGCTATTAGGAGGAAATCAAAGCGGCTAGCGGCCAGCCTGTGCCAGAGACTG CAACCGTGGGCAGATGTTGACCTGTGGAACTGGGCACACTGAAAGCTGGAGGAGAGGGAGGAGTGTCTGAGT GCACGGCTGCAGCATCAGATTACGCCCGGAGATCCAGAAATTTGTCCGAATGCAGGAGATCGGCCAGGAACT GTATGCCAAGATCAATTGACGTGTGTTCGCAGCCGAGTCCCAAAAGGCTCTGCCCTCAAGCCGGGTGGGGAAG ATCCTCTGCAGCCAGGAAGAACAAGAGCACTGAAAGCCAGCGGACCGCTTTCAGCGGATACCGGATTGCTGCACTG</p>

FIG. 13B

<p>ATCGGCAATCTGAGATCAGGGTGGATGGGGAGAAGAGGATTTCTGAGCGGTGGAGGGAGAAGAACTGGTGTTCGA CCACCTGGTGAACTCTGACTCCAAAGAAAGCCCGAATGGGTGACCATCGCCGAAATTTCTGGGCATCGATCGCG GGACGCTGATCGGAAACAGCTACTATGACCGGACATGGAGAGGAGCCGAGCAGAGCCCGAATCCACACCGATACT AACAGAAATAATTGTGAACACGCCGATCGCACCTGGTGGAAACAGCTAGCGCACTGAGCGCACTGGAGCGAGCA CGCCATGGTGAAGGCACCTGTCCAACGCCGAAGTCCAGATTTTGAATTCGCCGAGGGAGCAAAAGTGCAGGCAAT TCTTTGCCGATCTGGACGATGACGTCCACGCCAAGCTGGACCCCTGCATCTGCCCTGGACGACGACGCTTACT CCGAGGACACTCTGGTCAGACTGACCCGACGGATGCTGAGTGAATGGGTGGACCTGTATAACCCCGGCTGCAG GAGTTCGGAATTGAACCTAGCTGGACCCCAACCCACACCAAGAAATCCGAGAGCTGTGGGCAATCCAGCCGTCGA CCGGGTGTGAAACACAGTGGACAGATGGCTGGAAATCCGCAACAAGACTTGGGGCCGCCCAAGAGGGTCAATCA TTGACACGTCGGCAAGGCTTCGTCACTGAGAAACCGCTCGAGAAATGATGGGACATGAGAAAGGGCGCA GCCCGAAACCCAAAGCTGTTTCAAGGATGACAGAAAGCTGAAATGTCAGGGCAACCCAGTCCGAGCCGATCT GTGGAGTACCAGTCACTGTCAGAGACAGACTGCCAGTGGCTATTTGGGGTCCCAATTAACCTTTTCTAATAG TGAATGGACCAATCGTGCCACAGACAGGGGATCCACCAACAAGGGGAAATCTGGTCCCGCTGTGCC ATCGCTGTACCAGTCTAAGGGCAATACACCTTTGGCTATTTGGGCAAAAACACITCTATOGAAAGGGTCAAGTG TGAAGGAGCCGTGGAAACGACAGCAATGGTCACTGATACCGGCATGAGAAACACTGACTGACTCAAGAAGTTC ACCAAGCTGTGGTCGAGCGGTTTCAGAGAGCAACAATGATGAGGAATCGACGCCAAGCAATGGAATCCGT CGCTGGATGGCTAATGAGCTGAGGAGCCGCTGGCTCAGCACTTCCCATGGAAACCCAGTCAAGGJGT ACCGAGGCCAGCTGACAGCAAGGCTCGACGGCAATGGGATCAGTGGAAAGCTGAAATTTCTTTGATGGCGTG GGGAAATCCAGCTGGATAGAGGCCAATGCTATTGAGGCTGCAGTGAATGCAATCCCTCTGACTATGTGGCC GAAACACTGGCTGTCCGCTCAAACCTGAAACAGAGCCAGCCCAACGCAAGGCTCCCTCAGTGGAGAGATT CACCCGCAAGGATGCAAGCAATCGAGCACTGGAGTGGTGGCCAGAAAGATGGAAACTGAGCCCGCTGC TGAACCGAGGACCTGGAGATGACCGGGTGGTGTGATGCTAAGCTGGCACTGGGCTGGGAAATGGCAGTGCC CACAAGGAAACCAATGGCAACTGTCAAAGGTGAACCTGTCCCTCAGCTGTGATGCGGATATCGAACAAGC AAGTTCAAGGGCCCTGTGGTGTGCTCTGACCAAGAGCCCGGATTCGATCCCTAAGGAGGCTGGGCTGCAACCC TGAGAGACACATCAGGGTGAATGGGACACATGCTACGCCCGGGGACAATATGGACTGTTCCAGTGTCAAGAG GAAAGCATCGCATGAGGGGAGGATACCGCAGAGCTGGGCAAGCTCCACCAATGCTGGCTGTATAAAAATTAATT CCGGCAAGAAACCCGCAATTTGCCATGCTGAGGGTGTACCAACATCGATCTGCTGCCCTTATCGCAACCCAGGACCTGT TTAGGTGGAACTGAAGCCACAGCAATGTCCATGAGCCAGGCTGAGAGAAACTGGGCAACTGCTGGCAACT GGGATGCAGAAATATCTGGGATGGCTGGTGGATGACGAGCTGGTGGTGAATACATCTAAGATGGCCACTGA CCAGGTCAAAGCAGTGGAGGCCAACTGGGCACTACCCCGGATGGCCGGTGGATGGATCTTTTCCCGCTTA ACTGAGACTGAGCCCTGTCAGATGTCCAAAGGAGGGATCAAGAAAGACTCCGCTCCCGAATGCTTAAATCA TTGACAGACCAAGGATGGCTGCCCGCTGAACAAGCTGTCTGTGATGAAATGTCAACCGTCTGGGAGAGACT CTCTGGGACCGCTGGACTGGAGATACAGGCCAACCTGCTGCTGACTTGGAAAGGTGCAAGTAAGAATTC</p>	<p>1084</p>
<p>ligninase proteine</p>	

FIG. 13D

<p>Secuencias optimizadas en condiciones de mismos Terros (Agel-HA- NLS-Cas9- NLS-Iaa- EcoRI)</p>	<p>ACCGGTGCCACCAATGACCATACCGATGTTCCAGATTACCGCTTCGCCGAGAGAAAGGCCAAGGTGCGAAGCGGT CATGACTCAGAGCGGAGCGACGATTTCTTGCAGCATTTGGCATTGGCATTGACATGGGGCTAAGTACACTGGGGTGTCTA CGCACTGTTCCGACCGGGAGGAACCTGCCACAACCTGACAGCAAGGCCATGACCTGGTCAATGCCTGAGACAG GGCACAAGATACGTGTCAGGCAACAGAGAACCTCCGTCAGACACAGGCTGGGGACAGAAAGAGATATACCCCTGGCT AGGAACCTGGCATTTCTGGTGGTGCAGATAIGATAGAAACAGGAAAGAGGCTGACTCGGCCCAACGCTGATGAGAAATGGAA ACGAGACGGGAGGCCCTGTCCGCTGTGAAACGGGAGAGGCTACTTCGGCCCAACGCTGACCGGGGAGAATC TGAACCCCTCTGGAGAAATGTGAGAGCAGACGTTGTCGGCTCATCCCTGCCCTCAGCACATATTTTCCGGAAGTGC GCTCTGCTGAGCAGTGGAGAGATTCAACCGCAACATCAGCAATGTGAGAAAGTTCAGACCGGAGAAAGCCTAAC ATCCCGCGGATAAAGAGTTTCAATGAAATTTGCCGTGGCTGAAGGGCTGATTCAGACAGACCGGAGAAAGCCTAAC CCAGTCACTCTGAGCAACCTGAGGGCAACCGCCAATGTGCTGACAGGACTGGGGCAGATGGGCCACAAGCCTA GATCAGAAATTTTAAAGCAATCGAGGCCACCTGAAGAAAGATAGCCGCTGGCCAAAGATTAACGAAAGCAATTC GGAAGGACAGAGCGCTGGCTGACTGCTGGAAACCTGTCCAAATCGCAGCTCCGGCCAGAAAGATGTACTT CAATGCCCCCGACATCATGAAGATAGGGGCTGGAGGCTGATCGCTTCAAGAAACACTGGTGGGGCTTTTA AGTTCTTCAACCCAGCAAGGACCAAGCAACAGCATCTGGAACATCAACAGATTTGAACACAGCAAGAT ATCAATTGAGACTCTGTGCAACCTGGACCCAACAGAACCTGCCCTTACGAGATCAGAACAAATAGGCCGCA CCCCAGACTGACCTGGTGAACCCACACTGGCAACCTGACCCCGGAGTTCGGGAGATCTGGAAAGATCTACCGACAGGAAGA GTCCCTGGCATCAACGGACAGGACCACTGCCTACACTGGCTTACCACTGAGTTATGCACTGAGAGGCT TCGACAGGTCAAAAGCCCTGATCCATATGCTCTGAGGGCACTGGCTGCAGGCTCAAAAGCAATAGCTGACA TCCGCGCCACTGCTCTGGAGAACTGCAACGGAGGCCAGAAATGTGAAACCTTCCCTGGACTGGCCGACGGTAC TATCGGGAAACAGACGATGCCAAAGTCCGGCTGGTTCGACAACGCCGATGGACTGCTGGAGAGATCTGACCT GCAATCCCAATGAAAGAAAGATCCCTGCCCTGGTGGCCAAATTTCTGCAGACAGATGAAACCCACAGGCC AGAAATTTCTGGACGATCTGGCGAAACAGATTAAGGGCGGGAACHTGGCTAGCCGATGTGCACGGATC GACAGATGGGAAATCCCTCGGGAGGCTTAACTTCCCTACAATACCCGCTCAGTAAOGGAGGTGAACAA GCTGCCCGCAATGCCAGGATAAGAACTGCTGACAAATCAGAGATAGGGTGGCTGAGACTGCAGACTTCAATG CCGCTAACCTGGGCTGTGACGAGCAGAAAGAAAGTTCGCCAAATCTTTTAACTGGCTCAGTTCTACACCC TGATCGACAGAAATGTTCGGGATTTCTGCAACTACCCCTGCCGTCACCTGGAGAACCGCTGGAGAACCGCTGAGGATGACAA TCAAGGATGCTGTGATAATGGGAAACTGTGACAGCAGCACTGCACGAGGCTGCTGCAGAGACAGCTCGC CCAATCGATGGACTGGTGAAGGGCTGGACAGACAGGCTGGGAGATCGCAAGAGGGTCTCAACTGACAT TCAGACCAAGTCCATTTCTCCAACGGCATCGTGGACGTCAGCAATTTTGTGGAGGAAATAAAGTTGGAGTTTC CGCATCTGTGGCCGATCTGAAAGAAACAACCGGTCAACAGACAGATGCTGTCCGAGGCCGAAAGCTGGAA ACCAGATGGCTGATCAAAATGAGGGGATCAAGAAAGGCCAGCCGGGAACCTTGTCCCTACACCGGCGATAGGCT GGGTGAAGGGGGAGAAATCGACCAATTCGCCCGGAAAGCTGATCAAGGATGCCCGGGGAAATGTGTTTAAACG CTGAGCCATACTGATGCAAGCTCCCGGGCAACCAAGCTGAAAGAAAGAAATCAGCGATACAGTCTGTCAAGATC TGAAGGCCAATATCGGAAATGAGATCTTCAAAAACCTAGCAACACTGCTGCAATACCGCCGAGATGAGGACGGT</p>
---	---

FIG. 13E

<p>GTCACTAAGCTGCAGCAACCCATAGACTGAAATCTTTGATCTGGTGAATGAGCAGCAAGGACTGCGTGGCGG CACGCCCTGTTCTGGACGATGGCAGCGAAGCTCCGACGAGCTCCGAGCTGGAGCTGGGCAACAAGCCGCCGAAAC TCGCGTCAACGGGACACAGATCTGGATGATTAAGAACCTGGCCACAAGATCCGAGAGGAACCTGCAGAAATTTGGT GTAAGACAACAAATAGACTGCACTTTTCAGCCCGCTGCAACTAACGTTGCCGATGCAAAAGAAATCTGAAGCCTG AACTGGCCCCAGAAACAGCCCCACTTCGAGAAAGCCAGATATCCAGCCCAATGCCAGCCATTCCTATCGACGCCCTG TGCTCTTTCGCTGTGGGAGTGTGACCGCAAGCCGATCAGAAATGGATTTGACTACTCCTGGAATGGCAAGACCGTGTG CTGGACTGTATCCACAGAGCTGTGAGGTCAATTCACCTGACGCCAAGCCCAAGGAAAAAGTCAATTTTCGA TTCAGTGGCTATCTTTAAGGAAGGCACTCTACGCAGAGCAGTTCTGCTATGTTTACCCTGAACGAAAGAGATCTG GATTTGGATATGAGACACTGAATGCCAAAGCCGAAAGATGGCGGCTATTTGAGGTGAGTGGCAACAGCCCAAG GAGCTGCTGAAATGCTGCCCCCTTCTTAAGAAAGCCTGTGGGCGACCTGTGAGCCCAAGCTACTTACCCGGATC CTGAATAAGCCTGCATATGAGTTTCTGGCAAGGCCAGCTGTGCAGCCACTGAGCCGAGGAAATAAGACTGOC AGCCCTGTGGATGCTCTGGCTACTGTACCAAGTCCAAAGTCACTGATGAGCCTGTGAGCCTGTTCAATGGCTGCAACCGGAA ATCCCTGAAAAGGGGAGGAGCTGCTGAAACCCAAAGCTGTTCCAGCTGAAGGCTGGCTGAGAACTCTCCGATAGCCAGACT AGCTTCAAAGCTGAACGGGAGCCCTGACCCCTGCTGAAACAGGACTGGCTGAGAACTCTCCGATGCTGCAACCGGAA GGCAAGCCCTTTGGCAACCCCTGTTCCGGCGATGAGCTGACATCTAAGCTGGCTGGCAATTTGGAAACCGACCTGT GATGGGGATCTGGCTCATGCACAGTCCGGAGAGATTCAGCCTGCCGCAATCCAAACCCAAAGTGGAGGGT TCAGGATTAGGGGCAACAACTGTTTGGCAATGAGCTGTACCAAGTGCAGCCCAAGCTGCAAGCTGCAAGAAATCTGA GCTTCGCTCCGCTGGTCTAAATGTCGACTGGTCCAGGGGATCTGTTTAAACGAGCTGCAGCATGAAATACTGA CCGAGTGGGAGGCAAGTTCAATTACAAGCCCGATGTGACTCCTATGTCCGAAATGGCCAAAGTGGTCCGAGAG GAACAACCTGTCTACTGGAATGCTCCAGGGACAGAGGACGCGGTACGTGAGGGTCCGAGACAACATTCATCCCA GGCCAGTCACTGGTTTGAACAGTCAAGTGGAGAAATGGGCCATTAAGTCCCTCTGTCACTGCCAGCTTCCTTCAA UGTGGACAAACAGCTGAGTTTCAGAAAGCCAGTGGGAAACCCGAGCTGTCAAGAACTGTGGCCAGCCCAAGGAGCG AAATCTTCAATGAGAACTGGGCAATGCCAAGCATATCCGCTTTTGGTACAATTTGTTGTTGAGCAAGCAACAAGA TGAAACGAGTCTTACAACAATGTTGCTAAGAGTTAAGAATTC</p>	<p>1422</p>
<p>Longitud de proteína (sin NLS y marcos de afinidad)</p>	<p>NGG</p>
<p>Longitud de espaciador</p>	<p>35-39</p>
<p>Especie: <i>Legionella pneumophila</i> Paris</p>	

FIG. 13G

CAACCTGCTGGGCCACCTGTCCAAACCTGCAGTGGAGAATCTGCATAGGTATCTGGCCAAAATAATCCAAAGCAGTT
CGACGACAGACCTTCGGGAACGAAATTTCTCGGATGCTGAAGAAATTTCCGACACCTGAAAGGATCTCAGGAGA
GTCTGGCTGTGGCCAACTGTATCCAGCAGCTGGAAACAGTCTCAGGATACATCAGTATCTCTGGAGAATAACCCCCC
CTGAAATCACCAATCCACCCCTATGAGCCCGGACAAAACATGGCATGGAATAAGACCAAGCCCTGCTGTGAAC
CCCGAGAAGCTGAACAATCTGTACCCAAACTGGAGAAATCTGATCCCCCGGATCAATGACGCCCAACCCCTTCCCTG
GAGAAGTCTGGAAACATAACAAGCTGAGACAGAGAAACGCAATCAATTTCTCCAGTAACACAGGACGAGAAAGC
GGATAGCTACATCTCTGCAGAGATATCTGGACCTGAACAAAAGAATCGATTAAGTTCAAGTCAAGAAAGCAAGCTG
AGCTTCTGGACAGGCAACAGCTGCCTGCTACCTGATCGAGACACAGAAAGGATGGAAACTCACCTTCAA
TTCTAGTCTGGTCTGCTGTATCCAGATTCAGATTCAGATACAAACAAGAGAGGGAAGATGCCCTCAGGGGAT
CTGGTTCCGAATGCCITTTCACTGTGGAGCTGAGCAACATCAAATCCCTCACGCAACAGAGAATTCCTGCCCT
GCTGTGGGAGCAATCCTGAGCGGACTTCAATTAACAACAAGGATAAGTGGCCAAATTCAGAGATCTTTTGGGA
ACACCCACAAGATTGGGCGAACATCACTGAAAGCAAGTGTAAAGATCGAAGGAGCCCGGAAACAAGTGG
AAACGCTTTTAAAATCGACTACGAGGAAGCTCTGAATCAACCCAGAGCATTCAAAACAACAAGGCCCTGATCAAGA
TCAATTCAGACCAATTCCTGATATCAATTCAGGCCATCCAGTCAACCTGGGACATAACGACAGCCAGGCTCTGATCT
ACCACAACTCTTCTCACTGAGCCAGCTGTACTATCTCTGGAGACAAGAGGGATGGCTTTCATAAAAACCTGGC
TGGCCGTCACTTGTGAAATTAATCTGGCGGACCCAGAAAACCGAGATCGACCCAGAAAATTTCTCTATGCATCTAGGC
TGCCAGCAGACAGTGTGGCCCTTCGATGGCTCTGGCAGGATGATGCAAGCCCTGACGAGGCTGACGAGATCGCCA
TGGCTAAGTGGGACAGATCAACACATCTCTGATACTCAAGCTCTGATCCCAATCACTCCGAGCAGCAATC
GGTTCCGAATTTGAGGAGCTTCAAGAAATCAAGGGTCTCTAGTGACAAAACCTGGAGCAGGCCATCGAA
AAGCAGACATTCAGTGGGAGGAAAGTTCCAGAGAAATCAATTAACGCAAGTATGAATATCTGCCCTTACAAGGG
CGCTCAATTTGGCGGCGAGGGGAGATCGACCACTACCCAAAGTCCCTGTCTAAAAGCATTTCCGGCGTGT
CTTTAATCCGAAGTCAAATCTGATCTACTGTTCAAAGCCAGGGGAAATCCGGAGAAAAGGAGGAAACACTA
CGAATCTGTCTCCACTGTATCTGAAAACACCAAGTTCGGCACTGACACGTTCCGATATCAAGAAATTTATTTCT
CAGAACGTCTGTAATATCAAAAAGTACATTTCTTCCACCTGCTGACCCCAAGCAGCAGAAAGCCAGCAGGCA
CGCCCTGTTCTGGAATATGACGATGAGCATTCAAAACCAATACAAGTTTCTGATGTCTCAGCAGAAAGCCAG
AGTGAACGGCACACAGAAAATTCCTGGGGAAGCAGATCATGGATTTCTGTCCACTCTGGCAGATTTCTAACAAGCT
GCAGCTGGAGTTCAGCATCAAGCAGATTAACCCGAGGAAAGTGCAAGCCATAGAGAGTGTCTAAGCAGG
AACCCAAACTGGTCAAGATAGGCAGCAGATTTCCCTTCCACCGCTGACCCCAAGCAGCAGAAAGCCAGCAGGCA
GGCTGAAGGAGTTCCACAGTTAGTCAAGAACTGGATAACTCATGTTTATCAATCACCTGTATGCCAGCAGG
TGCACTGAACCCCGTCCGGAGCAAGGAAAATACAAAACCTAACATCTCCTCTACTCCACTGTTCACAGGATTT
CCCTGTATGTGAGCCGTTCTACCCCGTGTGGTCAAGGAAACCTTCGCAATCGGCTTTAGCGAGAAAGGACCC
TGTTCCGAGTCAAGCCCTCCAAAGGAGAACTGTTTACACTGTGAAACCTACAGTACTAAAATCTCTGGCG
AGTCACTGCAGAACTGCAGGCTAAGAGAAAGCAAGGCTGTACTCCCAATCAACAAAACCCCTGGCCCTG
GAGTTCCTGCACCAATTAATTTCAAGGAAATCGTGACACCCGAGTACTACCGTGTGCCATTTCACTCAACTC
TGAGATCTACACTAAAAGGAGAGCATCACCGTGAATAATCTGAAGGAAACCCATGCTGTGCTGCCGTC
TTCCGAGATTCAAAAAGAAACGTTGCTGGGATCTTTTAAACACACAAATCGCCCTGCCTGCTACTAAGGATTTGGGAG

FIG. 13H

<p>Longitud de proteína (sin NLS y marcador HA añadido)</p>	<p>AGGCTGTTCAACCAATCCAAATTTTCTGGCACTGAAGGCAACCCAGCTCCCAATCCTAAAGAGTTCAA TGAGTTC ATCCGCCAAGTACTTCTGAGCGACAACAATCCCAACTCCGATATCCCTAAACAATGGCCACAATA TCAAAGCCCCAG AAACATAAGGCCGTGGCAAGGCTTTAGCCTGCCAGTGA TCCCGGGAACGCTGGAACCAATGA TGGGCATTAG GGCAAAGACAATAAGGGACAGCCACTGTATCAGCTGCAGACAATGACGATAC TCCCAAGCATGGGCATCCAGA TTAACGAGGATCGCCTGGTGAACAAGGAAGTCTGTATGACGCTACAAGACACGAAATCTGAGCACTATCGAT GGGATTAACAATCCGAGGGACAGGCAATATGCCACATTCGACA ACTGGCTGACCCCTGCCGTGAGCACTTCAAG CCTGAGATCAATCAAGCTGGAATGAGCCCTCACTCTAAACCCGACGCTACATCAGAAATACACAGAGTCTGGC CGACTTCAATCAAACTATTGATGAGGCTCTGATGATCAAGCCCAAGTACTCAATTTGACCGATCCTCTGAAACATGCC AAATGAGATCGTGTGTAACAACAAGCTGTTCCGGGAATGAACTGAAAGCCTAGGGA TGGAAATAAGATCGTGA GCACTGGCAAGATTGTCAACCTACCGATTGAAAGCGACTCCACCCCAAGTGGATCCAGACCCCTGTATGTGACAC AGCTGAAAAAGCAGCCCTTAAGAAATTC</p>
<p>Longitud del espaciador</p>	<p>1372</p>
<p>PAM</p>	<p>NGG</p>
<p></p>	<p>34-36</p>
<p>Especie: <i>Trepionema denticola</i></p>	
<p>Secuenciación código de barras</p>	<p>TCGCCCTTG</p>
<p>ID</p>	<p>4</p>
<p>Secuenciación código de barras</p>	<p>NNNNNNNNNNNNNNNGUUGAGAGUUGGAAACAACGAGUUCAAAAUAGAAUUCACAAAAUCGGUC CCUUUGGACCGCCUCAUUGUGGAGCAUCAAGGCCUUAACAUGGUUAAGCCUUUUUUU</p>
<p>Secuencia de proteína Cas9 (de tipo 3A) (sin NLS añadido)</p>	<p>MKKEIKDYFLGLDVGTSVGWA VTDIDYKLLKANRDLWGMRCFETAETA EVRRLHRGARRRERRKRIKLIQEL FSQEIAKTDEGFFQRMKESPFYAEDKTI LQENTLFNDKDFADKTYHKA YPTINHLJKAWIENKVK PDPRLLYLACHNIIK KRGHFLFEDDFSENQFDSIQALFEYLRDMEVDIDADSQVKEILKDSSLKNSEKQSRNLKILGLKPSDKQKKAITN LISGNKINFADLYDNFDLKDAAEKNSISFSKDDFDALSDDLASILGDSPELLKAKAVVYNCVLSKVIGDEQYLSFAKVKI YEKHKTDLTKLKNVIKHFPKDYKVFQYKNEKNNNNYSGYVGVCKTKSKLIINNSVNDQEDFYKFLKTILSAKSEI KEVNDILTEJETGTLPKQISKNAEIPYQURKMELEKILSNAEKHFSFLKQKDEKGLSHSEKIIMLLTFKIPYYIGPINDN</p>

FIG. 13I

	<p>HKKFPDRCWVVKKEKSPSGKTPWNFFDHIDKKEKTAEAFTSRNFTCTLYGESVLPKSSLLYSEYTVLNEINNLQIII DGKNICDKLKQKIYEDLFFKYYKKITQKQISTFIKHEGICNKTDVEIILGIDKECTSSLSKSYIELKNIFGKQVDEISTKNML BEIIRWATIIYDEGEKTLKTKIKAEYGYCSDEQIKKII.NI.KFSGWGRLSRKFLFETVSEMPGSEPVNIITAMRETQNN LMELLSEFTFTENIKKNSGFBDAEKQFSYDGLVKPLFLSFKMLWQTLKLVKEISHITQAPPKIKFIEMAKGAELEP ARTKTRLKJLQDLYNNCKNDADAFSSSEIKDLKSIENEDNLR.RSDKLYLYTQLGKCMYCGKPIEIGHVFDTSNYDID HIYPOSKIKDDSIENRVLVCSSENKNEKEDYPLKSEJQSKQRFWNFLQRNNFISLEKLNRLTRA.TPISDDETA.KFIARQL VETRQA.TKVA.AK.VLEKMFPETK.IVYSKAEITVSMFRNKFDIVKCREINDPHAHDA.YLVNIVGNVYNTKFTNNPWNFI KEKRDNPKIAD'IYNY'K.VFDYDVK.RNNITA.WEKGK'TIITVKDML.KRNTPIY'TRQAACKKGELFNQITIMKKGLGQHP KKEGPFNSISK.YGGYNK.VSAA.YYTLIEYEBK.GNKRISI.ETIPLYLVKDIQKDDVL.KSYL.TDLLGKKEFKIL.VPKIKINSL L.KINGFPCHITGK.TNDSFL.RPA.VQFCCSNNEVL.YFKK.IIRFSEIRS.QREKIGK.TISPYEDLSFRSYIKENL.WKKT.KNDEIG EKEFYDLLQKKNL.EIYDMLL.TKHKDIYK.RPNSA.TIDIL.VKGKEKFK.SLIENQFEVILEILKLFSA.TRNV.SDLQ.HIGGS KYSGVAKIGNK.ISSL.DNCIL.IYQSITGIFEKRIDL.I.KV</p>
<p>Sequences homologous to domains of mRNAs (AgeI-HA- NLS-Cas9- NLS-tag- EcoRI)</p>	<p>ACCGGTGCCCA.CCA.TGTA.CCCAT.ACGATGTTCCAGATT.ACGCTTCGCCGAAGAAA.AAGCGCA.AGGTCG.AAGGTC CATGAAAA.AGAAA.TCAA.AGACTAC.TCCTGGGGCTGGATGTGGGACTGGGAGCGTGGGTGGGCTGTGACCG ATACTGACTACAA.ACTGCTGAA.GGCTAACCGAA.AAGAA.CCTGTGGGCA.TGAGATGCTTCGAGACAGCCGAA.ACT GCTGAGGTCCGAGACTGCACAGGGGAGCCAGCCGCA.ATCGAGCGGAA.AAGAA.AACGCATTA.AGCTGCTGC AGGAGCTGTTCTCTCAGGAAATCGCCAA.AACCGATGAGGGCTTCTTCAGAGAA.TGAAGAA.AGCCCTTTTACG CTGAGGCA.AA.ACAATCTCGAGAA.ACACTGTTCAT.TGACAA.GGATTTTGTGATA.AGACTTACCACAAA GCA.TATCCTACCA.TTATCATCTGATCAA.AGGCTTGGATTGAGAA.CA.AGGTGA.AACCGACCCCGACTGCTGTAC CTGGCATGTCACAA.CATCA.TTAA.GAAA.AGGGACA.FTTCTGTTTGAAGCGACTTCGATTCAGAGAA.TCAGTTT GATACCAGCATCCAGGCCACTGTTGAGTATCTGCCGAGGACATGGAGCTGGACATCGATGCCGACAGCCAGAA GGTCAA.AGAGATTTCTG.AAGGATAGCTCCCTGAA.AGAACTCTGAA.AACAGAGT.CGGCTGAA.TAAGATCCTGGGCG TGA.AGCCTTCCGACAA.CAGAA.GAA.AGCCATCA.CAA.ACCTGATTTCTGAA.ACA.AGATCA.ATTTCGCCGATCTGT ACGCAA.TCCAGATCTG.AAGGACCTG.AGAAA.ACTCAATCA.GCTTCTCCA.AGGACGATTTTGA.TGCACTGAGT GACGATCTGCCCTCAATTTCTGGCGACAGCTTTGAACTGCTGCTGA.AGGCCAA.AGCTGTCTATAA.CTGTCTGTG CTGATA.AGGTCA.TCGGGACGAGCAGTACCTGAGCTTCGCCA.AGGTGA.AA.TCTACGAA.AAGCACAAA.AACCGA TCTGACAA.AAGCTGAA.AA.ACGTGATCA.AGAA.ACAATTTCCCA.AGGACTACA.AGAA.AGGTCTTTGGATACA.ACAAGA ACGAGAAA.AACA.ACA.ACAATTA.CTCGGCTATGTGGAGTCTGTA.AGACCA.AGAGTA.AGAA.ACTGATCA.TTAA.C AACTGAGTCA.ACCAGG.AGATTTCTACA.AGTTTCTGAA.AACTATCCTGT.CAGCCAA.AGAGCGAGATCA.AGGAAGT GAA.TGACA.TCCTGACCGGAGATTGAA.AACTGGCA.CCTTCTGCCAA.AAGCAGATCTCTAAA.AGTA.ACGCAGAGATTC CTATCAGCTGAGGAAA.TGGAGCTGGA.AA.AGATCTCTCAATGCCGAA.AAGCA.CTCTCTTTTCTG.AAGCAGAA AGACGAA.AA.AGGACTGT.CACATAGCGA.AGAGATCA.TTATGTGCTGACCTTCA.AGATCCCTTACTATA.TTGGCCC AA.TCA.ACGATAA.TCACA.AGAA.TTCTTTCCCGACAGATGCTGGTGGTCA.AGAA.AGAA.TCCCTTCTGCGCA AGACCA.CCA.TGGA.ACTTCTTTGATCATATCGACA.AGGA.AA.AACAGCAGAGGCCCTCATTACTTCTAGGACCA ATTTTTCAC.TTACCTGCTGGGAGAGAGCGGCTCTGCTCAAGTCTAGTCTGCTACTCCGAA.TATACCGTGCIGAA</p>

FIG. 13J

CGAGATCAACAATCTGCAGATCATTATCGATGGCAAGAAATATTTGTGACATCAAGCTGAACACAGAAAGATCTACG
 AGGACCTGTTCAAGAAAGTACAAGAAATACCAGAAAGCAGATCAGCACCTTCAATCAAGCACGAAGGACATCTGC
 AACAAAACCGATGAAGTATCATCTGGGATTGACAAGGAAATGTACATCAAGCCTGAAAGCTCATCGAGCT
 GAAAACATTTTCGGCAAGCAGGTGGATGAGATCTCCACTAAGAATATGCTGGAGGAAATTAACAGATGGGCTA
 CCATACGACGAGGGGAAAGAACCACTCTGAAACAAAGATCAAGGACAGATACGGAAAGTATTGCTC
 CGACGACGAGATTAAGAAATCCTGAACCTGAAGTTCTCCGGTGGGGCGACTGCTCGGAAATTTCTGGAGA
 CAGTGAATAGTAAATGCCCGCTTCTCAGAACCTGTCAATATTTATCACGGCCATGAGGGAGACACAGAAACAATC
 TGA TGGAGCTGCTCTGAGTTCCACTTCCCGGACATTAAAGAAATCAATTTCTGGAATTCGAAGATGCCG
 AGAAGCAGTTTAGTTACGACGGCTGGTGAACCCACTGTTCTGAGTCCCTCAGTCAAGAAA TGTGTGGCAGA
 CCCTGAAAGCTGGTGAAGAGATTAGCCATATCACACAGGCCCCCTAAGAAAATTTTCATCGAAAATGGCAAAAG
 GGGCCGAGCTGGAAACCTGCTCGGACTAAGACCAGACTGAAAATCCTGCAGGATCTGTATAACAAATTTGTAAGAA
 CGATGCTGACGCCCTCAGCTCAGAGATCAAAGACCTGAGCGGAAAGATTGAGAACGAAAGATAATCTGAGGCTGC
 GCTCCGACAAAGCTGACTACTACTACTCAGCTGGGAAATGCATGTAATGTGGAAGGCCAATTGAGATCGGCC
 ACGTGTTCGATACCTCAAACCTAGGATAATGACCATATCTATCCCCAGAGCAAGATCAAAGACGATAGCATTTCCA
 ATCGGTGCTGGTGCAGCTCTGTACAAGAACAAAGGAGCAAGTACCCTGAAATCAGAGATCCAGAGC
 AAGCAGCGCGCTTCTGGAACTTTCTGCAGCAACAAATTTCTCTGGAGAGCTGAATAGACTGACAAGG
 GCCACTCAAATCAGTACGATGAGACAGCCAAAGTTTATTGTAAGCAGCTGGTGGAAACTCGCCAGGCTACCAA
 GGTGGCCGCTAAAGTCTTGGAAAAGATGTTCCCGGAGACAAAATCGTGTACAGCAAGCCGAGACTGCTCCCA
 TGTCCCGAACAAAGTTGATA TCGTGAAGTGCAGAGAAATTAACGATTTTCAACCATGCTCACGACGCATACCTGA
 ATATCGTGTCCGCAACGTGTATAATACCAAGTTCCACAAACAA TCCCTTGGAACTTTATCAAGGAGAAAAGAGAT
 AATCCAAAGATTGCTGACACCTCAACTACTATAAGGTGTTCCGATATGACGTCAAAGGAAACAATATCACAGCA
 TGGGAGAAAGGGAATAATTAACAACGTGAAGACATGCTGAAGAGAAACACACCAATCTACACTAGGCAGGC
 AGCTGTAAAGAAAGGGAGCTGTTCAA TCAGACCAATTAAGAAAGGACTGGGCCAGCACCCCTGAAAGAAAG
 AAGGACCTTTTCCAAATATCTCTAAATACGGCGGGTATAACAAGGTGAGCGCTGCA TACTATACACTGATTTGAGT
 ATGAGGAAAAGGGCAACAAAATCCGCTCCCTGGAAACTATTCCTCCCTGTACCTGGTGAAGATA TCCAGAAAGGAT
 CAGGACCTCTGAAGTCTTATCTGACAGACCTGCTGGGGAAGAAAGAGTTCAAAGATCCTGGTCCCAAGATCAA
 GATCAAACAGCCTGTAAGATCAATGGGTTTCTTGGCCATATTACAGGAAAACATAACGATAGTTTCTCTGCTGCG
 CCTCGCTGCAGTTTGTCTTCAACAATGAGTCCGTACTTCAAGAAAATTTATCCGGTTTCCGAAATCCCGC
 TCTCAGCGAGAGAAATCGGGAACAAATTAAGCCCA TACGAGGACCTGAGCTTCCGGTCA TATATCAAGGAGAA
 CCTGTGGAAGAAAACATAAGAACGATGAAATCGGAGAGAGGAAATTTACGACCTGCTGCAGAGAAAAAACCTG
 GAGATCTATGATATGCTGCTGACTAAGCACAAGACACCA TCTACAAAGAACGCCCTAATTTCTGCCACCAATTGAT
 ATCCTGGTGAAGGGAAAGAGAAATTCAAAAGCCTGATTATCGAAAACCCAGTTTGAAGTATCCTGGAGATCCT
 GAAAGCTGTTTCTGCAACACCGAATGTCAGTGACCTGAGCATATCGGAGGCAAGCAAGTACTCCGCGCTGGCCA
 AAA TCGGGAACAAGATCTCTAGTCTGGATA ACTGTATCCTGTACTATGAGTCCATCACCCGGCATCTTCGAGAAAC
 GGATCGACCTGCTGAAGGTGAAGAAATTC

FIG. 13K

<p>Longitud de proteína (sin NLS y marcador HA-afiliado)</p>	<p>1395</p>
<p>PAM</p>	<p>NYAAAT</p>
<p>Longitud del epítopo</p>	<p>30</p>
<p>Especie: Milifactor: alocis</p>	<p></p>
<p>Secuencia de código de barras</p>	<p>AAGACACT</p>
<p>ID</p>	<p>5</p>
<p>ARN guía primario</p>	<p>NNNNNNNNNNNNNGUUUGAGAGUAGGAAACUACA CGGUUCAAAUAAAGAAUUUUUUCUAAUCGGC CCAUUGGGCCCAUAUUGAUUGGAUGAAMCUUGCUUAGCGAGUUUUUUU MTKEYYLGLDVGTSVQWAVTDSQYNLCKPKKKDMWGIRLFESANTAKDRRLQRGNRRRLERKKQRIDLLOEIFSP EICKIDPFFIRLNE SRLHLEDKSNDFKYPLEFKDYSDIEYYKEFTPIFHLRKLHIESEKQDIRLJYLALHNIKTRQHFLI DGDLSAKQLPILDITFLSLQEEQNLSVLSSENOKDEYEEILKNRSIAKSEKVKKLNLFELSELEKEEKKAQSAVIE NFCKFVGNKGDVCKFLRVSKEELEIDSFSEGGKYEDDIVKNEEKVPEKVVYLFEQMKAMYDWNILVDILETEEYISF AKVKQYBKHKTNLRLRDILKYCTKDEYNRMFNDEKAGSYTAYVGLKKNKKYWIEKRNPEEYKSLGKLLD KIEPLKEDLEVLTMMIEECKNHTLLPIQKNKDNQVIPHQVHEVELKILENAKKVYSFLTETDKDGYSVVQKIESIFRFR IPYYVGPLSTRHQEKGSNVWVVRKPGREDRTPWNNMEEIDFEKSNENFTRMTNKCTYLIGEDVLPKHSLLYSKYMY LMELNNVKVRGKLPISLKQKVFEDLFENKSKVTGKNLLEYLQIQDKDIQDDLSGFDKDFKTSLSKSYLDFKKQIFGEE IEKESIQNMIEDIHKWITTYGNDKEMLRVIRANYSNQLTEEQMKITGFQYSGWGNFSKMFLLKGISGSDVSTGETFDIIT AMWETDNNLMQILSKKFTFMDNVEDFNKGVGKIDKITDYDSTVEMFLSPENKRAVWQTIQVABEIKKVMGCEPKKI FIEMARGGEKVKKRTKSRKAQLELYAAACEEDRELKFEIDRDERDFNSMKLFLYTFQFGKCMYSGDDIDINELIRGN SKWDRHIYQSKIKDDSIDNLYVNKTYNAKKSNEILLSEDIQKMHFWLSLLNKKLITKSKYDRLTRKGDFTDEEL SGFIARQLVETRQSTKAIAIDFKQIYSEVVYVKSLLVSDPRKPLNYLKSRRVNDYHIAKDAYLNVVGNYYNKKKFTS NPIQWMMKNRDTNYSLNKVFVHVDVINGEVIWEKCTYHEDTNTYDGGTLDRIRKIVERDNILYTEAYACEKGELEFNA TIQNKGNSTVSLKGLDVKKYGGYFSANTSYFSLIEFDKKGDRARHIIGVPIYIANMLHSPSAFLEYCEQKGYQNV RILVEKIKKNSLLIINGYPLRIRGENEVDTSFKRAIQKLQKQNYELVRNIEKFLKYEKKGNYPIDENRDHITHEKMN QLVEVLLSKMKKFNKGMADPSDRIEKSFKFKLELDLIDKINVINKMLNLLRCDNDTKADLSLIELPKNAGSFVYVK NTIGKSKILVNQSVTGLYENRREL</p>

FIG. 13L

<p>Secuencia de aminoácidos de la proteína (Alfa-HA- NLS-Cas9- NLS-laa- EcoRI)</p>	<p>A C C G G T G C C A C C A T G T A C C C A T A C G A T G T T C C A G A T T A C G C T T C C C G A A G A A A A G C G C M G G T C G A A G C G T C C A T G A C C A A G G A G T A T T A C C T G G G C T G G A T G T G G G A C C A A T T C C G T G G A T G G C A G T G G C A G T G A C C G A T T C T C A G T A C A A C C T G T G C A A G T T A A G A A A A G G A T A T G T G G G C A T C C G C T T C G A A A G C C C A C A C A C G C C A A G G A C C G G A G A C T G C A G A G A G G A A T A G G C C G A T G G A G C G G A A A A G C A G A G A A T T G A T C T G C T G C A G G A A A T C T T C T C C C A G A G A T C T G C A A G A T T G A O C C C A C T T T C T T A T C C G A C T G A C G A A T C C C G G C T G C A D C T G G A G G A C A A G T C T A A C C A T T T C A A T A C C C A C T T T A T T G A A A G G A C T A T T C T G A T A C G A G A C T A T A A A G A G T T C C C C A C C A T T T T C A C T G A G G A A G C A T C T G A T C G A G A G T G A G G A A A C A G A T A T C C G G C T G A T A C C T G G C C C T G C A C A A C A T C A T A A G A C C G A G G A C A T T T T C T G A T T G A C G G G A T C T G C A G C C C A A G C A G C T G A G C C C A T C C T G G A T A C A T T C C T G C T G C C C T G C A G G A G G A A C A G A A C C T G T C A G T G A G C C T G T C C G A A A T C A G A A G G A C G A G T A T G A G G A A A T T C T G A A A A C C C A G C A T G C C A A G T C C G A A A A G T G A A A A A G T G A A A A A G C T G A A G A A T C T G T T T G A G A T C T C A G A C G A G C T G G A A A A A G A G G A G A A G A G G C C A G A A G C C C A G C C C T G A T C G A G A A C T T C T G C A A G T T T A T C G T G G G A A A T A A G G G C G A T G T C T G T A A T T C C T G C G G G T G T C T A A G G A G G A A C T G G A G A T T C G A G A A C T T C T G C A A G T T T C C T G G A G G C C A A G T A C G A G G A C C A C A T C G T G A A A A C C T G G A G A A A A G T G C C T G A A A A G T G C C T A C C T G T T T G A C C A G A T G A A G G C A A T G T A T G A T T G G A A T A T T C T G T G C A T C C T G G A A C C C T G G A A C C C G A G G A A T A C A T C A G C T T C G C C A A A G T G A A C C A G T A T G A A A C A C A A G A C T A A C C T G C G G C T G C T G A G A C A T C A T T C T G A A A T A C T G C A C C A A G G A T G A G T A T A A T C G G A T G T T A C C G A C G A A G A A G C T G G C A C A C C G C A T A C C C G A T T C T A C A A T C C C T G G C A A G C T G C T G G A T A A A A C A A G A A G T A C T G G A T C C G A A A A G A G A A T C C C G A G G A G T T C T A C A A T C C C T G G C A A G C T G C T G G A T A A A A T T G A C C T C T G A A G G A G A C C T G G A A G T G C T G A C T A T G A T G A T C G A G G A T G A A G A A C C A C C C C T G C T G C C A A T T C A G A A A A T A A G G A C A A C G G C G T G A T C C C C A C C A G G T G C A T G G T A C C C A C C A G G A A A T C G A G A A A T G C C A A A A G T A C T A T T C C T C C T G A C C G A C A G A C A A G G A T G G G T A C T C A G T G G T C C A G A A A T C G A G A G C A T T T C A G G T T C C C A T C C C C T A C T A T G T G G G C C T G A G T A C C G G C A C C A G G A A A G G G A T C A A A C O G T G T G G A T G G T C A G A A A C C T G G C A G G G A G G A T C G C A T C A C C A T G G A A T A T G G A G A A T C A T T G A C T T T G A G A A G A G C A A C G A A A T T T C A T T A C A C G G A T G A C T A A C A A T G T A C A T A T C T G A T C G G G A A G A T G T C C T G C C C A A G C A T T C T C T G C T G A C A G T A A T A T A T G T G T G C T G A A T G A G C T G A C A A T G T G A A G G T C A G A G G A A A A A G C T G C C T A C A T C T C T G A A C A G A A G G T T C G A G A C C T G T T T G A A A C A A T C C A A A G T G A C T G A A A G A A T C T G C T G G A G T A C C T G C A G A T C C A G G A C A A G A T A T C C A G A T T G A C G A T C T G T C T G C C T C G A C A A G G A C T T C A A G A C C A G C C T G A A G A G C T A T C T G G A C T T C A A A A G C A G A T T T T G G G G A G G A A T C G A G A A G G A A G C A T T C A G A A C A T G A T C G A A G A T A T C A T T A A G T G G A T C A C C A T C T A O G G C A A T G A C A A G G A G A T G C T G A A A C G A G T G A T T C G G G C T A A T T A T A G C A A C C A G C T G A C A G A G G A A C A G A T G A A A A G A T C A C T G G A T T T C A G T A C A G T G G C T G G G G A A C T T C T C A A A G A T G T T T C T G A A G G G A T C A G C G G A T C C G A C G T G A G C A C C G C G A A A C A T T C G A C A T C A T T A C C G C A A T G T G G G A G A C A G A C A C A A T C T G A T C A G A T C C T G T C A A A A A G T T C A C C T T T A T G G A C A C G T G A G G A C T T C A A C A G C G G C A A G G T C G G G A A A T C G A C A A G A T T A C C A T A G C A C C G T G A A G G A A A T G T C C T G C C C T G A G A A C A A A G G G C C T C T G C A G A C C A T T C A G G T G G C T G A G G A T C A A G A A G T G A T G G G C T G G G C C A A A A A A G A T C T T T A T T G A A A T G C A C G G C G G G A A A G G T G A A A A A G G G A C A A A T C C G C A A G C C C A G C C T G C T G G A G C T G T A C G C C C T T G C C A O G G A A G A T T G T A G A G A A C T G A T C A A A G G A G T T G A G G A C C G G G A C G A G A G G A C T T C A A T A G C A</p>
--	--

FIG. 13M

	<p>TGAAGCTGTTCTGTACTATACCCAGTTCGGGAAATGTAATGTAATCCGGCGGACGACATCGGATATTAACGAGCTGA TTCGGCGCAATTCCTAAGTGGACCGGAGATCACATCTACCCCCAGGCAAAATTAAGGACGATTCCTCCGATAAACC TGGTCTGGTCAAATAAGACATATAATGCCAAAGTCCCATGAGCTGCTGAGGACATCCAGAAAGAATG CATTCATTCGGCTGAGCCTGCTGAACAATAAGTGTGATCACTAAAGCAAAGTACGACCCCTGACTCGAAMAGGG CGACTTACCCGATGAGGAACCTGAGTGGGTTTCATCCCTAGACAGCTGTTGGAACAAGGCAAGTCAAAGGCAA TCGCCGATATCTTCAAGCAGATCTACAGCTCCGAGGTGGTCTATGTGAAGAGCAGCCCTGCTGAGCGGACTTCAGGA AAAAGCCACTGAACCTACCTGAAAGTCTCCGGAGGTCATGATTAACCCATGCAAAAGACCCCTACTGAAACATT GTGGTGGGAAAGCTGTACAAACAATAAGTTTACCAGTAATCCCATCCAGTGGATGAAAGAATCCGGAACAAAA CTATAGCCCTGAACAAGGTGTTGAAACACGACGTGGTCAATTAACGGGAAAGTGTCTGGGAAAGTGCACATACC ATGAGGACACTAATACCTATGATGGAGGCCACTCTGGACCGAATCCGCAAGATTGGGAAAGCCGATAAACATTCCTG TACACCCGAGTACGCTTATTGTGAGAAAGGGGACTGTTTAAATGCACCATCCAGAACAAAATGGAAACTCCAC AGTCTCTGAAAGGCTGGACGTGAAAGGTAACGGGGATACITTCAGCGCCACACAAAGTTACTTCTCACT GATCGAATTGAGGACAAGAGGGGGATAGAGCAAGGCCACATCAITGGAGTGCCTATCTATATTGCAAAACATGC TGGAGCAITCTCCAAAGTCCCTCTCTGGAGTACTGGACACAGAGGGGTATCAGAAATGTTGGGATTCCTGGTCCGAGA AAATCAAAAAGAACAGCTGCTGATCAATTAATGGATACCCCTCTGGCAATTCGAGCGGAGAACGAAAGTGGATACT TCCTTTAAGAGGGCCATCCAGCTGAAGCTGAAAGCCAGAAAACATGAGCTGTCCGCAATATCGGAAAGTTCCTG GAAAATACGTGGAGAAAAGGGAACCTATCCAAATGACGAGAAATAGAGATCAATCACACATGAAAAGATGA ACCAGCTGTACGAGGTGCTGCTCCAAAATGAAAAGTTCAACAAAGAGGGCAATGGCCGACCCCTCTGATAGG ATCGAAAAGAGTAAGCCTAATTCATCAAGCTGGAGACCTGATCGATAAGATTAAATGTGATCAACAAAATGCT GAACCTGCTGGCTGTGACAAATGATACTAAGGCCGACCTGTCCTGATGAGCTCCCAAAAACCGCTGGGAGTTT CGTGGTCAAAAAGAAATACCCATCGGAAAGTCMAAAATCATCCTGGTGAATCAGAGGGGTACTGGACTGTACGGAGA ATAGACGGGAACCTGTAAAGATTC</p>
<p>Longitud de proteína (de tipo salivaje sin N.S. #residuo)</p>	<p>1365</p>
<p>FAM</p>	<p>NNNAAGC</p>
<p>Longitud del espesor</p>	<p>30-31</p>
<p>Especie: <i>Staphylococcus pseudintermedius</i></p>	
<p>Secuencia:</p>	<p>TCTCGGTC</p>

FIG. 13N

<p>sólo de barras</p>	<p>6</p>
<p>ID</p> <p>ARN guía química</p>	<p>NNNNNNNNNNNNNNNGUUUAGCACUAGAAAUGUUAAAGUUAAAACAAGCUUAAAAGCGGUCAAUGU AAAUUUUAUUAACAACCCUACUGUGUCAGUGGGUUUUUUU</p>
<p>Secuencia de proteína Cas3 (de tipo salvaje sin NLS añadida)</p>	<p>MGRKPYILSLDIGTGSVGYACMDKGFNVLKYHDRKDALGVYVDFDGAALTAQERRQFRTRRRRKNRRKRLLQELLAP LVQNPNFYQFORQFAWKNDNMDFKNKSLSEVLSFLGYESKYPITYHLQEALLKDEKDFPELJYMALYHLVKYRGGH FLFDHLKIENLTNNDNMHDFVELJETYENLNLIKLNLDYEKTKVYIEILKDNEMTKNDRAKRVKNMEKKLEQFSIMLL GLKFNKGLFNHADNAEBELKGANQSHTFADNYEENLTPELTVEQSEFIERANKIYLSLTLQDILKGGKSMAMSKVAAY DKFRNELKQVKDIVYKADSTRQPKKIFVSSKSKLKQYDAPNDQTFSSLCDFQYLIRPKQYSLLIKELKKIIPQDSE LYFAENDTLKVLNTTDNASIPMQINLYEAEITLNRQKQYHAETIDEMIEKVSLSIQFRIPYVVGPLVNDHTASKFGW MERKSNESIKPWNFDEVVDRSKSATQFIRRMTNKCSYLINEDVLPKNSLLYQEMEVLNELNATQIRLQTDPKNRKYRM MPQIKLFAVEHIFKKYKTVSHSKELEIMLNSNHRENFMNHGKLSIFGTQDDKFKASKLSSYQDMTKIFGDIEGKRAQI EEIIQWITIFEDKKILVQKLKECYPELTSKQINLKKLNYSYWGRLSEKLLTHAYQGHSHIELLRHSDENFMELTNDVY GFQNIKEENQVQSNKIQHDIANLITSPALKKGIWSTIKLVRELTIFGEPEKIIMEFATEDOQKQKQKSRKQLWDD NIKKNLKSYDEYKYHIDVANKLNNEQLQKEKLWLYSQNGKCMYSGQSIDLDAALSPNATKHVEVDHIFRPSFIKDD SIDNKVLVIKMNQTKGDQVPLQFIQPYERIA YWKSLNKAAGLSDSKLHKLMPKPEFTAMDKEGFIQRQLVETRQISV HVRDFLKEEYPNTKVIPIKAKMVSEFRKFDIPKIRQMNDAHHAIDA YLNGVVYHGAQLAYPNVDLDFDFNEKWEKV REKWKALGEPNTRKQKSRLEFFFKLEKMEVSOGERLISKIKLDMNHFKNYSRKLANIPOQFYNQTAVSPKTAELKYE SNKSNEVYKGLTPYQTYVVAIKSVNKKGKKEKMEYQIMIDHYVPDFYKFNQNGEKELALYLAQRENKDEVLDAQIVY SLNKGDLLYNNHPCYFVSRKEVINAKQFELTVEQQLSL YNVMMNKETNVEKLLIBYDFIAEKVINEYHHYLNLSKLE KRVRTFFSESNTIHEDFIKALDELFPKVVYASATRSRDKIGSRKNMSMTHRAFLOGKGDVKIAYTISISGLKTTKPKSLFKLA ESRNEL</p>
<p>Secuencias optimizadas en condiciones de mamíferos (AgeI-HIA- NLS-Cas9- NLS-taa- EcoRI)</p>	<p>ACCGGTGCCACCATGTACCCATAGGATGTTCCAGATTAAGCTTCGGCGGAGAAAAGCGCAAGGTGGAAGCGTGC CATGGGGAGGAAACCTTACATCTGTCTCTGGATATTGGAACCTGGTCCGTCCGGTACCGCTTGCATGGATAAAGG ATTC AACGTGTGAAGTACCACGACAAAGATGCCCTGGAGTGTATCTGTTCGACCGGCTCTGACTGCACACAGGA GGGAGACAGTTTATAGGACCTCCAGGGCCGGAAGAACCAGGAGAAATCAAAGCCCTGGCCCTGCTGCAGGAACCTGC TGGCACCCCTGCTGCAGAACCTAATTTCTACCCAGTTTCAGCGGCAGTTCCGCTGGAGAACGACAAATATGGATT TTAAGAACAAAGCCCTGCTGAGGTGCTGAGCTTCTGGGATATGAA TCCAAAGAAATACCCATACCCTACCCACC TGCAGGAGCCTGCTGCTGCTGAAAGACGAGAAAGTTTGATCCAGAACTGATCTACATGGCACTGATATCATCTGGTGA AATACAGAGGCCACTTCTGCTGATCATCTGAAATGCGAAACCTGACTAACAATGACAAATATGCACGATTTCCG TGGAGCTGATTTGAACCTATGAAACCTGAAACAATATCAAGCTGAACTCGGACATCGAGAAACCAAGTGTATC TATGAGTTCTGAAGACACAGAAATGACTAAGATGATAGAGCCAAAGGTCACAGAACATGGGAGAAAGAAC TGGACAGTTCTGATCATGCTGCTGGGCTGAAGTTCAATGAGGGAAAACCTGTTTAAACCCAGCCGATAATGCTG AGGAACTGAGGGGGCTAACCCAGAGCCATACATTTGACAGCAACTACGAGGAAAATCTGACTCCCTTCTGACTGACC</p>

FIG. 130

GTGGAACAGTCAGAGTTTATTGAAAGGGCCCAACAAAATCTATCTGAGCCTGACTCTGCAGGATATCCTGAAGGG
CAAGAAATCAATGGCTATGAGCAAAAGTGGCCGCTTACGACAAGTTCAGAAATGAGCTGAAACAGGTGAAGGACA
TTGTCTATAAGGCTGATTTCTACCGACACAGTTTCAAGAAATCTTTGTGAGCTCCAGAAAGTCTGAAGCAGT
ACGACGCAACTCCCAACGATCAGACCTTCTAGTCTGTGCCTGTTGACCACTCTGATTTCCGCCAAAGAAAC
AGTATAGCCTGCTGATCAAGGAGCTGAAGAAATCATTTCCCAAGACTCCGAACTGTACTTTGAGCGCAGAAAAT
GATACCTGTGTAAGGTGTAACACACAGCAATGCTAGCATCCCTATGCAGATTAACCTGTACGAGGCGAGA
AACCATCTGCGAAAATCAGCAGAAATATCACCGCCGAGATCACAGATGAGATGATTTGAAAAGGTGCTGTCTCTGA
TCCAGTTCGGCATTCOCATACTATGTGGGCCCCCTGGTCAACGACCATACAGCCAGTAAAGTTTGGATGGATGGAGC
GCAAAGTAAACGAAATCAAATCAAGCCTTGGAAATTCGACGAGGTGGTGGATCGAAAGTAAATCAGCCACTCAAGTTT
ATTAGGGCATGACCAACAAGTGTCTTACCTGATCAATGAGGATGTCGTGCCAAAANAACCTCTGTCTGTATCAG
GAGATGGAAGTCTGACGAACTGATGCCACACAGATCAGGCTGCAGACTGACCCCAANAACCGCAAGTACCG
AATGATGCCCCAGATTAAGCTGTTCGCTGTGGAGCACATCTTTAAGAAATATAAACCCGTACGCCATTCCAAAGTT
CCTGAAATTTATGCTGAACAGCAATCACAGGGGAACTTTATGAAATCATGGAGAAAAGCTGAGTATCTTCGGCA
CACAGGCGATAAAGAAATTTGCATCAAAGCTGTCAAAGTACCAGGACATGACTAAAATCTTCGGGGATAATTGAG
GGAAAGCGGCCAGATTTGAGGAAATCATTCAGTGGATCACCAATTTTGGAGCAAGAAAATCCTGGTGCAGAA
GCTGAAAGAGTGTCTATCCTGAACTGACATCAAGCAGATCAACAGCTGAAGAAAATGAAATTAATCTGGCTGGG
GGAGGCTGAGTGAGAAAGCTGCTGACTCACGCCATCAGGGCCATAGCATCAATGAACTGCTGGCCACTCCGATG
AGAAATTCATGGAAATCTGACCAACGACGCTGACGGTTCAGAAATTTTATCAAGAGAAACACAGGTCCAG
AGCAATAAGATCCAGCATCAGGATATTGCCAACCTGACTACCTCTCCCGCTGAAAGAAAGGCACTGGAGTACA
ATTAAGCTGGTGGGGAGCTGACTTCCAATTTTCGGGGAGCCTGAAAAGATCAATGAGGTTTGTACCCGAGGAC
CAGCAAAAGGCAAGAAACAGAAATCAAGAAAGCAGCTGTGGACGATAACATCAAGAAAATAAGCTGAAAA
GCGTGGACGAGTACAAATATATCATTTGATGTGCGCCAATAAGCTGAACAATGAGCAGCTGCAGCAGGAAAACTG
TGGCTGTAACCTGAGCCAGAACGGCAAGTGTATGATAGCGGCGAGTCCATCGACCTGGATGCCCTGCTGCCCC
AATGCTACCAAGCACTACGAGGTGGATCATATTTCCCTCGGAGCTTCAAGGACGATAGCATTTGACACAACAG
GTGCTGGTCAACAAGAAAATGAAATCAGACAAGGGCGATCAGGTGCCCTGCAAGTTCAATTCAGCAGCCTTACGA
GAGATCGCATAITGGAAAGAGCCTGAAACAAAGCCGGGCTGATCTGTGATAGTAAACTGCACAAGCTGATGAAAC
CAGATTCACACAGTATGGACAAAGGAGGCTTCAACAGGCGCAGCTGGTGGAGACTAGACAGATCAGCGTGCAT
GTCCGGGATTTTCTGAAAGAGGAAATACCCTAATACCAAGTATCCCAATGAAGGCCAATAATGGTGAAGGAGTT
CCGGAAGAAATTTGACATCCCAAGATTAGACAGATGAACGACGACCAACCAATGCCATCGATGCTTACCTGAATG
CGGTGCTATACAGGGGACACAGCTGGCCATCCCAACGCTGGAGCTGTTTCAATTTAAGTGGGAGAAAG
TCCGAGAAAAGTGGAAAGCCCTGGGAGATTCAAGCAAAAGCAGAAATCTCCGGGAACCTGTTCTTTTCAAGAA
CTGGAGAAAGTGGAAAGTCCAGGGGAGCGGCTGATCTCTAAGATCAAGCTGAGCATGAACCACTTCAAGAT
CAACTACTCCAGAAAGCTGGCCAAACATCCCTCAGCAGTITTAATATCAGACGCGAGTGTCTCCAAAAGACGCCGA
GCTGAAATACGAAATCAAGAGTAAATGAGGTGGTCTATAAGGGACTGACACCAATACCAGACTTATGTGGTCCG
CCATCAAAGCGTGAACAAAGGCAAGGAGAAAATGGAATACCAGATGATCGACCACTACGTTCCGATTTT
TATAATTCAGAAACGGCAATGAGAAAGGAACTGGCTCTGTACTGGCACAGAGGGGAGACAAGGACGAGTCT

FIG. 13P

	GGATGCTCAGATTGTCTATAGTCTGAAATAAGGGGGAATCGCTGTACATCAA CAATCA TCCTGCTATTTTCGTGTCACGCA AAGAGGTCA TCAACGCA AAGCAGTTTGAGCTGACCGTGGAA CACGACGCTGTCTCTGTACAA CCGTGA TGAA CAACAAGGAGACA AATGCGAAAAGCTGCTGATCGAGTATGACTTCATTCGCGAGAAAGTGATCA CCGAAATACC ACCATTATCTGAA TAGCAAGCTGAAAGAAAGCGAGTCCGGACCTTTTCTCAGAGAGCAACCCAGACACACGAG GACTTCA TCAAGGCCCTGGACGAGCTGTTAAGGTGCTCACCGGCATCCGCCACA AAGTCTGATAAAA TCGGGAGT CGCAAGAACAGCATGACTCATCGAGCCTTCCTGGGAAAAGGCAAGGACGTGAAGATTGCTTACACCTCCATCTCT GGA CTGAAAACA AACTAAACCTAAGAGTCTGTTAAGCTGGCGGAGTCAAGAAA CCGAACTGTTAAGAATTC
Longitud de proteína (sin NLS y marcos de HA añadido)	1334
PAM	NGG
Longitud del espaciador	29
Especie: <i>Lactobacillus johnsonii</i>	
Secuencia de código de barras	AATGTTCT
ID	7
ARN guía añadido	NNNNNNNNNNNNNNNNNNNGUUUUA GAUGGUGAAAACCA GAUUUAAAUAUCAAGCAAUGCAUCUUUUGAU GCAAAGUUUCAAUUUUGUCCCA CGUUUAUCGAGGGACUUUUUUUU
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadido)	MTKIKDDYIVGLDIGTDSCGWVAMNSNDIILKIQKTAIGSRLFEGGKSAERLLFRTHRRIKRRRWRLKLEEFFD PYMAEVDPPYFARLKESGLSPLDKRKTVSSVFP TSAEDKKFYDDYPTIYHLRYKLMTEDEKFDLREVYLAIHIIKRYR GNFLYNTSVKDFKASKIDVKSIEKLNELYENGLDLNVFNISNTABIEKVLKDKQIFKRDKYKKIAELFAIKTDNKEQ SKRIKDISKQYANA VLYKTRFDTIALKEISKDELSDWNFKLSDIDADSKFEALMGNLDENEAAILLTIKELFNEVTLNG IVEDGNTLSESMINKYNDHRDDLKLLKEVIENHIDRKKAKELALAYDL YVNNRHGQLLQAKKKLGGIKPRSKEDFYK VVKNLDDSRASKEIKKKIELDSFMPKQRTNANGVIPYQLQOLEDKIENQSKYYPFKKEINPVSSHLEAPYKLDLRI RFRVPYYVGPLISPNETKDIQTKKNQNFAMIRKBEGRITPWNFDOKVDRIESANKPIKRMTTKDTYLFGEDVLPANS LLYQKFTVLNELNNIRINGKRISVDLKQEIYENLFKKHTTVVKKLBNYLKENHNLVKVEIKGLADEKKFNSGLTYYN RFKNLNFQNDLKYRNDFEKIEWSTIFEDKSIYKEKLRISDWLNEKQNALSNIRLQGWGRLSKLLAQLHDHNGQ THIEQLWDSQNNFMQIVTQADFKDAIKANQNLLVATSVEDILNNA YTSPANKKAIROVIKVVDDIVKAASGKVPKQI

FIG. 13Q

<p>AIEFTRADENPKRSQTRGSKLQKVVYKDLSTELASKTIAEELNEAKDKKLVDQKYLLYFMQLORDAYTGEPIIDEIQ KYDIDHILPQSFIKDDALDNRVLVSRAVNINGSDNVVVKLFGNEMAANLGMTIRKMWEEWKNIGLISKTKYNNLLTD PDHINKYKSAGFIRRLQVETSQIKLVESTILQSRYPNTEIITVAKYNHYLREKFDLYKSREVNNDYHHAIDAYLSAICGNL LYQNYPNLRPFVYQYKFFSSDFDKKALFNKTRKFSFISOLLKNKSENSKEIAKKIKRA YQFKYMLVSRFETETRDQE MFKMTVYPRFSDHTVKA PRNLIPKMKMGMSPIYGGYTNNSDAYMVIVRIDKKGTEYKILGIPTREL VNLKKA EKED HYKSYLKEILTPRILYNKNGRDKKITSFEIVKSKIPYKQVIQDGDKFKPMLGSSTYYVNAKQLTSLTESMKAITNNFDK DSDENDALIKAYDEILDVKDKVLPFLDINKFREKLSHGREKFIKLSLEDKKDTILKVIIEGLHDNA VMTKIPTIGLSTPLG FMQFPNGVILSENAKLIYQSPTGLPKKSVKISDL</p>	<p>ACCGGTGCCACCA TGTACCCATACGATGTTCCAGATTAGGCTTGGCCGAAGAA AAGGCCAAGGTCGAAAGGTC CATGACAAA AATCAAAGACGACTACATCGTGGGACTGGACATCGGCACAGACTCCTCGGGTGGGTGGCTATGA ACAGCAATAATGACATCTGAACTGCAAGGCAAGACCGCAATCGGGTACGCCCTGTTCCGAGGGAGGGAAGAGC GCAGCTGAACGGAGACTGTTTCCACACACACAGGGCA TCAAACGACGGAGATGGGACTGAAGCTGCTGGA GGAGTCTTCCGACCCCTACATGGCAGAGGTGGATCTTCTTCCCGGCTGAAGGAATCTGGCCTGAGTCC ACTGGACAAAAGAAGACCGTGAAGCTCCATTGTGTTCCCCATCCGCGGAGGATAAAGATTTACCGAGATT ACCCTACAATCTACCATCTGAGGTATAACTGATGACTGAGGACGAAAGTTGATCTGGCCGGAAGTGTACCTGG CTATCCACCATATCAITTAAGTACCGAGGAACTTCTGTATATAACCAAGTGTGAAAGACTTCAAAGGCATCAAAGA TCGATGTCAAATCTAGTATCGAAGAAAGTGAACGAGCTGATGAAAATCTGGCCCTGGACCTGAAACGTTCA ACATTAGCAATACTGCCAGATCGAAAGGTTGCTGAAAGACAAGCAGATCTTCAAGCGGATAAAGTCAAAGAA ATTGCCGAGCTGTTTGCTATCAAACCGACAACAAGGAA CAGAGCAAGAGATCAAAGATATTTCCAAACAGGT GGCCAAATGCTGCTGGGTACAAGACCGAGTTCCGACACAACTCGTCTGAAAGAGATTTCCAAAGGCAACTGT CTGATTGGAACCTCAAACCTGTCAGACATCGATGCAGACAGCAAGTTTGAAGCCCTGATGGGAAACCTGGATGAG AATGAACAGGCCATCTGCTGACTATT AAGGAGCTGTTTAAACGAAAGTGACCCCTGAATGGAA TTGTGAGGACGG CAACACCCCTGAGCGAATCCAATGATCAAAGTACAA TGAATGATCACCCGGACGATCTGAAAGCTGCTGAAAGAAAGTGA TCGAAAATCAATTTGACAGAAAGAAAGCCAAAGGACTGGCCTGCGACTGGCTACGATCTGTATGTCAACAATAGGCAC GGACAGCTGCTCCAGGCTAAGAAAAGCTGGGCAAAATCAAGCCCCGCTCTAAGGAGACTTCTACAAAAGTGT CAACAA GAATCTGGACGATTCACGGGCAAGCAAGGAGATCAAAAAGAAATTTGAACTGGACAGCTTTATGCTTA AGCAGAGAA CCAACGCCAATGGGTGATCCCATACCACTGCAGCAGCTGGAGCTGGATAAGA TCAATCGAAAAC CAGTCTAAGTACTATCCAATCCCTGAAGGAGATTATCCCGTGTCAAGCCACCTGAAAGAGCCCCCTATAAGCTG GACGAACTGATCCGATTTCCGGTGCCTTACTATGTGCGGCCCTGATTTCTCTTAACCGAGAGTACCAAGGATATC CAGACAAA GAAAACCAGAAATTTGGCCTGGATGATTCGCAAAAGGAAAGGGGCAATCACACCTTGGAACTTTGA CCAGAAGGTGGATCGAATGAGAGGGCCAAATAAGTTCA TCAAACGGATGACTACCAAGGACTTACCTGTTTGG GGGAGATGTGCTGCCAGCTAACAGCCTGCTGATCAGAGTTCCACCGCTGAAACGAACTGAAACACATCCCGG ATTAATGGAAAAGAAATCTCCGTGGACCTGAAAGCAGGAGATCTACGAAACCTGTGTTAAGAAAACACACAACCTGT GACCGTCAAAGAACTGGAGAATTA TCTGAAGGAAAACCAATAATCTGTTGAAAGTCGAGATCAAAGGGGCTGGCCG ATGAAAAGAAATTCACAGCGGACTGACCCACATACAAIAGATTCAAAGAACCTGAAACA TCTTTGACAAACCAAGATT</p>
	<p>Sequence: optimizacja 5' sekwencja 3' region 5'-3'</p> <p>(AgeI-HA- NLS-Cas9- NLS-tar- EcoRI)</p>

FIG. 13R

<p>GACGATCTGAAAGTACAGGAACGATTCGAGAAAGATCATCGAATGGTCTACAAATTTTGAGGACAAGAGTATCTAC AAAGAAAGCTGAGGAGCATCGATTGGCTGACGAGAACGAGATTAAACGCTCTGTCTAATATCAAGACTGCACGGG GTGGGAAAGCTGAGTAAGAACTGCTGGCAGCAGCTGCACGACCAATAATGGCCAGACCAATCAATGAGCAGCTGT GGGATCCCAAGAACAAATTCATGACAGATTGTGACACAGCCGACCTTTAAAGATGCTATCCGAAAGGCCAAACCCAG AATCTGCTGGTGGCTACCTCAGTCGAGGACATCTGAAACAATGACATACACAACCCCGCAACAAAGAAAGCCAT CAGACAGGTCATCAAGTGGTGCAGGATATGTTGAAGGCGACCTCCGGAAGGTCCCAAAACAGATCCGCCATTG AGTTCACTAGGCAAGTGCACCAAAAATCCCAAGAGAAAGTCCAGGAGGCTCAAGCTGCAGAAAGTGTACAAG GACCTGAGCACTGAGCTGGCTCCAAAGCAATGCTGAGGAACTGAAAGCAATCAAAAGCAAGAAACTGGT GCAGGATAAGTACTACTGTACTTTATGCAAGCTGGGGGGACGCTATACAGGAGAGCCATAICAAATATCGATGA AATCCAGAAAGTACGATATCGACCACTTCTGCCACAGTCTTCATCAAGGACGATGCCCTGGACAACAGGGTGCT GGTGACCCGGCTGTGAACAATGGCAATCTGATAATGTGCTGTCGCAAGCTGTGGCAACGAGATGGCTGCA ATCTGGGATGACTATCAGGAAATGTGGGAGGATGGAAACAATCGGCCCTGATTAGCAAAACAAGTACAAC AATCTGCTGACTGATCCCGACCACTAACAAAGTATAAGATGCCGGTTTCATCGCCCGCAGCTGGTGGAGAC ATCACAGATCAACAAGCIGGTGACACTATCCTGCAGAGTCGCTACCCCTAACACTGAAATCAATTAAGGTGAAGGC TAAGTACAATCAATATCTGCGGGAGAAATTTGACCTGTATAAGAGCAGAGAATCAACGACTACCACCTAT TGAATGCAATCTGTCGCCCACTGCGGAAATCTGCTGTACCAGAACTATCCAATCTGCGGCCCTCTTTTGTATAC GGCCAGTATAAGAAATTCCTCTGTATCCTGACAAAGAGAAAGGCCATTTTAAACAACCCGGCAAGTCTCCCTT ATCTCAGCTGCTGAAACAAGAGTGAAACAAGCAAGCAAGGAAATCGCTAAGAACTGAAGGGGCATACCAATT CAAGTATATGCTGGTCTCGAGAGACTGAACCCGGACCCAGGAGATGTTCAAAATGACCGGTGACCCCGGTT CAGCCACGATACAGTCAAGGCTCTAGAACCTGATTCCAAAGAATGGCCATGTCCCTGACAATCACGGAG GCTATACAAACAATCTGACGCTACATGGTCACTCGTCCGCAATGATAAGAAAGGAACTGAGTATAAGATC CTGGCATCCAAACCCGGAACTGGTGAATCTGAAAAAGGCCGGAAGAGGAGGCCATTAACAAGCTACTGAA GGAGATCTGACACCAAGGATCTGTACAAACAATAAGGAAAGCCGATAAAAGATCACTCCCTCGAAATTG TGAATCTAAGATCCCTATAAGCAGGTCATCCAGGATGGGACAAAAGTTTATGCTGGGAAGTTCACAACTAC GTGATACCGCAAGCAGCTGACTGACACTGAGCCTGATCCATGAAAGCCATCACTAACAATTCGATAAGGACAG CGATGAGAAACGACGCTCTGATTAAGCCATACGATGAAATCTCTGGACAAAGTGATAGTATCTGCCACTGTTCCG ACATCAACAAGTTCGGGGAAGCTGCACAGTGGGGAGAAAAGTTTCATCAAGGCTGAGCCCTGGAGGACAAAA GGAATACCATCTGAAAGTGTGGAAAGGACTGGCATGATAAGCTGTCAAGCAAGATCCCTACTATTGGCCCTGTC CACACCACTGGGTTCAAGCAGTTTCCCAACGGCGGTGATCTGAGCGGAGAAATGCCAAACTGATCTACCAAGTCCCC CACCGGCTGTTCAAAAGTCAAGTGAAGATCAGCGGACCTGTAAAGAAATTC</p>	<p>1375</p> <p>NTAAA</p>
<p>Longitud de proteína (sin NLS y marcador His añadido)</p>	<p>PAM</p>

FIG. 13T

EcoRI)	<p>AGACAGAGATAAGGAGCTGGGAGCACCTGTGAAAGGAGTGGCCGACAAACACTCATGCTCTGCAGACCCGGCGGAT TTTAGGACACCCGCTGAGCTGGCACTGAATAAGTTGAAAGAGAGTGGACACATTCGAAACCAACCCAGCGGGCCGA CTATTCACATACCTTCAACCGCAAGGATCTGCAGCCGAGCTGAATCTGCTGTTGAAAGACAGAAAGAGTTCGG GAATCCCCACGTTCCGACGGGCTGAAAGAGAAATCGAGACACTGCTGATGACTCAGAGCCCTGCACCTGTCTG GCGATGCCGTGCAGAAAGATGCTGGGCAATGCACCTTTGACCAACAGAGCCCAAGGCAACCAAAACACCTAC ACAGCCGAGAGTTCCGTGTGGCTGACAAACAACTGGGCACTCTGAAACAGGGCAGTGGAGCGGCGCCCT GACTGACACCGAAAGAGCCACACTGATGGATGAGCCTTACAGGAAGTCTAACTGACTTATGCCCAAGCCTGGCA AGCTGTGGACCTGGACGATACTGCCCTCTTAAAGGCCCTGAGGTACGGGAAAGATAATGCAGAACCCAGCACCC CTGATGGAGATGAAAGGCCTATCACGCTATCTCCCGCCCTGGAAAAAGGGCCCTGAAGGCAAGAAATCTCC CCTGAACCTGATCTGAACTGCAGGATGAGATTGGGACCGCTTTTAGCCTGTTCAAGACTGACGAGGATATCAC CGGACGCCCTGAAAGACCGAGTGCAGCCGAAATCTGGAGGCCACTGCTGAAGCACATCAGTTTGTATAAATTCGT GCAGATTTCACTGAAGGCCCTGCGAGGATCGTCCCTCTGTGGAGCAGGGCAATCGGTACGACGAGGCCCTGCA CCGAGATCTACGGAGATCAITTAGGCAAGAAACACAGAAAGAAATCTATCTGCTGCAAAATTCAGGGAGTACTTTCCAAACTTC GAGATCCGGAAATCCAGTGGTCTGAGAGCTGTCTCACAGGCAAGAAAGTATCAACGGAGTCAACGGAGTGGTCAGAAAGGTA CCGCAGCCCTGTAGGATCCACATTTGAAACCCGACCGAAAGTGGAAAGTCCTTTAAAGACCCGCAAGGAAATCG AGAAGGACAGGAAGAGATAGAAAGATAGGAAAGTCTGCTGCAAAATTCAGGGAGTACTTTCCAAACTTC GTGGGCAACCCAGAGTAAAGACATCTGAAAGCTGGCCCTGTACGAGCAGCAGCAGCGGAAAGTGTCTGTATAG CGGAAAGAAATTAACCTGGCCGCTGAATAAGGGCTATGTGGAGATCGATCAACGGAGTGGTCAGAAAGTCC GAACATGGGACGATCTTTTCAACAATAAGTCTGGCTGCGGAGCGGAAACCAAGAACAAAGGAAATCAGACT CCTTACGAATATTTCAACGGGAAAGGCAATAGCCGAGAAATGGCAGGAGTTTAAAGCCCGCTGGAGACAAGCCG GTTCCACGAAAGCAAGAAACAGCGGATTTCTGTGCAGAAAGTTGACGAAAGTGAATCAAGAGAAACCTGA ATGACACCCGTCATCAACAGATTTCTGTGCCAGTTCTGGCTGATCACAATGCTGTGACCCGAAAGGGCAAAAC GCCGAGTCTTTGCAAGCAACGGCCAGATCACAATCTGCTGAGGGCTTTGCGGGCTGGGAAAGGTGAGAGCC AGGTTTGTGGCTATAAGAGATGAAACGCCCTTCGACGGAAAGACAATTTGATAAAGAAACTGGCAGGTGCTGCA CCAGAGGCACATTTCTCAGCCATGGGAGTTCTCCGCCAGAAAGTCCGACACTGCTGGCTGAAAACACTGAGCTCCA CGGAAAACAGAGTTCCGAGAGGCCGATACCCCAAGAAAGTCCATTAAGCAGGGCCCTAATCGCAAAATGTCCGGGCGAG GACCCGAGGCAAGTACAGTACGTCAACCCCTGTTCAATTAAGCAGGGCCCTAATCGCAAAATGTCCGGGCGAG GGACATATGGAGACTGTGAAATCAGTAAAGCCCTTCGACGGAAAGGCAATCAGCGTGTGAGATGCCACTGACCCCA GCTGAAAGCTGAAAGATCTGGAGAAGTGGTAAACCGGAAAGAGAGCCCAAGCTGTATGAAGCTCTGAAAGCA AGACTGGAGGCCCAAGAGGACGATCCAGCTAAGCAATTTGCGGAAAGTCCGAGCCCTTCAACAATATGACAAAGCCCGGCAA TCGGACACAGCAGGTGAAAGGCTGTACAGTGGAGCAGGTCCAGAAACTGGGGTGGTGCACAAACCTAATG GAAITGGCCGACAAACCTACAATCGTCCGGTGGATGTTCTGAGAAAGCCGGGAAAGTACTATCTGGTGCCTATCT ACTCTGGCAGGTGCCCAAGGAAATCCTGCCAGATAGAGCTGTGTCAGGGCAAGCAGAGGATTTGGACT GTGATGGACGATCTTTTCGAGTTAAGTTCTGCTGTACGCAACCGACCTGTCAAGCTGACAGCCAAAGAAAT GAAITTTCTGGGATATTTCCGTCACTGAACAGGGCAACTGGAGCCATCGATATTCGCACACATGACACTGATAGC</p>
--------	---

FIG. 13U

	ACCAAGGGAAAACGGGCATCTTTCAGTCTGTGGGGTCAAGACCCCGCTGAGTTTCCAGAAATATCAGATTGAC GAAC TGGGGAGGAGATCCGACCOCTGTCCGCTGAA GAACGACCAACCCGTCGGGTAA GAATTC
Longitud de proteína (sin NLS y marcador HA añadido)	1082
PAM	NNNGTAA
Longitud del residuo	30
Especie: Parvibaculum lavamentivorans	
Secuencia de código de barras	CGCCTTCC
ID	20
ARN guía químico	NNNNNNNNNNNNNGCUGGGGGAUUGGGGGA AUUGCCUUUUGGCAAGCAA AUUGACCCCUUGGUC GGGCUUGGCAUCCCAAGGUCAGCUGCCGGUUUAUUCGAAAAGCCCAAGCAGCGGGGCGCCU UUUUUU
Secuencia de proteína (con NLS y marcador HA añadido)	MERIFGDIGTTSIGFSVIDYSSTQSAQNIQRLGVRIFPEARDPDGTPLNQRRQKRMRRQLRRRIRRKALNETLHEA GFLPAYGSADWPVMADEPYELRRRGL EEGLSA YEFGRAIYHLAQHRHFKGRELEESDTPDPDV DDEKEAANERAAAT LKALKNEQTTLGAWLARRPPSDRKRGIHAHRNVVAEEFERLWEVQSKFHPALKSEEMRA RISDTIFAQRVFWRKNT LGEGRFMPGEP LCPKGSWLSQRRMLEKLNLAAGGNARPLDAEERDAILSKLQQQASMSWPGVRSALKALYKQR GEPGAEKSLKFNLELGGESKLLGNAL EAKLADMFGPDWPAHPRXQEIRHA VHERLWADYGETPDKKRVIILSEKDR KAHREAA NSFVDFGITGEQA AOLQALKLPTGWEPYSIPALNLF LAELKGERFGALVNGPDWEGWRRRTNFFHRNQ PTGEILDKLPSPASKEERERISQLRNPTVVRTQNELRKVYNNLIGLYGKPDRIEVRDVGKSKRREBBIQSGIRRNEKQ RKKA TEDLIKNGIANPSRDDIVEK WILWKEGQERCPYTDGQIGFNALFREGRYEVEHIWPRSRFDNSPRNKTLCKRDY NIEKGNRMPEAFGHDEDRWSAQIRLQGMVSAKGGTGMSPGKVKRFLAKTMPEDFAARQLNDRYAAKQILAQLK RLWPDMGPEAPVKVEAVTGGQVTAQLRKLWTLNNILADDDGEKTRADHRHA IDALTVACHTPGMTNKLRSRYWQLRD DPRAEKPALTPPWDTIRADA EKA YSEIVVSHRVRKKVSGPLHKETTYGDTGDIKTKSGTVRQFVYTRKIKIESLSKGELD BIRDPRIKBIVAAHVAGRGGDPKKAFPPYPCVSPGGPEIRK VRLTSKQQLNLMAQTNGYADLGSNHHIAJYRLPDGKA DFEIVSLFDSRRLAQRNPVQRTRADGASFVMSLAAGEAIMPEGSKGIWIVQGYWASGQYVLEFRD TDADHSTTTR PMPNPILKDDAKKVSIDPIGRVRFPSND

FIG. 13V

<p>Secuencias derivadas de posone de mamíferos (AgtI-HA- NLS-Cas9- NLS-taa- EcoRI)</p>	<p>ACCGTGGCCACCATGATCCCATACGATGTTCCAGATTACGGTTGGCGAAGAAAAGCCACAGGTCGGAAGGGTC CATGAGAGGATTTTCGGCTTTCATCGCATCGGCACAACAAGTATCGGATTCAGCGTGAATTGATTACAOTAGCACCCA GTCCGACGGCAATCCATGAGCGTGGCTGGCAATTTCCCTGAGGCAAGGACCCAGATGGGACCCCCCTGA ACCACAGGGGAGACAGAAACGGATGATGAGGGCGCACTGCGCACGGAGAAGATTCGGCCGAAAGGCACCTGAA TGAGACACTGCACGAAGCCGGCTTCTGCCAGCTTACGGGTCTGACAGTTGGCCCGTGTCTAGCCCGACCGAGCC TTATGAACTGGGGAGAGGGACTGGAGAAAGGCTGAGTCTACGATTTCCGACGGGCAATCTATCATCTGG CCCAGCACCGGCATTTTAAAGCCAGAGAACTGGAGAACTCCGATACACCCGACCCCTGATGTGGACGATGAGAAG GAGCCGCTAACGAGAGAGCAGCCACTGTAAAGCCCTGAAAATGAAACAGACCACTGGGAGCATGGCTGG CCGCCGACCCCTTCTGACCCGCAAGCGAGGAATCCACCCCATAGBAACGTGCTCCGTGAGGATTCGAGCCG CTGTGGAAAGTGCAGTCCAAGTTTCAACCCCGCTGAAATCTGAGGAAATTCGGGCAAGAAATCAGTGAATACAATT TTCGCCAGAGCCCTGTGTTTGGCGCAAGAACAATCTGGGAGAGTGCAGATTCATGCCTGGCCAAACCACTGTGT CCCAAGGGTCCCTGGCTCTCAAGCAGCGGAGATGCTGGAGAACTGAAACAATCTGGCTATCGCAGGGGGAA TGCTAGGCCACTGATGCAGAGAAACCGGACGCCATTTCTGAGTAACTGCAGCAGCAGGCCACGATGCTCCTGGC CAGGGTGGGTCACTCTGAAGGCATCTGTAACAACAGAGAGGCGAGCCGGGCTGAAAAGAGCCCTGAAATTC AACTGGAGCTGGAGGGGAAATCCAGCTCTGGGAAATGCCCTGGAGGCTAACTGGCAGATAATGTTTGGCC TGACTGGCCAGCTACCCCGAAGCAGGAGATCCGGCACCGCAGTGCATGAAACGGCTGTGGCTGCAGATTACG CCGAGACCCCGACAAGAAAGATCATCTCTGCCGAGAGGATGAAAAGCTCATCGGGAAGCCGCTGCA AACTCTTTCTGGCAGACTTTGGAAATTACTGGGAGCAGGCAAGCTCAGCTGCAAGCCCTGAAAGCTGCCAACCCGGC TGGAAACCTTATAGCATCCACAGCACTGAACCTGTCTCTGGCCGAGCTGGAAAAGGGGAGAGGTTTGGAGCCCT GGTGAATGGACCTGATGGCAAGGCTGGAGGCGCACAACTCCCGCACCAATCAGCCTACTGGGGAGATCC TGGACAAGCTGCCAAGTCCCGCTCAAAAGAGGAAGGGAACCGCAATTAGCCAGCTGGCNAACCAACCCGTGGTC CGAACACAGAAATGACTGAGAAAGGTGCTCAACAATCTGATCGGGCTGTATGGAAAACCCGATCGGAATCCGGAT TGAAGTGGCCCGGACGTCGGGAAAGTCCAAAGAGAAAGGGAGGAATCCAGTCTGGCAATCGACGGAAACGAG AAGCAGAAAGAAAGCCACTGAAGATCTGATCAAAAACGGAAATTCCTAGCCCGGACCGATGTGGAGA AGTGGATCCTGTGGAAAGAGGGCCAGGAAGATGCCCATACACCCGGGACCCAGATTGGCTTCAAATGCCCTOTTT AGAGAGGCAGATATGAGGTGGAACACATCTGGCCCTCGCTCGAAGTTTGTATACAGCCCAAGGAATAAGAC ACTGTGTCGCAAGGACGTGAACATCGAGAAGGAAATAGGATGCCTTTGGAGGCAATTTGGCCATGACGAAAGATC GGTGGAGGCCCATCCAGAATAGACTGCAAGGCAATGCTGTCAAGCCAAAGGGGAACTGGATGAGCCCGGAA GGTCAAACGCTTCTGGCTAAGACCATGCCCTGAGGATTTTGCAGCCCGGCAAGCTGAAACGACACAGATAACGCTGC AAAGCAGATCCTGGCCAGCTGAAAGGCTGTGGCCAGACATGGACCTGAGGCTCCAGTGAAGGTCGAAAGCAG TGACTGGACAGGTCAACCCGACGCTGGCAAACTGTGGACTCTGAAACAAATTTCTGGCTGACGATGGGGAGAA ACCAGAGCAGATCAAGGCCACTGCCATCGACTGACCTGACCTGCACTCATCCTGGATGACCAACAA GCTGAGCAGGTAATGGCAGCTGCCGACCGATCCACGAGCAGAGAAAGCCAGCTGACTCCACCCCTGGATACCA TCCGCGCCAGCGCTGAGAAAGCCGCTGTGAAATTTGTGCTGATCCAGCCGGGTGAGAAAGAAAGTCAAGCGGGCCCA CTGCAATAGGAGACTACTACGGGGATACAGGACTGACATTAAGACCAATTCGGGCAATATAGACAGTTCGT</p>
--	---

FIG. 13W

	GACCAGGAAGAAAATCGAGTCACTGAGCAAGGGGGAGCTGGATGAAATTCGGGACCCCGAATCAAAAGAAATT GTGGCAGCTCAGCTCGCAGGACGAGGAGGCAACCCCAAGAAGCCCTTCCCTCCAACCCCTGTGTCTCCCGGA GGCCCTGAGATCCGGAAAGTCCAGACTGACCAGTAAACAGCAGCTGAACCTGATGGCCACACAGGAAATGGATA CCCTGACCTGGCTCCAAACCATATCGCAATCTACCGGCTCCCGATGGAAAGCCGACTTCGAGATTGTGTC ACTGTTGATGCTAGCAGAAGGCTGGCACAGAGAAATCCAAATCGTGCAGAGGACACGAGCAGACGGAGCCAGCT TCGTCAATGTCCTCGCAGCCGGAGAGGCCATCATGATTCGCCAAGCTCAAAGAAGGGATCTGGATTGTGCAG GGAGTCTGGGCAAGCCGACAGGTGCTGAGAGGGACACCGATGCTGACCACTACCAACTACCCGCTCAT GCCAAACCCCATCCTGAGGACGATGCCAAGAAGTGAGTATCGATCCTATTGGCCGATCCGGCCATCAAAATG ACTAAGAAATTC
Longitud de proteína (sin NLE y marcos de HA añadidos)	1037
PAM	NNNCAT
Longitud del separador	30
Especie: <i>Staphylococcus aureus</i> subsp. <i>Aureus</i> - <i>double-check</i>	
Secuenciación código de barras	CTATGCGT
ID	21
ARN guía numérico	NNNNNNNNNNNNNNNNNNNGUUUUAGUACUCUGGAAACAGAAUCUAAAACAAGGGCAAAAUGCCGGUG UUUAUCUCUGCAACUUGUUGCGGAGAUUUUUU
Secuencia de proteína Cas2 (de tipo salvaje sin NLS añadidas)	MKRNYYLGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRSRKRGARRLKRERRHRJQRVKKLLFDYNLLTD HSELSGINPYEARVKGLSQKLSSEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERL KKDGEVRSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRTYEGPGEQSPFGWKDIKEWYEMLMG HCTYFPEELRSVKYAYNADLYNALNDLNLLVITRDENEKLEYEYKQIENYFKQKKPTLKQIAKEILVNEEDIKGYR VTSTGKPEFTNLKVYHDIKDITARKEIENAEELDQIAKILTIYOSSEDIQEELTNLNSELTQEBIEQISNLKGYTGTGHNLSL KAINLILDELWHTNDNQIAIFNRLLKLVPKKVDLSQQKEIPTTLVDDFILSPVYKRSFIQSIKVINAIKKYGLPNDIIEELAR EKNSKDAQKMINEMQKRNROTNERIEEIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHJIP RSVDFNSFNNKVLVKQEENSKKGNRTPFOYLSSSDSKISYETFKKHILNLAKGKRISKTKEYLLEERDINRFSVQK

FIG. 13X

	<p>DEINRLVDTRYATRGLMNLRSYFRVNNLDVVKVKSINGFTSFLRRKWKFKKERNKGYKHHAEADALIIANADFIKKE WKLLDKAKVMENOMFEKQAESMPEIETEYKEJIFTHQIKHIKDFKDYKYSHRVDKKNRELINDTLIYSTRKDD KGNTLIVNNLNGLYDKDNDKUKLINSPEKLLMYHHIDPQTYQKLLKIMEQYGDEKNPLKYVYEETGNVLTKYSKK DNGPVIKKIKYYGNKLNALHDITDDYPNRNKVVKLSLKPYRFDVYLDNGVYKFTVYKNDLVIKKENYEVNSKCYE EAKLKKISNQAEFIASFVNNDLIKINGELVYRVIGVNNDLINRIEVNMDITYREYLENNMNDKRRPPRIIKTIASKTQSIKK YSTDILGNLYEVKSKKHPQIKKK</p>
<p>Secuencias distribuidas en clones de mamíferos (Agel-HA- NLS-Cas9- NLS-taa- EcoR)</p>	<p>ACCGTGCCACCATGTAACCATACGATGTTCCAGATTACGGTTCCGGCAAGAAANAAGCCGAAGGTGCAAGCGTTC CATGAAAAGGAACTACTTCTGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTTATTGACTATGAAA CAAGGGACGTGATCGACGCAGGGCTCAGACTGTTCAAGGAGCCAAAGTGGAAACAATGAGGGACGGAGAAG CAAGAGGGAGCCAGCGCCTGAAACGACGGAGAGGCACAGAAATCCAGAGGTGAAGAAACTGCTGTTCCGAT TACAACCTGCTGACCGACCAATCTGAGCTGAGTGGAAATTAATCTTATGAGCCAGGGTGAAGGGCCTGAGTCCAG AAGCTGTACAGAGGAAGATTTTCCGACGCTGTGCTGCACTGGCTAAGCCCGAGGAGTGCATTAACGTCAAATGA GGTGGAAGAGGACACCGCAACAGCTGTCTAGAAAGGACAGATCTCACGCAATAGCAAAGCTCTGGAAAGAG AAGTATGTCGAGAGCTGCAGCTGGAACGGCTGAAGAGATGGCAGAGTGCAGAGGGTCAAATTAATAGGTTCAA GACAACGGACTACGTCAAAGAGCCAAAGCAGCTGCTGAAAGTGCAGAAAGCTTACCAACAGCTGGATCAGAGCT TCATGGATACTTATATCGACCTGCTGGAGACTGGAGAACCTACTATGAGGGACCAAGGAGAGGGACCCCTTC GGTGGAAAGACATCAAAGGAATGGTACGAGATGCTGATGGACATTTGCCAGAAAGCTGAGAAAG CGTCAAGTACGCTTATAACGCAGATCTGTACAACGCCCTGATGACCTGAAACACTGGAACAACCTGGTCAACACCAGGATGA AAAACGAGAAACTGGATACTATGAGAAGTTCCAGATCATCGAAACCGTGTTTAAGCAGAAAGAAAAGCCTACAC TGAACAGATTGCTAAGGAGATCCTGGTCAACGAAAGAGGACATCAAGGGCTACCGGGTGAACAAGCAAGCACTGGAAA ACCAGAGTTCACCAATCTGAAGTGTATCAGGATTTAAGGACATCACAGCAGGAAAGAAATCAATTGAGAACG CCGAACTGCTGGATCAGATTGCTAAGTCCCTGACTATCTACGAGACTCCGAGGACATCCAGGAGAGCTGACTA ACCTGAACAGCGGACTGACCCAGGAAGAGATCGAACAGATTAATGTAATCTGAAGGGGTACACCGGAACACACAA CCTGTCCCTGAAGGCTATCAAATCTGATTTCTGGATGAGCTGTGGCATACAAACGACAAATCAGATTCCTTTAA CCGGCTGAAGCTGGTCCCAAAAAGGTGGACCTGAGTCAAGCAGAAAGAGATCCCAACCACTGGTGGACGAT TCAATCTGTCAACCGTGGTCAAGCGGAGCTTCAACAGAGCATCAAAGTGTCAACGCCATCATCAAGAAAGTACG GCTGCCCAATGATATCATTATCGAGCTGGCTAGGGAGAGAAACAGCAAGGACCGCACAGAGATGATCAATGAG ATGCAGAAAGGAAACCGGCAAGCAATGAACGCAATGAAAGGATTAACCGAACTACCGGAAAGAGAAACGCAA AGTACCTGATTGAAAATCAAAGCTCCAGGATATGCAAGGAGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCC TGGAGGACCTGCTGAAACATCCATTCACACTCAGGCTGATCATATTATCCCAAGAAAGCTGCTCGACCAAT CCTTTAACAAGGCTGGTCAAGCAGAAAGAACTCTAAAAGGGCAATAGGACTCCTTTCCAGTACCTG TCTAGTTTCAAGATCTCTTACGAAACCTTTAAAAGCACAATCTGAAATCTGGCCAAAGGAAAGGGCCGC ATCAGCAAGAGCAAAAAGGAGTACCTGCTGGAAGAGCGGGACATCAACAGATTTCTCGTCCAGAGGATTTTAT TAAACGGAAATCTGGTGGACACAAGATACGCTACTCCGGGCTGATGATGATCTGCTGGATCCCTATTTCCGGGTGAA CAATCTGGATGTAAGTCAAGTCCATCAACGGCGGGTTCACATCTTTTTCIGAGGGCAATGGAAAGTTTAAAAA</p>

FIG. 13Y

	<p>GGAGCGCAACAAAAGGGTACAAGCACCATGCCGAAGATGCTCTGATTA TCGCAAAATGCCGACTTCATCTTTAAAGG AGTGGA AAAAGCTGGCAAAGCCAAAGAAAGTGTATGGAGA ACCAGATGTTTCGAAGAGAAACAGGCCGAAATCTAT GCCCGAAAATCGAGACAGAACAGGAGTACAAGGAGATTTTCATCACTCCTCACCAGATCAAGCAATACAAGGATT TCAAGGACTACAAGTACTCACCGGGTGGATAAAAACCCAAACAGAGAGCTGATCAA TGACACCCCTGTATAGT ACAAGAAAAGCCGATAAGGGGAATACCCCTGATTTGTGAACAA TCTGAAACGGACTGTACGACAAAGATAATGACA AGCTGAAAAGCTGATCAACAAAAGTCCCAGAAAGCTGCTGATGTACCAACCATGATCCTCAGACATATCAGAAA CTGAAGCTGATTTATGGAGCAGTACGGGACGAGAAAGAACCCACTGTATAAGTACTATGAAAGAGACTGGGAACTA CCTGACCAAGTATAGCAAAAAGGATAATGCCCCCTGATCAAGAAAGATCAAGTACTATGGAAACAAAGCTGAATG CCGATCTGGACATCACAGACGATTACCCCTAACAGTCCGACAAAGGTGGTCAAGCTGTCAAGTCAACAAAGGAGAACT TCGATGCTATCTGGACAAACCGGGTGTATAAATTTGTGACTGTCAAGAA TCTGGATGTCA TCAAAAAGGAGAACT ACTATGAAAGTGAATAGCAAGTGTCTACGAAGAGGCTAAAAAGCTGAAAAGATTAGCAACCCAGCAGATTTCAATC GCCTCCTTTTACAACAACGACCTGATTAAGATCAATGGCGA ACTGATAGGGTCA TCCGGGTGAACAATGATCTG CTGAAACCGCATTTGAAAGTGAATGATGATGACATCACITACCGGAGATATCTGAAAACATGAA TGAATAGCGGCC CCTCGAATTA TCAAAAACAATTCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACA TTTCTGGGAAAACCTG TATGAGGTGAAGAGCAAAAAGCAACCTCAGATTA TCAAAAAGGGCTAAGAA TTC</p>
Longitud de proteína (sin MILS y marcador HA añadido)	1053
PAM	NNGRRT
Longitud del espaciador	30
Especie: <i>Mycoplasma gallisepticum</i> str. F	
Secuenciación código de barras	CGCTATGT
ID	23
AR(y guía química)	<p>NNNNNNNNNNNNNNNGUUUUUAGCACUGUACAAGAAAUUUGUGUGUUA AAAAUAACGCCGUGUUAAU GCAGCUGCCGCAUCCGCCAGAGCAUUUAUGCUCUGGCCUUUUUUU</p>
Secuencia de proteína Cas5	<p>MNNSIKSKPEVTIGLDLGVGSVGAIVDNETNIIHHLGSRKLFQAQKTAEDRRSFRGVRRLIRRRKYKLRFRVNLJWKYN SYTFGFKNKEDILNYYQEQQKLHNTVLNLKSEALNAKIDPKALS WILHDY LKNRGHFYEDNRDFNYYPTKELAKYFDK</p>

FIG. 13Z

<p>(GE 100 salivare em NLS antidade)</p>	<p>YGVYKGIIDSKEDNDNKLEBELTKYKFSNKHWHLBEVKKVLSNQTLGLPEKPKEEVESLFSYVRNYSEPGSINSVSPYGI YHLDEKBEKVQKYNINWDKTIGKCNIFPDEYRAPKNSPIAMIPNEINELSTIRSYSIYLTJGWFINOEFKAYLNKLLDL LJKTNGEKPIDARQFKLREETIAESIGKLEKDVENEELKEKEDHKWKLKGLKLNNGKIQYVNDLSSLAKFVHKLKQ HLKDLFLEDDQYATLTKINFLQSLFVYLGKHLRYNSRVDSANLKEFSDSNKLFERILQKQKDGIFKLEQTKDFDKDEKIL AQTHSLSTKAMLLAITRMNTNLDNEDNQKNNDKGNWFENKFDQKIDITKNNLSLKQNKRYLDRFRINDAILSP GVKRLREATKVFNAIKQFSEEDYTKVVIELARELSEKELENTKNYKKIKKNGDKISEGLKALOISEDEIKDILKSP TKSYKFLWLQDHDHYPYSLKEJAFDDIFTKTEKFEIDHIIPYSISFDDSSNNKLLVLAESNQAKSNQTPYEFISSNAGIK WEDYEAYCRKFKDGSLLDSTQRSKFKAKMMKTDTSKDYDIGFLARNLNDTRYATIVFRDALEDYANNHLVEDKP MFKVVCINGSVTSFLRKNPDPSSYAKKDRDKNIHHA VDAJISIFSNETKTLFNQLTFADYKLFKNTDGSWKKIDPKT GVVTEVDENWKQIRVRNQVSEIAK VIEKYIQDSNIERKARYSRKIENKTNISLFNDTVSYAKKVGYEDQIKRKNLKT DIIESAKENKNSKVKQFVYRKLNVVSLNNDKLADELFAEKEDILMYRANPWVINLAQIFNEYTENKLIKSONVFEK YMLDLTKEFPEKFSFLVKSMLRNKTAIHYDDKKNIVHRIKRLKMLSELKENKLSNVIRSKNQSGTKLSYQDTINSLA LMIRSIDPTAKKQYIRVPLNTLNHLGDHDFDLHINMDAYLKKPKPVYKANEIGDEYKPPWRVLTSGTLJHKKDK KLMYSSFQNLNDVIEIKNLJETEYKENDSDSKKKKANRFLMTLSTILNDYILLDAKDNFDILGLSKNRIDEILNSKLG LDKIYK</p>
<p>Sequências otimizadas e cofinadas de mimiferos (AgeI/HA- NLS-Cas9- NLS-tad- EcoRI)</p>	<p>ACCGTGCCACCATGTA CCCATAGCATGTTCCAGATTACGGCTCGCGGAGAAAAAGCGCAAGGTCCGAAGCGTGC CATGAACAATAGCATCAAA TCTAAACCTOAGTGAACCATCGGGCTGGAGCGGGAGTGGGAAGCGGAGCGGTGGG CAATCGTGGATAACGAAACAACATCATTCACCATCTGGCTCCAGCTGTTTCTCAGGCCAAGACTGCTGAGG ATCGGAGATCTTTCGGGGGTGAGGGGCTGATCCGAGGGAGAAAATACAAGCTGAAAAGGATTCGTCAA'ICTG ATTTGGAAGTACACAGETAATTCGGCTTCAAGACAAAGAGGACATCTCTGAA'AA'ATTA TCAGGAGCAGCAGAA GCTGCAAAATACCGTGTGAACCTGAAA TCAGAGCCACTGAA TGCC'AGATCGATCTCTAAAGC'ACTGAGCTGGA TTCGACGACTACCTGAGAACAGAGGGCCATTTTATGAGGACAAATAGGGATTTCAACGTTGCCAAACAAGG AGCTGGCCAAAGTACTTCGATAAGTACGGGTACTACAAGGGAATCA TTGACAGCAAGGAGGACAAATGATAACA CTGGAAGAGAGCTGACAAAGTACAA ATCTCCAATAAGCACTGGCTGGAGAGGOTGAAGAAAGTCTGTCTAA CCAGACTGGCCTGCCAGAAAAGTTTAAAGAGATAGTACTGTTGACTGTTGACTGTTGACTGTTGACTGTTGACTG GCCAGGGAGCATCAACTCTGTCAAGTCCCTACGGGATCTACCATCTGGACGAAAAGAGGAAAGGTGGTCCAG AGTACAACAACATCTGGGATAAGACAATCGAAAGTGAACAATCTTCTCCTJAGGAGTATAGAGCTCCCAAGAA CAGTCTATCGCAATGATTTTCAA TGAA'ATCAACGAGCTGTCCACAATCAGGTCA'ATCAGCATCTACCTGACTGG CTGTTCA'ATA'CAAGGAGTTCAAGAAAGCTACCTGAAAGCTGCTGATCTGCTGATCAAAACCAACGGAG AGAAGCCAAITGACGCAAGGAGTTCAAGAACTGGCGGAAAGAGACAAATCGGAAAAGCATTTGCCAAAGAGAC ACTGAAGGATGTGGAGAATGAGAGA AACTGGAAAGGAGGACCA CAAGTGA'AACTGAAGGACTGAAGCTG AATACCAAGCGCAAAATCCAGTACAACGATCTGAGCTCCCTGGCTAAGTTTGTGCACAAACTGAAGCAGCATCTG AAACTGGATTTCTCTGGAGGACCAAGTATGCAACACTGGACAAGATCAATTTCC'ITGCCAGTCCCTGTTTGTGTAC CTGGCAAGC'CTGAGATA'ITCCAATAGGGTGAATCTG'CCAACTG'AGGAAATTTTCGGACTCTAACAAACTG TTEGAGCCCATCCTGCAGAAACAGAAAGGATGGGCTGTTCAGCTGTTTGAACAGACTGACAAAGACGATGAGAA</p>

FIG. 13AA

<p>GATCTGGCCAGACATAGTCTGTCAACTAAGCCATGCTGGCTATTACCCGGATGACAAATCTGGACAA CGATGAGGACCAACCAAAACAATGACAAGGGCTGGAAATTTGAGGCCATCAAAAATCTCGATCAGAAGTTTA TCGACATCACCAAGAAAACAACCAACCTGAGCCCTGAACAGAAATAAGCGCTACCTGGACGATCGATTCAACAAC GATGCTATTCTGTCCCTGGGGTGAAGCGAACTCTGGGGAGGCAACCAAGGCTTTAATGCCAATCTGAAACAG TTCTCTGAAGAGTACGACGTGACAAAGGTGGTCAACGACTGGCTCGGACGCTGAGCGAAGAGAAAGAACTGGG GAACCAAAAGAACTACAAGAACTGATCAAGAAACCGGGACAAGAATTAGTAGGGCTGAAAGCACTGGG ATCTCAGAAGA TGAGATCAAGAGACTTTCTGAAGATCCCACTAAATCATACAAGTTCTGCTGTGGCTGCAGCAG GACCACA TCGA TCCCTTATAGCCTGAGGAGA TCGCCCTTCGACGATA TTTTACCAAAACAGAAAGTTCCGAGATC GACCATATCATTCCTACAGCATTTCTTCGACGATTTCTAGTTCAAACAAGCTGCTGGTGTGCTGAAAGTAATC AGGCAAA GTCAA CCAAGACTCCTTATGAGTTTCATCAGCTCCGAAACCGCAGCCATTAAGTGGAAAGATTACGAG GCCTATTCCCGCAAGTTCAAGGATGGGACTCTAGTCTGCTGGACCCAGCCAGCCAGGGTCCAGAAATTCGCCAA AATGATGAAAACCGATACCTCAAAGCAAGTACGACATCGGATTTCTGGCTCGAAATCTGAACGATCTCGGTACGG AACCA TTTGT TCCGGGACGCTGGAGGACTATGCTAATAACCACTGGTCGAGGACAACTTGAACGATTCCTCTTACGCCAAGAAAGATAG GGTCTGATCAA TGGGTCGGTGAOCTCTTCTCTCGGGAAGAACTTTGACGATTCCTCAATTTTCAGCAACGAGACAAGACTCTGTCAA AGCAAAGAA TATCCACCATGCTGTGGATGCAAGTATCATCTCAATTTTCAGCAACGAGACAAGACTCTGTCAA CCAGCTGACTCAGTTTCTGTA CTATAAACTGTTCAGAAACCGA TGGACCGA TGGCAGCTGGAGAAATCGACCCCTAAGA CAGGGTGGTCACTGAAGTGACCGACGAGAA TTTGGAAGCAGATTAAGGTGCGCAACCCAGGTGAGCGAAATCGCC AAAGTCA TTTGAGA AGTATCCAGGATAGCAACAATCGAAAGAAAGGCTAGGTATTCGCCAAAATCGAGAA TAA GACTAACATTTCCCTGTTTAA TGA CACCGTGTACTCTGCCAAGAAATGCGGCTATGAGGATCAGATCAAAGGAAA GAACCTGA AAAACCTGGACATTCACGAA TCTGCTAAAGAGAA TAA GAA CAGTAAAGTGAACCGGCA GTTTGTCT ACAGAAAGCTGGTGA TGTCAAGCTGCTGAAATAAGATAAGCTGGCAGACCTGTTCCCGAAAAGAGGATATC CTGATGATAGGGCCAA TCCATGGGTCA TCAACTGGCTGAGCAGATTTTCAATGAA TACACTGAGAA CAA GAA AATCAAGTCCCGAAGCTGTTTGAAAATA TATGCTGGA CCTGACCAAGAGTTCCTCCGAGAGTTCAAGGAGTT TCTGGTGAAGTCCATGCTGAGAAA CAAGACCCGCA TCACTACGAGGATAAGAAAACAATTTGCCATCGAATCA AACGGCTGAAGATGCTGAGTTCAGAACTGAAAGAGATTAAGCTGTACGTTGATCA TTAGGTCTAAGAAATCAG AGTGGGACCAAACTGTCA TACCAGGATACAATCAACAGCCCTGGCCCTGATGATTA TCGCCAGCATCGACCCCTACT GCTAAGAA CAGTATA TCGAGTCCCACTGATAACCTGAACTGCACTGACCTGGGAGATCATGACTTTGATCTCCAC AATATGGATGCTTACCTGAGAA ACCAAATTCGTGAAGTATCTGAAAGCAAACGAAA TCGGCGAGGAGTACAA GCCCTGGAGGGTCCCTGACATCTGGCACTCTGCTGATCCA TAAAGAGGATAAGAAACTGATGTACATCAGTCCCT CCAAGATCTGAACGACCTGATCCGAAATTAAGATCTGATCGAAACCGAGTATAAGAGAAACGACGATTCCTGATA GTAAAGAAAAGAAAAGGCAACCCGCTTCTGATGACCCCTGAGCAATACTCTGAA TGACTACA TTTCTGCTGGAC GCCAAGGATAACTTCGACATCCCTGGGCTGTCAAAAATCCGATCGATGAGATTTCTGAAACAGTACGCTGGGACT GGACAAGA TTTGTGAAA TAA GAA TTC</p>	<p> 269</p>
<p>Longitudinal profile</p>	<p>ns</p>

FIG. 13BB

Sein NLS y mezclador (HA añadido)	
PAM	NGGAT
Longitud del espaciador	30
Especie: Campylobacter lari CF89-12	
Secuenciación: código de barras	CTGTGGCG
ID	24
ARN guía químico:	<p>NNNNNNNNNNNNNGUUUUAGUCUCUGAAAAGAGACUAAAADAAGUGGUUUUUUGGUAUCCAGG CAGGGUUACAUAUCCUUUAAAACCAUUAAAUAUUAAAACUAGGUUGUAUCAACUUAGUUUUUUU</p>
Secuencia de proteína Cas9 (se tipo salvaje sin NLS añadido)	<p>MRILGFDIGINSIGWAFVENDELKDCGVRIKTAENPKNESLALPRNARSRRRLKRRKARLIAIKRILAKELKLNK DYVAADGELPKAYEGSLASVYELRYKALTONLETKDLARVILHIAKHGRYMNKNEKSNDAKGGKILSALKNNALK LENYQSVGEYFYKEFTQK YKNTKNFIKIRNTKDNYNVNCVLSDDLKELKLEKQKQEPGYNYSEDFINEILKVAFFQR PLKDFSHLVGACTFFEBEKRAKNSYSAWEFVALTKIINEIKSLEKISGEIVPTQTTINEVNLILDKGSITYKKFRSCINLH ESISFKSLKYDKENAENAKLIDFRKLVFVKKALGVHLSRQELDQJSTHITLIKDNVYKLVLEKYNLSNEQINLLEIEF NDYINLSFKALGMILPLMREGKRYDEACBIANLKPVTYDEKKDFLPFCDSIFAHLSNPVYVNRVRAISEYRKYVNLNALLK YGK VHKIHLELARDVGLSKAREKIEKEQENQAVNAWALKECENIGLKASAKNILKLLWKEQKEICITYSGNKISIE HLKDEKALEVDHIYPYRSFDDSPINKVLVFTKENQEKLNKTPPEAFGKNIEKWSKIQTLAQNLPYKKNKILDENFKD KQQEDFISRNLDTRYJATLIAKYTKEYLNFLLSNENANLKSQEGSKIHVQTSIGMLTSVLRHTWGFDDKDRNNH LHHLDAIIVAYSTNSIIKAFSDFRKNQELLKARFYAKELTSDNYKHQVKFFPFKSFREKILSKIDEIFVSKPPKRRARR ALHKDTFFHSENKIIDKCSYNSKEGLQIALSCGRVRKIGTKYVENDTIVVDIFKQNKFYAIPYAMDFAIGLIPNKIVIT GKDKMNNPKQWQTIIDESYEFCSLYKNDLILQKKNMQEPFAYYNDFSISSICVEKHDNKFENLTSNQKLLFSNAK EGSYKVESLGIQNLKVFEKYIITPLGDKIKADFPRENISLKTSSKYYGLR</p>
Secuencias optimizadas en codones de mamíferos (Agel-HA-	<p>ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTAACGCTTGGCCGGAAGAAAAGCGCAAGGTCGAAAGCGTC CATGAGGATTC'TGGGGTTTGACATGGCATTAAACAGCATCGGGTGGGCTTTGTGGAGAACGACGAACTGAAGG ACTGGGGAGTGCGGATCTTCAAAAGCGCGAAGACCCAAAATAAGGAAGCCCTGGCACTGCCCGGAGAAAT GCACGAGCTCCAGGCGCGACTGAAACGGAGAAAGCCCGGCTGATCGCTATTAAGAGAATCCTGGCCAAAGA GCTGAAGCTGAACTCAAGGACTATGTGCGAGCTGATGAGAGCTGCCAAAGCCCTACGAAGGATCCCTGGCAT CTGTGTACGAGCTGCGGTAAGGCCCTGACACAGAACCTGGAAACTAAAGATCTGGCCAGAGTATCCTGTGCAC</p>

FIG. 13CC

<p>NLS-Cas9- NLS-<i>taa</i>- EcoRI)</p>	<p>ATTGCTAAGCATAGGGGGTACATGAACAAGAAAGAGAAATCAAACGACGCTAAGAAAGAAAGATCCCTGA GGCTCTGAAAACAAATGCACGTGAGCTGGAGA ACTACGAGAGCTGGCGAATACTCTACAAGGAGTCTTT CAGAAATACAAGAAAACAAAGAACTTCATCAAGTCCGCACTAAGGATAATTTACAAACAATTGCGTGTCT GTCTAGTGACCTGGAAAAGAGTGAAGTGTATCTCGGAAAACAGAAAGGAGTTCGGCTACAACACTCTCTGAAG ATTTTCATCAACGAGATTCTGAAGTGGCTTCTTTCAGGGCCCTGAAAGGACTTCAGTCCGCTGGAGTTTGTGGCTCTGACCAAGA /GCACCTTTCTTTGAGGAAGAGAAAAGGGCCTGTAGACAGCTACTCTGCTGGAGTTTGTGGCTCTGACCAAGA TCATTAACGAGATCAAGAGCCTGGAGAAGATCAGGGCGGAAATTTGCGCAAACCCAGACAATCAACGAGTCCCTG AATCTGATCCTGGACAAGGGGTCTATCACTACAAGAAATTCAGAAGTTGTATCAATCTGCATGAGATATCAGC TTCAAGAGCCTGAGTGTAAAGAAAACGCCGAGATGCTAAACTGATCGACTCCGCAAGCTGGTGGAGTT TAAGAAAGCCCTGGGATCCACAGCTGTCCCGCAGGAATGGATCAGATCTCCAATCACTCATATCACCTGATTA GGACAACGTGAAGTGAACCCGCTCTGGAGAAATACAACCTGAGTAATGAACAGATCAACAATCTGCTGGAAA TTGAGTTCAACGATATACAACCTGAGCTTCAAGGCCCTGGAAATGATTTCTGCCACTGATGGCGAGGGCAAAC GATACGACGAGCCCTCGAGATGCCAACTCTGAAACTAAGACCCTGGACGAGAAAGAAATTTCTTCCAGCA TTTTGTGATTCATTTCCGCCACGAGCTGTCTAACCCCGTGTCAATAGGGCTATCAGCGAATACCGCAAGGTGC TGAACGCCTGCTGAAGAAATATGGAAGGTCCACAAAATTCATCTGAGCTGGCTCGGACGTTGGCCCTGAAAG AAGAAAGCACGAGAGATCGAAAAGAGCAGAAAGGAAAACAGGCCCTGTAATGCATGGCCCTGAAAGGAAT CGGAGAAATTTGGCCTGAGGCCAGCCAAAGACATCTCTGAAACTGAAAGCTGTGAAAAGAAAGAAAGGAGAT CTGTACTACTCCGGAATAAGATCTCTATTGAGCACCTGAAGATGAAAAGGCCCTGGAGGTGGAACCATATCTA CCCCTATTCTAGGAGTTCCGACGATTTCTTTTACAACAAGTGTGGTGTTCACCAAGGAAATCAGGAGAAACT GAACAAGACACCTTTCCGAGGCCTTTGGCAAGAAATATGAAAATAATGGAGCAAGATCCAGACCCCTGGCTCAGAACC TGCCATACAAGAAAAGATAAGATTCTGGACGAGAACTTCAAAGATAAGCAGCAGGAGGACTTTATCTCTCGA AATCTGAACGACACCCGGTATATCGCTACACTGATTGCAAAAATACACAAGGAGTATCTGAACCTTCTGCTGCTG AGCGAAAATGAGAACGCCAAATCTGAAGAGTGGCAAAAAGGGTCAAAGATCCACGTGCAGACTATTAGCGGGA TGGTGACCTCCCTGAGGCACACATGGGGTGTGACAAAAGGATCGCAACAATCACTGCACCATGCACCTGG ATGCCATCAATGTGGCTACAGTCAAATTCAAATCAATTAAGGCTTTTCAGCGAATTTCCGGAAAACCCAGGAGCTGC TGAAGGCCAGATTCTACGCTAAGAACTGACTTCCGATAACTATAAACATCAGGTCAGTTCTTTGAGCCCTTCA AGAGTTTTAGAGAAAATACTCTGTCAAAGATCGACGAGATTTTCGTGTCCAAACCACTCGAAAAGCGAGCTAGG CCGGCATGCACAAGGATACCTTTCAATCTGAAACAAGATCAATTGACAAGTGCAGTACAACCTCAAAGGAAAGG CCTGCAGATTGCCCTGAGCTGTGGAAGATGAGGAAAATCGGCACTAAGTATGTCGGAATGATACCATCTGTA GGGTGCACATTTCAAAGCAGAAACAAGTTTACGCTATCCCAATCTACCGAAATGAAATTTGCCCCTGGGGATCC TGCCCAATAAGATCTGTGATTACTGGAAAAGATAAGAAACAATAACCCCAACAGTGGCAGACCATTTGACGAATCA TAGAGTTCTGCTTAGCCCTGATAAGAAATGACCTGATCTCTGTGCAGAAAAGAAACATGCAGGAACCTGAGTTC GCCTACTATAACGATTTTCAAATCAGCACATCAAGCTATTGTTGGGAAAACACGACAACAGTTCGAAAATCTG ACTAGCAACCAGAAAGTCTGTTTCCCAATGCAAAAGAGGGCTCTGTGAAGTCCGAAAGTCTGGGGATCCAGAA CCTGAAAAGTGTTCGAGAAAGTACATCAATACCCCTGGGAGATAAATTAAGGCTGACTTTTACGCTCGAGAAAA CATCAGCCTGAAAACCCAGTAAAAGTATGGCCTGAGGTAAGAATTC</p>
---	--

FIG. 13DD

Longitud de proteína (sin NLS y marcador HA añadido)	1003
PAM	NNGGGYA
Longitud de espaciador	29
Especie: <i>Streptococcus pyogenes</i> SF370	
Secuenciación código de barras	CCAGTTAG
ID	25
ARN guía (almacenado)	NNNNNNNNNNNNNNNGUUUUAAGAGCUAGAAUAAGCAAGUAAAAUAAGGCUAAGUCCGUUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU
Secuencia de proteína Cas9 (de tipo salvaje sin NLS añadido)	MDKYSIGLDIGTNSVQWAVITDEYKVPSSKFKVLGNTDRHSIKNLIIGALLFDSGETAETRLKRTARRRYTRRKNR ICYLQEIFSNEMAKVDDSFHRLSESLVEEDKKHERHPIFGNIVDEVAHYEYPTTYHLRKKLVDSSTDKADRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVDFKLIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLF GNLIAISLGLTPNFKSNFLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLKALVRQQLPEKYEIFFDQSKNGYAGYIDGGAHQEEFYKFIKPILEKMDGTPELLVKNRE DLLRRQRFTFDNGSIPHQHLGELHAILRRQEDFYFLKDNREKIEKILTRIPYVYVGLARNSRFAWMTTRKSEETTPW NFEVVDKGAQAQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVYVTEGMRKPAFLSGEQKKAIVDLLPK TNRKVTVKQLKEDYFKIECFDSVEISGVEDRFNASLOTYHDLKIIKDKDFLDNEEDILEDIVLTLTFEDREMIEE RLKTYAHLFDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKA QVSGQDLSLHEHIANLAGSPAIKKGILOTVKVVDLTKVMGRHKPENIVEMARENQTTQKQKNSRERMKRIEIGIK ELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGS DNVPSEEVVKKMKNYWRQLNAKLITQRKFDNLTKAERGGSELKAGPIKQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKSLLVSDFRKDFQYKVRINNYHHAHDAYLNAVGTALIKKYPKLESEFVYGDYKYYDVRKM IAKSEQEKATAKYFFYSNIMNFFKTEITLANGEIRKRPIETNGETGEIVWDKGRDFATVRKVLSPQVNVKKTVEVQ TGGFSKESILPKRNSDKLIARKDWDPKKYGGFDSPTVAYSVLVAKVEKGSKKLKSVELLGITIMERSSEFKNPID FLEAKGYKEVKKDLIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFVYLASHYEKLGKSPEDNEQQLF VEQHKHYLDEIIEQISEFSKRVILADANLDKVLVSAYNKHRDKPIREQAENIHLFTLTNLGAPAAFKYFDITDIRKRYTST

FIG. 13EE

<p>Secuencias clonadas en copias de material (AgeI-HA- NLS-Cas9- NLS-1aa- EcoRI)</p>	<p>KEVLDATLHQISITGLYETRIDLSQLGGD ACCGGTGGCAACCATGTACCCATACGATGTTDCAGATTACGGCTTCCGGGAAAGAAAAGGGCAAGGTGGAAGCGGTC CGACAGAAAGTACAGCATCGGCCCTGGACATCGGCACCAACTGTGTGGCTGGCCGTGATCACCCGACCGAGTACA AGGTGCCCAAGAAATTCAGGTGCTGGGCAACCCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCC CTGCTGTTCCGACAGCGGCAACAGCCGAGGCCACCGGCTGAAGAGAACCGCCAGAGAGATACACCCAGAC GGAGAAACCGGATCTGCTACTGCAAGAGATCTTCAGCAACGAGATGGCCAAAGGTGGACGACAGCTTCTCCAC AGACTGGAAGAGTCTCTGGTGAAGAGGATTAAGAAGCAGAGCGGCAACCCATCTTCGGCAACAATCGTGGGA CGAGTGGCTAACACGAAAGTACCCCAACATCTACCACCTGAGAAAGAACTGGTGGACAGCACCGACAAGG CCGAAGCTGGCTGATCTACTGGCCCTGGCCACATGATCAAGTTCGGGGCCACTTCCCTGATCGAGGGGGGACC TGAACCCGACAACAGGACGTTGGAACAAGCTGTTCAATCCAGCTGGTGCAGACCTACAACCCAGCTGTTCCGAGGAA AACCCCATCAACGCCAGCGCCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGACGACAGCGGTGGA AAATCTGATCGCCAGCTGCCCGGAGAGAGAAATGGCCCTGTTCCGCAACCTGATTCGCCGTGAGCCCTGCGGCT GACCCCAACTTCAAGAGCAACTTCGACTGGGGAGGATGCCAAACTGCAGCTGAGCAAGCAAGGACACCTACCGA ACGACTGACAACTGTGGCCAGATCGGGACCAAGTACCGCACTGTTCGCGCCGCAAGAACCTGTCCG ACGCCATCTGCTGAGCGACATCTGAGTGAACCCGAGATCACCAAGCCCCCTGAGCGCCCTCTATGATCA AGAGATCGACGAGCACCAACAGGACCTGACCTGCTGAAGCTCTGTGGCCAGCAAGCTGAGCAAGTAC AAAGAGATTTCTCGACAGAGCAAGAACCGCTACCGCGCTACATTCAGCGGGGAGCCAGCCAGGAAGAGTTT CTACAGTTCAACAAGCCCATCTCGAAAGATGGACGGCAACCGAGAACTGCTCGTGAAGCTGACAGAGAG ACCTGCTGGGAGCAAGCGGACCTTGGACAACGGCAGCATCCCCCAACCAATCCACCTGGGAGACTGCAACGCCC ATTCGGGGGGCAGGAAGATTTTACCCATTCCTGAAGGACAACCCGGAAAGAATCGAGAAAGTCTGACCTTC CGCATCCCTACTACGTGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCCTGGATGCCAGAAAGAGCGGATGA AACCATCAACCCCTGAACTTCGAGGAAAGTGGTGGACAAGGGCGCTTCGCCCCAGAGCTTCAATCGAGCGGATGA CCAACITCGATAAGAACCTGCCAACGAGAAAGTGTGCTGCCCAAGCAAGCCTGCTTACGAGTACITTCACCCGTGT ATACGAGCTGACCAAGTGAATACGTGACCGGAGGAAATGAGAAAGCCCGCCCTTCCCTGAGCGGCGAGCAGAA AAAGCCCATCGTGGACCTGTGTTCAAGACCAACCGGAAAGTGAACCGTGAAGCAGCTGAAGAGGACTACTTCA AGAAATCGAGTCTCGACTCCGTGGAAATCTCCGGGGTGGAAAGATCGGTTCAACGCCCTCCCTGGGCACATACC ACGATCTGCTGAAATATCAAGGACAAGGACTTCCCTGGACAAATGAGGAAAGAGGACATTCCTGGAAGATATC GTGCTGACCTGACACTGTTGAGACAGAGAGATGATCGGAGGAAACGGCTGAACCTATGCCCAACCTGTTCGA CGACAAAGTGAAGAGCAGCTGAAGCGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAAC GGATCCGGGACAAGAGTCCGGCAAGCAATCCCTGGATTCCTGAAAGTCGACGGCTGCGCAACAGAAACTT CATGCAGCTGATCCACGACGACGCTGACCTTTAAGAGGACAATCCAGAAGCCCAAGGTGTCCGGCCAGGGCGG ATAGCCCTGACCGAGCACATCGCCAACTCTGGCCAGCCCGCCGCAATTAAGAGGGCATCCCTGACAGCTGAAG GTGGTGGACGACTCGTGAAGTGAATGGCCGGCACAAGCCGAGACATCGTGTATCGAAATGCCACAGAGAGA ACCAGACCAACCAAGAGGACAGAAACAGCGCGGAGAGAAATGAAGCGGATCGGAGAGGGGATCAAAAGGCT GGGCAGCCAGATCTGAAAGAAACACCCCGTGGAAACAACCCAGCTGCAGAACGAGAGCTGTACTACTAC</p>
--	--

FIG. 13FF

	<p>TGCAGAA TGGGGGGA TATGT ACGTGGACAGAACTGGACATCAA CCGGCTGTCCGACTACDATTGGACCAT ATCGTCCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAA CAAGTGTCTGACCAAGAAAGGCAAGAAACCGGGG CAAGAGCGACAACGTGCCCTCCGAGAGGTCCTGAAAGATGAGAACTACTGCGGCGCAGCTGCTGAAACGGCCA AGCTGATTACCCAGAGAAAGTTCCGACAA TGTGACCAA GCGCGAGAGAGCGGCTGAGCGAACTGGATAAGGCC GGTTTCA TCAAGAGACAGCTGGTGGAA CCGGACAGTCAACAAGCACTGGGACACAGATCTGGACTCCCGGAT GAACACTAAGTACGACGAGATGACAAGCTGATCCGGGAAGTGAAGTGA TCA CCGTGAAGTCCAAGCTGGTGT CCGATTTCCGGAGGATTTCCAGTTTTACAAGTGGCGGAGATCAACA ACTACCA CCGCCACGACGCGCTACC TGAAGCGCGTGTGGAA CCGGCCCTGATCAAAAAGTACCTAAGCTGGAAAGCGAGTTCGTGTACGGCGGACTAC AAGGTGTA CGACGTCCGGAAGATGATCGCCAAGAGCGGAGCAAGAAATCGGCAAGGCTACCGCCAAAGTACTTCTT CTACAGCAACA TCA TGA ACTTTTCAAGACCGAGATTACCTGGCCAAGGGGAGATCCGGAAAGCGGCCCTCTGAT CGAGACAAA CCGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGATTTTGCCACCGTGCAGAAAGTGTCTGA GCA TGGCCCAAGTGAATATCGTGA AAAAGACCGAGGTGCA GACAGCGGCTTCAGCAAAGAGTCTATCCCTGCC AAGAGGAACAGGATAAGCTGATCGCCAGA AAGAAGGACTGGACCTAAGAAGTACGGGCGCTTGGACAAGCC CCACCGTGGCCTATTCTGTCTGGTGGTGGCCAAAGTGGAAAGGGCAAGTCCCATCGACTGGAAGCCAAAGGG GAGCTGCTGGGATCACCATCATGGAAAGAGCAAGCTGCCTAA GTACTCCCTGTTCGAGCTGGAAACCGGCCGGA CTACA AAGAGTGA AAGGACCTGATCATCAAGTGGAAAGGGCAAGTCCAAAGAACTGAAAGTGTGAAA AGAGAATGCTGGCTCTCCGGCGA ACTGCAGAA GGGAAACGAACTGGCCCTGCCCTCCAATATGTGA ACTTC CTGACTGGCCAGCCACTATGAGAAAGCTGAAGGGCTCCCGAGGATAA TGAGCAGAAACACAGCTGTTGTGGA ACAACAAGCACTA CCTGACGAGATCATCGAGAGATCAAGGAGTTCGCAAGAGAGTGA TCCCTGGGCGACG CTAATCTGGACAAAGTGTCTCCGCTACAACAAGCACCGGATAAGCCCATCAGAGAGCGCGCGGAGATATC ATCCACCTGTTTACCTGACCAATCTGGGAGCCCTGCGGCCCTCAAGTACTTTGACACCA CCA TCGACCCGGAAG AGGTACACCA GCA CCAAGAGGTGCTGGACCGCCACCTGATCCACGAGCATCACCGGCCCTGTACGAGACACG GATCGACCTGTCTCAGCTGGGAGGGGACAGCCCCAAGAAGAGAGAAGGTGGAGGCCAGCTAAGAAATTC</p>
<p>Longitud de proteína (sin NLS y marcador HA añadido)</p>	<p>1368</p>
<p>PAM</p>	<p>NGG</p>
<p>Longitud de espaciador.</p>	<p>30</p>
<p>Especie: <i>Streptococcus thermophilus</i> LMD9 (CRISPR-1)</p>	

FIG. 13GG

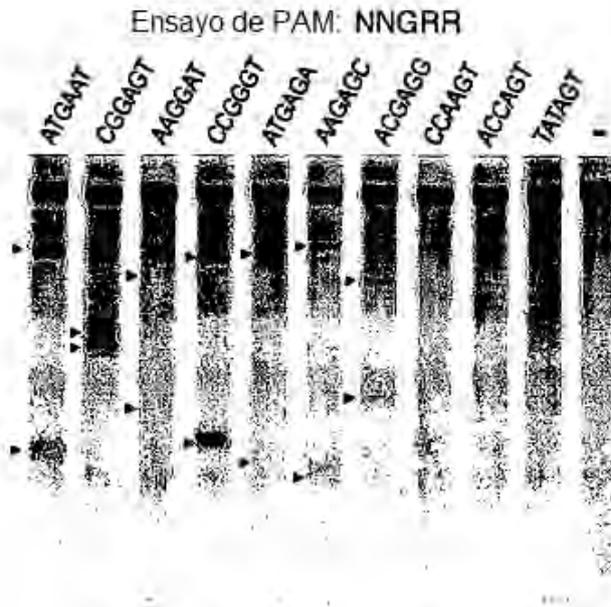
Secuencia de aminoácidos	TGCGAATG
ID	26
ARN guía número	NNNNNNNNNNNNNNNGUUUUGUACUCUCUAA GAUUUAGAAA CUUGCAGAA GCUACAAGAUAAAGG CUUCAUGCCGAAAUCAA CACCCUGUCAUUUU AUGGCAGGGUGUUU
Secuencia de proteína Cas9 (de 100 salivale en NLS anclada)	MSDLVLGLDIGSVGVGILNKVTGEIHKNSRIFPAQAENNLVRRTRQRRRLARRKRRRLNRLNRLFEESGLITDFT KISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDASDDGNSSVGDYAQIVKENSQOLETKTPGOIQLERY QTYGQLRGDFTVEKDGKKHRLINVFPTSA YRSEALRILQTOQEFNPQITDEFINRYLEILTGKRY YHGGPNEKSRITDY GRVRTSGEITLDFNFGILIGKCTFY PDEFRAAKASYTAQEPNLLNDELNLLTVPTETKLSKEQKNQIINYVKNKAMGPA KLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAVVLTNTEREGIQEAL EHEF ADGFSQKQVDELVQFRKANSSIFGKGVHNFVSKLMMELJPEL YETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTE EIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVEMARETNEDDEKAIQKIQKANRDEKDAAMLKAAANQYNGKAEELP HSVFHGHKQLATKIRLWHQQGERCLYTGKTIHDLNNSNQFEVDHILPLSITPDDSLANKVLYYATANQEKGGQRTPY QALDSMDDA WSFRELKAFVRESK TLSNKKKEYLLTEEDSKFDVRKFFIERNLVDTRYASRVVLTNTEREGIQEAL EHEF KVSVVRGGFTSQLRRHWGIEKTRDTHHHAVDALIAASSQJLWKQKNLTVSYSEDQLDIETGELISDDEYKESV FKAPYQHFDLTKSKEFEDSILFSYQVDSKFNKISDAIYATRQAKVQKDKADETYVLGKIKDIYTQDGYDAFMKIY KDKSKFLMYRHPQTEFEKVIPILENYPNKQINEKGVPCNPFLLK YKEEHGYIRK YSKKGNQPEIKSLKYVDSKLG N HIDIYTPKDSNNKVVYLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGYKISQEKVNDIKKKEGVDSSEFKFT LYKNDLLL VKDTETKEQQLFRFLSRTPKQKHYYVELKPYDKQKFEQGEALIKV LGNVANSQQCKKGLGKSNISYK V RTDVVLGNQHIHKNEGDKPKLM
Secuencias optimizadas en: - secuencias de mamíferos - (AgeI-HIA- - NLS-Cas9- - NLS-taa- - EcoRI)	ACGGTGCACCATGTACCCATACGATGTTCCAGATTACGGTTCGGCAGAAAAGCGCAAGGTGGAAGCGGTC CTCCGACCTGGTACTTGGACTGGATATGGTATCGGTTCGGTGGGAGTGGGATTCCTCAACAAGGTCAACGGGGA GATCATTCACAAGAACTCGCGGATCTCCCGCAGCTCAGGCTGAGAA CA ACTTGGTCCGGAGAACGAAATAGGC AGGCAAGGACTGGCAGGAGGAGAAACA CAGGAGAGTCCGATTGAA CCGGCTGTTCCAGGAGTCCGGTGT GATCACCGACTTACGAAAATCTCGAATTAACCTTAATCCCTATCAGCTTCGGGTGAAAAGGCTGACACGAACT TCCGAAATGAGGAACTTTTCATCGCGCTGAAAACATGGTCAAGCACAAGGGATTTCCCTACCTCGATGACCCCTC GGATGACGGAAATTCCTCAGTAGGAGATTATGCACAGATCGTGAAAGAGAACTCAMAAGCAACTGGAAACA AAG ACACCGGGCAGATCCAACTTGAAGATACAGACATACGGACAGCTCAGAGGATTTTACGGTGGAGAGGA CGGTAAAAGCAGACACTAATACGTAATTCACCGTCCGGCTACAGATCCGAAAGGCTCCGCATCTTCAGAC TCACAAGGATTCACACCGCAATTA CTGATGAGTTCATCAACCGCTATTTGGAAATCTTGACCGGAAAGCGCAA GTATTA TCAATGGCCGGGTAATGAGAAATCCAGAACATATTCAGCCGCAATACAGAACTCCGGGGAAACCTTGG ATAACTCTTGGTATTGGAAAGTGCACCTTTACCCGAGGAGATTTCAGAGCGGCCAAAGCGGTCATACA CAGCACAAGGATTTAAATCTCTTGAATGATTGMACTTACCGGTCCCGCAGGAGACAAGAGAGCTCTCCAAG AGCAAAAAGAACCAATCATCACTACGTCAGAAAGAGGCTATGGGGCCAGCGAAAGCTTCAAAGTATATC

FIG. 13HH

<p>GCTAAACCTCTCAGCTGTGATGTGGGGACATCAAAGGGTACCGAATCGACAATCGGAAAGTGGGAAAGCGGAAATTICA CACGTTTTGAAGCATACGAAAGATGAAACGTTGGAAACACTGGAACTGGACATTTGAGCAGATGACCGGGAAACCGCTCG ACAAACTGGCATAAGTCTCAGCTTGAATACTGAAACGAGAGGGAAATCCAAAGAGGCCCTTGAACATGAGTTCGCC GATGGATCGTTACAGCAGAGCAAGTGTGCAATTCGCAAGCGCAATAGCTCCATCTTCGGGAA GGGATGGCAAACTTTCCGGTCAAATCAATGAGTGGAGTTGATCCCAAGACTTTATGAGACTCCGAGGAGCAANT GACGATCTTGACCGGCTTGGGGAAACAGAAACGCAAGCTCATCGAAACAAACTAATGATTTGAGAAAT TGCTGACGGAAAGAAATCTATATCCGGTAGTGGAAATCGGTAAGACAAAGCAGCAAAATCGTGAACCGCGGG ATCAAGGAATAATGGTACCTTTGATAACATCGTAATTTGAAATGGCTAGAGAGACGAAACGAAATGACGAGAA GGCAATCCAGAAGATOCAGAAAGCCAAAGGATGAAAGATGACAGGATGAAAGCTTAAAGCGGCAACCAATAC AATGGAAAGCGGAGCTGCCCAATTCAGTGTTCACGGTCAATAACAGTTGGCGAACCAAGATCCGACTCTGGCAT CAGCAGGTTGAGGGTGTCTCTACACCGGAAAGACTATCTCCATCCATGACTTGTGATTAACAATTCGAAACAGTTT GAAGTGGATGATAATCTGGCCCTGTCAATCACTTTGACGACTCGCTTGGCAACAAGGTGCTGTGTACGCAACG GCAAATCAAGGAAAGGCCAAGCGGACTCCGTAACAAGCGCTCGACTCAATGGAAGATGGCTGGTCAATCCGGCA GCTGAAAGGGTTCGTACGGGAGGCAAGACACTGAGCAACAATAAGAAAGATATCTGCTGACAGAGGAGGAC ATCTGGAAATTCGATGTCAGGAAGTTCATCGAGCGGAAATCTTTGCGACACTCGCTACCGTTCCAGAGTAGTA CTGAAACCGCTCCAGGACAATTTAGAGGGCAAAATTTGACAAGAGGTTCAGTGGTGAGAGGCGCAATTCAC ATCCAACTCCGCGACATTTGGGCAATCGAAAGAGACCGGGACACATATCACCATCATCCGGTGGACCGCTGA TTAATGGCCCTTCGTCCAGTTGAAATCTCTGGAAAGAGCAAGAAACAGCTGGTGTGATTCGGAGCATCAGC TTTTGGACATCGAAACCGGGGAGCTGATTTCCGACGATGATAACAAGATCGGTGTTAAGGCACCAATCAGG ATTTCTGTGGACACCGCTGAAGAGCAAGAGTTTGAGGACAGCATCTCTTTTCGTACCAAGTGGACTCGAAGTTTA ATCGCAAGATTTCAAGCCCAATCTAGCGGACGAGGCAAGGAGGCGCAAGATAAAGCAGATGAAAC CTACGTCCTTGGTAAATCAAGGACATCTACACTAGACGGGTACGATCCGTTCAAGAAATCTACAAGAGG ATAAGTCGAAGTTTCTCATGTAACGGCCAGATCCACAGACTTTCCGAAAGAGTCAATTTGGAGACT ACCTAACAAAGCAATCAACGAGAAAGGAAAGAAAGTCCGTCAGACCCCTTCTGAGTACAAAGAGGAC GGTATATCCGCAATACTCAAGAAAGGAAATGGGCTGAGATTAAGTCCGTTAAGTATTACGACTCAAAGTTG GGTAACCACTGACATTAACCCGAAAGACTCCAAACAAGTCTGGATTAAGTCCGTTAAGTATTACGAGCCCTTCAATTCGA GATGTGATTTAATAAGACGACCGGCAATATGAGATCTTGGATCAATACCGAGACCTTCAATTCGA GGGACGGGCACTTATAAGATTTCAAGAGAAAGTCAACGACATCAAGAAAGGAAAGGAGGGTCCGATTCAGATT GGAGTTCAAATTCACCCCTTCAAAACGACCTCTGTTGTAAAGGACACAGAAACGAAAGGACGACTCT TTCGGTTCCTCACCACGATGCCAAACAATAACATACCGTCGAACTTAAACCTTACGTAAGCAAAAGTTT AAGGGGAGAGGCACTGATCAAGTATTGGGTACGTAGCCATAGCCGACAGTGTAAAGAAAGGCTGGGAAA GTCCAAATATCTCGATCTAATAAGTACGAAACAGATGATTTGGGAAACCAAGCATACTCAAAATAAGGGGGGATA AACCCAAACTCGATTTCAAGCCCAAGAAAGAGAAAGGTTGGAGGCCACGCTAAGAAATTC</p>	<p>1120</p>
	<p>Longitud de proteína</p>

FIG. 13II

(sin NLS y marcador HA añadido)	
PAM	NNAGAA
Longitud del espaciador	30



Secuencias de espaciador para PAMs ensayadas

<u>Espaciador</u>	<u>PAM</u>
GCCCGGTGGAAGTGGTAGCC	ATGAAAT
GTTGAAGATGAAGCCCAGAG	CGGAGT
GCTCCGACGAGGTGGCCATC	AAGGAT
GCACCATCTCTCGTGGTACC	CCGGGT
GGTGGAACTGGTAGCCATGA	ATGAGA
GCCATGAATGAGACCGACCCA	AAGAGC
GCATCCTCGTGGCACTCCG	ACGAGG
GCAGAGCGGAGTGCTGTTCTC	CCAAGT
GGTCGGTCTCATTATGGCT	ACCAGT
GCAATAAAAGGTGCTATTGC	TATAGT

FIG. 14

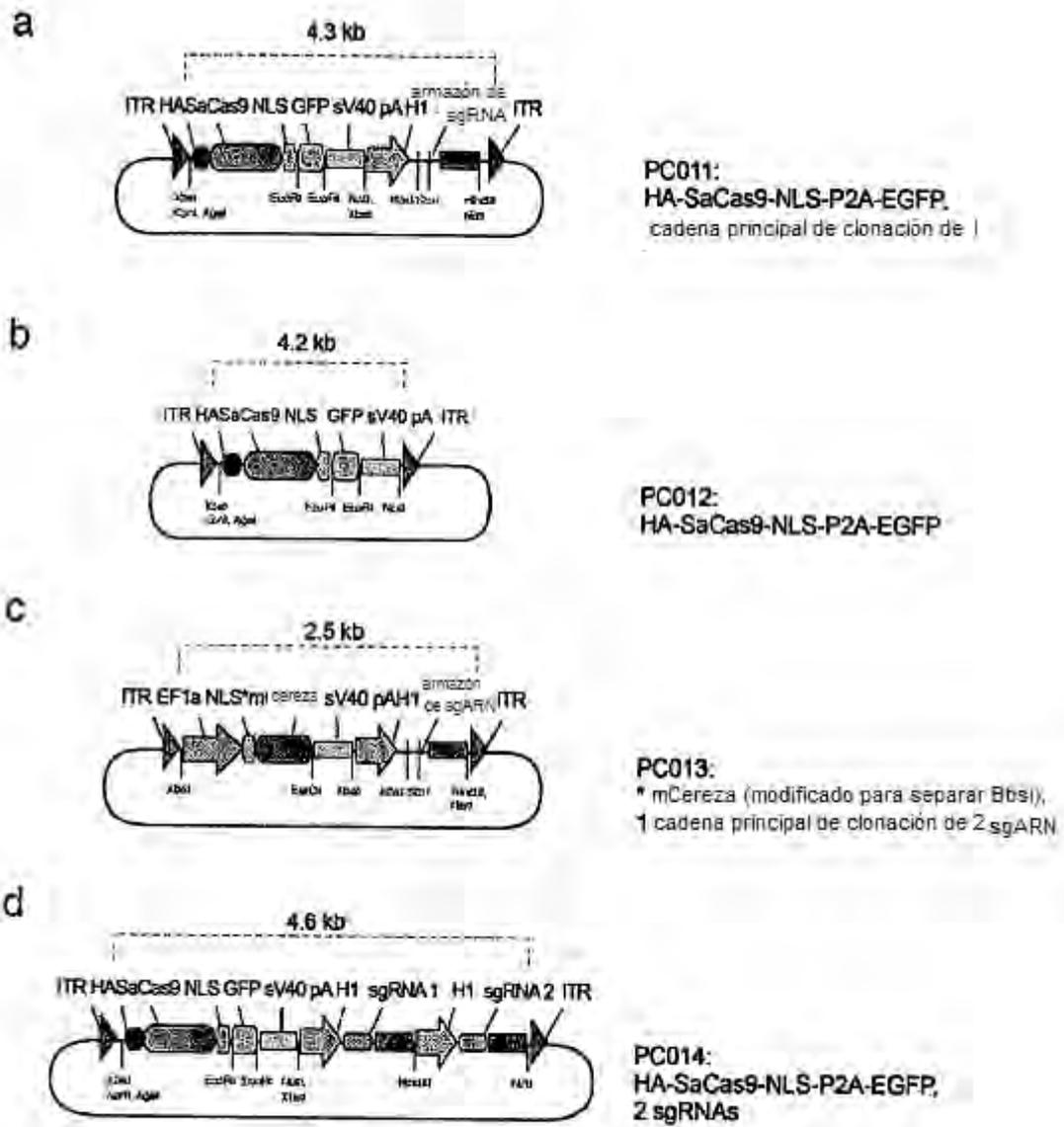


FIG. 15

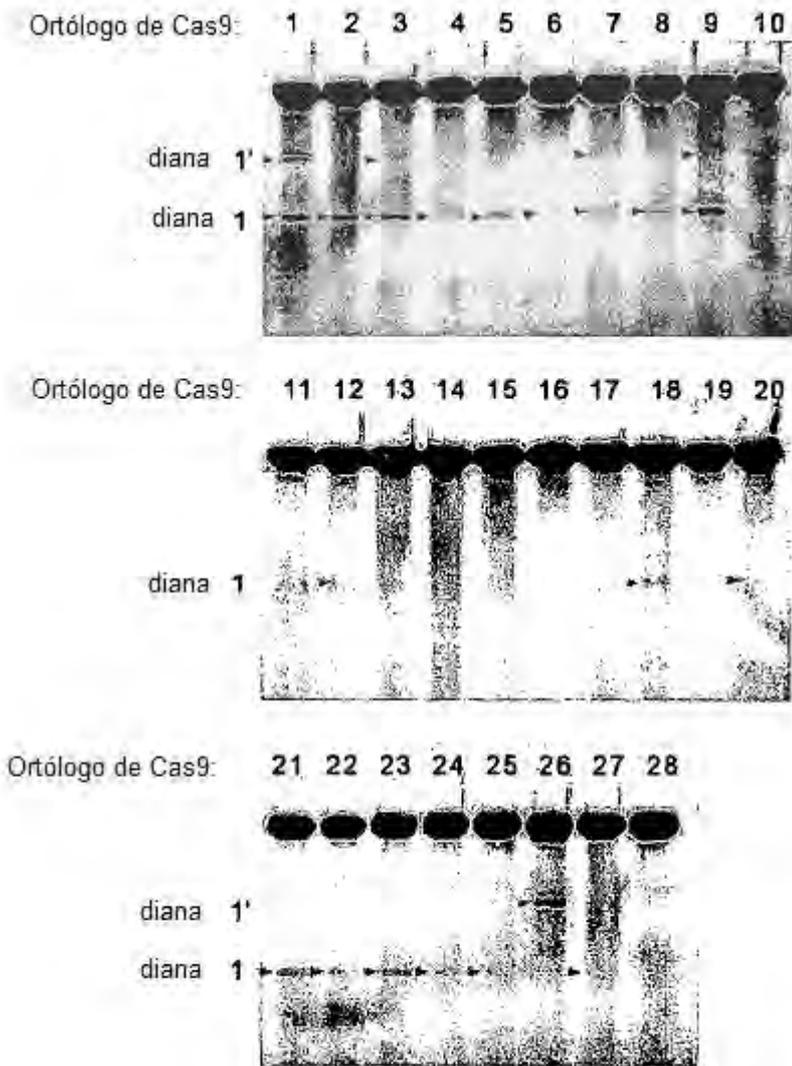
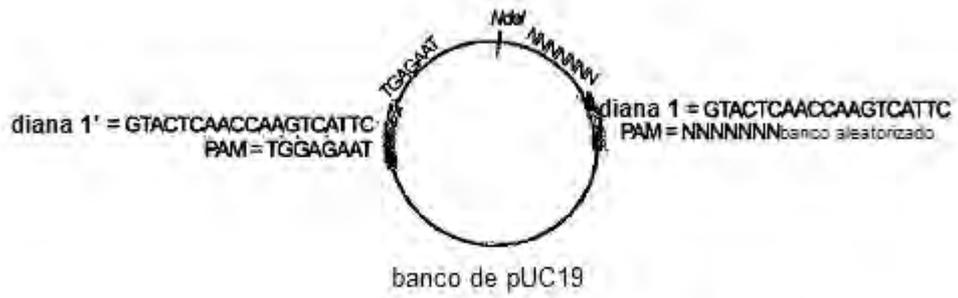


FIG. 16

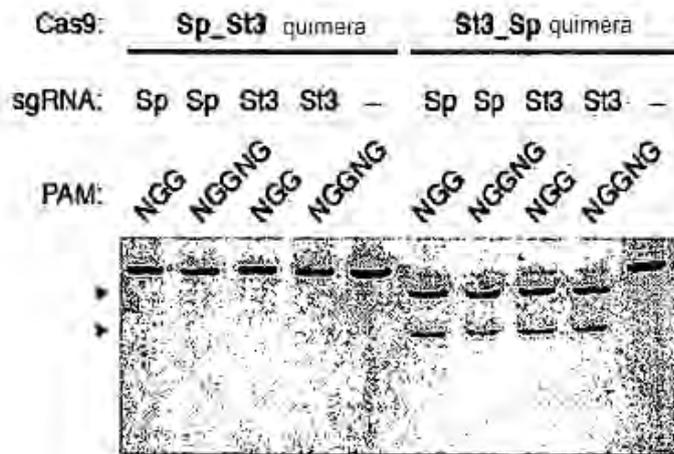
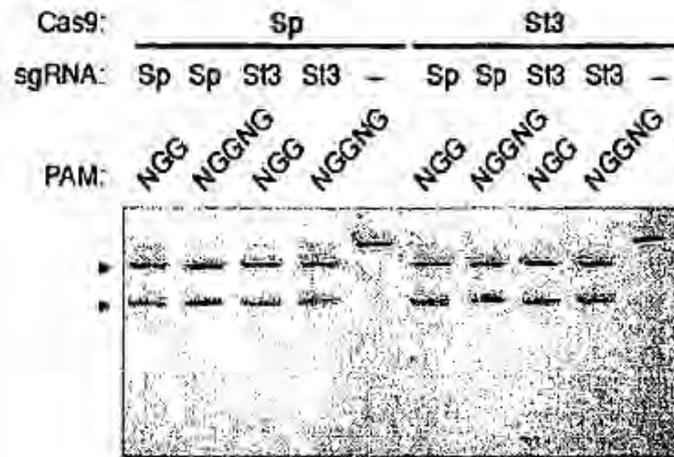


FIG. 17

Sábado, 15 de junio de 2013

Servidor de la web de CRISPRs

Trabajos relacionados

CRISPRs (12/2012)

Sumario de datos

Items	14
Items por página	10
Mostrar	1-14
Página	1 de 2
CRISPRs	14
CRISPRs con espaciadores	14
CRISPRs con repeticiones	14

Estadísticas de datos

Actualizadas: 2013-07-23

Contenido: 1000000000

Home | Datos CRISPRs | Datos CASs | News | Contact Us | Estadísticas | Usar

[Home](#) | [Datos CRISPRs](#) | [Datos CASs](#) | [News](#) | [Contact Us](#) | [Estadísticas](#) | [Usar](#)

Welcome to CRISPRs web site. This site is primarily for CRISPR databases and activities. It is a part of the CRISPR project in the Laboratory of Molecular Biology and Biotechnology, University of Zaragoza. CRISPRs present in the database. It is a part of the CRISPR project in the Laboratory of Molecular Biology and Biotechnology, University of Zaragoza. This web site is for CRISPRs and CASs. It is a part of the CRISPR project in the Laboratory of Molecular Biology and Biotechnology, University of Zaragoza.

Por favor baten:
 CRISPR Finder: una herramienta para identificar repeticiones palindrómicas cortas en rdnomo y regularmente interespaaciadas.
 Núcleo Acids Res, 21 mayo 2007
 La base de datos de CRISPR y las herramientas para exhibir CRISPRs y para generar diccionarios de espaciadores y repeticiones.
 BNC Bioinformatics, 23 de mayo del 2007, 6(1): 177.
 CRISPRfinder es un sitio web para comparar repeticiones palindrómicas cortas en rdnomo y regularmente interespaaciadas.

Navegación

Inicio

CRISPRs database

CRISPRs finder

CRISPRs comparison

Herramientas

CRISPR Finder

CRISPRs database

CRISPRs comparison

CRISPRs finder

CRISPRs database

Enlaces CPMS

CRISPRs

CRISPRs

CRISPRs

Propósito general de este sitio

Este sitio proporciona herramientas para ayudar a encontrar las estructuras de CRISPRs en genomas microbianos publicados o en secuencias proporcionadas por el usuario. De manera importante, este sitio no se centra en el entorno genético o metadato asociado con estructuras de CRISPR, las denominadas CAS (estructuras asociadas a CRISPR). La razón de esto es que la mayoría de las estructuras de CRISPR no están asociadas con genes CAS, de modo que CAS no puede utilizarse como una herramienta de búsqueda de CRISPR. Un segundo aspecto importante es que hemos tratado aquí de proporcionar una herramienta que no pierda incluso las estructuras más pequeñas de CRISPR (que consisten en solo espaciador) y CRISPR pudiera ser ligeramente diferente de estructuras canónicas. Mientras que la identificación de las estructuras grandes es relativamente fácil y directa, y se ha conseguido mediante una diversidad de métodos, la búsqueda de un CRISPR muy pequeño es mucho más ambigua. Como resultado, en la base de datos está contenido un número de CRISPR "cuestionable" muy pequeño. Muchas de estas no son CRISPRs verdaderos y necesitan ser investigados cuidadosamente. Se espera que tales candidatos se investigaran, por ejemplo, tipando múltiples copias de la misma especie. Algunos grandes, si existen, necesitan ser investigados, y la naturaleza del locus de CRISPR sería entonces confirmada.

Descripción del sitio:

Con el fin de abordar este objetivo general, aquí está disponible un cierto número de herramientas. Una de ellas es la propia base de datos CRISPR en la que se han procesado genomas microbianos en búsquedas de estructuras CRISPRs. Esta base de datos puede ser examinada de una manera intuitiva. El segundo objetivo es CRISPR Finder. Puede utilizar esta página para analizar sus propios datos.

FIG. 18

FIG. 19A

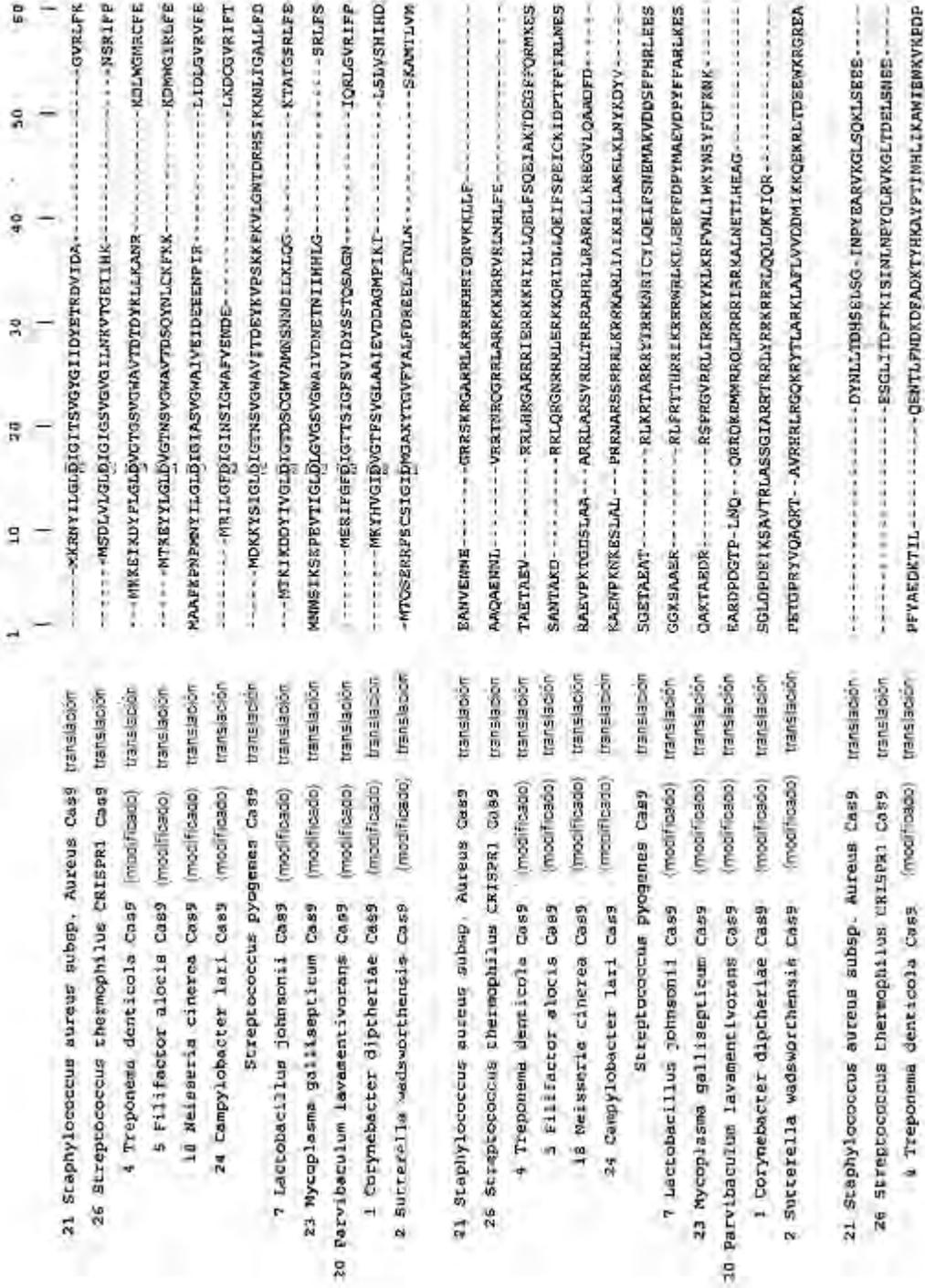


FIG. 19B

3	<i>Filiifactor alocis</i> Cas9 (modificado)	translación	RLALRDKSNQ-----EKYPLFEIKYSDIEYKKEFTTEFLRKHILIESBEKQOI
10	<i>Melisseria cinerea</i> Cas9 (modificado)	translación	-----ENC-----LTKSEPTNMLBAALDRKLTPLB-----
24	<i>Campylobacter lari</i> Cas9 (modificado)	translación	-----ADGRLKAYDGLASVVELRYKALTONBTKD-----
	<i>Streptococcus pyogenes</i> Cas9	translación	FLVREDKK-----HERHPICGIVDQVAVHENVPTIYHLRKEIWDSTQADL
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translación	GLSLPDKRK-----TVSSIVPPTSADKKKRYDQYPTIYHLRYKLTWEDBKFDL
23	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	translación	-----EDTLNHYQSQKELNVTNLKLSDALNAKIDPKA-----
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translación	-----FLPAYGADPVMVADEPYKLRKRRGLESLAVS-----
1	<i>Corynebacter diptheriae</i> Cas9 (modificado)	translación	-----QGWPIZLEDSYDPIYPMKVAELMAASYIADBEROE-
2	<i>Streptococcus pyogenes</i> Cas9 (modificado)	translación	LSGLLKRQYRPNADGDLTEENHIVADVPAHPAFSTVFSVRSLSLABQWREPTANTSR
21	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> Cas9	translación	--FSALLHLAKRGRVYVY-----EVEDTGRKELATKREQIS
25	<i>Streptococcus thermophilus crispus</i> Cas9	translación	--LFTALKVYKIRGISLD-----DASDDGNSVGDYAGLV
4	<i>Treponema denticola</i> Cas9 (modificado)	translación	RLLYLADNLIKKEGHELFEBDFDSK-----QVEFSDAL-FEYLABMEVDIDADSQK-
5	<i>Filiifactor alocis</i> Cas9 (modificado)	translación	RLIYLAHLMIKTRGHFLIDDDIQSAX-----QLRFEILDTE-LLSQEFGNLSVLSSEWCK
18	<i>Melisseria cinerea</i> Cas9 (modificado)	translación	--WSAVLEHLKXHRGYLSQR-----KNBGETADKELGALAKGV
24	<i>Campylobacter lari</i> Cas9 (modificado)	translación	--LARVFLHIAHRGYVM-----KNBKSHDAAKKGALLSAL
	<i>Streptococcus pyogenes</i> Cas9	translación	RLIYLAARMIKPRGHFATIGDLAFDQNSVDKLTGLAVOT-YNOLFEENPINALSGVDAKA
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translación	REVYLAIRHILIKYRGRFLYNTSVKDFKASKIDVRSSTKRNILYRMLGLDLRNVFNSIT
21	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	translación	--LSENIHDVYKXNRGHPYEDR-----RDFNYVPTRELAKZYFDKY
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translación	--PGRALYHLAQRHFKGRB-----LEESDTPDFDVEDRBNV
1	<i>Corynebacter diptheriae</i> Cas9 (modificado)	translación	-KLSVALRHTARWRQWRYAKVSS-----LYLPDQSPDAPKAIIRBEI
2	<i>Streptococcus pyogenes</i> Cas9 (modificado)	translación	VERPLGDPPNTPADKKEFIEFVABG-----LIDRTEKXAYQBALSTL
21	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> Cas9	translación	R-HSKALBEEIYVA-----
26	<i>Streptococcus thermophilus crispus</i> Cas9	translación	XENSKOLETKTNG-----
4	<i>Treponema denticola</i> Cas9 (modificado)	translación	--VKEIIRKSSLLKNSKOSRLWKJLGLK--PSDKOKKAITNLLS-----GNKINPAUL
5	<i>Filiifactor alocis</i> Cas9 (modificado)	translación	DEYVETILMKRS LANSKRYKLLKXNLPETSDLEUBAKKAKQASVAVIENPCPIVQINGDVCNT
18	<i>Melisseria cinerea</i> Cas9 (modificado)	translación	ADVYHALQIQD-----
24	<i>Campylobacter lari</i> Cas9 (modificado)	translación	NNNALKLENYQSYG-----
	<i>Streptococcus pyogenes</i> Cas9	translación	ILSALUSKSRRLRNLTALQDGERKNGLFGMLIALSLGLTWFPSN-----
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translación	AETKLVKDKQTFKRRKXKXITARI,PAIKTNKQSKIKDKISQVAVAVLQVYTRFDITIA
21	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	translación	GYTKGIQSDKEDND-----

FIG. 19C

20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translación	NERAARTLKALKMBO-----
1	<i>Corynebacter diptheriae</i> Cas9 (modificado)	translación	KRASQQVPTATVIG-----
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	translación	RAMANVLTGLRIMG-----
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translación	-----
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translación	-----
4	<i>Treponema denticola</i> Cas9 (modificado)	translación	YUNFDLKDAXKNSISFSKIDFD-ALSDDIARSILDSPFELLEAKAVYNCVLSKVIGD--
5	<i>Filifactor alocis</i> Cas9 (modificado)	translación	LR-VSKBELEIDSFSEKYEEDIVRNLEEKVPEKVVYLFROMKANYDMILVDILET--
18	<i>Neisseria cinerea</i> Cas9 (modificado)	translación	-----
24	<i>Campylobacter lari</i> Cas9 (modificado)	translación	-----
	<i>Streptococcus pyogenes</i> Cas9	translación	-----
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translación	-----
23	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	translación	-----
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translación	-----
1	<i>Corynebacter diptheriae</i> Cas9 (modificado)	translación	-----
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	translación	-----
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translación	-----
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translación	-----
4	<i>Treponema denticola</i> Cas9 (modificado)	translación	-----
5	<i>Filifactor alocis</i> Cas9 (modificado)	translación	-----
18	<i>Neisseria cinerea</i> Cas9 (modificado)	translación	-----
24	<i>Campylobacter lari</i> Cas9 (modificado)	translación	-----
	<i>Streptococcus pyogenes</i> Cas9	translación	-----
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translación	-----
23	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	translación	-----
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translación	-----
1	<i>Corynebacter diptheriae</i> Cas9 (modificado)	translación	-----
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	translación	-----
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translación	-----
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translación	-----
4	<i>Treponema denticola</i> Cas9 (modificado)	translación	-----
5	<i>Filifactor alocis</i> Cas9 (modificado)	translación	-----
18	<i>Neisseria cinerea</i> Cas9 (modificado)	translación	-----
24	<i>Campylobacter lari</i> Cas9 (modificado)	translación	-----
	<i>Streptococcus pyogenes</i> Cas9	translación	-----
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translación	-----
23	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	translación	-----
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translación	-----
1	<i>Corynebacter diptheriae</i> Cas9 (modificado)	translación	-----
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	translación	-----
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translación	-----
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translación	-----

FIG. 19D

1	<i>Treponema denticola</i>	Cas9	(modificado)	translación	CSTVSKKRLIIMHVNQEDFYKELTKILSAKSRIKLVNIDICT-----EISTOTFLPRQISAKS
5	<i>Filifactor aloclisis</i>	Cas9	(modificado)	translación	LAKIKRKKTWIIEKRPBEEFYKSLCKLQIADNIEELKEDLEVLDWKNIEKCKNHTLLEIQNKAD
18	<i>Neisseria cinerea</i>	Cas9	(modificado)	translación	-----NGRGDY
24	<i>Campylobacter lari</i>	Cas9	(modificado)	translación	-----VTKDHY
	Streptococcus pyogenes Cas9				OSKMGVAYIEDGGASQBERFYKELKPLLSNMDSRELLVKEK-----REDLRLKRLKRTFD
7	<i>Lactobacillus johnsonii</i>	Cas9	(modificado)	translación	LLQAKKKGKIKRPSKEDFYKVVNKVLDSBASKEIKKKIELDS-----EMKQKQNTNA
21	<i>Mycoplasma gallisepticum</i>	Cas9	(modificado)	translación	-----LPRKPKSEY
20	<i>Parvibaculum lavamentivorans</i>	Cas9	(modificado)	translación	-----
1	<i>Corynebacter diptheriae</i>	Cas9	(modificado)	translación	-----
2	<i>Sutterella wadsworthensis</i>	Cas9	(modificado)	translación	-----NLSHLQLRAERWYFNAP
21	<i>Staphylococcus aureus</i>	subsp. Aureus	Cas9	translación	IMRFKTSYVVRKAKLKYQKAYH-QLDQS-----PIDDVIDLLEBTRTYVEDQGR
26	<i>Streptococcus thermophilus</i>	CRISPR1	Cas9	translación	IMWFPTSKYRSEALRILIQOQSNPQITDE-----PIMRYLAILTCKRKYVINGKGN
4	<i>Treponema denticola</i>	Cas9	(modificado)	translación	NARTPYQIRKMBLEKILSNMGRHFFSLKQKD---EKGLSHSEKIMLLTFKIPYYIGPIN
5	<i>Filifactor aloclisis</i>	Cas9	(modificado)	translación	MGVLPHQVHEVELKILRNKAKYYSFLTETD---KQGYVWQKTESIFRFRIFYYVQPLS
18	<i>Neisseria cinerea</i>	Cas9	(modificado)	translación	SIFTFNRKDLQRELMKLFKQKQKZFGNPHYSD-----GLKKEGIFLIMATORPAL---
24	<i>Campylobacter lari</i>	Cas9	(modificado)	translación	MKCVLSSDLKEMKILKLEKQKSGYNSGD-----PINEILKVAFYOR-----
	Streptococcus pyogenes Cas9				MGSIPROTHICELHAILKQBDYFPFLKDN-----SEKTEKILTFRIFYYVEDLA
7	<i>Lactobacillus johnsonii</i>	Cas9	(modificado)	translación	XGUTPYOQLQELDKLIEHQSRYYFPIKBITPVSYSHLKEAPKLDLIRFRVYYVGLLI
23	<i>Mycoplasma gallisepticum</i>	Cas9	(modificado)	translación	ESLUSYVYVRYNSIEGSGINSVSPYQIYHLDK-----EGKYVQKINRTWDTKTKCKIFPD
20	<i>Parvibaculum lavamentivorans</i>	Cas9	(modificado)	translación	GIHARRHWVAEERFKUMQVQSNFHPALASE-----MRAKISDYLFAQRVFWB---
1	<i>Corynebacter diptheriae</i>	Cas9	(modificado)	translación	GGVLSABIQDSYASBSIQLSICRMQIEGQELY-----RKLIIDVVPAES-----
2	<i>Sutterella wadsworthensis</i>	Cas9	(modificado)	translación	DTMKDRGWERPKLTVRAFRFPHPAKDNKQHLLELLEKQIENSEDIETLCTLDPWRTI
21	<i>Staphylococcus aureus</i>	subsp. Aureus	Cas9	translación	---GSPFG-----MKDIKHYDMLNGHCHTYPESR
26	<i>Streptococcus thermophilus</i>	CRISPR1	Cas9	translación	EKSRATDYGR-----YRTSGEILDNIFGILIGKACTFPDDE
4	<i>Treponema denticola</i>	Cas9	(modificado)	translación	DPHKKEFFEDRCH---VVKKEKSPSKITPWHFFDRIDKEKTAQAFITSKNTNCTYLWGE
5	<i>Filifactor aloclisis</i>	Cas9	(modificado)	translación	TRHQEKGSN-VN---MVRKP-GREDRIYPMWBEIIDFEKSNENFTIRMTNKTLYLIGE
18	<i>Neisseria cinerea</i>	Cas9	(modificado)	translación	-----SGDAVQRMLGHCTEFPTE
24	<i>Campylobacter lari</i>	Cas9	(modificado)	translación	-----PLXDPSHUIVGACTFPESR
	Streptococcus pyogenes Cas9				AGNSRFAMFTR-----KSEETIPMNFBEVVDKGSAAQSFTIRMTNFDKRLDNE
7	<i>Lactobacillus johnsonii</i>	Cas9	(modificado)	translación	SPNESTKDIQTKKQKQNFAMNIRKESGRITDWNFDQKYVDRISSAMKFKIRVITTKDTLPLGE

FIG. 19E

23	<i>Mycoplasma gallisepticum</i> Cas9	(modificado)	translación-INMFHEINELSTIRSYIYLQW
20	<i>Parvibaculum lavamentivorans</i> Cas3	(modificado)	translación+KNTLRECRFMPGG
1	<i>Corynebacter diptheriae</i> Cas9	(modificado)	translación-PKCSASSRVGKQHPLOPGK
2	<i>Sutterella wadsworthensis</i> Cas9	(modificado)	translaciónPVPYEQWNRREP.....DDOTULLSPKLTQVGEIKVTSARUTSAGFTLAFPAEILER
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translación	translación	IESVRYAVNADLYMADLMLVITRDENSEKLEYEKFOIEEMVFRKQKK---PTLAKLIA
26	<i>Streptococcus thermophilus</i> CRISPRI Cas9	translación	translación	FRAAKASVTAQRFKGLNDGAMVTF--TEFKKLSKEQKNQIINYVAKENAKMGAFLKFXVIA
4	<i>Treponema denticola</i> Cas9	(modificado)	translación	SVLPKSSLLYSEYVVLHBEINVLQIEIDKRNICDINKIKQIYEDLFRKRYK---ITQKQIS
5	<i>Fillifactor alocis</i> Cas9	(modificado)	translación	DVLPKNSLLYSKYMVLHQLN--NVKVRGKLP--TSLKQKVFEDLFEKSK---VTGKNDL
18	<i>Neisseria cinerea</i> Cas9	(modificado)	translación	PRANKTTIAERFVWLFSLM--MRLIDGSSERPLDTRATJMDPEYKSKLTYAQARK
24	<i>Campylobacter lari</i> Cas9	(modificado)	translación	KDACKNSYSAWEPVALTKIIG-EIKSLKRTSGEIVFTQTNELVMLTIDKGSITVYKFRS
	<i>Streptococcus pyogenes</i> Cas9	translación	translación	KVLPKHSLLIYFTVYVHLKAVYVTECHKRPAPLSCQKKAIVDLPYTRKRVTVKQLK
7	<i>Lactobacillus johnsonii</i> Cas9	(modificado)	translación	DVLPANSLLYKFTVVLHQLN---IRINGKRSIVDLKQEIYEMPKRHTT----VTVYKLL
23	<i>Mycoplasma gallisepticum</i> Cas9	(modificado)	translación	FINDGFKAYLKNKLDLILLIKTWESKPIDBQFKLREPTIASSIGKFLKQVEMEEKERK
20	<i>Parvibaculum lavamentivorans</i> Cas3	(modificado)	translación	PLCPKGSNLSGQRSMLEKLNMLAAGNAGPLDARERKQNTLSKLDQDASMEWPEVRSALK
1	<i>Corynebacter diptheriae</i> Cas9	(modificado)	translación	HELLKASDAPQRYRTAALIGLVAVVDGKRTILSYSEKMLYFDHLVHLTPKSEPRWVPIA
2	<i>Sutterella wadsworthensis</i> Cas9	(modificado)	translación	ETDRKSRVAVVGNHSPLEPLAVQ--SYALQMAFDNSKALQRYALBALAGSKSNKUTSARTKTA
24	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translación	translación	KRIILVNEDEKRG--YRVT--STGKPSPTMLAVYHDIQD--ITA--RKEIIEAELLEQDI
26	<i>Streptococcus thermophilus</i> CRISPRI Cas9	translación	translación	KLLSCQVADIRKG--YRID--KSKGASLHTFPAVTRMKT--LET--LDIQQMDRSTLQKL
4	<i>Treponema denticola</i> Cas9	(modificado)	translación	TEFKHGGICRKS--DEVIILGIDNCSFTSLSKSYIELKX-IPGK--QVDEISTKMLBEI
5	<i>Fillifactor alocis</i> Cas9	(modificado)	translación	STLQTDKRDIDT--DQLS--GEDKDFKTSLSYLPFKKQIFGE--ETEKESIQRMTEDI
18	<i>Neisseria cinerea</i> Cas9	(modificado)	translación	LLDLIDTPFPKQ---LRYGKDAEASTIMEMKAYHAIKSR-ALEK--EQLKDKKSPMLSP
24	<i>Campylobacter lari</i> Cas9	(modificado)	translación	CINLHESISPKS--LKYDRENESYAKLIDFKKLVFPKX-AIG-----VHSLSR
	<i>Streptococcus pyogenes</i> Cas9	translación	translación	EDYFKKIECPDS--VNTS--GVEDRFPASLCTPHDLKXIKDKD-FLDNEENEDILDDI
7	<i>Lactobacillus johnsonii</i> Cas9	(modificado)	translación	ENYLKERNHLVK---VBIKGLADEKRFNSGICITTYNFRMLITEFN-QIDDLKXRNDFEKI
23	<i>Mycoplasma gallisepticum</i> Cas9	(modificado)	translación	EDHKWKLKGLKLTWYSKIIQYMDLUSLAKPVIKLLKQHLKADFLLED-QYATLDKINPLQSL
20	<i>Parvibaculum lavamentivorans</i> Cas9	(modificado)	translación	ALYKORGEPGAKK-LAPRIELGSKLGNALBAKAKADMPSPDWPFAHPRKQEIIEAVHIE
1	<i>Corynebacter diptheriae</i> Cas9	(modificado)	translación	BILGIDRISQLIG-----TATHTDQDSGACARPPHTDWRSEVNSIAPLVQWVKYASAL
2	<i>Sutterella wadsworthensis</i> Cas9	(modificado)	translación	LKNCIGGQWVKT--FLDCARVYREADDNAKVGLEWFDWADGGLERSDLIIPPKKKILFLVY
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translación	translación	AKIITTYO--SSEMIQSELTW--LHSELTOERTQISNKGVT-----

FIG. 19F

25	<i>Streptococcus thermophilus</i>	CRISP1_Cas9	translación	AVVLTLP--ERAGTORALRERADRSFSQKQVDRVLAVTRKANS-----	
4	<i>Treponema denticola</i>	Cas9 (modificado)	translación	IRWATYDQEQKTIKTKIKAEYGYCSDIOIKILNLXFSGWTRLSRK-----	
5	Fillifactor	alocis_Cas9 (modificado)	translación	TKWITY--SNDXELKRVTRASYSNDLTFBQMKKITGFOYSCWGNFSRM-----	
18	<i>Neisseria cinerea</i>	Cas9 (modificado)	translación	ELQOEGYAFELFKTDEDITGRUKOKRQVREILKALKXHSFDXP-----	
34	<i>Campylobacter lari</i>	Cas9 (modificado)	translación	QELQDISTRHUTTRWVKLVLEKYNLSNEQINHLIEIFBNQV-----	
	<i>Streptococcus pyogenes</i>	Cas9	translación	VLTLLTFE----DREMTESLKYVAHFQDKVMKQIKRRRYGWRLSRK-----	
7	<i>Lactobacillus johnsonii</i>	Cas9 (modificado)	translación	IEWSTIFE----DKSIYKENLSIDWLNKQTHALSMTIRLQWQ-----	
23	<i>Mycoplasma gallisepticum</i>	Cas9 (modificado)	translación	EYVLGKHLRYSNVDSANLAEFSDSKLFTENLQKQKQKQGLFKLFEQTRKDD-----	
36	<i>Parvibaculum lavamentivorans</i>	Cas9 (modificado)	translación	RLWADYGETPOKKRVILISEKDKAKHRAKAAKSEVADPOITGEOAQLQA-----	
1	<i>Corynebacter diphtheriae</i>	Cas9 (modificado)	translación	ROHNVKALSNABVEDFDSFGAKVQAPFADLDDDDVHAKLDSLHP-----	
2	<i>Sutterella wadsworthensis</i>	Cas9 (modificado)	translación	ANILQDETQOKFIDRIRKQIKGRETVAERCAURTSVRKSKGGGFIAYNTAQYRNVN-----	
21	<i>Staphylococcus aureus</i>	subsp. <i>Aureus</i>	Cas9	translación	-----GTRRLSKAINLIQDEGHTHD-----NQIALEN-----
25	<i>Streptococcus thermophilus</i>	CRISP1_Cas9	translación	-----SIFQKQWHFYSVKLMLKELIPELPEYFSE-----EQMTIIT-----	
4	<i>Treponema denticola</i>	Cas9 (modificado)	translación	-----FLRTVTSERSEFSBPVMIITAMRETON-----NDMLRLLSSEFTFTB-----	
5	Fillifactor	alocis_Cas9 (modificado)	translación	-----FLKISGSDVSTGETPDIITAMMETOR-----KLMQILSKKRETPHD-----	
18	<i>Neisseria cinerea</i>	Cas9 (modificado)	translación	-----VQISLKAERIVPLMDSQGRYDE-----ACTEYIG-----	
24	<i>Campylobacter lari</i>	Cas9 (modificado)	translación	-----INLSFVALGIMILPLMRGSKRYDE-----ACEINLKLK-----	
	<i>Streptococcus pyogenes</i>	Cas9	translación	-----LINGERHOKSKRTIILDFLSDSEANRN-----FMQLIHDS-----L-----	
7	<i>Lactobacillus johnsonii</i>	Cas9 (modificado)	translación	-----RLSKKLLALDRDENGQTTIEOLMDSQV-----NPMQYVTCQAB-----	
23	<i>Mycoplasma gallisepticum</i>	Cas9 (modificado)	translación	-----EKILAQTHSLSTKMLLAETRMVLDGDEDMQKNDKQNNKHFALKMFQDKKFTD-----	
36	<i>Parvibaculum lavamentivorans</i>	Cas9 (modificado)	translación	-----LKLPTGWEYYSIPALNLPFLAELEKSERFG-----ALVNGSDWISGMRKTFEPR-----	
1	<i>Corynebacter diphtheriae</i>	Cas9 (modificado)	translación	-----VGRNAVSEDTLYRITRRMLSDQVD-----	
2	<i>Sutterella wadsworthensis</i>	Cas9 (modificado)	translación	KLPRNAGKELLTTRDRVAETAPRTLAANLIGLSDGKSKFANPFSLQDFYTLIETEVSGPS-----	
21	<i>Staphylococcus aureus</i>	subsp. <i>Aureus</i>	Cas9	translación	RLKLVPKVLDLQGNKRIPTLVDDFELSPVVKRSPFIQSKRVINAIKKYQ---LPMQIIL-----
25	<i>Streptococcus thermophilus</i>	CRISP1_Cas9	translación	RLGKQATTSNSNKKYIDKLLTETRYDPVAKSVKQAKRTVRAIAIKNEYQ---DFQWIVL-----	
4	<i>Treponema denticola</i>	Cas9 (modificado)	translación	NWIKTHGFFDAEKQFSDGLVYKPLTLPSPYVKKILMQLKLVKRETSHTQY---PPKKIPI-----	
5	Fillifactor	alocis_Cas9 (modificado)	translación	NVEDPNSGKVGKIDKITYDSTVWEMPLSPEKBAWQTYQVAREYKAVGCC---EPKKIPI-----	
18	<i>Neisseria cinerea</i>	Cas9 (modificado)	translación	---DHVGNKATEEKIYLPPIPADRTRRPVLVLAQSQAQKRVINGVYVRYG---SPAKIHT-----	
34	<i>Campylobacter lari</i>	Cas9 (modificado)	translación	---PKTYDEKQDELPAFPQPSIYAHUSNRYVYMBALSRYRKLVMALLKLYG---KVRKIHU-----	
	<i>Streptococcus pyogenes</i>	Cas9	translación	YPRKEDIQKAVSQGKDSLHHTANLAGSFAFKKQTLQTVKVVDSLVKVMGRH--KPRKIVL-----	

FIG. 19G

7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	transición	FKDAIARAKNDILVATSVEDILNNAITSPANKKAIQVIRKVVDDI VKNASGK - VPKDIAI
23	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	transición	YTKXNHMELSKKKKAYLDRPTMDAILSPGKRIILRENTKVFPAIKQPSRBYDVTKVVI
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	transición	HQPTGEILDKLPSPKXERERISQLRNPTVVRTQNRKRVVKNLIGLYC --- KDRERI
1	<i>Corynebacterium diptheriae</i> Cas9 (modificado)	transición	LYTARLQHPGI EPSWTPTFRIGEPVQFAVDRVKTVSRMLSEAKTKWG --- APERVII
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	transición	ATTLVAEENAHMPTIKOAVINGETVRAQACSLCPAETASFPDGLVRRLLVDR -QANVEIAK
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9 (modificado)	transición	ELAR-----EKSKD -ACKINIMQKRNQVWERIEETIRTTGEM-----ASY
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9 (modificado)	transición	EMAR-----ETREDD -EKAIQIQLIMK - DEKDAAMIKAANOYNGKELPHSPVPHRQ
4	<i>Treponema denticola</i> Cas9 (modificado)	transición	EMAKG - ELSPAR -TKVRLKILDGLYNNCKNDADAFSEIKDLGKRIHG --- DNLLAL
5	<i>S. factor</i> <i>alocis</i> Cas9 (modificado)	transición	EMARQ --- EKVKR -TKSRKQALLLEIAACREDCEITKBTGR --- DERDP
18	<i>Neisseria cinerea</i> Cas9 (modificado)	transición	ETAREV --- GSKFD -RKEIETKQESNRKDRKESAKKFBYFPFVVG -----EEXEK
24	<i>Campylobacter lari</i> Cas9 (modificado)	transición	ELADV --- GSKKA -REKIEKQEN --- QVNHMLKCECEIGL -----KASAK
Streptococcus pyogenes Cas9			
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	transición	EMAEEN --- QTTQK -QKHSREMKRIEERIKEGSQLKEHEVER -----TQL
23	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	transición	EPTDRA --- DEMPKR -SOTRGSKLOKLVKDLSTELASKTIAELNERTK -----DKKL
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	transición	ELARELSEKLEEN -TKYKXITKNSCKISEGLKALGISEDEIKD -----ILKSPY
1	<i>Corynebacterium diptheriae</i> Cas9 (modificado)	transición	EWCRWG --- ESKRE -RBEIQSIRANEKORKKATDILKNGIANPS -----RD
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	transición	EHVRE - EVTEKR -AREMDCMHRRAABNAKULFORMQEKLVQCK -----PSBA
Streptococcus pyogenes Cas9			
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9 (modificado)	transición	RVSTDTQSKAVDFSRGIVDVSIFVEENKFPFSAVADLKKKRYKDKNLS --EAEKLETRW
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9 (modificado)	transición	LIEIXLHDMQKGCYSLALPLEDLNMPFVVEVDRIIPRSVSPNS -----FNNKVLV
4	<i>Treponema denticola</i> Cas9 (modificado)	transición	LATKIRLAKQGERCLYTKTISTHDAIMNSQPEVDIPLSLITFDS --- LANKVLV
5	<i>S. factor</i> <i>alocis</i> Cas9 (modificado)	transición	RSKXLYLVTOLSKCHVCGKFEIGHVFD -TSNYVDIITYPQSKIKDS --- ISNRVLV
18	<i>Neisseria cinerea</i> Cas9 (modificado)	transición	NSMMLFLYTOPCKMWSGDDIDINELRNSKWRDRIIYQSKIKDS --- IDMLVLV
24	<i>Campylobacter lari</i> Cas9 (modificado)	transición	DIKLRLVQGHKCKLYSKXKINLQRLNE -KGVEIDQLALPFRSTWDS --- FNNKVLV
Streptococcus pyogenes Cas9			
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	transición	NIDKLLAKRQKEICLYSNKISIEHLKD -EKALVVDIIVPSRSFDS -----FINKVLV
23	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	transición	QNRKLYLVQNGRDMYVQELNRLSD --- YDVIDIIVPOSFLKDS --- IDNRVLT
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	transición	VQDKYLVFMQLGRDAYTQEPINDEIQX --- YDIDITLPOSFIKDA --- LDWRVLV
1	<i>Corynebacterium diptheriae</i> Cas9 (modificado)	transición	KSYFPLMTQQDRIIDPYSLEIAFDPTTKETFDIITIPYSISFDS --- SSNKLVV
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	transición	DVBEILNREGERCPYTCDDIIGFALPR -EGRYVEEIMPRSRSPNS --- PRNKTLC
Streptococcus pyogenes Cas9			
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9 (modificado)	transición	DLRWYSQRKQCCACVCSPTFS -----NSRMNIVPRAGQESTN --- TRZVLVA
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9 (modificado)	transición	IINNRKXASRGTCTPTCDRLANGS -----EIDITLPRSEIEDARDIVNAPPELI

FIG. 19H

21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translaci3n	KOBNSKQKQRTPPZYLSSDGKISZETPKHIIINLA-KGKGRISKTKKBYELLEBRDINF
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translaci3n	YATANGKQQRTPYQALSDMDAWSERELKAFVR---EKKTLJSHKKKEXLITTEEDISK
4	<i>Treponema denticola</i> Cas9 (modificado)	translaci3n	CSSCHNKKSDKYPFKSEIQSGKQGNFNFLQ-----KSNFISLEKUNLITRATYISD
5	<i>Pfllifactor aloclis</i> Cas9 (modificado)	translaci3n	YKTYNAKSNW-ILSSEDLQKZNSFPUSLIL-----NKKLITKSKYDRUTRKGDPFTD
18	<i>Neisseria cinerea</i> Cas9 (modificado)	translaci3n	LGSENGKGNQTPYBYFNGKDNHRENDCEFPAR-----VETSRFPKSKQRILQK--PD
24	<i>Campylobacter lari</i> Cas9 (modificado)	translaci3n	PTKNSDEKLKTPPRAPG--KNLEKSNLQIL-----AQN--LPHYKKNKILDEN--FK
	<i>Streptococcus pyogenes</i> Cas9	translaci3n	RSDKNKSKSDRYDSEVYVCKKRYKQJLJL-----AKLITQRKFDNLTNAERKGL
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translaci3n	SPAYHNSKSDMPYVKLFCSTENAAKNGHTIRKQWEE--KNTGLISKTKYNNLITDPDHLN
83	<i>Mycoplasm gallisepticum</i> Cas9 (modificado)	translaci3n	LARSNQAKSNQTPYEFISSGNAGTKMEDYEAICHAFKDGSSLSLSDTQSKKFAKNNKTD
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translaci3n	AKVNHLEKGNRPFRAPDIDSDKWSLQIQR-----LQGNVSAKGGIGMSPGKVAKE
1	<i>Corynebacter diphtheriae</i> Cas9 (modificado)	translaci3n	WCHRCKOSKHTFPRAIMAKNTSISQSYVKEVETTERHWYTDCHRSSTDYKXPTKAWERE
2	<i>Succarella wadsworthensis</i> Cas9 (modificado)	translaci3n	YASRGNQLKKNQRYLSDLKSNYRMSIPTSTNIAAITAEIEDVVTLQOQTRHLAFDOLL
23	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translaci3n	F8VQKD--FIRNLVDTRYATRGMLNLSRYFVNRLD-----
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translaci3n	FQVMKK--FIERNLVDTRYASRVVLSNALQERERAIKID-----
4	<i>Treponema denticola</i> Cas9 (modificado)	translaci3n	DETAK---PIARQLVETRQATKYAAKVEKMPETRIY-----
5	<i>Pfllifactor aloclis</i> Cas9 (modificado)	translaci3n	SELG---FIARQLVETRQATKYAIDPKIY--SSIVY-----
18	<i>Neisseria cinerea</i> Cas9 (modificado)	translaci3n	E---DG--FKERNLADTRYINTFCQFYADHMLLTKG-----
24	<i>Campylobacter lari</i> Cas9 (modificado)	translaci3n	DKQGED--FISRNLDTPYIATLAKYTKELKFLLSSEMMAL-----
	<i>Streptococcus pyogenes</i> Cas9	translaci3n	SELDKAG--FIKRLVETRQITKHWAOILDSRNWTKYDENQK-----
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translaci3n	KYASAG--FIKRLVETISQILKLVSTILQSRYPNTEIL-----
23	<i>Mycoplasm gallisepticum</i> Cas9 (modificado)	translaci3n	TSSKYDIDGLARINIDTRYATIIVFDALQEDYAKNHELVEDKP-----
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translaci3n	LAKTNPEDFAARQLADTRYAAKOTLADUKRLMPDQWGFENP-----
1	<i>Corynebacter diphtheriae</i> Cas9 (modificado)	translaci3n	QKATNDEEDIDAKSSVAMAMKELRERVQHFASHGCTT-----
2	<i>Succarella wadsworthensis</i> Cas9 (modificado)	translaci3n	NEHEDQCVSRHALFDQDQSEADAVLGLLQATQRTKVTQIDRIKPLAUKTPELEIQNCKX
21	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9	translaci3n	-----VKYKSLNGEFTSFLRR--KWKFKX-----SKMKYKHTAEDALIANAD
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translaci3n	-----TKVSVVRQGFSQLRR---RWGTEK-----TRDT-YHHHNDALIIAASE
4	<i>Treponema denticola</i> Cas9 (modificado)	translaci3n	-----YSKATVSMFRNK-----FDIVK-----CREINDFHHRHDAYLKIYVG
5	<i>Pfllifactor aloclis</i> Cas9 (modificado)	translaci3n	-----YVKSLSVSDPKKXK-----LNYLX-----SRVNDYIDKADAYLNFVVC
18	<i>Neisseria cinerea</i> Cas9 (modificado)	translaci3n	-----KRVYFASNGQITNLDRG---RWGLRK-----VHASNDRHNLQDNYVYACST
24	<i>Campylobacter lari</i> Cas9 (modificado)	translaci3n	-----KSBEGKSKIEHWGTTSGMLTSYLRN---TWGFDK-----KDRNHLNHALQALIVAYST

FIG. 19I

	Streptococcus pyogenes Cas9	translación	-----LIREVRYVTLTKSLKLSVDFPK----DFDFYK-----VREINWVIMHAIPIADYIMAVVE
7	Lactobacillus johnsonii Cas9 (modificado)	translación	-----TVKAKYHHILIE---SEFDYK-----SREYNDYHRIIDAYLEIICC
23	Mycoplasma gallisepticum Cas9 (modificado)	translación	-----MFKVVCINGSVTSFLKXNPIHSSYAK-----KDRDANIRHVAIDASTIISIPS
20	Parvibaculum lavamentivorans Cas9 (modificado)	translación	-----VKKAVTQVMTACLKLEKLTMLRILADOC---EKTRAGHRAHRAIDAUTVACTH
1	Corynebacterief diptheriae Cas9 (modificado)	translación	-----VRYVROSLHTKRAHRASGISCKLKFDFDVGKSRDRDRIBULADAAVIAFTS
3	Sutterella wadsworthensis Cas9 (modificado)	translación	TTWNR.HIQAATIVSDAMKRELKLAQHQDFEK-----POIQSIASHSIDALCSFANG
21	Staphylococcus aureus subsp. aureus Cas9	translación	FIFENKKLILAKKAKK-----EMQWFERKQASMEELIST-EGEYKE
26	Streptococcus thermophilus CRISPR1 Cas9	translación	QLNLWKQKINTLVSYIS-----EQQJLADIETGELISDDEYKESYPKA
4	Treponema denticola Cas9 (modificado)	translación	WVXKTKETMPNPNZIK-----EKRONPKIADTYNYVYKVFYDVKRW
5	Pilifactor allocis Cas9 (modificado)	translación	WYWNKKEFSRTIQMK-----KWRD-----TMSLNKKVZEHQVIVN
18	Neisseria cinerea Cas9 (modificado)	translación	IANQQKIS-REVTYKE-----WVAFDQATIDKSTGRVILHOKALIPPO
24	Campylobacter lari Cas9 (modificado)	translación	NSIKAFS-DFKQNGE-----LLEAK-----
	Streptococcus pyogenes Cas9	translación	TALIKKYPKLSSEFVYGDYK-----VYDVAWHTAKSDQSIGKATKAYEZYSTNIRKFFRTE
7	Lactobacillus johnsonii Cas9 (modificado)	translación	MELYQYDMLRIRFPYK-----DYKKSDDPKDKALPKNTRKFSFI
23	Mycoplasma gallisepticum Cas9 (modificado)	translación	WPKTLFRQUTOPADYKLFK-----HTDGSWKKLDFKTVYVTEIDEMKACIRVRKQUSE
20	Parvibaculum lavamentivorans Cas9 (modificado)	translación	PGNTRKLSRYHQLADD-----
1	Corynebacterief diptheriae Cas9 (modificado)	translación	DYVAVETLAVRSELRQEQ-----AHRQAPQWRFEFTGKDAHRRAARV
3	Sutterella wadsworthensis Cas9 (modificado)	translación	SADAERDQNGFDYDQKTVLGHLYPQSESVIRLQAKPQREKSNFDSVAIFKGGIYAEQLP
21	Staphylococcus aureus subsp. aureus Cas9	translación	IFTPHQIHTKDFKD-KYSHRVDDKREHLELDEIYSTR-----KDRKNTLIVNNLM
26	Streptococcus thermophilus CRISPR1 Cas9	translación	PYQHFVDLKSKEFEDSILFSYQASKFNKLSIDATIYATROAKVQKDKADFTYVLSKTK
4	Treponema denticola Cas9 (modificado)	translación	WIDAMPK-----CKTIITVYKQKLEKATPIYTRQAKCKAGGEGFRCTIMKQDL
5	Pilifactor allocis Cas9 (modificado)	translación	GEVIVKERTYHEDTNT--LVDQSTLWRIRKIVERDNIYATEYAVCTKGLFNATIQNKG
18	Neisseria cinerea Cas9 (modificado)	translación	PWEFFAQSVNLRVFGK-PDQKPSFEADTPSKJRTLLASKLSEKSEVAVHKVYVTPLFISR
24	Campylobacter lari Cas9 (modificado)	translación	---FYAKSLTSDNY-----XHQVYKFFPKSFRKILSRIDEIYVSK
	Streptococcus pyogenes Cas9	translación	ITLANGEETKSPULSTNGETGSIWMDKGRDFATVRKVLSPQVWIVKKTLYQGGFSKES
7	Lactobacillus johnsonii Cas9 (modificado)	translación	SOLAKKSEKSEKLEAK-----SLKLRAYKQFYKQYSRETETRDQEMFKYVYVPRFS
23	Mycoplasma gallisepticum Cas9 (modificado)	translación	IADVTBKYGQSNIBHKAVYSHKISNKTNLSLFDNVTVYSKAKVGYEDQIKRKLKTLDDH
20	Parvibaculum lavamentivorans Cas9 (modificado)	translación	-----PRASKPALTEPMDITRADAEEKAVSEIWSH
1	Corynebacterief diptheriae Cas9 (modificado)	translación	WQQRKLSALLTEDLR-----IDRWVVMNRVILRGGNGSARKFTTKLSKVKLSS
3	Sutterella wadsworthensis Cas9 (modificado)	translación	ZFTLAKETWIGYETLQAKGERCCDIAZVSKQKPELJLMLIAFFENKPVGDLASAHATVRIK

FIG. 19J

21	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> Cas9	translación	GLYRD-IDIKLKLKMSPEKLLMVRHDP-----QTYQK-LALL
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translación	DLXTQGGYDAEMKLYKDKSAPLVRHDP-----QTFEKVLEPI
4	<i>Treponema denticola</i> Cas9 (modificado)	translación	GGHDLUKKGGPESNLSKYG-----YNAKVS-----AAYVTLEIYE
5	<i>Filifactor aloclis</i> Cas9 (modificado)	translación	KSTVSLKXG--LDVKKKGG-----YFSAM-----TSYPSLEIPE
14	<i>Weissella cinerea</i> Cas9 (modificado)	translación	APNRKMSGGHNETVKSAAKALDEGISYLA-----VPLTOLAKLD
24	<i>Campylobacter lari</i> Cas9 (modificado)	translación	PPRKRARALAKYDTPHSENKIIDKCSVN-----VAYSVLVVAK
	<i>Streptococcus pyogenes</i> Cas9	translación	ILPKRMSDKLTARKKDWDPKAYGGDSPT-----DAYWIVRID
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translación	HDTVKAPRNLIPKKGMSPDYGGYTKNS-----DAWVIVRID
21	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	translación	SSAKENKSKVYRQPVYRKLWVSLANDKLAIDLFAEKKDIWVYRNPWVWIKLAEQIEME
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translación	KVRKKVSDPLHKETTYGDTGTDIKTKSGT-----
1	<i>Corynebacter diphtheriae</i> Cas9 (modificado)	translación	QLSVSDIDKASSEALNCALTRPSPDPKE-----
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	translación	KPAYEFLAKALOPLSABRFLAALLDALVCTSRKSLMSLFPYRANGKSKKRDVWLKPK
21	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> Cas9	translación	MBOYGD-----EK-----NPELYYVETQHYVTKYSKKDKNGPVIKIKIYGNLMAHLDI
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translación	LBNYPKQINERKGEVPCNPFKLYKBERB-YIRKYSKKGSPETIKSLKYDSDKLGRRHIDI
4	<i>Treponema denticola</i> Cas9 (modificado)	translación	EK-QMKIRSLGTFIP-----LYLVKDIQNDQVILAKSYFLDMLGKKEFKILVPEKIKINSLLKI
5	<i>Filifactor aloclis</i> Cas9 (modificado)	translación	DKKEDRAHHITGVP-----IYIANMLESPSAFLEYCF-QXGYGNVRIQVEKIKNSILLI
14	<i>Weissella cinerea</i> Cas9 (modificado)	translación	LEKMWNRERKLYEALKARIEAKHDDPAKAFAPFYKYDKAGNRTOOVFAVVEQVQKTI
24	<i>Campylobacter lari</i> Cas9 (modificado)	translación	-----SKBGLQIALSCGRVKKI
	<i>Streptococcus pyogenes</i> Cas9	translación	VEKGSKKA-----LKAVKELUGITIMERSSEKKNPFDLFRANGYKVEKQDLJLKL
7	<i>Lactobacillus johnsonii</i> Cas9 (modificado)	translación	KKKGTSYKLLQ-----IPTRELVLKKAASKEDHYAKSLKELUTPRILYNKNGKDKKI
23	<i>Mycoplasma gallisepticum</i> Cas9 (modificado)	translación	YTBKKLKSQWVPEKYHLDTKEPPEPSEFLVKSMLRNKTAIIYDQKKNIVHRTKRIEM
20	<i>Parvibaculum lavamentivorans</i> Cas9 (modificado)	translación	-----YRQPVTRKKIETSLSKGELQSIIRDPAIKBI
1	<i>Corynebacter diphtheriae</i> Cas9 (modificado)	translación	-----GLIPEAMPBRHIVNCTHVYAGDNIQLP PVSAGSIALRGGYAE
2	<i>Sutterella wadsworthensis</i> Cas9 (modificado)	translación	LFQLKLELKEKSTFLKMSLTLPVKQDWLALICDSFLADAPGKPCSADSLTSLARITWR
21	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> Cas9	translación	TDDYPSRBNXVVKLSLAPYRFDVIYLDN--GVYKPVTVVKMLDVIKKE-----
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9	translación	TPK--DSNNXVVLQSVSDWADVYPRKTSKYELIGLKYADLQFENGT-----
4	<i>Treponema denticola</i> Cas9 (modificado)	translación	NGFPCHITGKTNDSEFLKRPVQPCSSMNEVLYEKIIRPSEIKSQREKIGKTIISPVEDLS
5	<i>Filifactor aloclis</i> Cas9 (modificado)	translación	NGYFLIRGENEVDTSFKRAIQKLQDKKNYELVANEISKFLBKVVEK-----
14	<i>Weissella cinerea</i> Cas9 (modificado)	translación	QWVYHNINGIADNATIVRIVDVEKGGKYVLRVPSYQVARGILPDR-----

FIG. 19K

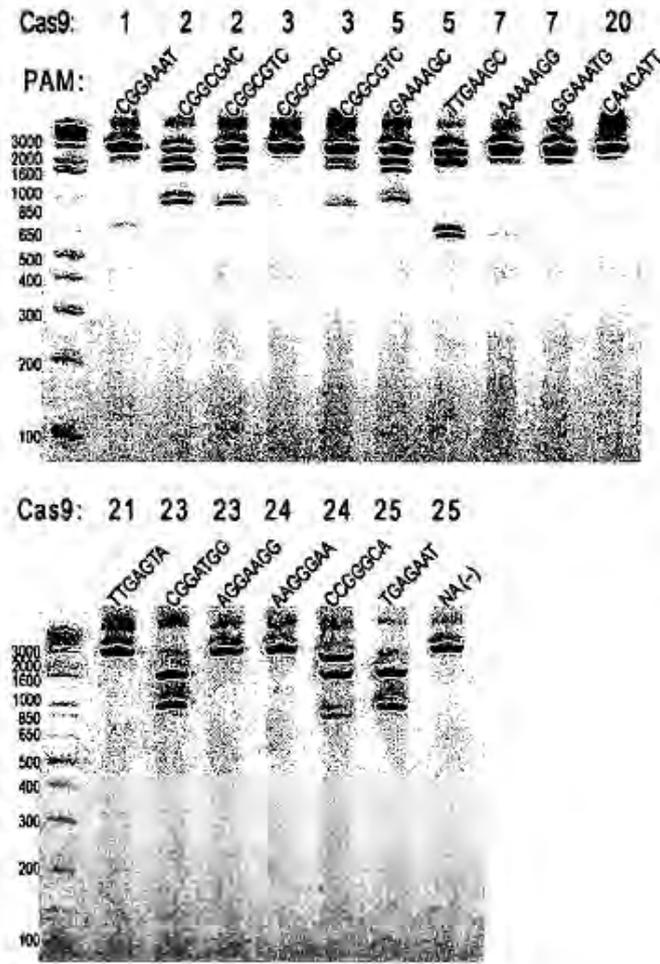
24	<i>Campylobacter lari</i> Cas9	(modificado)	translación	Q-----TKYVEHDTIVVDFIKOMKPKVAIPYVMDPFLGILPMK-----
	<i>Streptococcus pyogenes</i> Cas9		translación	RYSLFELERORXKMLASAGBLQKGNELALPSKYWNFLYLASHYEK-----
7	<i>Lactobacillus johnsonii</i> Cas9	(modificado)	translación	TSPEYVKKTYKYQYIQGGDKKPKHLSSYYVYRACKOULTUSYEMKA-----
23	<i>Mycoplasma gallisepticum</i> Cas9	(modificado)	translación	LSSELUKXKLSNVIIRSKNOSYKLSYQDTEINSEALMIMRSIDPTAKKQYEVVPAK----
20	<i>Parvibaculum lavamentivorans</i> Cas9	(modificado)	translación	VAAGVARGSDPKKAPFPYPCVSRGGCFEAKVELTSKQOQNTMAGTQ-----
	<i>Corynebacter diphtheriae</i> Cas9	(modificado)	translación	LCSSPHIARVYKITSGRKPAFAMLRUYTIDLAPYRMDLFSVELXP-----
	<i>Sutterella wadsworthensis</i> Cas9	(modificado)	translación	PVMDLAHAPVRRBFSLPAIDNPSGGPRAIRTNLFGNELVYVRAINAKNKYCFASNG-----
31	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9		translación	-----MYVAVSKCVERAKKUKKISNOAEFLASFYVNDLIXNG-----GLYRVYIG-VN
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9		translación	-----GTVKISQKYNIDIKKKEGVDSDFEFKFLYKNDLALAKDTSTKBOGDFRFLSRTH
4	<i>Treponema denticola</i> Cas9	(modificado)	translación	FRSYIKENLWKTANDRIGK&PFDLLQKKNLEIYDMLTSEKDTYKRRPNSAYIDILV
5	<i>Filifactor alocis</i> Cas9	(modificado)	translación	-----KGNYPIDENRDAIT-----IERKHQLYEVLISRMKK..FNKKGNAQDSDRIE
18	<i>Neisseria cinerea</i> Cas9	(modificado)	translación	-----AVVOGKDERD-----KTYMIDQSPFKFVLYANLILKUTAK--KNEFEGYFVSLNR
24	<i>Campylobacter lari</i> Cas9	(modificado)	translación	-----IIVTCKDNNPKQKQFIDESYRPFPSLYKNDLILLQKKNMOEPEFAVYNDPFI
	<i>Streptococcus pyogenes</i> Cas9		translación	-----LRGSPEDNEQQLFVQMKHYLQELIIEQISEPFSKRYILADANLQVLSAYM
7	<i>Lactobacillus johnsonii</i> Cas9	(modificado)	translación	-----ITWNFDRSDENDALISATDIBLQVDPKYLPLPFDIN-----KFRBKLIH
23	<i>Mycoplasma gallisepticum</i> Cas9	(modificado)	translación	-----TLNLHUGDRDFDGHMDAYLAKPKPVYLYKAKHIGDEYKPRVLTSTQTLIHKKD
20	<i>Parvibaculum lavamentivorans</i> Cas9	(modificado)	translación	-----KCYADLQGSNHHIATYRUPDGNADPELWSLPDASERLQGR---NPIVQIRADGGA
	<i>Corynebacter diphtheriae</i> Cas9	(modificado)	translación	-----QTMSHRQAEKKLADALATGNAYVIGWLVYDDLVYDTSKLIATQVQKVAEAEI
	<i>Sutterella wadsworthensis</i> Cas9	(modificado)	translación	-----SNVDHSGILLFNEIQHENLTGCGRPITSAQVTPMSERKRYVAEDNLSIKTRFQTE
23	<i>Staphylococcus aureus</i> subsp. <i>Aureus</i> Cas9		translación	NDLNLRIEVIMEDITYRBYLEIM-----MDKR-----PFRILIKTIASKYQIRKYS
26	<i>Streptococcus thermophilus</i> CRISPR1 Cas9		translación	PKQKHYYELKPYDQKQFQGGAL-----LXVLQVANSQGGCKGCGKENSISYKVR
4	<i>Treponema denticola</i> Cas9	(modificado)	translación	KGRKFKSLI--IENQFVTLLEIL-----KLPSTYRH-VSDLQHLGGSKYSQVAKIG
5	<i>Filifactor alocis</i> Cas9	(modificado)	translación	KSKPKFKLSLDLIDKINVIKNU-----MLKCDMDTKADLSLIELPKMKGSPVVK
18	<i>Neisseria cinerea</i> Cas9	(modificado)	translación	ATGALDIRTHD-----TDSYKGMGIFCSYGVYKALSFOKYQ
24	<i>Campylobacter lari</i> Cas9	(modificado)	translación	STSSICEKHENKFNELTSNQKL-----LFSNAKESGSKVESLQTONKRYFEKYI
	<i>Streptococcus pyogenes</i> Cas9		translación	KRRDKPREDAGENTIHLPVTRHU-----GAPAAFKYPTOTTIDRK---RYTSKLEVL
7	<i>Lactobacillus johnsonii</i> Cas9	(modificado)	translación	SGREKFKLSLEDKDTILKAVLE-----GLHQN---AVMTKIPITIGLSTPLGPNQFP
23	<i>Mycoplasma gallisepticum</i> Cas9	(modificado)	translación	PKLWYTSFQKNDVIVETKHLLE-----TEYKENDSDSKKKNKAVRPLMTLELTL
20	<i>Parvibaculum lavamentivorans</i> Cas9	(modificado)	translación	SPVMSLAGEADINIFEDSKGIW-----IVQGVWASQVWLERDIDADHSYTTTRPM
	<i>Corynebacter diphtheriae</i> Cas9	(modificado)	translación	TIKRRVDTGPTSPSKLRPLQW-----SKESIKKESAPEDSKIIDRPGMLPVAVN

FIG. 19L

2	<i>Sustercella medswort(hemii) Cas9</i> (modificado)	translacion	CRRYVRYVEITTFIQASHIPQCSYRNNAITSPILSLPASFKYCKRNSCFQKAVGTBLSBELLDQP
21	<i>Scaphylococcus aureus subsp. Aureus Cas9</i>	translacion	TDIDLRLVYEVKSKKIIPOIIRKGG-----
26	<i>Streptococcus thermophilus CRISPR1 Cas9</i>	translacion	TDVLRGHNHIIKNSGDRFNLDI-----
4	<i>Treponema denticola Cas9</i> (modificado)	translacion	WKISSDMLIILYYSITGIFPKRIDILAY-----
5	<i>Flitfactor allocis Cas9</i> (modificado)	translacion	KMTIGKSKIIIVNOSVYGLYENRRL-----
18	<i>Weissberia cinerea Cas9</i> (modificado)	translacion	FDRLGKRLRP-----CPLKXKPPVR-----
24	<i>Campylobacter lari Cas9</i> (modificado)	translacion	ITPLGDKIKADQFRENISLTKTKYGLR-----
	<i>Streptococcus pyogenes Cas9</i>	translacion	DATLIFGISTGLYRTRDLSGLGGD-----
1	<i>Lactobacillus johnsonii Cas9</i> (modificado)	translacion	NGVTLSSNAKLIYQPTGLPKKSVKISDL-----
23	<i>Mycoplasma gallisepticum Cas9</i> (modificado)	translacion	NDVTLGDAKDNEDIGLSENRIDELLSKSLGDKTYK-
20	<i>Parvibaculum lavementivorans Cas9</i> (modificado)	translacion	QWPIKADDAKKVSLDZGRVRFSDN-----
3	<i>Corynebacter diptheriae Cas9</i> (modificado)	translacion	LFSQGRVTVVRRDELGRVRLLESTALPLVTKKVG-----
2	<i>Sustercella medswort(hemii) Cas9</i> (modificado)	translacion	RSBIFIERKGNRAKHTRPRYYTVWSSNKKKNESYRWVSKS



FIG. 20



Cas9 nº especie	PAM putativo	diana pUC19
1 <i>Corynebacter diphtheriae</i>		AAAAGGGAAATAGGGCGACA
2 <i>Sutterella wadsworthensis</i>	NGGNNNN	CATTCTGAGAATAGTGTATG GCGACCGAGTTGCTCTTGCC
2 <i>Sutterella wadsworthensis</i>	NGGNNNN	CATTCTGAGAATAGTGTATG GCGACCGAGTTGCTCTTGCC
3 <i>Legionella pneumophila</i> _Paris		
3 <i>Legionella pneumophila</i> _Paris		
5 <i>Filifactor alocis</i>	NNNAAGC	TTGAGTACTACCAGTCACA CTCTTCCTTTTCAATATTA
5 <i>Filifactor alocis</i>		
7 <i>Lactobacillus johnsonii</i>	NNAAANN	TAAATGCTTCAATAATATTG AAAGGGAATAAGGGCGACAC
7 <i>Lactobacillus johnsonii</i>		
20 <i>Parvibaculum lavamentivorans</i>	NNNCATN	AAAGGAAGAGTATGAGTATT CTATTCTCAGAATGACTTGG
21 <i>Staphylococcus aureus</i> subsp. Aureus		
23 <i>Mycoplasma gallisepticum</i> str. F	NNGANgg	AGTCACAGAAAAGCATCTTA GTTTGTGGTGAGCAAAAAC
23 <i>Mycoplasma gallisepticum</i> str. F		
24 <i>Campylobacter lari</i> CF89-12	NNGGGNA	GAAGGCCAAAATGCCCAAAA GGTATTATCCCGTATTGACG
24 <i>Campylobacter lari</i> CF89-12		
25 SpCas9		GGGATCAACCAAGTCATTG
25 SpCas9		

FIG. 21

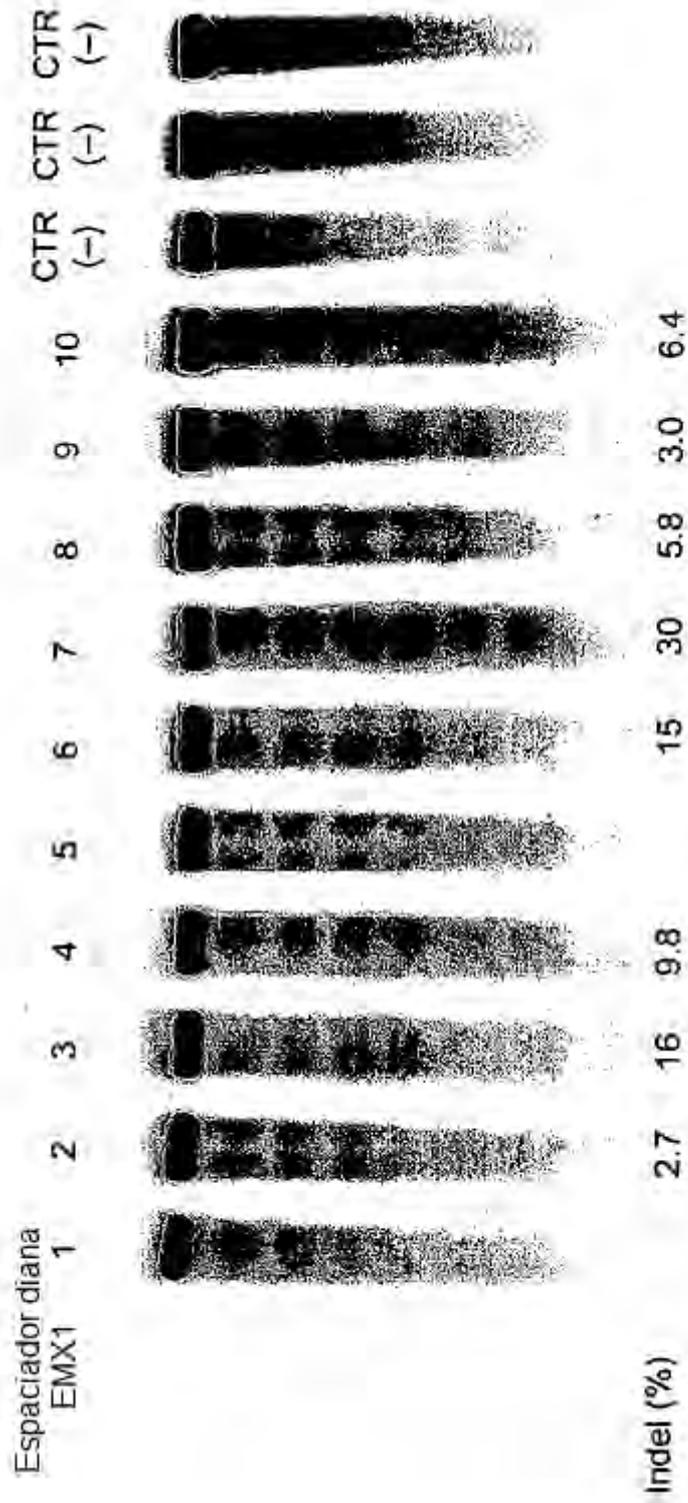


FIG. 22

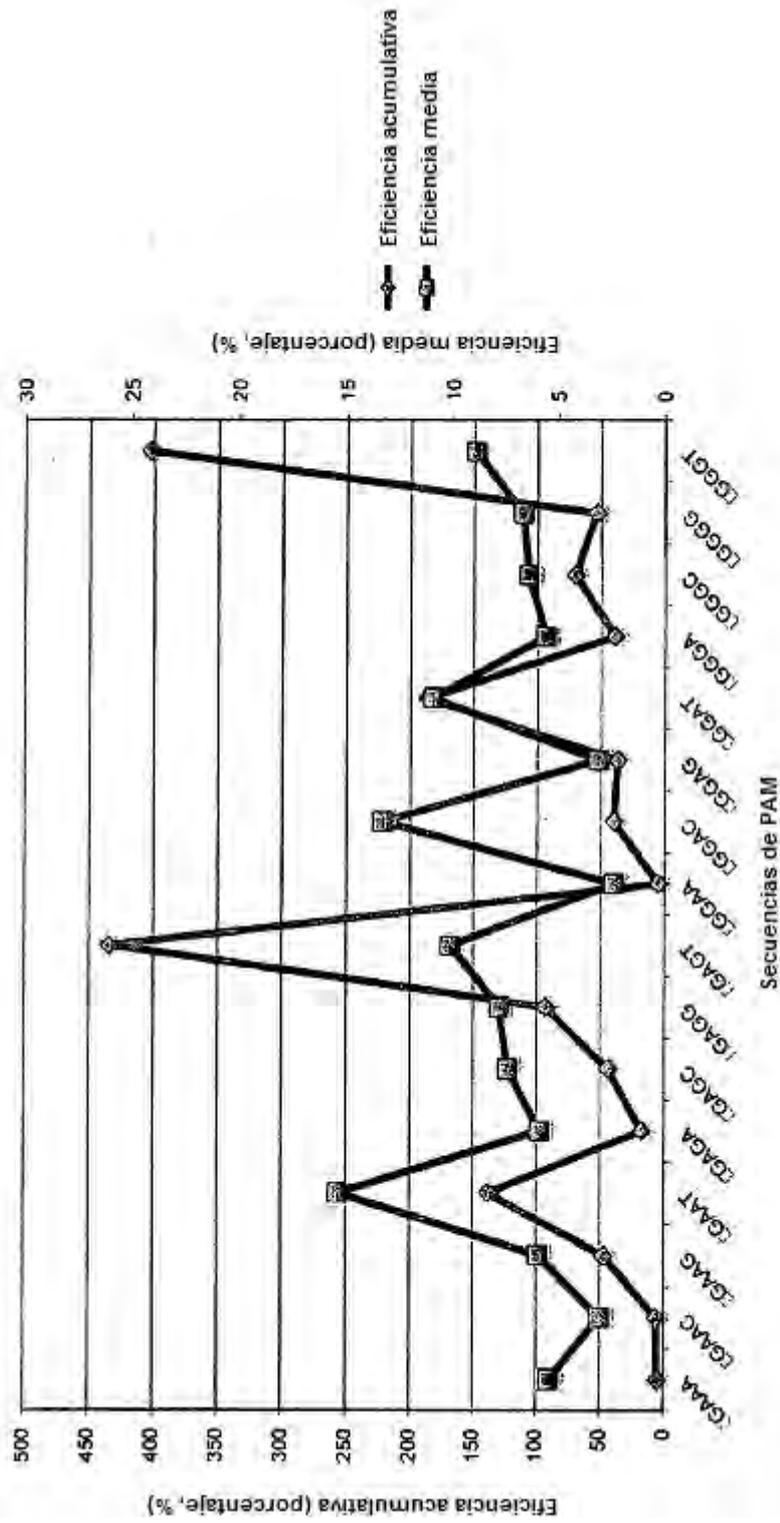


FIG. 23

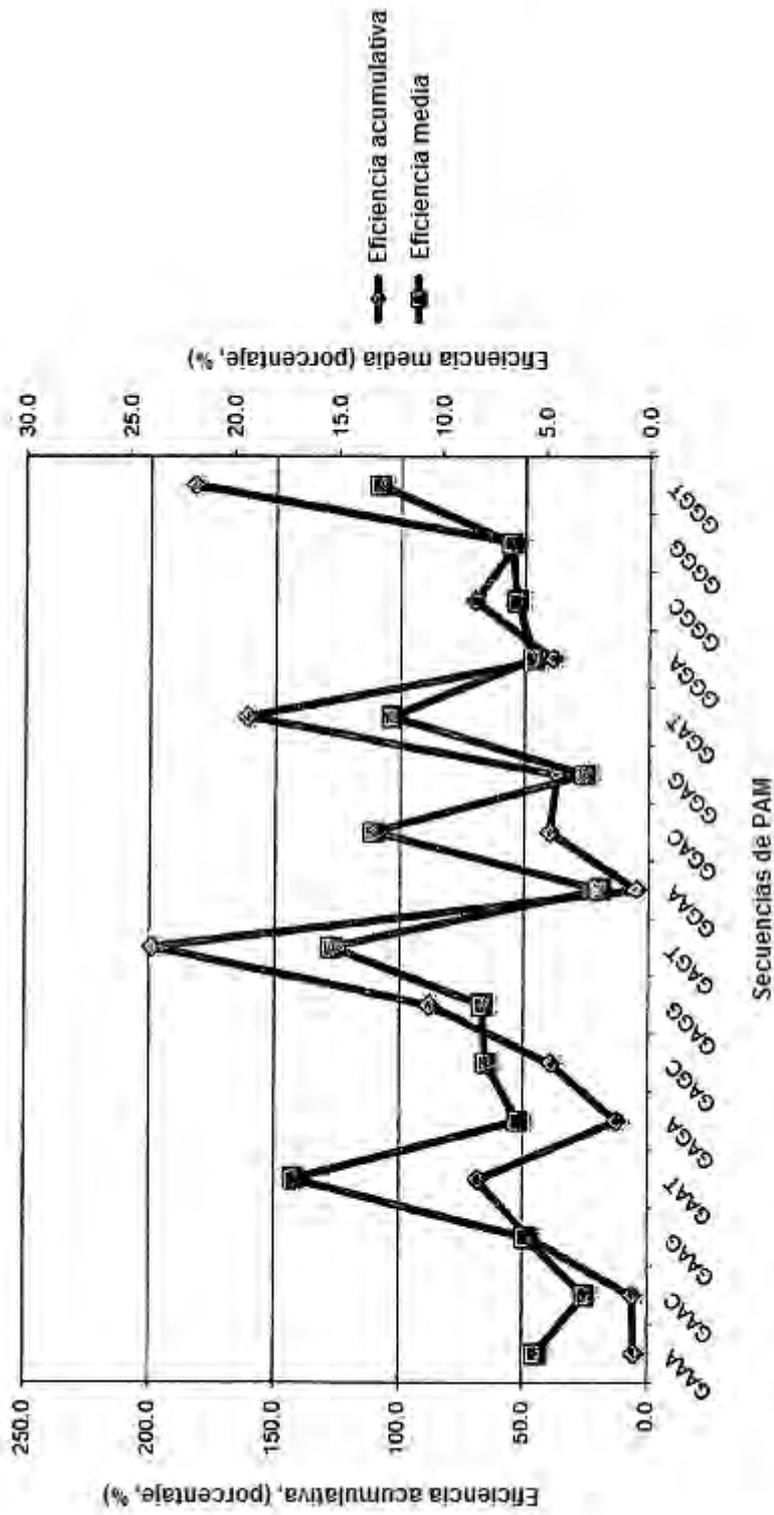


FIG. 25

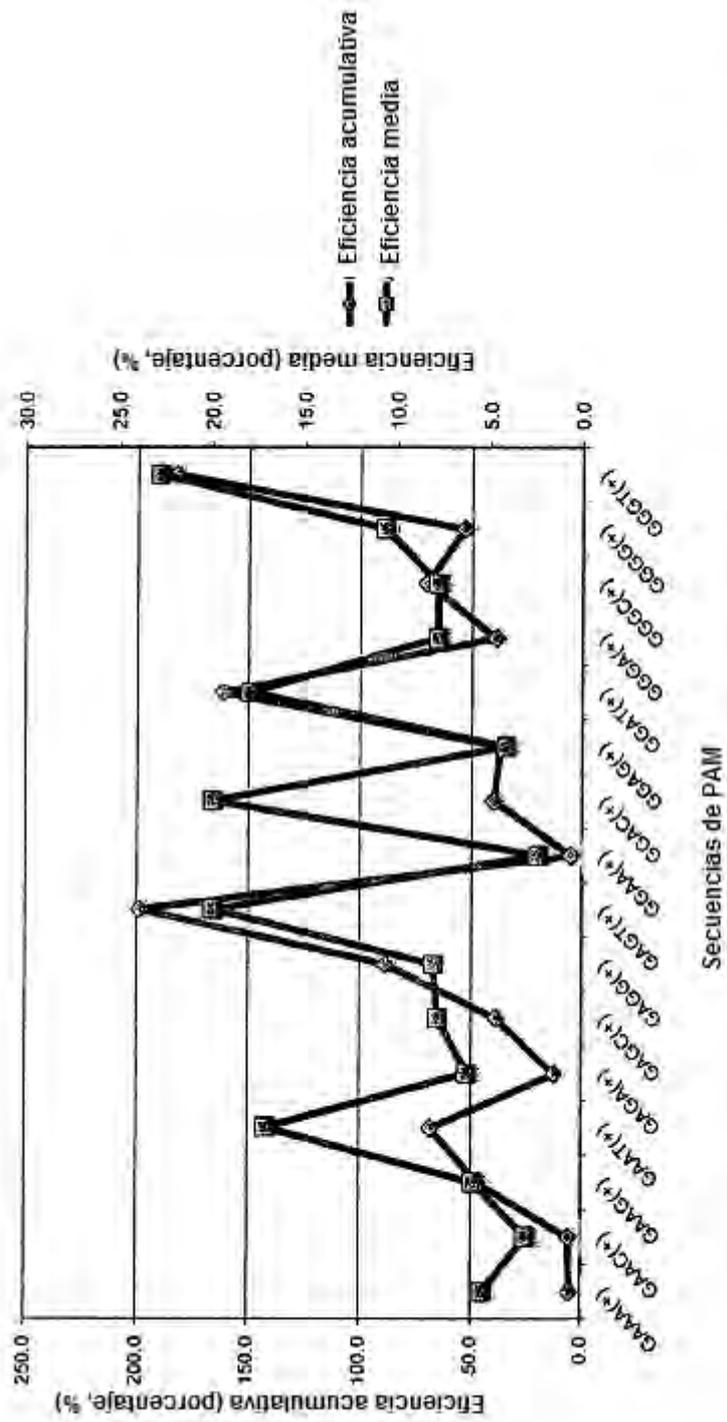


FIG. 26

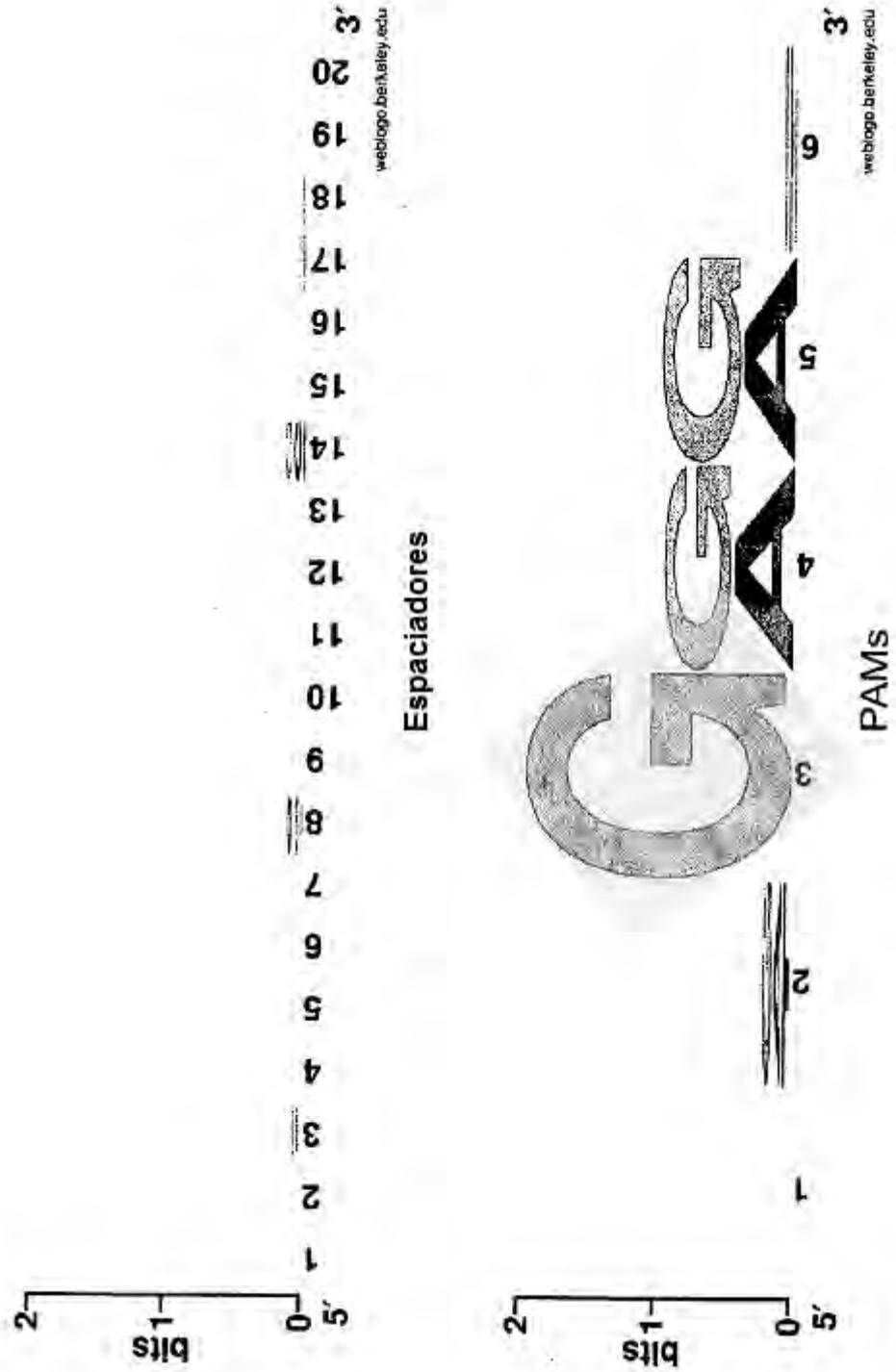


FIG. 27