

MEMORIA DESCRIPTIVA

Campo de la invención

La invención se relaciona con el campo de la biotecnología, específicamente con enzimas novedosas, nucleasas Cas de sistemas CRISPR-Cas, utilizadas en el corte de ADN y edición del genoma de distintos organismos. Esta técnica se puede utilizar en el futuro de la terapia génica de enfermedades humanas hereditarias, así como también en la edición del genoma de otros organismos.

Antecedentes de la invención

La modificación de secuencias de ADN es uno de los problemas que actualmente aquejan al campo de la biotecnología moderna. La edición y modificación de los genomas de organismos eucariontes y procariontes, así como también la manipulación de ADN *in vitro*, requiere la introducción focalizada de roturas de cadena doble en las secuencias de ADN.

Actualmente, para resolver este problema, se utilizan las siguientes técnicas: sistemas de nucleasa artificial, que contienen dominios tipo dedos de zinc; sistemas TALEN; y sistemas bacterianos CRISPR-Cas. Las primeras dos técnicas requieren una complicada optimización de la secuencia de aminoácidos de la nucleasa para que reconozca una secuencia de ADN específica. Por otra parte, cuando se trata de sistemas CRISPR-Cas, las estructuras que reconocen una diana de ADN no son proteínas, sino secuencias cortas de ARN guía. El corte de una diana de ADN

específica no requiere la síntesis *de novo* de nucleasa ni de su gen, sino que se produce mediante ARN guía complementario de la secuencia diana, lo que convierte a los sistemas CRISPR-Cas en medios convenientes y eficaces para el corte de distintas secuencias de ADN. La técnica permite el corte simultáneo de ADN en varias regiones mediante el uso de ARN guía de diferentes secuencias. Este método se utiliza, además, en la modificación simultánea de distintos genes en organismos eucariontes.

Por su naturaleza, los sistemas CRISPR-Cas corresponden a sistemas inmunitarios procariontes, que permiten realizar una introducción altamente específica de rupturas en el material genético viral (Mojica F. J. M. *et al.*, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements (Intervención de secuencias de repeticiones procariontes a espacios regulares derivadas de elementos genéticos foráneos), *Journal of Molecular Evolution* 2005, vol. 60, número 2, pág. 174 a 182). La abreviatura "CRISPR-Cas" corresponde a la sigla en inglés de "Repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas y Genes asociados a CRISPR" (Jansen R. *et al.*, Identification of genes that are associated with DNA repeats in prokaryotes (Identificación de genes asociados a repeticiones de ADN en procariontes), *Molecular Microbiology* 2002, vol. 43, número 6, pág. 1565 a 1575). Todos los sistemas CRISPR-Cas se componen de casetes y genes que codifican diferentes proteínas Cas (Jansen R. *et al.*, *Molecular Microbiology* 2002, vol. 43, número 6, pág. 1565 a 1575). Los casetes de CRISPR consisten en separadores, cada uno

con una secuencia nucleotídica única y repeticiones palindrómicas repetidas (Jansen R. et al., Molecular Microbiology 2002, vol. 43, número 6, pág. 1565 a 1575). La transcripción de casetes de CRISPR, la que precede al su procesamiento, produce la formación de secuencia de ARNcr guía, el cual, junto a proteínas Cas, forma un complejo efector (Brouns S. J. J. et al., Small CRISPR RNAs guide antiviral defense in prokaryotes (Pequeñas moléculas de ARN de CRISPR guían la defensa antiviral en procariontes), Science 2008, vol. 321, número 5891, pág. 960 a 964). Debido al apareamiento complementario entre el ARNcr y el sitio diana del ADN, el cual se denomina "protoespaciador", la nucleasa Cas reconoce una diana de ADN e introduce en el lugar, de manera altamente específica, una ruptura.

Los sistemas CRISPR-Cas con una única proteína efectora se agrupan en seis tipos diferentes (tipos I a VI), dependiendo de las proteínas Cas incluidas en los sistemas. En el año 2013, se propuso, por primera vez, el uso del sistema CRISPR-Cas9 tipo II para la edición del ADN genómico de células humanas (Cong L. et al., Multiplex genome engineering using CRISPR/Cas systems (Ingeniería múltiplex del genoma con sistemas CRISPR-Cas), Science 2013, feb 15;339(6121):819-23). El sistema CRISPR-Cas9 tipo II se caracteriza por su composición y mecanismo de actividad simples (es decir, su funcionamiento requiere la formación de un complejo efector compuesto únicamente de una proteína Cas9 y dos moléculas cortas de ARN: ARNcr y ARN trazador (ARNtracr). El ARN trazador se aparea en forma complementaria con una región de ARNcr, la cual se

origina a partir de una repetición CRISPR, para dar lugar a una estructura secundaria necesaria para la unión de moléculas de ARN guía con el efector Cas. La determinación de la secuencia de las moléculas de ARN guía es un paso importante en la caracterización de ortólogos Cas que no se han estudiado previamente. La proteína efectora Cas9 es una ADN endonucleasa dependiente del ARN, que presenta dos dominios de nucleasa (HNH y RuvC), los cuales introducen rupturas en las cadenas complementarias del ADN diana, produciendo así una ruptura de ADN de cadena doble (Deltcheva E. *et al.*, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III (Maduración de ARN CRISPR mediante un pequeña secuencia de ARN transcodificado y factor huésped RNasa III), Nature 2011, volumen 471, número 7340, pág. 602).

Hasta el momento, se sabe que distintas nucleasas CRISPR-Cas pueden introducir, en forma focalizada y específica, rupturas de cadena doble en el ADN. La tecnología de CRISPR-Cas9 es una de las técnicas más avanzadas y de rápido desarrollo para la introducción de rupturas en el ADN de distintos organismos, que van desde cadenas bacterianas hasta células humanas, la cual ofrece, además, aplicación *in vitro* (Song M., The CRISPR/Cas9 system: Their delivery, *in vivo* and *ex vivo* applications and clinical development by startups (Sistema CRISPR-Cas9: introducción, aplicaciones *in vivo* y *ex vivo*, y desarrollo clínico para fases iniciales), Biotechnol Progress 2017, Jul;33(4):1035-1045).

El complejo ribonucleico efector, compuesto de Cas9 y un dúplex ARNcr/ARNtracr, requiere la presencia de PAM (patrón

ajustado protoespaciador) en una diana de ADN para el reconocimiento y la hidrólisis posterior del ADN, además de una complementariedad de separador-protoespaciador de ARNcr (Mojica F. J.M. *et al.*, 2009). El patrón PAM es una secuencia estrictamente definida de varios nucleótidos ubicados en sistemas tipo II adyacentes al extremo 3' del protoespaciador de una cadena desfasada, o separados de él por varios nucleótidos. En ausencia del patrón PAM, no se produce la hidrólisis de los enlaces de ADN, que precede a la formación de una ruptura de cadena doble. La necesidad de la presencia de una secuencia de PAM en una diana aumenta la especificidad de reconocimiento, pero al mismo tiempo, impone restricciones a la selección de las regiones de ADN diana para introducir una ruptura. Así, la presencia de una secuencia de PAM deseada, que flanquea la diana de ADN por el extremo 3', es una característica que limita el uso de sistemas CRISPR-Cas en cualquier sitio del ADN.

Para desarrollar su actividad, diferentes proteínas CRISPR-Cas utilizan diferentes secuencias de PAM únicas. El uso de proteínas CRISPR-Cas con diferentes secuencias de PAM novedosas es necesaria para permitir la modificación de cualquier región del ADN, tanto *in vitro* como en el genoma de organismos vivos. La modificación de genomas eucariontes requiere, además, el uso de nucleasas de tamaño pequeño, que permitan la introducción, mediada por AAV, de sistemas CRISPR-Cas a las células.

A pesar de que se conoce una serie de técnicas de corte de ADN y la modificación de una secuencia de ADN genómico, sigue habiendo

necesidad de un medio eficaz y novedoso de modificación del ADN de distintos organismos y en sitios estrictamente específicos de una secuencia de ADN.

Resumen de la invención

El objetivo de la presente invención es proporcionar un medio novedoso de modificación de una secuencia de ADN genómico de organismos unicelulares o multicelulares mediante el uso de sistemas CRISPR-Cas9. Los sistemas actuales tienen un uso limitado debido a que debe haber presente una secuencia de PAM específica en el extremo 3' de la región del ADN que se desea modificar. La búsqueda de enzimas Cas9 novedosas con otras secuencias de PAM ampliará el rango de medios disponibles para la formación de rupturas de cadena doble en sitios estrictamente específicos de moléculas de ADN en distintos organismos. Para resolver este problema, los autores caracterizaron la nucleasa CRISPR tipo II PpCas9, ya predicha para el microorganismo *Pasteurella pneumotropica* (*P. pneumotropica*), la cual permitiría introducir modificaciones dirigidas al genoma tanto de los organismos ya mencionados como de otros. La presente invención presenta las siguientes características esenciales: (a) una pequeña secuencia de PAM, diferente de otras secuencias de PAM conocidas; (b) un tamaño relativamente pequeño de la proteína PpCas9 caracterizada, la cual corresponde a 1055 residuos de aminoácidos (a.a.r.).

Dicho problema queda resuelto mediante el uso de una proteína, que comprende la secuencia de aminoácidos de SEQ ID NO: 1 o una

secuencia de aminoácidos al menos 95 % idéntica a la secuencia de aminoácidos de SEQ ID NO: 1, y se diferencia de la SEQ ID NO: 1 únicamente en residuos de aminoácidos no conservados para obtener, en la molécula de ADN, una ruptura de cadena doble ubicada **inmediatamente** antes de la secuencia nucleotídica 5'-NNNN(A/G)TT-3' en dicha molécula de ADN. En algunas modalidades de la invención, este uso es caracterizado porque la ruptura de cadena doble en la molécula de ADN se produce a una temperatura de 35 °C a 45 °C. En algunas modalidades de la invención, este uso es caracterizado porque la ruptura de cadena doble se produce en el ADN genómico de una célula de mamífero. En algunas modalidades de la invención, este uso es caracterizado porque la formación de la ruptura de cadena doble en la molécula de ADN produce la modificación del ADN genómico de dicha célula de mamífero.

Dicho problema se resuelve, además, a través de un método de modificación de una secuencia de ADN genómico en una célula de un organismo unicelular o multicelular, que comprende la introducción a dicha célula del organismo de una cantidad efectiva de: a) una proteína, que comprende la secuencia de aminoácidos de SEQ ID NO: 1, o un ácido nucleico, que codifica la proteína que comprende la secuencia de aminoácidos de SEQ ID NO: 1, y b) un ARN guía, que comprende una secuencia que forma un dúplex con la secuencia nucleotídica de una región del ADN genómico del organismo, la cual se encuentra directamente adyacente a la secuencia nucleotídica 5'-NNNN(A/G)TT-3' e interactúa con dicha proteína luego de la formación del dúplex, o una secuencia de ADN que codifica dicho

ARN guía; donde la interacción de dicha proteína con el ARN guía y la secuencia nucleotídica 5'-NNNN(A/G)TT-3' produce la formación de una ruptura de cadena doble en la secuencia de ADN genómico inmediatamente adyacente a la secuencia 5'-NNNN(A/G)TT-3'.

En algunas modalidades de la invención, el método es caracterizado porque comprende, además, la introducción de una secuencia exógena de ADN simultáneamente con el ARN guía. En algunas modalidades de la invención, el método es caracterizado porque dicha célula es una célula de mamífero.

Se puede utilizar una mezcla de ARNcr y ARN trazador (ARNtracr), la cual forma un complejo con una región de ADN diana y proteína PpCas9, como ARN guía. En modalidades preferidas de la invención, se puede utilizar un ARN híbrido construido a base de ARNcr y ARN trazador como ARN guía. Las personas versadas en el arte conocen métodos de construcción de ARN guía híbrido (Hsu PD, *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases (Especificidad de diana de ADN en nucleasas Cas9 guiada por ARN), Nature Biotechnology 2013, Sep;31(9):827-32). En los siguientes ejemplos, se divulga uno de los métodos de construcción de ARN híbrido.

La invención se puede utilizar tanto para el corte *in vitro* de ADN diana como para la modificación del genoma de algunos organismos vivos. El ADN genómico se puede modificar directamente, es decir, cortando el ADN genómico en un sitio correspondiente,

así como también insertando una secuencia exógena de ADN a través de reparación homóloga.

Se puede utilizar, como secuencia exógena de ADN, cualquier región de ADN de cadena doble o cadena simple del genoma de un organismo distinto al utilizado para administración (o una composición de dichas regiones entre sí y con otros fragmentos de ADN), donde dicha región (o composición de regiones) se debería integrar al lugar de una ruptura de cadena doble en el ADN diana, provocada por la nucleasa PpCas9. En algunas modalidades de la invención, se puede utilizar, como secuencia exógena de ADN, una región de ADN de cadena doble del ADN genómica de un organismo, con la cual se introduce la proteína PpCas9, modificada adicionalmente mediante mutaciones (sustitución de nucleótidos), así como también mediante inserciones o deleciones de uno o más nucleótidos.

El resultado técnico de la presente invención consiste en el aumento de la versatilidad de los sistemas CRISPR-Cas9 disponibles, lo que permite el uso de nucleasa Cas9 para el corte de ADN genómico o plásmido en un número mayor de sitios específicos y en condiciones específicas. La nucleasa novedosa se puede utilizar en células bacterianas, de mamíferos y de otros organismos.

Breve descripción de las figuras

Figura 1. Esquema del locus del sistema CRISPR-PpCas9. La repetición directa (DR) es una región repetida periódicamente, que forma parte de un casete de CRISPR.

Figura 2. Selección *in vitro* de PAM. Esquema del experimento.

Figura 3. Corte de nucleasa PpCas9 de fragmentos de repertorio de 7N a diferentes temperaturas de reacción.

Figura 4. (A) Análisis de los resultados de la selección *in vitro* de la nucleasa PpCas9 mediante el uso del cálculo del logaritmo de cambio de proporción de cada nucleótido específico en cada posición de PAM (FC). (B) Logo de PAM de la nucleasa PpCas9. Se indica la aparición de adenina, citosina, timina y guanina correspondiente a cada posición. La altura de las letras corresponde a la aparición de un nucleótido en una posición determinada de la secuencia de PAM.

Figura 5. Verificación del efecto de sustituciones de un solo nucleótido en la posición de PAM 1 sobre la eficacia del corte de la diana de ADN mediante nucleasa PpCas9.

Figura 6. Verificación de la importancia de las posiciones nucleotídicas en la secuencia de PAM de PpCas9.

Figura 7. Verificación del efecto de la sustitución de A a G en la posición de PAM 5 sobre la eficacia del corte de la diana de ADN mediante nucleasa PpCas9.

Figura 8. Verificación del efecto de sustituciones de un solo nucleótido en la posición de PAM 7 sobre la eficacia del corte de la diana de ADN mediante nucleasa PpCas9.

Figura 9. Corte de distintos sitios de ADN mediante el uso de la proteína PpCas9. Los carriles 1 y 2 corresponden a controles positivos.

Figura 10. Verificación del reconocimiento de la secuencia de PAM CAGCATT mediante el uso de nucleasa PpCas9. Los carriles 1 y 2 corresponden a controles positivos.

Figura 11. Diagrama de la herramienta de corte de ADN PpCas9.

Figura 12. Experimento de corte de una diana de ADN. Se utilizaron moléculas de ARN guía híbrido de diferentes longitudes.

Figura 13. Alineación de secuencias de aminoácidos de las proteínas PpCas9 y Cas9 de *Staphylococcus aureus* con el software BLASTp (parámetros predeterminados) del National Center for Biotechnology Information (NCBI).

Figura 14. Modificación del ADN genómico de células humanas con PpCas9. (A) es el esquema del experimento destinado a determinar la eficacia de la modificación del ADN genómico de células humanas con un plásmido portador de PpCas9; (B) corresponde a los resultados del análisis de inserciones y deleciones de nucleótidos en la secuencia de sitios diana del ADN genómico de células humanas (arriba: se aplicaron productos de reacción con endonucleasa I de T7 a una electroforesis en gel de agarosa;

abajo: ejemplos de inserciones y deleciones formadas por la nucleasa PpCas9 en el gen EMX1, las cuales se determinaron mediante secuenciación de alto rendimiento).

Divulgación detallada de la invención

En la descripción de la presente invención, los términos "incluye(n)" e "incluido(s)/a(s)" se deben interpretar como "que incluye(n), entre otros...". Dichos términos no se deben interpretar como "se compone(n) únicamente de...". A menos que se definan por separado, los términos técnicos y científicos de esta aplicación tienen los significados comunes, generalmente aceptados en la literatura científica y técnica.

En el presente documento, el término "homología porcentual de dos secuencias" es equivalente al término "identidad porcentual de dos secuencias". La identidad de una secuencia se determina de acuerdo con una secuencia de referencia. En el arte, son conocidos algunos algoritmos de análisis de secuencias, como BLAST, descrito en Altschul et al., Journal of Molecular Biology 215, pág. 403 a 10 (1990). Para fines de la presente invención, la determinación del nivel de identidad y la similitud entre secuencias nucleotídicas y secuencias de aminoácidos, se puede utilizar la comparación de secuencias nucleotídicas y de aminoácidos, que se realiza con el paquete de software BLAST del National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/blast>) mediante alineación con brechas de acuerdo con parámetros estándar. La identidad porcentual de dos secuencias se determina

mediante el número de posiciones de aminoácidos idénticos en estas dos secuencias, considerando el número de brechas y la longitud de cada brecha que se ingresará para la comparación óptima de las dos secuencias mediante alineación. La identidad porcentual es igual al número de aminoácidos idénticos en posiciones determinadas en que se considera la alineación de secuencia dividido por el número total de posiciones y multiplicado por 100.

El término "hibrida específicamente" se refiere a la asociación entre dos moléculas de ácido nucleico de cadena simple o secuencias suficientemente complementarias, lo cual permite este tipo de hibridación en condiciones predeterminadas que se aplican normalmente en el arte.

La frase "una ruptura de cadena doble ubicada inmediatamente antes de la secuencia nucleotídica de PAM" significa que la ruptura de cadena doble en la secuencia de ADN diana se produce a una distancia de 0 a 25 nucleótidos antes de la secuencia nucleotídica de PAM.

Una secuencia exógena de ADN introducida simultáneamente con una secuencia de ARN guía se refiere a una secuencia de ADN preparada específicamente para la modificación específica de una secuencia de ADN diana de cadena doble en el sitio de la ruptura determinado mediante la especificidad del ARN guía. Por ejemplo, esta modificación puede corresponder a la inserción o delección de ciertos nucleótidos en el sitio de la ruptura del ADN diana. El ADN exógeno puede corresponder ya sea a una región ADN de otro

organismo o a una región ADN del mismo organismo que el del ADN diana.

Una proteína que comprende una secuencia específica de aminoácidos se refiere a una proteína con secuencia de aminoácidos compuesta de dicha secuencia de aminoácidos y, posiblemente, otras secuencias unidas, por medio de enlaces peptídicos, a dicha secuencia de aminoácidos. Un ejemplo de otras secuencias puede ser una señal de localización nuclear (NLS) u otras secuencias que brindan mayor funcionalidad a secuencia de aminoácidos.

Una secuencia exógena de ADN introducida simultáneamente con una secuencia de ARN guía se refiere a una secuencia de ADN preparada específicamente para la modificación específica de una secuencia de ADN diana de cadena doble en el sitio de la ruptura determinado mediante la especificidad del ARN guía. Por ejemplo, esta modificación puede corresponder a la inserción o delección de ciertos nucleótidos en el sitio de la ruptura del ADN diana. El ADN exógeno puede corresponder ya sea a una región ADN de otro organismo o a una región ADN del mismo organismo que el del ADN diana.

Una cantidad efectiva de proteína y ARN introducida en la célula se refiere a una cantidad de proteína y ARN tal que, al introducirse a dicha célula, tiene la capacidad de formar un complejo funcional, es decir, un complejo que se une específicamente a una secuencia de ADN diana y produce en ella una ruptura de cadena doble en el sitio determinado a través del ARN

guía y la secuencia de PAM en el ADN. La eficacia de este proceso se puede evaluar analizando ADN diana aislado de dicha célula mediante técnicas convencionales que conocen las personas versadas.

Se puede depositar en una célula una proteína y una secuencia de ARN mediante distintas técnicas. Por ejemplo, una proteína se puede depositar como un plásmido de ADN, que codifica un gen de esta proteína como un ARNm para la traducción de esta proteína en el citoplasma celular o como un complejo ribonucleoproteico que incluye esta proteína y una secuencia de ARN guía. La introducción se puede realizar mediante distintas técnicas que conocen las personas versadas.

El ácido nucleico que codifica los componentes del sistema se puede introducir en una célula directa o indirectamente del siguiente modo: por medio de una transfección o transformación de células a través de métodos que conocen las personas versadas, mediante el uso de un virus recombinante, por medio de manipulaciones en la célula, como microinyección de ADN, etc.

Se puede introducir un complejo ribonucleico compuesto de una nucleasa y moléculas de ARN guía y ADN exógeno (si es necesario) mediante la transfección de los complejos a una célula o mediante la introducción mecánica del complejo a la célula (por ejemplo, mediante microinyección).

Una molécula de ácido nucleico que codifica la proteína introducida en una célula se puede integrar al cromosoma o puede

ser ADN de replicación extracromosómica. En algunas modalidades, para que se produzca la expresión efectiva del gen de la proteína con ADN introducido en una célula, es necesario modificar la secuencia de dicho ADN de acuerdo con el tipo de célula para optimizar los codones de expresión, lo cual se debe a frecuencias disímiles de aparición de codones sinónimos en las regiones de codificación del genoma de distintos organismos. La optimización de codones es necesaria para aumentar la expresión en células animales, vegetales, fúngicas o de microorganismos.

Para que una proteína que presenta una secuencia al menos 95 % idéntica con la secuencia de aminoácidos de SEQ ID NO: 1 funcione en una célula eucarionte, es necesario que esta proteína termine en el núcleo de esta célula. Por lo tanto, en algunas modalidades de la invención, se usa una proteína con una secuencia al menos 95 % idéntica con la secuencia de aminoácidos de SEQ ID NO: 1, la cual está modificada adicionalmente en uno o ambos extremos mediante la adición de una o más señales de localización nuclear para producir rupturas de cadena doble en ADN diana. Por ejemplo, se puede utilizar una señal de localización nuclear del virus SV40. Para que la introducción en el núcleo sea eficaz, la señal de localización nuclear se puede separar de la secuencia proteica principal mediante una secuencia espaciadora (por ejemplo, la descrita en Shen B, *et al.*, "Generation of gene-modified mice via Cas9/RNA-mediated gene targeting" (Generación ratones modificados genéticamente a través de modificación genética dirigida mediada por Cas9/ARN), Cell Research 2013, May;23(5):720-3). Además, en

otras modalidades, se puede utilizar otra señal de localización nuclear o un método alternativo de introducción de dicha proteína al núcleo celular.

La presente invención contempla el uso de una proteína del organismo *P. pneumotropica*, el cual es homólogo a las proteínas Cas9 caracterizadas previamente, para introducir rupturas de cadena doble en las moléculas de ADN en posiciones estrictamente específicas. El uso de nucleasas de CRISPR para introducir modificaciones dirigidas al genoma tiene una serie de ventajas. En primer lugar, la especificidad de la actividad del sistema se determina mediante una secuencia de ARNcr, la cual permite el uso de un tipo de nucleasa en todos los locus diana. En segundo lugar, la técnica permite la introducción en una célula de varias moléculas de ARN guía, que complementan diferentes genes diana, de una sola vez, lo que permite modificar simultáneamente varios genes a la vez.

La nucleasa PpCas9 es una nucleasa Cas que se encuentra en la cepa de *Pasteurella pneumotropica* ATCC 35149, un patógeno de roedor que vive en los pulmones de los animales. El sistema CRISPR-Cas9 de *Pasteurella pneumotropica* (*P. pneumotropica*) (de ahora en adelante, denominado "sistema CRISPR-PpCas9") pertenece a los sistemas CRISPR-Cas tipo II-C y se compone de un casete de CRISPR que porta cuatro repeticiones directas (DR) con la secuencia 5'ATTATAGCACTGCGAAATGAAAAAGGGAGCTACAAC3', interespaciada por las secuencias de separadores únicos. Ninguno de los separadores del sistema coincide en la secuencia con los

bacteriófagos ni plásmidos que se conocen en la actualidad, hecho que no permite determinar la secuencia de PAM de PpCas9 de interés mediante análisis bioinformático. Adyacente al casete de CRISPR, se encuentra el gen para la proteína efectora Cas9 PpCas9, así como también los genes para las proteínas Cas1 y Cas2, que participan en la adaptación e integración de nuevos separadores. Próximo a los genes Cas, se descubrió una secuencia que complementa parcialmente repeticiones directas y que se pliega en una estructura secundaria característica, la cual se considera el ARN trazador (ARNtracr) (figura 1).

Para predecir la dirección de transcripción del casete de CRISPR, hizo falta el conocimiento de la arquitectura característica del complejo proteico ARN-Cas de los sistemas tipo II-C: El ARNpre-cr se transcribe en dirección opuesta a la de los genes Cas (figura 1).

Así, el análisis de la secuencia del locus de PpCas9 permitió predecir las secuencias de las moléculas de ARN trazador y guía (tabla 1).

Tabla 1. Secuencias de ARN guía del sistema CRISPR-PpCas9, las cuales se determinaron mediante métodos bioinformáticos. La negrita indica la secuencia de repetición directa (DR).

Nombre	Secuencia
PpCas9	5' GCGAAATGAAAACGUUGUUACAAUAAGAGAUGAAUUUCUCGCAAAGCTCUGCCU
trRNA	CUUGAAAUUUCGGUUUCAAGAGGCAUCUUUUU-3'

PpCas9	5'-xxxxxxxxxxxxxxxxxxxxxx GUUGUAGCUCCCUUUUUCAUUUCGC -3'
crRNA	

Para verificar la actividad de la nucleasa PpCas9 y determinar la secuencia de PAM de PpCas9 de interés, realizamos experimentos de recreación de la reacción de corte de ADN *in vitro*. Para determinar la secuencia de PAM de la proteína PpCas9, se empleó el corte *in vitro* de repertorios de PAM de cadena doble. Por ello, fue necesario obtener todos los componentes del complejo efector de PpCas9 del siguiente modo: Moléculas de ARN guía y una nucleasa en forma recombinante. Determinación de la secuencia de ARN guía permitió sintetizar moléculas de ARNcr y ARNtracr *in vitro*. La síntesis se llevó a cabo con el kit de síntesis de ARN HiScribe T7 de NEB. Los repertorios de ADN de cadena doble fueron 374 fragmentos de pares de bases (bp), que comprenden una secuencia de protoespaciador flanqueada por siete nucleótidos aleatorios (5'-NNNNNNN-3') del extremo 3': 5'-

cccggggtaccacgagagatggttgaaatcatctttctcgtgggcatccttgatggccacctcgt
cggaagtgccacgaggatgacagcaatgccaatgctgggggggctcttctgagaacgagctctgc
tgcttgacacggccaggacggccaacaccaaccagaacttgggagaacagcactccgctctgggct
tcatcttcaactcgtcgactccctgcaaacacaaagaaagagcatgttaaaataggatctacatca
cgtaacctgtcttagaagaggctagatactgcaattcaaggaccttatctcctttcattgagcac**N**
NNNNNNaactccatcta ccagcctactctcttatctctgtggtatt -3'

Para cortar esta diana, se utilizaron moléculas de ARN guía de la siguiente secuencia:

ARNtracr:

5' GCGAAATGAAAAACGUUGUUACAAUAAGAGAUGAAUUUCUCGCAAAGCTCUGCCUCUUGAAAUU
UCGGUUUCAAGAGGCAUCUUUUU

y ARNcr: 5'

uauccuccuuucauugagcacGUUGUAGCUCCCUUUUUUCAUUUCGC.

La negrita indica la secuencia de ARNcr complementaria al protoespaciador (secuencia de ADN diana).

Para producir una proteína PpCas9 recombinante, se clonó el gen correspondiente en el plásmido pET21a. Se utilizó ADN sintetizado de Integrated DNA Technologies (IDT) como el ADN codificador del gen. Se optimizaron los codones de la secuencia para excluir codones raros descubiertos en el genoma de *P. pneumotropica*. Se transformaron células Rosetta de *E. coli* con el plásmido resultante pET21a-6xHis-PpCas9.

Se diluyeron 500 µL de cultivo durante la noche en 500 mL de medio LB (Luria-Bertani), y las células se cultivaron a 37 °C hasta obtenerse una densidad óptica de 0,6 Ru. La síntesis de la proteína objetivo se provocó mediante la adición de IPTG (isopropil-β-D-1-tiogalactopiranosido) a una concentración de 1 mM; luego, las células se incubaron a 20 °C durante 6 horas. Posteriormente, las células se centrifugaron a 5000 g durante 30 minutos, y los precipitados celulares resultantes se congelaron a -20 °C.

Los precipitados se descongelaron en hielo durante 30 minutos, se volvieron a poner en suspensión en 15 mL de buffer de lisis (50

mM de Tris-HCl a pH 8, 500 mM de NaCl, 1 mM de β -mercaptoetanol, 10 mM de imidazol) complementado con 15 mg de lisozima, y se volvieron a incubar en hielo durante 30 minutos. Posteriormente, las células se perturbaron con ondas sonoras durante 30 minutos y se centrifugaron durante 40 minutos a 16000 g. El supernadante resultante se hizo pasar por un filtro de 0,2 μ m y se aplicó a una columna HisTrap HP de 1 mL (GE Healthcare) a 1 mL/min.

Se efectuó una cromatografía con el cromatógrafo AKTA FPLC (GE Healthcare) a 1 mL/min. La columna con la proteína aplicada se lavó con 20 mL de buffer de lisis complementado con 30 mM de imidazol, luego de lo cual la proteína se removió con buffer de lisis complementado con 300 mM de imidazol.

Luego, la fracción proteica obtenida en el curso de la cromatografía de afinidad se hizo pasar por una columna de filtración con gel (24 mL) Superdex 200 10/300 GL equilibrada con el siguiente buffer: 50 mM de Tris-HCl a pH 8, 500 mM de NaCl, 1 mM de DTT (ditiotreitól). Con un concentrador Amicon (con un filtro de 30 kDa), se concentraron fracciones correspondientes a la forma monomérica de la proteína PpCas9 hasta llegar a 3 mg/mL, luego de lo cual la proteína purificada se almacenó a -80 °C en un buffer con glicerol al 10 %.

La reacción *in vitro* de corte de los repertorios lineales de PAM se llevó a cabo en un volumen de 20 μ L en las siguientes condiciones. La mezcla de reacción se compuso de 1X buffer CutSmart (NEB), 5 mM de DTT, 100 nM de repertorio de PAM, 2 μ M de

ARNtr/ARNcr, 400 nM de proteína PpCas9. Como control, se prepararon muestras sin ARN de manera similar. Las muestras se incubaron a diferentes temperaturas y se analizaron mediante electroforesis en gel de agarosa al 2 %. En caso de reconocimiento correcto y corte específico de ADN con proteína PpCas9, se deberían generar dos fragmentos de ADN de aproximadamente 326 y 48 pares de bases (véase la figura 2).

Los resultados del experimento demostraron que PpCas9 tiene actividad de nucleasa y corta una fracción de los fragmentos de repertorio de PAM. El gradiente de temperatura (figura 3) demostró que la proteína está activa en el rango de temperatura de 35 a 45 °C. Luego, se utilizó en el estudio una temperatura de 42 °C como temperatura de trabajo.

La reacción de corte de repertorio se repitió en las condiciones seleccionadas. Los productos de reacción se aplicaron a gel de agarosa al 1,5 % y se sometieron a electroforesis. Los fragmentos de ADN sin cortar de 374 bp de longitud se extrajeron del gel y se prepararon para secuenciación de alto rendimiento con el kit NextUltra II de NEB. Las muestras se secuenciaron en la plataforma Illumina; luego, se llevó a cabo el análisis de las secuencias a través de métodos bioinformáticos: determinamos la diferencia de aparición de nucleótidos en posiciones individuales de PAM (NNNNNNN) en comparación con la muestra de control a través del método descrito en Maxwell CS et al., A detailed cell-free transcription-translation-based assay to decipher CRISPR protospacer-adjacent motifs (Ensayo detallado sin células a base

de transcripción-traducción para descifrar patrones CRISPR adyacentes a protoespaciadores), Methods 2018, julio 1;143:48-57). Además, se construyó un logo de PAM para analizar los resultados (figura 4).

Ambos métodos de análisis de datos (figura 4) indican la importancia de las posiciones 5, 6 y 7 de PAM. Así, el análisis *in vitro* permitió establecer la siguiente secuencia putativa de PAM para PpCas9: NNNNATT. Sin embargo, esta secuencia es únicamente putativa considerando los resultados inexactos obtenidos a través de métodos de selección para determinar PAM.

A este respecto, se verificó la importancia de posiciones individuales de secuencia de PAM para determinar la secuencia con mayor precisión. Con este fin, se efectuaron reacciones *in vitro* de corte de fragmentos de ADN con una diana de ADN 5'-atctcctttcattgagcac-3' flanqueada por la secuencia de PAM CAAC**ATT** (o sus derivados):

5'-
cccggggtaccacgagagatggtggaaatcatctttctcgtgggcatccttgatggccacctcgt
cggaagtgccacgaggatgacagcaatgccaatgctgggggggctcttctgagaacgagctctgc
tgctgacacggccaggacggccaacaccaaccagaacttgggagaaacagcactccgctctgggct
tcatcttcaactcgtcgactccctgcaaacacaaagaaagagcatgttaaaataggatctacatca
cgtaacctgtcttagaagaggctagataactgcaattcaaggaccttatctcctttcattgagcacC
AAC**ATT**aactccatcta ccagcctactctcttatctctgtggtatt- 3'

Todas las reacciones de corte de ADN se efectuaron en las siguientes condiciones:

1x buffer CutSmart

400 nM de PpCas9

20 nM de ADN

2 μ M de ARNcr

2 μ M de ARNtracr

Tiempo de incubación: 30 minutos; temperatura de reacción: 42 °C.

La sustitución de la posición 1 de PAM por las cuatro variantes posibles de nucleótidos no influyó en el rendimiento de la actividad proteica (figura 5).

La importancia predicha de las posiciones 5 y 6 se confirmó en forma experimental mediante sustituciones de nucleótidos individuales (purina por pirimidina y viceversa) en cada una de las posiciones de PAM. Cuando las sustituciones se produjeron en las posiciones 5 y 6, la proteína prácticamente cesó su actividad. Cuando la sustitución se produjo en la posición 7, el rendimiento de la actividad de PpCas9 disminuyó el doble, hecho que refleja los menores requisitos para el nucleótido en esta posición (figura 6). Así, según los resultados de la selección *in vitro* de PAM de nucleasa PpCas9, los nucleótidos con mayor probabilidad de producirse en la posición 5 de PAM son adenina o guanina, hecho que se confirmó experimentalmente (figura 7). La sustitución de A por G no redujo la eficacia de corte del fragmento.

Según los resultados de la selección *in vitro*, se deben reconocer con mayor eficacia los fragmentos con "T" o "S" en la posición 7. Se efectuaron experimentos adicionales para verificar definitivamente la importancia de los nucleótidos en esta posición. Los resultados de las pruebas *in vitro* demostraron que la sustitución del nucleótido "T" en la posición 7 por A o G redujo el rendimiento de corte en 40 a 50 % (figura 8). Así, la posición 7 del PAM se conserva menos en comparación con las posiciones 5 y 6: las purinas en la posición 7 reducen el rendimiento de reconocimiento, pero no impiden que la proteína PpCas9 introduzca rupturas de cadena doble en el ADN.

Los resultados del estudio fueron los siguientes: El PAM reconocido por la nucleasa PpCas9 corresponde a la siguiente fórmula: 5'-NNNN(A/G)TT-3'. La posición 7 se conserva menos.

Los siguientes ejemplos de modalidades del método se proporcionan con el fin de divulgar las características de la presente invención y no se deben interpretar como que limitan de alguna manera el ámbito de la invención.

Ejemplo 1. Prueba de la actividad de la proteína PpCas9 en el corte de varias dianas de ADN.

Para poder confirmar la capacidad de PpCas9 de reconocer distintas secuencias de ADN flanqueadas por la secuencia 5'-NNNN(A/G)TT-3', se efectuaron experimentos de corte *in vitro* de dianas de ADN provenientes de una secuencia génica de *grin2b* humano (véase la tabla 2).

Tabla 2. Dianas de ADN provenientes del gen GRIN2B humano.

Secuencia	PAM						
TATCTCCTTTTCATTGAGCAC	C	A	A	A	C	C	C
CAGCTGAAGTAATGTTAGAG	C	C	A	C	A	T	T
AATAAGAAAAACATTATTAT	C	A	C	C	A	T	T
GGGGCTATAAGTACACAAGC	C	C	T	G	C	A	T
CGTTCTCAGAAGAGCCCCC	C	A	G	C	A	T	T
CCCACGAGAAAGATGATTTC	C	A	C	C	A	T	C

Se utilizó un fragmento de PCR del gen *grin2b* que porta sitios de reconocimiento (tabla 2) supuestamente reconocibles por PpCas9 de acuerdo con la secuencia de consenso del PAM 5'-NNNN(A/G)TT-3' como diana en la reacción de corte. Se sintetizaron moléculas de ARNcr que dirigen PpCas9 a estos sitios a fin de reconocer estas secuencias.

Las reacciones de corte se efectuaron en condiciones seleccionadas para PpCas9; el resultado se muestra en la figura 9. La figura 9 demuestra que la enzima PpCas9 cortó satisfactoriamente tres de las cuatro dianas con PAM adecuado.

La diana del carril 6 tuvo una secuencia de PAM CAGCATT, la cual, según las predicciones basadas en los resultados del análisis de agotamiento, debería ser reconocida eficazmente por la proteína. Sin embargo, en este experimento, no se produjo el reconocimiento de este fragmento.

Por lo tanto, el PAM CAGCATT se verificó adicionalmente en otra diana protoespaciadora restringida al mismo PAM (figura 10). En este caso, el PAM fue reconocido eficazmente, lo que produjo el corte del ADN. Así, la proteína tiene algunas preferencias adicionales con respecto a la secuencia diana de ADN. Posiblemente, las preferencias se relacionan con la estructura secundaria de ADN.

Así, los estudios demostraron la presencia de actividad de nucleasa en PpCas9 y permitieron, además, determinar su secuencia de PAM y verificar las secuencias de las moléculas de ARN guía.

El complejo ribonucleoproteico de PpCas9 introduce específicamente rupturas en dianas restringidas al PAM 5'-NNNN(A/G)TT -3' desde el extremo 5' del protoespaciador. En la figura 11, se muestra el esquema del complejo PpCas9/ARN.

Ejemplo 2. Uso de ARN guía híbrido para el corte de una diana de ADN.

El ARNsg es una forma de ARN guía, que corresponde a una fusión de ARNtracr (ARN trazador) y ARNcr. Para seleccionar el ARNsg óptimo, se construyeron tres variantes de esta secuencia, cuya longitud fue diferente a la del dúplex ARNcr-ARNtracr. Las moléculas de ARN se sintetizaron *in vitro*, y se efectuaron con ellas experimentos de corte de diana de ADN (figura 12).

Se utilizaron las siguientes secuencias de ARN como moléculas de ARN híbridas:

1 - ARNsg1, 25 DR:

UAUCUCCUUUCAUUGAGCACGUUGUAGCUCCCUUUUUUCAUUUCGCGAAAGCGAAAUGAAAAACGUU
 GUUACAAUAAGAGAUGAAUUUCUCGCAAAGCTCTGCCUCUUGAAAUUUCGGUUUCAAGAGGCAUCU
 UUUU

2 - ARNsg2, 36 DR:

UAUCUCCUUUCAUUGAGCACGUUGUAGCUCCCUUUUUUCAUUUCGCGAGUGCUAUAUGAAAAUUUAU
 AGCACUGCGAAAUGAAAAACGUUGUUACAAUAAGAGAUGAAUUUCUCGCAAAGCUCUGCCUCUUGA
 AAUUUCGGUUUCAAGAGGCAUCUUUUU

La negrita indica una secuencia de 20 nucleótidos, que permite el apareamiento con la diana de ADN (fracción variable de ARNsg). Además, en el experimento, se utilizó una muestra de control sin ARN y un control positivo, el cual correspondió al corte de la diana con ARNcr + ARNtr.

Se utilizó una secuencia con el sitio de reconocimiento 5' tatctcctttcattgagcac 3', con la secuencia de consenso correspondiente PAMCAACATT como diana de ADN: 5'-
 cccggggtaccacggagagatggtggaaatcatctttctcgtgggcatccttgatggccacctcgt
 cggaagtgccacgaggatgacagcaatgccaatgctgggggggctcttctgagaacgagctctgc
 tgcttgacacggccaggacggccaacaccaaccagaacttgggagaacagcactccgctctgggct
 tcatcttcaactcgtcgactccctgcaaacacaaagaaagagcatgttaaaataggatctacatca
 cgtaacctgtcttagaagaggctagatactgcaattcaaggacct**tatctcctttcattgagcac**C
 AACATTcaactccat ctaccagcctactctcttatctctgttatt - 3'

La negrita indica el sitio de reconocimiento; las letras mayúsculas corresponden a PAM.

La reacción se produjo en las siguientes condiciones: La concentración de una secuencia de ADN con PAM (CAACATT) fue 20 nM; la concentración de proteína fue 400 nM; la concentración de ARN fue 2 μ M; el tiempo de incubación fue 30 minutos; la temperatura de incubación fue 37 °C.

Se descubrió que las moléculas de ARNsg1 y ARNsg2 seleccionadas fueron tan eficaces como las secuencias de ARNcr y ARNtracr nativas: El corte se produjo en más del 80 % de las dianas de ADN (figura 12).

Estas variantes de ARN híbrido pueden permitir el corte de cualquier otra molécula de ADN diana luego de modificar la secuencia que se aparee directamente con la diana de ADN.

Ejemplo 3. Proteínas Cas9 de organismos estrechamente relacionados con *P. pneumotropica*.

A la fecha, no se han caracterizado enzimas CRISPR-Cas9 en *P. pneumotropica*. La proteína Cas9 de *Staphylococcus aureus*, la cual tiene un tamaño similar, es idéntica a PpCas9 en un 28 % (en la figura 13, el grado de identidad se calculó con el software BLASTp (parámetros predeterminados)). Otras proteínas Cas9 presentan un grado similar de identidad (no aparece en la figura).

Así, la proteína PpCas9 tiene diferencias importantes en su secuencia de aminoácidos con respecto a otras proteínas Cas9 estudiadas hasta ahora.

Las personas versadas en ingeniería genética sobrentenderán que la variante de secuencia proteica de PpCas9 obtenida y caracterizada por el solicitante en la presente descripción se puede modificar sin cambiar la función de la proteína misma (por ejemplo, mediante mutagénesis dirigida de residuos de aminoácidos que no influyen directamente en la actividad funcional (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 1989, CSH Press, pág. 15.3 a 15.108)). Específicamente, las personas versadas reconocerán que se puede modificar residuos de aminoácidos no conservados sin influir en los residuos responsables de la funcionalidad proteica (determinación de la función o estructura de la proteína). Algunos ejemplos de estas modificaciones incluyen las sustituciones de residuos de aminoácidos no conservados por residuos homólogos. En la figura 12, se muestran algunas de las regiones con residuos de aminoácidos no conservados. En algunas modalidades de la invención, es posible el uso de una proteína, que comprenda una secuencia de aminoácidos al menos 95 % idéntica a la secuencia de aminoácidos de SEQ ID NO: 1 y que se diferencie de la SEQ ID NO: 1 únicamente en residuos de aminoácidos no conservados para obtener, en la molécula de ADN, una ruptura de cadena doble ubicada inmediatamente antes de la secuencia nucleotídica 5'-NNNN(A/G)TT-3' en dicha molécula de ADN. Se puede obtener proteínas homólogas mediante mutagénesis (por ejemplo, mutagénesis dirigida al sitio o mediada por PCR) de las moléculas de ácido nucleico correspondientes, para luego someter a pruebas a la proteína Cas9 codificada modificada de la conservación de sus

funciones de acuerdo con los análisis funcionales descritos en el presente documento.

Ejemplo 4. Modificación del ADN genómico de células humanas con PpCas9.

Para modificar el ADN genómico de células humanas, se clonó el gen de nucleasa PpCas9 en un vector plásmido eucarionte bajo el control de un promotor de citomegalovirus (CMV). En los extremos 5' y 3' del gen PpCas9, se agregaron secuencias que codifican señales de localización nuclear asegurándose de que se produjera la introducción de nucleasa al núcleo celular. La secuencia de ARNsg se introdujo en el vector bajo el control de promotor U6 mediante clonación. Para la prueba de la actividad del sistema, se utilizaron moléculas de ARNsg con una secuencia complementaria al ADN diana de 20 y 24 nucleótidos de longitud. Se utilizó, como control positivo, un plásmido similar portador de un sistema de modificación de ADN genómico a base de SpCas9, conocido en el estado del arte. Para evaluar la efectividad de la transfección, los plásmidos portaron, además, el gen de la proteína verde fluorescente (GFP). Se utilizaron las siguientes regiones de ADN genómico humano como dianas de ADN (tabla 3).

Tabla 3. Dianas de ADN de genes EMX1 y GRIN2B humanos.

Nucleasa	Nombre del sitio	Secuencia diana	PAM
PpCas9	EMX1.1 sg20	GCCCTTCCTCCTCCAGCTTC	GTT

PpCas9	EMX1.1 sg24	TCAGGCCCTTCCTCCTCCAGCTTC	GTT
FpCas9	EMX1.2 sg20	GGAGGTGACATCGATGTCCT	ATT
FpCas9	EMX1.2 sg24	CATTGGAGGTGACATCGATGTCCT	ATT
PpCas9	GRIN2B1.1 sg20	CAGCTGAAGTAATGTTAGAG	ATT
PpCas9	GRIN2B1.1 sg24	TTAGCAGCTGAAGTAATGTTAGAG	ATT
PpCas9	GRIN2B1.2 sg20	AATAAGAAAAACATTATTAT	ATT
PpCas9	GRIN2B1.2 sg24	ATAAAATAAGAAAAACATTATTAT	ATT
SpCas9	EMX1 sg20	GAGTCCGAGCAGAAGAAGAA	GGG
SpCas9	GRIN2B sg20	ACCTTTTATTGCCTTGTTCA	AGG

Los sitios EMX1.1 y EMX1.2 fueron dos sitios de modificación diferentes del gen EMX1; asimismo, GRIN2B1.1 y GRIN2B1.2 fueron dos sitios de modificación diferentes del gen GRIN2B.

Las dianas de ADN se flanquearon por el extremo 3' con las secuencias de PAM de PpCas9 5'-NNNNRTT-3' o de SpCas9 5'-NGG-3'.

Para lograr la actividad efectiva de la nucleasa PpCas9 en células eucariontes, es necesario importar la proteína al núcleo de una célula eucarionte por medio de una señal de localización nuclear del antígeno T de SV40 (Lanford *et al.*, Cell 1986, 46: 575-582) unido a la secuencia de PpCas9 a través de una secuencia espaciadora descrita en Shen B. *et al.*, "Generation of gene-

modified mice via Cas9/RNA-mediated gene targeting", Cell Research 2013, May;23(5):720-3 o sin la secuencia espaciadora.

En el ejemplo, la secuencia completa de aminoácidos de la nucleasa transportada al interior del núcleo de las células humanas fue la siguiente:

MAPKKKRKVGIIHGVPAAEQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVA
KTGESLALSRRRLARSSRRLIKRRAERLKKAKRLLKAEKILHSIDEKLPINWQLRVKGLKEKLERQ
EWAAVLLHLHSKHRGYLSQRKNEGKSDNKELGALLSGIASNHQMLQSSEYRTPAEIIVKKFQVEEGH
IRNQRGSYTHTF SRLDLLAEMELLFQRQAELGNSYTTTLLLENLTALLMWQKPALAGDAILKMLGK
CTFEPSEYKAAKNSYSAERFVWLTKLNNLRILENGTERALNDNERFALLEQPYEKSCLTYAQVRAM
LALSDNAIFKGVRYLGEDKKTVESKTTLIEMKFYHQIRKTLGSAELKKEWNEKGNLSDLLDEIGTA
FSLYKTDDDICRYLEGKLPERVLNALLLENLNFDFKFIQLSLKALHQILPLMLQGQRYDEAVSAIYGD
HYGKKSTETTRLLPTIPADEIRNPVVLRTLTQARKVINAVVRLYGSPARIHIETAREVGKSYQDRK
KLEKQQEDNRKQRESAVKKFKEMFPHFVGEPKGDILKMRLYELQQAKCLYSGKSLELHRLLEKGY
VEVDHALPFSRTWDDSFNNKVLVLNENQNKGNLTPYEWLDGKNNSERWQHFFVVRVQTSGFSAKK
QRILNHKLDEKGFIERNLNDTRYVARFLCNFIADNMLLVGKGKRNVFASNGQITALLRHRWGLQKV
REQNDRHHALDAVVVACSTVAMQQKITRFVRYNEGNVFSGERIDRETGEIIPLHFPSPAFFKENV
EIRIFSENPKLELENRLPDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSARKLNEGLSVLKVP
LTQLKLSDLERMVNRDREIALYESLKARLEQFGNDPAKAFAPFYKKGGALVKAVRLEQTQKSGVL
VRDGNGVADNASMVRVDVFTKGGKYFLVPIYTWQVAKGILPNRAATQGKDENDWDIMDEMATFQFS
LCQNDLIKLVTKKKTIFGYFNGLNRATSNINIKEHDLDKSKGKLGIIYLEVGVKLAISLEKYQVDEL
GKNIRPCRPTKRQHVRFKRPAATKKAGQAKKKK

La secuencia del plásmido utilizado en este experimento fue la siguiente:

gagggcctatttcccatgattccttcataatttgcatatacgatacaaggctgtagagagat
aattggaattaatttgactgtaaacacaaagatatttagtacaaaatacgtgacgtagaaagtaata
atctcttgggtagtttgagtttttaaattatgttttaaattggactatcatatgcttaccgtaac
ttgaaagtatttcgatttcttggtttatatatcttggtgaaaggacgaaacaccgXXXXXXXXXX
XXXXXXXXXXXXXXXXGTTGTAGCTCCCTTTTTCATTTTCGCGAAAGCGAAATGAAAAACGTTGTTACAA
TAAGAGATGAATTTCTCGCAAAGCTCTGCCTCTTGAAATTTTCGGTTTCAAGAGGCATCTTTTTTgc
tTCTCATGTCCAATATGACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATT
ACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTacataacttacggtaaatggcccg
cctggctgaccgcccacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacg
ccaatagggactttccattgacgtcaatgggtggagtatttacggtaaaactgcccacttggcagta
catcaagtgtatcatatgccaagtccgccccctattgacgtcaatgacggtaaatggcccgctgg
cattatgcccagttacatgaccttacgggactttcctacttggcagttacatctacgtatttagtcac
gctattaccatgggtgatgcggttttggcagttacaccaatgggctggatagcggtttgactcacgg
ggatttccaagtctccacccattgacgtcaatgggagtgttttggcaccaaaatcaacgggac
tttccaaaatgtcgtaataacccccgccccgttgacgcaaattgggctgtaggcgtgtacgggtgggag
gtctatataagcAGAGCTCGTTTAGTGAACCGTCAGAATTAATTCAGATCGATCTACCaccgcccac
cATGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGAAC**CAGAATAA**
TCCGCTTAACTACATTCTTGGGCTGGATT**TGGGAATTGCGAGTATAGGCTGGGCGGTGGTTGAGAT**
CGATGAAGAGAGTAGTCCGATACGCCTTATCGACGTTGGAGTTAGGACGTTTCGAGAGGGCGGAGGT
CGCCAAGACCGGTGAGAGCTTGGCCCTCAGCCGGCGGCTCGCTCGATCTAGTCGCAGGCTTATAAA
GAGGAGGGCTGAGCGCCTTAAGAAAGCTAAGAGGCTCCTTAAGGCAGAAAAAATTCTGCATAGTAT
CGACGAAAAGCTGCCGATAAATGTTTGGCAGCTCCGAGTAAAAGGGCTGAAGGAAAAATTGGAAAG
GCAGGAGTGGGCGGCGGTACTGCTTCATCTCTCCAAGCACCGGGGCTATCTGTCTCAGCGAAAAAA
CGAAGGTAAGTCAGACAACAAGGAGCTGGGCGCACTTTTGTCCGGGATAGCGTCAAATCATCAGAT
GCTCCAATCAAGTGAGTATCGGACCCCTGCGGAGATCGCCGTTAAAAAGTTTCAAGTTGAGGAGGG
CCACATCAGAAATCAGAGGGGGTCTTACACCCATACGTTCTCTAGACTCGACCTCCTTGCGGAAAT
GGAACCTCCTGTTTCAGCGCCAGGCGGAGCTTGGTAACTCCTACACGTCCACTACCCTCCTGGAAAA

CCTGACAGCCCTGCTGATGTGGCAGAAGCCCGCTTTGGCGGGGGATGCCATCCTGAAGATGCTGGG
TAAATGCACCTTTGAGCCGTCAGAATATAAAGCCGCCAAGAATAGTTACTCTGCGGAGCGATTTGT
TTGGTTGACAAAGTTGAATAACCTGCGCATCCTGGAGAACGGTACCGAGCGCGCACTCAATGATAA
TGAGCGCTTCGCCCTCCTGGAACAGCCCTACGAGAAGTCCAAGCTCACCTACGCCCAAGTCAGAGC
CATGCTGGCTCTTAGTGACAACGCGATTTTTTAAGGGCGTGCGATACTTGGGCGAGGATAAGAAAAC
CGTAGAGTCAAAAACGACTCTGATCGAGATGAAATTCTATCACCAAATTAGAAAGACCCTCGGTTC
TGCCGAGCTGAAAAGGAATGGAACGAACCTTAAGGGTAACAGCGACCTGCTCGATGAAATCGGTAC
CGCATTTAGCCTTTATAAACGGACGACGACATCTGCCGATATTTGGAGGGGAAGCTCCCAGAGCG
AGTATTGAATGCACTCCTTGAGAACCTTAATTTTGACAAGTTCATTCAGCTGTCCCTCAAAGCACT
GCATCAAATCCTCCCACTTATGCTGCAAGGACAACGATACGACGAAGCCGTCAGCGCGATATATGG
AGATCATTACGGAAAAAGTCCACCGAGACCACACGACTGCTTCCTACGATCCCCGCCGATGAGAT
CAGAAATCCCGTAGTCCTTCGAACACTTACTCAGGCTAGGAAGGTGATTAATGCGGTAGTTAGGTT
GTATGGATCTCCGGCACGGATACATATAGAAACAGCTCGCGAAGTGGGTAAATCTTACCAAGACCG
CAAGAAATTGGAGAAACAACAGGAGGATAACCGAAAGCAACGAGAATCTGCCGTTAAAAAGTTTAA
GGAAATGTTTCCTCACTTTGTAGGAGAACCGAAGGGTAAAGATATCTTGAAAATGCGGTTGTACGA
GTTGCAGCAAGCTAAGTGTCTCTATAGCGGCAAGAGTTTGGAATTGCACCGCCTCCTGGAGAAAGG
CTACGTGGAAGTAGACCATGCGCTCCCGTTTTCCCGAACCTGGGATGATTCTTTCAATAACAAAGT
CCTTGTGCTGGCAAATGAGAACCAGAACAAAGGAAATCTGACTCCTTATGAGTGGTTGGATGGCAA
GAATAATTCTGAGCGGTGGCAACATTTGTTGTCCGCGTCCAAACGTCAGGGTTCAGCTATGCTAA
GAAACAAAGGATCCTCAATCACAAGCTCGACGAGAAAGGATTCATAGAACGAAATTTGAATGACAC
TAGGTATGTGGCTCGATTTCTCTGCAATTTTATTGCTGACAATATGCTCCTCGTTGGGAAGGGAAA
GCGGAATGTTTTTGCATCAAATGGGCAGATAACGGCGCTCTTGAGACATAGATGGGGGCTGCAAAA
GGTGAGAGAGCAAAATGATAGACATCACGCCCTGGATGCCGTTGTAGTCGCCTGTTCAACGGTTGC
GATGCAGCAAAAGATCACTCGGTTCGTTAGGTATAACGAAGGGAACGTTTTTTAGTGGAGAGCGCAT
AGATCGGGAAACAGGCGAAATCATCCCTTTGCATTTCCCAAGTCCTTGGGCTTTTTTCAAAGAGAA
TGTGGAAATAAGGATATTCAGTGAAAACCCTAAGTTGGAGCTTGAGAATCGGTTGCCCGATTATCC
CCAGTACAATCATGAGTGGGTTC AACCGCTGTTTCGTATCCCGCATGCCAACCCGAAAGATGACCGG

GCAGGGTCACATGGAGACTGTGAAATCTGCAAAGAGACTTAATGAGGGCCTGTCAGTGTTGAAGGT
GCCCTTGACTCAACTGAAATTGAGCGACCTCGAGCGCATGGTAAACCGCGATAGAGAAATCGCACT
TTATGAGAGTCTGAAGGCGCGATTGGAACAATTCGGTAATGATCCGGCAAAGGCTTTCGCTGAGCC
ATTCTACAAGAAGGGTGGAGCGCTGGTTAAGGCTGTCCGACTCGAACAGACACAAAAGTCAGGGGT
CTTGGTCAGAGATGGTAACGGGGTTGCCGACAACGCCTCCATGGTACGAGTAGATGTTTTTCACGAA
AGGAGGAAAATACTTTCTGGTACCTATCTATACCTGGCAAGTTGCCAAGGGAATACTCCCGAATAG
GGCGGCGACCCAGGGAAAGGATGAAAACGACTGGGATATAATGGATGAAATGGCTACGTTTCAGTT
TAGCTTGTGCCAGAATGACCTCATAAACTGGTAACCAAAAAAAGACTATATTCGGGTATTTCAA
TGGCCTTAATCGGGCAACTTCCAATATCAACATCAAGGAACATGATCTGGATAAGAGCAAGGGAAA
GCTTGGTATCTATCTCGAAGTTGGAGTCAAGCTCGCTATTTCCCTCGAGAAATATCAAGTAGATGA
ACTGGGAAAGAATATACGGCCATGCCGGCCCCACAAAAGACAACACGTACGGTTCAAAAGGCCGGC
GGCCACGAAAAAGGCCGGCCAGGCAAAAAGAAAAAGGGATCCTACCCATACGATGTTCCAGATTA
CGCTTATCCCTACGACGTGCCTGATTATGCATACCCATATGATGTCCCCGACTATGCCGGCGCAAC
AAACTTCTCTCTGCTGAAACAAGCCGGAGATGTGGAAGAGAATCCTGGACCGgtgagcaagggcgga
ggagctgttcaccgggggtggtgcccattcctggtcgagctggacggcgacgtaaacggccacaagtt
cagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatctgcac
caccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgctt
cagccgctaccccgaccacatgaagcagcagcacttcttcaagtccgcatgcccgaaggctacgt
ccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgcccagggtgaagttcga
gggcgacaccctggtgaaccgcatcgagctgaagggtcacttcaaggaggacggcaacatcct
ggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaa
cggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgacca
ctaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcac
ccagtccgacctgagcaaagaccccaacgagaagcgcgatcacatgggtcctgctggagttcgtgac
cgccgcccgggatcactctcggcattggacgagctgtacaagTAA

En la secuencia del plásmido, se distinguieron las siguientes fracciones: el promotor de U6 (primera región; letras mayúsculas), la secuencia complementaria al protoespaciador ("XXX-XXX"), la fracción conservada de ARNsg (tercera región; letras mayúsculas), el gen PpCas9 (destacado en negrita) y el gen GFP (última región; letras mayúsculas).

Se transfectaron plásmidos con PpCas9 o SpCas9 al cultivo de células HEK293T humanas con el reactivo Lipofectamine 2000. Luego de 72 horas de la transfección, las células se lisaron, y los lisados resultantes se sometieron a PCR para generar regiones que incluyan sitios de modificación de dianas de ADN genómico. Los fragmentos de PCR resultantes se sometieron a una reacción *in vitro* con endonucleasa I de T7 para determinar la frecuencia de inserciones y deleciones en los sitios diana de ADN genómico. Los productos de reacción se aplicaron a gel de agarosa y se sometieron a electroforesis. La figura 14A demuestra que la proteína PpCas9 introduce activamente modificaciones en los genes EMX1 y GRIN2b, con una eficacia similar a la de la nucleasa SpCas9 descrita en el arte previo.

Este experimento demostró que, para modificar efectivamente ADN genómico, la proteína PpCas9 requiere moléculas de ARNsg alargadas en comparación con las de SpCas9. En el ejemplo, la eficacia de las modificaciones genéticas es mayor cuando se utiliza ARNsg con una secuencia complementaria a la diana de ADN con 24 nucleótidos de longitud (en comparación con 20 nucleótidos de longitud).

A través de una secuenciación de alto rendimiento, se confirmaron las modificaciones introducidas en los sitios de ADN diana. La figura 14B muestra ejemplos de modificaciones detectables realizadas a la secuencia nucleotídica del gen EMX1.

La introducción en forma de un complejo ribonucleico se puede emplear, además, para introducir proteína NLS_PpCas9-NLS en células humanas. Esto se lleva a cabo mediante la incubación de una forma recombinante de señal de localización nuclear de PpCas9 con moléculas de ARN guía en el buffer CutSmart (NEB). La proteína recombinante se produce a partir de células productoras bacterianas mediante la purificación de la primera a través de cromatografía de afinidad (NiNTA, Qiagen) con exclusión por tamaño (Superdex 200).

La proteína se mezcla con moléculas de ARN en una relación de 1:2 (PpCas9 NLS:ARNsg); la mezcla se incuba durante 10 minutos a temperatura ambiente, para luego transfectarse a las células.

A continuación, el ADN extraído de allí se analiza en busca de inserciones o deleciones en el sitio de ADN diana (según la descripción anterior).

La nucleasa PpCas9 de la bacteria ***Pasteurella pneumotropica*** caracterizada en la presente invención se puede introducir, para modificar el ADN, en células de distintos orígenes a través de métodos estándar que conocen las personas versadas. La nucleasa PpCas9 tiene una serie de ventajas con respecto a proteínas Cas9 caracterizadas previamente.

La nucleasa PpCas9 tiene un PAM corto de dos letras, distinto de otras nucleasas Cas conocidas, necesario para que el sistema funcione. La invención ha demostrado que la presencia de un PAM corto (RTT) ubicado a 4 nucleótidos del protoespaciador es suficiente para que la nucleasa PpCas9 funcione satisfactoriamente *in vivo*.

La gran cantidad de nucleasas Cas de menor tamaño conocidas hasta ahora, capaces de introducir rupturas de cadena doble en el ADN, tienen secuencias complejas de PAM de varias letras, lo que restringe las opciones de secuencias adecuadas para corte. Entre las nucleasas Cas estudiadas hasta ahora, que reconocen PAM cortos, únicamente las nucleasas PpCas9 puede reconocer secuencias flanqueadas por el patrón de RTT.

La segunda ventaja de la nucleasa PpCas9 es el pequeño tamaño de la proteína (1055 a.a.r.). A la fecha, es la única proteína de tamaño pequeño estudiada que tiene una secuencia de PAM de RTT de tres letras.

La nucleasa PpCas9 es una novedosa nucleasa Cas de pequeño tamaño, con un PAM corto y fácil de usar que se diferencia de las secuencias de PAM de otras nucleasas conocidas en la actualidad. La proteína PpCas9 corta con alta eficacia distintas dianas de ADN, incluido ADN genómico de células humanas a 37 °C y se puede convertir en la base de una nueva herramienta de edición genómica.

Aunque la invención se ha descrito aludiéndose a las modalidades divulgadas, las personas versadas en el arte

sobrentenderán que las modalidades específicas descritas en detalle se proporcionan con el fin de ilustrar la presente invención y no se deben interpretar como que limitan de alguna manera el ámbito de la invención. Se sobrentiende que se puede incorporar diferentes modificaciones sin desviarse del espíritu de la presente invención.

REIVINDICACIONES

1. Uso de una proteína, que comprende la secuencia de aminoácidos de SEQ ID NO: 1 o una secuencia de aminoácidos al menos 95 % idéntica a la secuencia de aminoácidos de SEQ ID NO: 1, y se diferencia de la SEQ ID NO: 1 únicamente en residuos de aminoácidos no conservados para obtener, en la molécula de ADN, una ruptura de cadena doble ubicada inmediatamente antes de la secuencia nucleotídica 5'-NNNN(A/G)TT-3' en dicha molécula de ADN.

2. El uso según la reivindicación 1, CARACTERIZADO porque la ruptura de cadena doble en la molécula de ADN se produce a una temperatura de 35 °C a 45 °C.

3. El uso de la proteína según la reivindicación 1, CARACTERIZADO porque la proteína comprende la secuencia de aminoácidos de SEQ ID NO: 1.

4. El uso según la reivindicación 1, CARACTERIZADO porque la ruptura de cadena doble de la molécula de ADN se produce en el ADN genómico de una célula de mamífero.

5. El uso según la reivindicación 4, CARACTERIZADO porque la ruptura de cadena doble de la molécula de ADN produce la modificación del ADN genómico de dicha célula de mamífero.

6. Un método de modificación de una secuencia de ADN genómico en una célula de un organismo unicelular o multicelular, que comprende el ADN genómico, CARACTERIZADO porque incluye la introducción a dicha célula del organismo de una cantidad efectiva

de: a) una proteína, que comprende la secuencia de aminoácidos de SEQ ID NO: 1, o un ácido nucleico, que codifica la proteína que comprende la secuencia de aminoácidos de SEQ ID NO: 1, y b) un ARN guía, que comprende una secuencia que forma un dúplex con la secuencia nucleotídica de una región del ADN genómico del organismo, la cual se encuentra directamente adyacente a la secuencia nucleotídica 5'-NNNN(A/G)TT-3' e interactúa con dicha proteína luego de la formación del dúplex, o una secuencia de ADN que codifica dicho ARN guía;

donde la interacción de dicha proteína con el ARN guía y la secuencia nucleotídica 5'-NNNN(A/G)TT-3' produce la formación de una ruptura de cadena doble en la secuencia de ADN genómico inmediatamente adyacente a la secuencia 5'-NNNN(A/G)TT-3'.

7. El método según la reivindicación 6, CARACTERIZADO porque comprende, además, la introducción de una secuencia exógena de ADN simultáneamente con el ARN guía.

8. El método según la reivindicación 6, CARACTERIZADO porque dicha célula es una célula de mamífero.

USO DE PROTEÍNA CAS9 DE LA BACTERIA *PASTEURELLA PNEUMOTROPICA*

RESUMEN

La presente invención describe una novedosa nucleasa bacteriana del sistema CRISPR-Cas9 de la bacteria *P. pneumotropica*, así como también su uso para producir rupturas de cadena doble estrictamente específicas en una molécula de ADN. Esta nucleasa tiene propiedades inusuales y se puede utilizar como herramienta de modificación de la secuencia de ADN genómico en la célula de un organismo unicelular o multicelular. Así, aumenta la versatilidad de los sistemas CRISPR-Cas9 disponibles, hecho que permite el uso de distintas variantes de nucleasas Cas9 para el corte de ADN genómico o plásmido de distintos organismos en un número mayor de sitios específicos y/o en condiciones específicas.