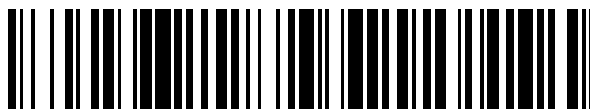


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 768 976**

51 Int. Cl.:

C12N 9/22

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.04.2016 PCT/EP2016/058442**

87 Fecha y número de publicación internacional: **20.10.2016 WO16166340**

96 Fecha de presentación y número de la solicitud europea: **15.04.2016 E 16724290 (8)**

97 Fecha y número de publicación de la concesión europea: **11.12.2019 EP 3283625**

54 Título: **Edición de genoma mediada por nucleasa**

30 Prioridad:

16.04.2015 GB 201506509

18.12.2015 US 201562269143 P

24.03.2016 US 201662312724 P

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

24.06.2020

73 Titular/es:

WAGENINGEN UNIVERSITEIT (100.0%)

Droevendaalsesteeg 4

6708 PB Wageningen, NL

72 Inventor/es:

VAN DER OOST, JOHN

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 768 976 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

Edición de genoma mediada por nucleasa

5 Campo de la invención

La invención se refiere al campo de herramientas, procedimientos y técnicas de ingeniería genética para la edición de genomas o genes. Dicha edición o manipulación de secuencias de polinucleótidos, incluidas secuencias de genes estructurales o de control, tiene aplicación en muchos campos de la salud y la biotecnología, por ejemplo en

10 tratamientos de terapia génica de seres humanos o animales, el cultivo de plantas y la cría de animales y la mejora de organismos industriales, por ejemplo mediante la alteración de enzimas y rutas metabólicas, particularmente microorganismos; también en los sectores de la biología sintética y la producción de biocombustibles a partir de algas, por ejemplo. Además, la invención se refiere también a herramientas y procedimientos de investigación para su uso en investigación científica básica que implica genética molecular.

15 Antecedentes de la invención

Las nucleasas específicas del sitio pueden permitir la generación de roturas bicatenarias (DSB) en posiciones seleccionadas a lo largo de una cadena de ADN. En un organismo de interés, esto permite que las DSB se realicen

20 en posiciones predeterminadas en el genoma. La creación de dichas roturas por medio de nucleasas específicas del sitio hace que la maquinaria de reparación celular endógena se readapte para insertar, eliminar o modificar el ADN en posiciones deseadas en el genoma de interés. La escisión dirigida de ADN mediada por nucleasas específicas del sitio es, por lo tanto, una herramienta de investigación básica importante que ha facilitado la determinación funcional y la anotación de genes específicos, pero, entre otras cosas, también ha permitido la mutación, la adición,

25 la sustitución o la modificación dirigida de genes en organismos de importancia agrícola, industrial o comercial. Como la base genética de fenotipos orgánicos tanto deseados como no deseados se revela a través de la secuenciación del ADN, la capacidad para generar alteraciones dirigidas en loci genómicos específicos es fundamental para la ingeniería genética de rasgos útiles y en el desarrollo de tratamientos clínicos para enfermedades con una base genética.

30 Otros enfoques de nucleasas específicas del sitio implican roturas de ácido nucleico diana monocatenario, ya sea individualmente o en combinación.

Durante el decenio pasado se han desarrollado una serie de herramientas moleculares para permitir la ingeniería genética específica en general, y para la edición especializada de genomas eucariotas en particular. Inicialmente, se desarrollaron nucleasas con dedos de zinc (ZFN), seguidas de nucleasas efectoras de tipo activador de la transcripción (TALEN). Recientemente, se ha producido una revolución mediante el desarrollo de la nucleasa Cas9 asociada a CRISPR, como una alternativa muy eficaz, genérica y barata para la cirugía genómica especializada en una serie de células eucariotas (desde levaduras y plantas hasta peces cebras y seres humanos) (revisado por Van der Oost 2013, Science 339: 768-770 y Charpentier y Doudna, 2013, Nature 495: 50-51).

40

Se han descubierto y aislado muchas nucleasas específicas del sitio útiles a partir de procariotas. Al igual que los eucariotas, los organismos procariotas poseen un conjunto variable de sistemas de defensa para protegerse contra los virus. Las estrategias de defensa que protegen a su huésped microbiano contra el ADN invasor se basan principalmente en sistemas de inmunidad generales (innatos), tales como las, bien conocidas, enzimas de restricción.

45

Un descubrimiento reciente importante en este sector ha sido la demostración de un sistema inmunitario específico (adaptativo) en bacterias y arqueas. Este sistema inmunitario adaptativo consiste en repeticiones palindrómicas interesparadas regularmente agrupadas (CRISPR) y genes Cas asociados a CRISPR que codifican las proteínas Cas. El sistema CRISPR-Cas utiliza pequeños ARN de CRISPR que guían las proteínas Cas efectoras a los ácidos nucleicos invasores complementarios, eventualmente neutralizando la invasión. Se distinguen dos clases de complejos efectoras de Cas: complejos de múltiples subunidades (por ejemplo, Cascade de *E. coli*) y sistemas de una única proteína (por ejemplo Cas9 de *Streptococcus pyogenes*) (Van der Oost *et al.*, 2014, Nature Rev. Microbiol. 12: 479-492).

50

55

Los análisis moleculares de CRISPR-Cas han proporcionado la base para el desarrollo de herramientas de ingeniería del genoma. Cas9 es un complejo efector CRISPR-Cas relativamente sencillo que puede expresarse funcionalmente en una amplia serie de células procariotas y eucariotas. Es importante indicar que la guía de ARN de Cas9 se puede manipular fácilmente para que se dirija específicamente a cualquier secuencia de interés. Aunque también es posible ajustar la especificidad para un determinado gen diana con el sistema TALEN, un inconveniente de este sistema es que esto requiere una laboriosa ingeniería de proteínas. En el caso de Cas9, solo se debe generar y clonar un oligonucleótido corto, lo que ahorra tiempo y dinero. Las aplicaciones del sistema Cas9 incluyen ingeniería genética general (alteración, reparación e integración de genes), control de la expresión génica (estimulación y silenciamiento) y marcado de genes (toma de imágenes). La coexpresión de Cas9 con diferentes guías permite la multiplexación, por ejemplo, generar múltiples inactivaciones génicas simultáneamente.

60

65

El sistema CRISPR-Cas permite la escisión específica de la diana del ADN genómico guiada por la nucleasa Cas9 en complejo con un ARN guía (ARNg) que se une complementariamente a una secuencia objetivo de 20 nucleótidos. Por lo tanto, la alteración de la secuencia del ARNg permite que la endonucleasa Cas9 se programe para cortar ADN bicatenario en sitios complementarios al ARN guía de 20 pares de bases. El sistema Cas9 se ha utilizado para modificar genomas en múltiples células y organismos.

En comparación con sistemas alternativos de edición de genoma (nucleasas de dedo de cinc, TALEN), la ingeniería genética por medio de Cas9 es muy eficaz, barata y rápida.

A pesar de estos desarrollos, el sistema Cas9 todavía tiene algunos inconvenientes prácticos. En primer lugar, basándose en un mecanismo intrínseco de discriminación propia/discriminación no propia, el Cas9 requiere un motivo de secuencia (motivo adyacente protoespaciador, PAM) en la región flanqueante adyacente a la secuencia diana. El requerimiento del PAM impone una limitación de diseño significativa al sistema de endonucleasa, que excluye potenciales sitios diana.

En segundo lugar, aunque las nucleasas guiadas por ARN tales como Cas9 incorporan ARN guía que se dirigen a la escisión de sitios diana específicos y, por lo tanto, muestran una reducción en la actividad significativa fuera de la diana observada en la mayor parte de las otras nucleasas disponibles, se produce aún un determinado nivel de escisión fuera de la diana (Pattanayak *et al.*, 2013, Nat. Biotechnol. 31: 839-843), es decir, la escisión de secuencias genómicas que difieren de la secuencia diana prevista en uno o más nucleótidos. Generalmente, se requieren 15-17 nucleótidos para el emparejamiento de bases con una diana complementaria de 20 nucleótidos; se ha planteado la hipótesis de la tolerancia a los emparejamientos incorrectos para explicar los problemas fuera de la diana de los que se ha informado. La especificidad imperfecta de la unión específica del sitio modificado mediante ingeniería genética puede conducir a la inserción, modificación o eliminación no deseada de loci genómicos durante un evento de direccionamiento génico, que se ha asociado con toxicidad celular. Las consecuencias de dichos eventos de escisión fuera de la diana que dan como resultado alteraciones no deseadas de los loci genómicos distintos de la diana deseada pueden ser extremadamente graves en un contexto clínico.

La escisión específica de secuencia del sitio diana de la nucleasa prevista en ausencia de actividad de escisión fuera de la diana, o con solo una actividad mínima de fondo, es un requisito previo para la manipulación genómica de alta eficacia en aplicaciones de investigación básicas y especialmente para evitar la escisión de genes que no se deseaban escindir durante las modificaciones genómicas dirigidas asociadas con aplicaciones clínicas de las tecnologías de endonucleasas específicas del sitio, en particular debido a que las roturas bicatenarias resultantes dan como resultado modificaciones genómicas estables y heredables.

A pesar de que se prestó mucha atención a abordar estas características no deseadas del sistema Cas9, hasta la fecha siguen ampliamente sin resolverse.

La especificidad imprecisa, en particular, sigue siendo una dificultad y solo se ha abordado parcialmente expandiendo la secuencia diana que se desea reconocer mediante dímeros de Cas9 catalíticamente inactivado fusionado al dominio de nucleasa de FokI (dCas9-FokI) (Guilinger *et al.*, 2014, Nat. Biotechnol. 32: 577-582). Además, se ha demostrado que las variantes de nickasa modificadas genéticamente de Cas9 (en las que uno de los dos sitios de nucleasa está alterado) facilitan la reparación dirigida por homología en genomas eucariotas con una especificidad aumentada y una actividad reducida fuera de la diana (Ran *et al.*, 2013, Cell 154: 1380-1389. También, Mali *et al.*, 2013, Nat. Biotechnol. 31: 833-838).

El documento WO 2015/035139 describe composiciones, procedimientos, sistemas y kits para controlar la actividad y/o mejorar la especificidad de las endonucleasas programables por ARN, tales como Cas9. Por ejemplo, los ARN guía (ARNg) están diseñados para que existan en un estado "activado" o "desactivado", lo que controla la actividad de unión y, por lo tanto, de escisión de las endonucleasas programables por ARN. También se describen ARNg de detección de ARNm que modulan la actividad de las endonucleasas programables por ARN, en función de la presencia o ausencia de un ARNm diana. Se describen algunos ARNg que modulan la actividad de una endonucleasa programable por ARN basándose en la presencia o ausencia de un ADN extendido (ADNx).

Otro enfoque para mitigar la actividad fuera de la diana se ha centrado en el desarrollo de paquetes informáticos para asistir al proceso de diseño de ARN guía mediante la realización de búsquedas exhaustivas de secuencias diana frente a secuencias de referencia genómicas, lo que permite la selección de secuencias diana con efectos de escisión mínimos fuera de la diana (Naito *et al.*, 2015, Bioinformatics 31: 1120-1123). Sin embargo, esto simplemente permite una exploración eficaz del espacio de secuencia diana disponible para el diseño de secuencias guía en lugar de abordar directamente las limitaciones inherentes de CRISPR-Cas9 como herramienta de edición de genoma.

Por lo tanto, las nucleasas disponibles actualmente, incluidos los sistemas CRISPR-Cas9, no se encuentran en su estado actual de desarrollo en una forma necesariamente adecuada para la mayor parte de las aplicaciones clínicas o, de hecho, muchas otras aplicaciones de edición de genoma sensibles a la diana. Existe una necesidad continua

de herramientas de edición de genoma con mayor especificidad y fiabilidad inherentes de las que están disponibles actualmente en la técnica.

Schunder *et al.* proporcionaron la primera indicación de un sistema CRISPR/Cas funcional en *Francisella tularensis* (Schunder *et al.*, 2013, International Journal of Medical Microbiology 303: 51-60). Sin embargo, hasta la fecha la estructura y la funcionalidad del sistema siguen siendo inciertas.

Posteriormente, Vestergaard *et al.* proporcionaron una clasificación de todos los sistemas inmunitarios adaptativos CRISPR conocidos de Archaea basada principalmente en sus secuencias de proteínas Cas concatenadas, en el que Cas_Cpf1 se identificó como un único sistema de interferencia de proteínas que carece de Cas3, Cas5, Cas7 y Cas8, que recuerda a Cas9 en sistemas bacterianos tipo II a pesar de no parecer compartir ningún dominio estructural (Vestergaard *et al.*, 2014, RNA biology 11.2 (2014): 156-167).

Sumario de la invención

Tratando de superar determinadas desventajas prácticas asociadas con los sistemas Cas9, los inventores proporcionan una nucleasa novedosa (Cpf1) no relacionada con Cas9 para su aplicación como herramienta de edición de genes. Se ha descubierto que Cpf1 tiene características mecanicistas excepcionalmente ventajosas tales como un dominio de nucleasa único y un motivo PAM aguas arriba y se puede aplicar como una herramienta mejorada para la edición especializada del genoma en general, y para reparar trastornos genéticos de células madre humanas. Además, la nucleasa Cpf1 puede funcionar como parte de un sistema de ingeniería multiplex para microorganismos.

En consecuencia, la presente invención proporciona un vector de expresión tal como se establece en las reivindicaciones 1 a 9, 11, 12, 14 o 15; un sistema vector tal como se establece en las reivindicaciones 10 a 12; y una célula huésped tal como se establece en la reivindicación 13.

En el presente documento se describe un polipéptido aislado o un fragmento del mismo, que comprende la secuencia de aminoácidos SEQ ID NO: 1 o una secuencia con al menos 60% de identidad con el mismo, y que tiene una actividad de nucleasa.

El polipéptido o el fragmento puede comprender una secuencia de aminoácidos con al menos el 75%; preferentemente al menos el 85%; de forma más preferida al menos el 90%; de forma incluso más preferida al menos el 95% de SEQ ID NO: 1.

Además de la SEQ ID NO: 1 de referencia, también se describe cualquier secuencia variante que tenga el porcentaje de identidad definido con la misma. Dicho porcentaje de identidad incluye cualquiera de los siguientes: se divulga una secuencia de aminoácidos o ácidos nucleicos de referencia y secuencias con al menos un determinado porcentaje de identidad, por ejemplo al menos el 60%, pudiendo ser diferente opcionalmente el porcentaje de identidad. Por ejemplo: un porcentaje de identidad que se selecciona entre uno de los siguientes: al menos el 60%, al menos el 61%, al menos el 62%, al menos el 63%, al menos el 64%, al menos el 65%, al menos el 66%, al menos el 67%, al menos el 68%, al menos el 69%, al menos el 70%, al menos el 71%, al menos el 72%, al menos el 73%, al menos el 74%, al menos el 75%, al menos el 76%, al menos el 77%, al menos el 78%, al menos el 79%, al menos el 80%, al menos el 81%, al menos el 82%, al menos el 83%, al menos el 84%, al menos el 85%, al menos el 86%, al menos el 87%, al menos el 88%, al menos el 89%, al menos el 90%, al menos el 91%, al menos el 92%, al menos el 93%, al menos el 94%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99%, al menos el 99,5% o al menos el 99,8%. Dicha identidad de secuencia con una secuencia de aminoácidos SEQ ID NO: 1 es una función del número de posiciones idénticas compartidas por las secuencias en una ventana de comparación seleccionada, teniendo en cuenta el número de espacios que deben introducirse para un alineamiento óptimo de las dos secuencias y la longitud de cada espacio.

En cualquier aspecto de la presente invención, los residuos de aminoácidos pueden sustituirse de forma conservativa o no conservativa. Las sustituciones conservativas de aminoácidos se refieren a aquellas en las que los residuos de aminoácidos se sustituyen por otros residuos de aminoácidos con propiedades químicas similares (por ejemplo, carga o hidrofobicidad) y, por lo tanto, no alteran las propiedades funcionales del polipéptido resultante. De forma similar, el lector experto apreciará que las secuencias de ácido nucleico pueden sustituirse de forma conservativa o no conservativa sin afectar a la función del polipéptido. Los ácidos nucleicos modificados conservativamente son aquellos sustituidos por ácidos nucleicos que codifican variantes idénticas o funcionalmente idénticas de las secuencias de aminoácidos. El lector experto apreciará que cada codón en un ácido nucleico (excepto AUG y UGG; típicamente los únicos codones para metionina o triptófano, respectivamente) pueden modificarse para producir una molécula funcionalmente idéntica. En consecuencia, cada variación silenciosa (es decir, un codón sinónimo) de un polinucleótido o un polipéptido, que codifica un polipéptido de la presente invención, está implícita en cada secuencia de polipéptidos descrita.

Se describe un polipéptido o un fragmento que tiene actividad de nucleasa y que comprende el motivo de secuencia de aminoácidos: FQIYN. Este corresponde a los residuos 786-790 de SEQ ID NO: 1.

También se describe un polipéptido o un fragmento que tiene actividad de nucleasa y que comprende el motivo de secuencia de aminoácidos: FQIYNK. Este corresponde a los residuos 786-791 de SEQ ID NO: 1.

- 5 Además se describe un polipéptido o un fragmento que tiene actividad de nucleasa y que comprende el motivo de secuencia de aminoácidos: FQIYNKD. Este corresponde a los residuos 786-792 de la SEQ ID NO: 1.

Además se describe un polipéptido o un fragmento que tiene actividad de nucleasa y que comprende el motivo de secuencia de aminoácidos: $X^1X^2X^3X^4X^5FQIYNKDX^6X^7$, correspondiente a los residuos 781-794 de SEQ ID NO: 1, en la que X^1 es uno de entre G o K, X^2 es uno de entre K, S o D, X^3 es uno de entre L o I, X^4 es uno de entre Y o F, X^5 es uno de entre L o M, X^6 es uno de entre F o Y y X^7 es uno de entre S, A o V.

10 También se describe un polipéptido o un fragmento que tiene actividad de nucleasa y que comprende el motivo de secuencia de aminoácidos: GKLYLFQIYNKDFS. Este corresponde a los residuos 781-794 de SEQ ID NO: 1.

15 El motivo de secuencia de aminoácidos puede comprender en vez de los mismos residuos seleccionados de entre 784-794, 785-794, 786-794, 787-794, 788-794 o 789-794 de SEQ ID NO: 1. El motivo puede seleccionarse de entre los residuos 783-793, 783-792, 783-791, 783-790, 783-789 o 783-788 de SEQ ID NO: 1. Además, el motivo puede seleccionarse de entre los residuos 784-793, 785-792 o 786-790 de SEQ ID NO: 1.

20 Aquí se describe una versión catalíticamente inactiva de Cpf1 en la que el dominio RuvC puede comprender:

- (a) un residuo Glu (E) y un motivo corto Glu-Ile-Asp (GID); o
- 25 (b) un residuo Glu (E), y un motivo corto Gly-Ile-Asp (GID); o
- (c) un residuo Glu (E) y un motivo corto Glu-Ile-Asp (EID); o
- (d) un residuo Glu (E) y un motivo corto Ser-Ile-Asp (SID); o
- 30 (e) el motivo de secuencia de aminoácidos: $X^8IDRGER$ en la que X^8 es uno de entre G o S; o
- (f) el motivo de secuencia de aminoácidos: DANGAY; o
- 35 (g) el motivo de secuencia de aminoácidos: EX^9LN en la que X^9 es uno de entre D, N o E; o
- (h) el motivo de secuencia de aminoácidos: EDLN.

40 Un polipéptido o un fragmento descrito en el presente documento puede definirse tanto en términos de la secuencia de referencia SEQ ID NO: 1 como de cualquier porcentaje de variante con respecto a la misma, en combinación con cualquiera de los motivos de aminoácidos mencionados anteriormente como características esenciales.

La proteína o el polipéptido descrito en el presente documento puede tener un dominio RuvC (nucleasa); preferentemente el dominio RuvC puede comprender:

- 45 (a) un motivo corto GID; o
- (b) un motivo corto SID;
- 50 (c) un residuo Glu (E) y un motivo corto GID.

El dominio RuvC puede comprender un residuo Glu (E) y un motivo corto SID.

55 Cuando el dominio RuvC comprende un residuo Glu (E) y un motivo corto GID o SID, el residuo D (aspartato) del motivo puede ser un residuo catalítico.

El dominio RuvC puede comprender el motivo de secuencia de aminoácidos $X^8IDRGER$ en el que X^8 es uno de entre G o S. Por ejemplo, la proteína o el polipéptido puede tener un dominio RuvC (nucleasa), en el que el dominio RuvC comprende el motivo de secuencia de aminoácidos SIDRGER.

60 Cuando el dominio RuvC comprende un motivo de secuencia de aminoácidos GIDRGER o SIDRGER, el residuo D (aspartato) del motivo puede ser un residuo catalítico.

65 La proteína o el polipéptido puede tener un dominio RuvC (nucleasa), en el que el dominio RuvC puede comprender el motivo de secuencia de aminoácidos DANGAY.

Cuando el dominio RuvC comprende un motivo de secuencia de aminoácidos DANGAY, el residuo D (aspartato) del motivo puede ser un residuo catalítico.

La proteína o el polipéptido puede tener un dominio RuvC (nucleasa), en el que el dominio RuvC puede comprender el motivo de secuencia de aminoácidos: EX⁹LN en el que X⁹ es uno de entre D, N o E. Por ejemplo, la proteína o el polipéptido puede tener un dominio RuvC (nucleasa), en el que el dominio RuvC comprende el motivo de secuencia de aminoácidos: EDLN. Cuando el dominio RuvC comprende un motivo de secuencia de aminoácidos EDLN, ENLN o EELN, el residuo E (glutamato) del motivo puede ser un residuo catalítico.

El polipéptido o el fragmento descrito en el presente documento puede tener un dominio RuvC (nucleasa) que comprende un residuo Glu (E) y los motivos de secuencia de aminoácidos SID y DANGAY.

Opcionalmente, el polipéptido o el fragmento puede tener un dominio RuvC (nucleasa) que comprende un residuo Glu (E) y los motivos de secuencia de aminoácidos SID y EDLN.

Opcionalmente, el polipéptido o el fragmento puede tener un dominio RuvC (nucleasa) que comprende un residuo Glu (E), y los motivos de secuencia de aminoácidos SID, DANGAY y EDLN.

Opcionalmente, el dominio RuvC (nucleasa) puede comprender el motivo de secuencia de aminoácidos: X⁸IDRGER en el que X⁸ es uno de entre G o S, y el motivo de secuencia de aminoácidos DANGAY.

Opcionalmente, el dominio RuvC (nucleasa) puede comprender el motivo de secuencia de aminoácidos: X⁸IDRGER en el que X⁸ es uno de entre G o S, y el motivo de la secuencia de aminoácidos: EX⁹LN en el que X⁹ es uno de entre D, N o E.

Opcionalmente, el dominio RuvC (nucleasa) puede comprender el motivo de secuencia de aminoácidos: X⁸IDRGER en el que X⁸ es uno de entre G o S, y el motivo de la secuencia de aminoácidos: EDLN.

Opcionalmente, el dominio RuvC (nucleasa) puede comprender el motivo de secuencia de aminoácidos: X⁸IDRGER en el que X⁸ es uno de entre G o S, y el motivo de secuencia de aminoácidos: DANGAY y el motivo de secuencia de aminoácidos: EX⁹LN en el que X⁹ es uno de entre D, N o E.

Opcionalmente, el dominio RuvC (nucleasa) puede comprender el motivo de secuencia de aminoácidos: X⁸IDRGER en el que X⁸ es uno de entre G o S, y los motivos de la secuencia de aminoácidos DANGAY y EDLN.

Preferentemente, el dominio RuvC (nucleasa) comprende los motivos de secuencia de aminoácidos: SIDRGER, DANGAY y EDLN.

En otros aspectos, el polipéptido o el fragmento puede tener un motivo rico en arginina.

El motivo rico en arginina puede comprender el motivo de secuencia de aminoácidos: X¹⁰YX¹¹X¹²X¹³LX¹⁴X¹⁵X¹⁶EX¹⁷X¹⁸X¹⁹X²⁰X²¹ARX²²X²³, en el que X¹⁰ es uno de entre D o N, X¹¹ es uno de entre R, Q o H, X¹² es uno de entre K, E, S o D, X¹³ es uno de entre A, K o L, X¹⁴ es uno de entre D, N o A, X¹⁵ es uno de entre V, N, Q, K o A, X¹⁶ es uno de entre R, K o I, X¹⁷ es uno de entre Y, K o I, X¹⁸ es uno de entre D o E, X¹⁹ es uno de entre N, R o M, X²⁰ es uno de entre K, V, F o D, X²¹ es uno de entre E, A, D o S, X²² es uno de entre R, Q o K y X²³ es uno de entre N, A, S o D.

El motivo rico en arginina puede comprender el motivo de secuencia de aminoácidos: DYRKALDVREYDNKEARN, DYQKKLDNREKERVAAARQA, DYREKLNQREIEMKDARQS, DYHSLDDKKEKERFEARQN o NYHDKLAAIEKDRDSARKD.

Un polipéptido o un fragmento descrito en el presente documento puede tener un dominio RuvC (nucleasa) que comprende un residuo Glu (E), y los motivos de la secuencia de aminoácidos Ser-Ile-Asp (SID), DANGAY y el motivo de la secuencia de aminoácidos EDLN. Preferentemente, el dominio RuvC (nucleasa) comprenderá el motivo de secuencia de aminoácidos: X⁸IDRGER en el que X⁸ es uno de entre G o S, y los motivos de la secuencia de aminoácidos DANGAY y EDLN. Más preferentemente, el dominio RuvC (nucleasa) comprenderá los motivos de secuencia de aminoácidos: SIDRGER, DANGAY y EDLN.

Un polipéptido o un fragmento descrito en el presente documento preferentemente no comprende un dominio HNH (nucleasa). Adicionalmente o alternativamente, un polipéptido o un fragmento tal como se describe en el presente documento no comprende un lóbulo de reconocimiento que está típicamente presente en Cas9.

Determinados polipéptidos o fragmentos descritos en el presente documento pueden tener actividad de nucleasa que viene proporcionada por un único sitio en el polipéptido.

Otros polipéptidos o fragmentos descritos en el presente documento pueden comprender además un dominio de dedo de zinc, aunque el sitio de unión a metal (típicamente 4 aminoácidos, Cys y/o His) no está completo en todas las variantes de Cpf1.

- 5 Los polipéptidos o los fragmentos descritos en el presente documento pueden tener una actividad de nucleasa que es una escisión monocatenaria, por ejemplo actividad de nickasa.

10 Preferentemente, se pueden utilizar dos subunidades de Cpf1 en una disposición dimerica en la que los dominios de nucleasa de cada una de las dos subunidades escinden cadenas de ADN individuales. Preferentemente, dicho dímero puede ser un homodímero en el que los dominios similares a RuvC de cada una de las dos subunidades escinden cadenas de ADN individuales. Alternativamente, los polipéptidos Cpf1 pueden modificarse genéticamente para que contengan más de un dominio de nucleasa, nativo o de otro tipo, que permita la escisión de ambas cadenas de ADN.

- 15 El polipéptido o los fragmentos de la invención tienen preferentemente afinidad de unión por una molécula de ARN guía.

20 En otros aspectos, un polipéptido o un fragmento descrito en el presente documento puede tener un ARN guía que comprende una secuencia sustancialmente complementaria a una secuencia comprendida en una cadena de ácido nucleico diana.

25 Un polipéptido o fragmento descrito en el presente documento tiene preferentemente afinidad de unión por un motivo de secuencia de polinucleótidos presente en una cadena de ácido nucleico diana. Este motivo de secuencia generalmente se conoce como secuencia de motivo adyacente protoespaciador (PAM). Preferentemente, el motivo de la secuencia de nucleótidos es de al menos 3 residuos contiguos de ácido nucleico.

30 El PAM está ubicado en la diana (adyacente al protoespaciador). Típicamente, el dominio SEED del ARN guía (la región que es la responsable más probable del emparejamiento de bases inicial de la guía/diana) es complementario a la secuencia de ácido nucleico diana. Preferentemente, la parte SEED de la guía no tolera los emparejamientos incorrectos.

35 Para mejorar aún más los polipéptidos o los fragmentos descritos en el presente documento, se pueden añadir aminoácidos adicionales, preferentemente por medio de una fusión al extremo N-terminal o C-terminal. La secuencia de aminoácidos adicional puede tener actividad de modificación, visualización, activación de la transcripción o represión de la transcripción de ácidos nucleicos o cromatina y preferentemente se fusiona traduccionalmente a través de la expresión en sistemas de expresión de proteínas naturales o artificiales, o se une covalentemente mediante una etapa de síntesis química a la, al menos una, subunidad; preferentemente el, al menos un, resto funcional está fusionado o unido a al menos la región del extremo N-terminal y/o la región del extremo C-terminal.

40 La secuencia de aminoácidos adicional que tiene actividad de modificación, activación, represión o visualización de ácidos nucleicos o cromatina puede ser una proteína; opcionalmente seleccionada de entre una helicasa, una nucleasa, una nucleasa-helicasa, una ADN metiltransferasa (por ejemplo, Dam) o ADN desmetilasa, una histona metiltransferasa, una histona desmetilasa, una acetilasa, una desacetilasa, una fosfatasa, una quinasa, un (co-)activador de la transcripción, una subunidad de ARN polimerasa, un represor de la transcripción, una proteína de unión al ADN, una proteína de estructuración de ADN, una proteína marcadora, una proteína reportera, una proteína fluorescente, una proteína de unión a ligando (por ejemplo, mCherry o una proteína de unión a metales pesados), un péptido señal (por ejemplo, secuencia de señal TAT), una secuencia de localización subcelular (por ejemplo, secuencia de localización nuclear) o un epítopo de anticuerpo.

50 Cuando la proteína es una nucleasa, puede ser una seleccionada de entre una endonucleasa de restricción de tipo II tal como FokI, o una porción mutante o activa de la misma. Preferentemente, un complejo de proteínas de la invención puede fusionarse al dominio N-terminal de FokI y otro complejo de proteínas de la invención puede fusionarse al dominio C-terminal de FokI. Estos dos complejos de proteínas pueden utilizarse juntos (en una configuración dimerica) para lograr un corte bicatenario específico del locus ventajoso en un ácido nucleico, por lo que la ubicación del corte en el material genético depende del diseño y la elección del usuario, por ejemplo de forma guiada por el componente de ARN (definido y descrito más adelante) y se debe a la presencia de una secuencia denominada "motivo adyacente protoespaciador" (PAM) en la cadena de ácido nucleico diana (también descrita con más detalle más adelante).

60 Una proteína o un polipéptido descrito en el presente documento puede tener una secuencia de aminoácidos adicional que es una endonucleasa de restricción modificada, por ejemplo FokI. La modificación se encuentra preferentemente en el dominio catalítico.

65 El FokI modificado puede ser KKR Sharkey o ELD Sharkey, que está fusionado con la proteína Cpf1. En una aplicación preferida de estos complejos, dos de estos complejos (KKR Sharkey y ELD Sharkey) pueden estar juntos en combinación. Un par heterodímero de complejos de proteínas que emplean FokI modificado de forma diferente

tiene una ventaja particular en el corte bicatenario dirigido de ácido nucleico. Si se utilizan homodímeros, entonces es posible que se produzcan más escisiones en sitios no diana debido a la actividad no específica. Un enfoque heterodímero aumenta ventajosamente la fidelidad de la escisión en una muestra de material.

Ventajosamente, las modificaciones anteriores pueden permitir a un usuario seleccionar de forma predeterminada un locus genético preciso que se desea escindir, marcar o alterar de otra forma de alguna manera, por ejemplo por metilación, utilizando cualquiera de los ácidos nucleicos o cromatina modificando, visualizando, activando la transcripción o reprimiendo la transcripción de entidades definidas en el presente documento. La otra parte componente del sistema es una molécula de ARN que actúa como guía para dirigir los complejos de la invención al locus correcto en el ADN o ARN que se pretende modificar, cortar o marcar.

Un polipéptido o un fragmento descrito en el presente documento está preferentemente unido a un ARN guía y a un ácido nucleico diana. De esta forma, se forma un complejo que proporciona actividad de nucleasa de cadena de ADN dirigida, en el que se escinde un locus diana deseado.

Se describe un polinucleótido que comprende una secuencia de polinucleótidos que codifica un polipéptido o un fragmento tal como se ha definido en el presente documento anteriormente.

También se describe un vector de expresión que comprende un polinucleótido tal como se ha mencionado anteriormente.

Además se describe un vector de expresión tal como se ha definido anteriormente, que comprende además una secuencia de nucleótidos que codifica un ARN guía que tiene una complementariedad sustancial con una secuencia deseada presente en la cadena de ácido nucleico diana. El ARN guía en el estado nativo es un único ARN que consiste en un ARNcr.

Un vector de expresión tal como se describe es preferentemente un vector vírico, por ejemplo adenovirus o virus adenoasociado (AAV).

También se describe una célula huésped transformada para expresar un polipéptido o un fragmento tal como se ha descrito anteriormente en el presente documento.

Típicamente, el ADN del vector de expresión puede suministrarse a la célula huésped por transformación, electroporación o virus (AAV). Además, el ARN puede suministrarse a una célula huésped mediante inyección o electroporación. Las proteínas pueden suministrarse a las células mediante electroporación, etiquetas de péptidos (VIH). Una célula huésped tal como se ha descrito anteriormente en el presente documento puede transformarse adicionalmente para que contenga un ARN guía que comprende una secuencia sustancialmente complementaria a una secuencia comprendida en una cadena de ácido nucleico diana presente en la célula huésped.

En el presente documento se describe cualquier célula huésped transformada con un vector de expresión tal como se ha descrito anteriormente en el presente documento.

También se describe un procedimiento para escindir una cadena de ácido nucleico diana en un locus específico, que comprende exponer el ácido nucleico diana a un polipéptido o un fragmento tal como se define en el presente documento, y con una molécula de ARN guía que comprende una secuencia sustancialmente complementaria a una secuencia comprendida en la cadena de ácido nucleico diana.

Además se describe un procedimiento para escindir una cadena de ácido nucleico diana en un locus específico en el genoma de una célula de un organismo, que comprende transformar la célula con un vector de expresión tal como se describe en el presente documento, y transformar la célula con un vector que expresa un ARN guía que comprende una secuencia sustancialmente complementaria a una secuencia comprendida en una cadena de ácido nucleico diana.

También se describe un procedimiento para escindir una cadena de ácido nucleico diana en un locus específico en el genoma de una célula de un organismo, que comprende transformar la célula con un vector de expresión tal como se describe en el presente documento.

Además, se describe un procedimiento de edición de genes de unión de extremos no homólogos que comprende (a) transformar la célula con un vector de expresión tal como se describe en el presente documento y transformar la célula con un vector que expresa un ARN guía que comprende una secuencia sustancialmente complementaria a una secuencia comprendida en una cadena de ácido nucleico diana; o (b) transformar la célula con un vector de expresión tal como se describe en el presente documento. Los polipéptidos pueden modificarse o utilizarse para causar roturas bicatenarias.

También se describe un procedimiento de edición de genes de unión a extremos homólogos que comprende (a) transformar la célula con un vector de expresión tal como se describe en el presente documento, y transformar la

célula con un vector que expresa un ARN guía que comprende una secuencia sustancialmente complementaria a una secuencia comprendida en una cadena de ácido nucleico diana; o (b) transformar la célula con un vector de expresión tal como se describe en el presente documento; para crear una rotura bicatenaria en un locus deseado en el material genético, y exponer el material genético a una secuencia de polinucleótidos que tiene regiones terminales complementarias a las regiones finales rotas del material genético.

Descripción detallada

La proteína de secuencia de aminoácidos SEQ ID NO: 1 es una proteína grande (aproximadamente 1300 aminoácidos) que contiene un dominio de nucleasa similar a RuvC homólogo a los respectivos dominios de Cas9 y el elemento transponible ORF-B, junto con una región rica en arginina similar a aquella en Cas9 y dedo de cinc (ausente en Cas9 pero compartido con ORF-B), pero carece del dominio de nucleasa HNH que está presente en todas las proteínas Cas9.

La invención se describirá ahora en detalle con referencia a los ejemplos y a los dibujos en los que:

La figura 1 muestra la estructura del dominio de la nucleasa novedosa CRISPR-Cas, Cpf1. Se muestran tres dominios de nucleasa RuvC, un dedo de cinc y un dominio rico en arginina que permite la interacción con guía de ARN y diana de ADN.

La figura 2 muestra los resultados de un análisis *in silico* del motivo adyacente protoespaciador conservado (PAM). El panel A muestra un Weblogo basado en flancos de 5' de protoespaciadores representados en la tabla 1. El panel B muestra un Weblogo basado en flancos de 3' de protoespaciadores representados en la tabla 1.

La figura 3 muestra los resultados de un alineamiento múltiple de la familia de proteínas Cpf1. Cada secuencia está marcada con el número identificador de GenBank (GI) y el nombre sistemático de un organismo. La estructura secundaria pronosticada (SS) se muestra sombreada. Los residuos del sitio activo de los dominios similares a RuvC se muestran en negrita y con doble subrayado. La hélice de puente potencial se muestra sombreada y con un subrayado simple. La secuencia de aminoácidos FQIYN también se indica en negrita, con sombreado y subrayado discontinuo.

Ejemplo 1 - Nucleasas novedosas para la edición de genes

Ejemplos específicos son (1) Cpf1 asociado a CRISPR de la bacteria marina *Francisella novicida* (Fn-Cpf1), y (2) Cpf1 asociado a CRISPR del arqueón *Methanomethylophilus alvus* cepa Mx1201 (Mal-Cpf1) que reside en el intestino humano.

Sin que los inventores deseen vincularse a ninguna teoría particular, Cpf1 reconoce el ARNcr en una forma específica de secuencia, después de lo cual se produce la escisión del segmento de ARN bicatenario y, finalmente, la formación de un complejo efector que consiste en Cpf1 y una única guía de ARNcr. Cpf1 puede funcionar como un dímero, escindiendo los dominios similares a RuvC de cada una de las dos subunidades cadenas de ADN individuales. Alternativamente, Cpf1 puede contener más de un dominio de nucleasa que permite la escisión de ambas cadenas de ADN. Alternativamente, uno o más dominios RuvC de Cpf1 pueden mostrar una flexibilidad inusual que permite la escisión de ambas cadenas.

Los ejemplos siguientes se realizaron en paralelo para las variantes de proteínas bacterianas Fno-Cpf1 y de arqueas Mal-Cpf1:

La clonación se lleva a cabo en la totalidad del locus CRISPR, incluido el operón *cas* (*cpf1-cas4-cas1-cas2*), la región líder, la matriz CRISPR y las regiones flanqueantes (aproximadamente 10 kb) en un vector de copia baja (por ejemplo, pACYC184) en una cepa de *E. coli* K12; no se conocen detalles sobre la maduración de la guía, que puede ser similar a la de Cas9 (traARNcr/RNasall), o puede ser similar a la de Cascade (ribonucleasa similar a Cas6, aunque esa no es parte de los operones *cpf1*), o puede ser única. Otros materiales y procedimientos detallados se proporcionan en Sapranaukas *et al.*, 2011, Nucleic Acids Res. 39: 9275-9282.

Se utilizaron procedimientos estándar para optimizar posibilidades de producción funcional de proteínas de las proteínas Cpf1 seleccionadas en *E. coli*: (i) realizando un diseño de armonización de codones para ajustar secuencias de nucleótidos *cpf1* (véase Angov *et al.*, 2008, PLoS One 3, e2189); (ii) incluyendo la etiqueta strepII N-terminal o C-terminal, que permitirá una purificación por afinidad; (iii) clonando un gen sintético en el vector de expresión T7 (por ejemplo pET24d) y transformando el plásmido en cepa de no producción de *E. coli* (por ejemplo JM109, que carece del gen de la ARN polimerasa T7), (iv) transfiriendo el plásmido por medio de la segunda transformación a la cepa de producción de *E. coli* (por ejemplo, BL21 (DE3), que contiene el gen de la ARN polimerasa T7 bajo el control del promotor ramnosa, que permite un ajuste preciso de la expresión, (v) variando condiciones de expresión (medio, concentración de inductor, tiempo de inducción), (vi) utilizando condiciones óptimas para el cultivo a escala de litro, después de lo cual las células se recolectan y se alteran mecánicamente para obtener extracto desprovisto de células (pequeños volúmenes por sonicación; grandes volúmenes por French

Press), (vii) separando membranas y fracciones solubles, y realizando una purificación por afinidad utilizando resina de estreptactina, (viii) analizando fracciones relevantes por SDS-PAGE, y almacenando la proteína pura para análisis posteriores.

5 Además de lo anterior, adicionalmente, el gen de ARNcr predicho se secuencía, o se crea un gen de ARN de guía única (ARNsg), por ejemplo mediante la adición de 4 bucles sintéticos de nucleótidos (Jinek *et al.*, 2012, Science 337: 816-821); genes de ARN que residen en el mismo plásmido que el gen *cpf1*, o en un plásmido separado.

10 Adicionalmente, se produce un mutante Cpf1 catalíticamente inactivo (el sitio activo RuvC contiene glutamato conservado (E), así como un motivo GID).

Adicionalmente, se produce un mutante Cpf1 catalíticamente inactivo (el sitio activo RuvC contiene glutamato conservado (E), así como motivo SID).

15 Además, las fusiones N-terminales o C-terminales están hechas del mutante Cpf1 con el dominio nucleasa FokI con enlazadores de conexión diferentes (tal como se describe para Cas9; véase Guilinger *et al.*, 2014, Nat. Biotecnología 32: 577-82).

20 Ejemplo 2 - Caracterización bioquímica de nucleasas Cpf1

Estos experimentos caracterizan la vigilancia de la guía y la escisión de la diana. El sistema CRISPR es un sistema inmunitario adaptativo en bacterias y arqueas. Las matrices CRISPR consisten en repeticiones idénticas (por ejemplo 30 pb) y espaciadores variables (por ejemplo 35 pb). La naturaleza adaptativa del sistema CRISPR se basa en la adquisición regular de nuevos separadores, a menudo correspondientes a fragmentos (protoespaciadores) derivados de virus. La adquisición generalmente depende de la selección de un protoespaciador sobre la base de la presencia de un motivo adyacente de protoespaciador (PAM). La presencia de este motivo es crucial para la eventual interferencia del complejo efector asociado a CRISPR (por ejemplo Cas9) con su guía de ARNcr. El motivo PAM permite la discriminación propia frente a la discriminación no propia: las secuencias diana potenciales (es decir, complementarias a la secuencia guía de ARNcr) residen tanto en el genoma del huésped (la matriz CRISPR propia) como en el genoma del invasor (el protoespaciador no propio); la presencia del protoespaciador en el ADN invasor activa el complejo efector para unirlos de forma escalonada; cuando se produce un emparejamiento perfecto de bases entre la secuencia del protoespaciador inmediatamente adyacente al PAM (la denominada secuencia semilla), las bases se emparejan como una cremallera, que eventualmente conduce a un estado de Cas9 para catalizar la escisión de las cadenas de ADN diana (véase Jinek *et al.*, 2012, Science 337: 816-821; también Gasiunas *et al.*, 2012, PNAS 109: E2579-E2586).

40 Análisis *in silico* del PAM asociado a Cpf1 por análisis BLAST de los espaciadores CRISPR de los loci de *cpf1*. El análisis BLAST de algunos espaciadores muestra varias secuencias homólogas (90-100% de identidad), (tabla 1). Los éxitos más prometedores se refieren a secuencias idénticas de genes víricos en general, y genes de profagos en particular. Los profagos se derivan a partir de virus lisogénicos, cuyos genomas se han integrado en el genoma de las bacterias. Como es el caso con los virus eucariotas, la gama de huéspedes de los virus procariotas es a menudo bastante limitada; por lo tanto, cuando se encuentra el profago coincidente en una bacteria que está estrechamente relacionada con la bacteria que tiene la secuencia espaciadora correspondiente en su matriz CRISPR, esto proporciona cierta confianza de que es un verdadero éxito. En otras palabras, puede ser que el profago se asemeje a un virus que ha intentado infectar la bacteria que contiene CRISPR, pero la invasión ha dado como resultado la adquisición de espaciadores y la inmunidad vírica de esta última bacteria.

50 Tabla 1. Resultados de BLAST con espaciadores CRISPR asociados a *cpf1* FnU112 como secuencias de consulta. Las secuencias de nucleótidos de ambos espaciadores (arriba) y protoespaciadores están sombreadas; los flancos 5' y 3' de los protoespaciadores no están sombreados; Herramienta: Diana CRISPR (bioanalysis.otago.ac.nz/CRISPRTarget/). Consulta: toda la matriz CRISPR de subespecies de *Francisella novicida*. Base de datos diana: Genbank-NT. Apertura de hueco -10, Extensión -2; Emparejamiento correcto de nucleótidos 1, Emparejamiento incorrecto de nucleótidos -1; Valor de E 1; Tamaño de palabra 7; Puntuación de corte 20; Protoespaciador flanqueante del extremo 3' 8 pb; Protoespaciador flanqueante del extremo 5' 8 pb.

55

Nº de espaciador de subespecies de Fn	Huésped del profago, número de acceso del gen diana		Alineamiento del espaciador de subespecies de Fn con protoespaciador (más flancos de 8 nt en ambos lados)
<i>Francisella novicida</i> U112 Nº 1	<i>Francisella novicida</i> 3523, hipo prot AEE26301.1	espaciador protoespaciador	5' <u>AGATTAAAAGCTAATCTATCTTCTTGAG</u> 5' ATAATTTAAGATTAAAAGCTAATCTATTTCTTGAGATCTGACC
<i>Francisella novicida</i> U112 Nº 2	<i>Francisella novicida</i> 3523, secuencia intergénica en profago	espaciador protoespaciador	5' <u>TAGCGATTATCAAGGTCATTTT</u> 5' CTAATTTATAGCGATTATCAAGGTCATTTTAAAAAGTT
<i>Francisella novicida</i> Fx1 Nº 1	<i>Francisella novicida</i> 3523, hipo prot AEE26295.1, "proteína del tubo de cola principal del fago"	espaciador protoespaciador	5' <u>ATCGATTATTACTTAACTGCAGCTCTTAC</u> 5' AATGTTCAATCGATTATTACTTAACTGCAGCTCTTACCTCGATGG
<i>Francisella novicida</i> FTG Nº 1	<i>Francisella novicida</i> 3523, hipo prot YP_0058240 59.1	espaciador protoespaciador	5' <u>CCCACAAATACTACAAAAATAACTTAA</u> oo 5' ATTTTGTGGCTCCAAATACTACAAAAATAACTTAAACTTTGAA
<i>Francisella novicida</i> GA99-3549 Nº 1	<i>Francisella novicida</i> 3523, hipo prot FN3523_100 9, "placa base_J"	espaciador protoespaciador	5' <u>ATTGTCAAAACATAACGAGCTGCTTCAATAAT</u> o o o 5' GGTCTTTTACGTTATTACATAACGAGCTGCTTCAATAATCTTAGCAA

El análisis de las secuencias que flanquean los protoespaciadores en los genes de profago dio como resultado un motivo conservado rico en T; curiosamente, este motivo no reside aguas abajo del protoespaciador (como en el sistema Cas9), sino aguas arriba. Aunque no desean vincularse a ninguna teoría particular, los inventores encuentran que Cpf1 de la invención requiere un motivo similar a PAM (3-4 nucleótidos) para unirse a una molécula de ADN diana que es complementaria a la guía, tiene una secuencia semilla (8-10 nucleótidos) en la que no se permiten emparejamientos incorrectos, y tiene un único sitio de nucleasa que permite mellar la cadena de ADN diana emparejada con la base.

Los motivos PAM de Cpf1 y las variantes de la invención también se caracterizaron utilizando el enfoque de Jiang *et al.*, 2013, Nat. Biotecnología 31: 233-239). Se utilizaron dos derivados de *E. coli* BL21 (DE3), inicialmente transformados con un plásmido diana o con un plásmido no diana; dos plásmidos diana variantes utilizados tienen una parte similar (marcador GFP, marcador KmR, origen de replicación) y una parte variable con secuencia diana (protoespaciador) con un PAM degenerado asociado (5-8 nucleótidos variables) aguas arriba o aguas abajo del protoespaciador); a continuación, esta cepa se transformó con un plásmido de expresión Cpf1 (incluye diseño CRISPR con ARN de guía única (ARNsg, marcador CmR); el cribado de transformantes se realizó en placas con cloranfenicol (Cm) (no kanamicina (Km)), y cribado de colonias no fluorescentes, que indica la pérdida de plásmido diana. A medida que se pierdan los plásmidos con los PAM correctos, se realizó DNA Deep Seq de productos de PCR apropiados de la totalidad del conjunto de plásmido diana, antes y después de la transformación. Las diferencias revelan el PAM (Bikard *et al.*, 2013, Nucleic Acids Res. 41: 7429-7437).

Las firmas PAM se confirmaron mediante caracterización *in vitro* de la actividad de escisión de BsCas9/ARNsg; Los ensayos revelan condiciones óptimas (temperatura, tampón/pH, sal, metales).

La presencia de una secuencia semilla en el PAM se estableció según procedimientos descritos por Jinek *et al.*, 2012, Science 337: 816-821.

Ejemplo 3 - Ingeniería genética bacteriana

Realización de ingeniería genética de alto rendimiento del genoma bacteriano con variantes de nucleasa. Sin desear vincularse a ninguna teoría particular, los inventores esperan que los complejos Cpf1/guía de la invención permitan

un direccionamiento específico del ADN genómico. El direccionamiento multiplex se puede establecer mediante el uso de un diseño CRISPR junto con un ARNcr coincidente.

Los experimentos proporcionan la aplicación de Cpf1 y variantes de la invención. Cas9 se analiza en paralelo como referencia.

Se realiza la activación/desactivación génica (inserción/alteración de cualquier secuencia). La cepa huésped *E. coli* K12 (LacZ +, GFP-) se modificó genéticamente de la forma siguiente: el gen que codifica una variante de la proteína verde fluorescente (GFPuv) se inserta en el gen lacZ, dando como resultado un fenotipo claro (LacZ-, GFP+). El gen *cpf1* se introdujo en un plásmido (o derivados de esos plásmidos), junto con un fragmento que permite la recombinación homóloga de la secuencia diana. Se seleccionó una secuencia diana (protoespaciadora), con una secuencia PAM apropiada ubicada de forma adyacente; una guía correspondiente diseñada, que consiste en el ARNcr (con espaciador complementario al protoespaciador diana) y el gen ARNcr (adaptado del procedimiento descrito para Cas9 por Jiang *et al.* (2013a) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31: 233-239).

El silenciamiento de la expresión génica (utilizando Cas9 inactivado catalíticamente, fue tal como se describe: dCas9 derivado de Spy-Cas9; (Bikard *et al.*, 2013, Nucleic Acids Res. 41: 7429-7437; Qi *et al.*, 2013, Cell 152: 1173-1183); mediante la unión al promotor (sitio de unión a ARN polimerasa) del gen diana, o de genes diana utilizando un enfoque multiplex (utilizando un diseño CRISPR).

Activación de expresión génica; como anteriormente (silenciamiento); sitio de unión aguas arriba de la unión de ARN polimerasa, con Cas9 fusionado al dominio de activación (tal como se ha descrito para Spy-Cas9) (Bikard *et al.*, 2013, Nucleic Acids Res. 41: 7429-7437).

La fusión de Cpf1 inactivado y el dominio de nucleasa FokI (descrito en el ejemplo 1) se compararon con un Cpf1 activo en diferentes configuraciones experimentales. Esto requirió dos interacciones simultáneas de guías y dianas, lo que da como resultado una mejora importante de la escisión en el sitio deseado.

Ejemplo 4 - Ingeniería genética de células madre humanas

La edición dirigida de mutaciones genéticas que causan enfermedades sería un tratamiento elegante y eficaz para enfermedades genéticas. Los sistemas de edición de genes recientemente descubiertos, tales como Cas9, permiten el direccionamiento específico de mutaciones que causan enfermedades en el genoma, y pueden utilizarse para reparar funcionalmente o deshabilitar permanentemente genes mutados. La eficacia de los sistemas de edición de genes se ha demostrado en un entorno de laboratorio, y ahora se utilizan de forma rutinaria en la edición del genoma de una amplia diversidad de tipos de células de muchas especies diferentes, incluidos los seres humanos. Sin embargo, a pesar del éxito de estos sistemas en el entorno de investigación, la aplicación clínica de los sistemas de edición de genes se ve obstaculizada por la falta de un sistema de suministro adecuado para introducir tecnologías de edición de genes en las células de los pacientes de una forma segura, transitoria y eficaz. Varios laboratorios están trabajando en el desarrollo de vectores víricos recombinantes que se puedan utilizar para suministrar sistemas de edición de genes a las células del paciente, pero la expresión prolongada de CRISPR/Cas9, por ejemplo, a partir de dichos vectores aumentará la probabilidad de efectos fuera de la diana y, por lo tanto, no es ideal. El suministro intracelular de proteínas de edición de genes recombinantes y ARN de CRISPR sintético sería un procedimiento eficaz, no integrador y transitorio para la aplicación de la tecnología de edición de genes en células de pacientes.

Recientemente se ha desarrollado un procedimiento novedoso que permite la transducción de proteínas nativas en prácticamente cualquier tipo de célula (D'Astolfo *et al.*, 2015, Cell, 161: 674-690). Esta tecnología, denominada iTOP, por Transducción por Osmocitosis y Propanobetaina inducida (*induced Transduction by Osmocytosis and Propanebetaine*), se basa en una combinación de compuestos de molécula pequeña, que desencadenan la absorción y la liberación intracelular de proteínas nativas. iTOP es altamente eficaz, logra de forma rutinaria eficacias de transducción > 90% de las células y funciona en una amplia diversidad de tipos de células primarias. Se ha demostrado que la transducción mediada por iTOP de la proteína Cas9 recombinante y ARNsg transcrito *in vitro* permite una edición de genes altamente eficaz en tipos de células difíciles de transfectar, incluidas las células madre humanas. Tras la transducción iTOP-CRISPR/Cas9, se ha informado de > 70% de direccionamiento de genes bialélicos en células ES humanas sin la necesidad de selección con fármacos de células transducidas.

Las ventajas clave de iTOP sobre las tecnologías existentes son: (i) la capacidad de transducir células primarias (madre) con proteína nativa con una eficacia muy alta, (ii) la naturaleza transitoria no integradora de la edición de genes mediada por proteínas, que asegura la seguridad y minimiza efectos fuera de la diana, y (iii) el control estricto de la dosis y el momento de la proteína suministrada. Hemos demostrado que iTOP-CRISPR/Cas9 es una herramienta eficaz para modificar una gran diversidad de tipos de células (del paciente) primarias. Sin embargo, debido a problemas de solubilidad de proteínas y de tamaño, la producción de Cas9 recombinante está obstaculizando la adopción (clínica) a gran escala de este sistema. Cpf1 podría resolver estos problemas y allanar el camino para el desarrollo de nuevos tratamientos para tratar enfermedades genéticas.

La tecnología iTOP se utilizará para permitir el suministro intracelular eficaz de Cpf1 en células madre humanas. La ventaja de iTOP es su enfoque altamente flexible. En primer lugar, la hipertonidad mediada por NaCl induce la absorción intracelular de proteínas a través de un proceso denominado macropinocitosis (D'Astolfo *op. cit.*). En segundo lugar, un compuesto de transducción de propanobetaína (NDSB-201 o ácido gamma-aminobutírico (GABA) u otros desencadena la liberación intracelular de proteínas a partir de las vesículas de macropinosomas. Además de estos compuestos, se añaden osmoprotectores tales como glicerol y glicina para ayudar a las células a hacer frente al estrés hipertónico inducido por NaCl. Variando la concentración de NaCl, la concentración y el tipo de compuesto de transducción y/o la concentración y el tipo de osmoprotectores, el sistema iTOP puede adaptarse y optimizarse para que cumpla con los requisitos específicos de la proteína de carga y/o las células diana. Los parámetros iTOP se optimizaron para permitir la edición eficaz de genes de células madre embrionarias humanas (hESC), direccionando al gen endógeno WDR85 por Cpf1 (equipado con una señal de localización nuclear N-terminal o C-terminal (NLS)), tal como se ha mostrado recientemente para Cas9.

En el listado de secuencias siguiente, los residuos de aminoácidos Glu Xaa Asp (subrayado simple) son el motivo GID de un dominio RuvC. Por lo tanto, en la SEQ ID NO: 1, el residuo Xaa puede ser I.

Los residuos de aminoácidos Ile Asp Arg Gly Glu Arg (subrayado doble) incluyen los residuos IDR de un dominio RuvC.

Los residuos de aminoácidos Phe Glu Asp (subrayado triple) incluyen el residuo E que forma parte de los residuos de sitio activo de un dominio RuvC.

Ejemplo 5 - Alineamiento múltiple de proteínas Cpf1

La figura 3 muestra los resultados de un alineamiento múltiple de proteínas Cpf1. El alineamiento se construyó utilizando el programa MUSCLE y se modificó manualmente sobre la base de alineamientos por pares locales PSI-BLAST y el resultado de HHpred. Cada secuencia está etiquetada con el número de Identificador de GenBank (GI) y la denominación sistemática de un organismo. El análisis de cinco secuencias en este trabajo está etiquetado con los números respectivos. La estructura secundaria (SS) fue predicha por Jpred y se muestra sombreada. El CONSENSO se calculó para cada columna de alineamiento ajustando la puntuación de la suma de pares dentro de la columna entre los de una columna homogénea (el mismo residuo en todas las secuencias alineadas) y una columna aleatoria con un corte de homogeneidad de 0,8. Los residuos de sitio activo de los dominios similares a RuvC se muestran en negrita y con doble subrayado. La hélice de puente potencial se muestra sombreada y con un subrayado simple. La secuencia de aminoácidos FQIYN también se indica en negrita, con sombreado y subrayado de puntos.

Listado de secuencias

<110> Wageningen Universiteit

<120> Cpf1 Nuclease

<130> RAW/P223284WO

<160> 36

<170> PatentIn versión 3.5

<210> 1

<211> 1304

<212> PRT

<213> Secuencia artificial

<220>

<223> Secuencia consenso de Cpf1

<220>

<221> misc_feature

<222> (439)..(439)

<223> Xaa puede ser cualquier aminoácido natural

<220>

<221> misc_feature

<222> (504)..(504)

<223> Xaa puede ser cualquier aminoácido natural

```

<220>
<221> misc_feature
<222> (521)..(521)
5  <223> Xaa puede ser cualquier aminoácido natural

<220>
<221> misc_feature
<222> (539)..(539)
10 <223> Xaa puede ser cualquier aminoácido natural

<220>
<221> misc_feature
<222> (800)..(800)
15 <223> Xaa puede ser cualquier aminoácido natural

<400> 1
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1          5          10          15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
          20          25          30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
          35          40          45

```

ES 2 768 976 T3

Lys	Ala	Lys	Gln	Ile	Ile	Asp	Lys	Tyr	His	Gln	Phe	Phe	Ile	Glu	Glu
50						55					60				
Ile	Leu	Ser	Ser	Val	Cys	Ile	Ser	Glu	Asp	Leu	Leu	Gln	Asn	Tyr	Ser
65					70					75					80
Asp	Val	Tyr	Phe	Lys	Leu	Lys	Lys	Ser	Asp	Asp	Asp	Asn	Leu	Gln	Lys
				85					90					95	
Asp	Phe	Lys	Ser	Ala	Lys	Asp	Thr	Ile	Lys	Lys	Gln	Ile	Ser	Glu	Tyr
			100					105						110	
Ile	Lys	Asp	Ser	Glu	Lys	Phe	Lys	Asn	Leu	Phe	Asn	Gln	Asn	Leu	Ile
		115						120					125		
Asp	Ala	Lys	Lys	Gly	Gln	Glu	Ser	Asp	Leu	Ile	Leu	Trp	Leu	Lys	Gln
	130					135					140				
Ser	Lys	Asp	Asn	Gly	Ile	Glu	Leu	Phe	Lys	Ala	Asn	Ser	Asp	Ile	Thr
145					150					155					160
Asp	Ile	Asp	Glu	Ala	Leu	Glu	Ile	Ile	Lys	Ser	Phe	Lys	Gly	Trp	Thr
				165					170					175	
Thr	Tyr	Phe	Lys	Gly	Phe	His	Glu	Asn	Arg	Lys	Asn	Val	Tyr	Ser	Ser
			180					185					190		
Asp	Asp	Ile	Pro	Thr	Ser	Ile	Ile	Tyr	Arg	Ile	Val	Asp	Asp	Asn	Leu
		195					200					205			
Pro	Lys	Phe	Leu	Glu	Asn	Lys	Ala	Lys	Tyr	Glu	Ser	Leu	Lys	Asp	Lys
	210					215					220				
Ala	Pro	Glu	Ala	Ile	Asn	Tyr	Glu	Gln	Ile	Lys	Lys	Asp	Leu	Ala	Glu
225					230					235					240
Glu	Leu	Thr	Phe	Asp	Ile	Asp	Tyr	Lys	Thr	Ser	Glu	Val	Asn	Gln	Arg
				245					250					255	
Val	Phe	Ser	Leu	Asp	Glu	Val	Phe	Glu	Ile	Ala	Asn	Phe	Asn	Asn	Tyr
			260					265					270		

ES 2 768 976 T3

Leu	Asn	Gln	Ser	Gly	Ile	Thr	Lys	Phe	Asn	Thr	Ile	Ile	Gly	Gly	Lys
		275					280					285			
Phe	Val	Asn	Gly	Glu	Asn	Thr	Lys	Arg	Lys	Gly	Ile	Asn	Glu	Tyr	Ile
	290					295					300				
Asn	Leu	Tyr	Ser	Gln	Gln	Ile	Asn	Asp	Lys	Thr	Leu	Lys	Lys	Tyr	Lys
305					310					315					320
Met	Ser	Val	Leu	Phe	Lys	Gln	Ile	Leu	Ser	Asp	Thr	Glu	Ser	Lys	Ser
				325					330					335	
Phe	Val	Ile	Asp	Lys	Leu	Glu	Asp	Asp	Ser	Asp	Val	Val	Thr	Thr	Met
			340					345					350		
Gln	Ser	Phe	Tyr	Glu	Gln	Ile	Ala	Ala	Phe	Lys	Thr	Val	Glu	Glu	Lys
		355					360					365			
Ser	Ile	Lys	Glu	Thr	Leu	Ser	Leu	Leu	Phe	Asp	Asp	Leu	Lys	Ala	Gln
	370					375					380				
Lys	Leu	Asp	Leu	Ser	Lys	Ile	Tyr	Phe	Lys	Asn	Asp	Lys	Ser	Leu	Thr
385					390					395					400
Asp	Leu	Ser	Gln	Gln	Val	Phe	Asp	Asp	Tyr	Ser	Val	Ile	Gly	Thr	Ala
			405						410					415	
Val	Leu	Glu	Tyr	Ile	Thr	Gln	Gln	Val	Ala	Pro	Lys	Asn	Leu	Asp	Asn
			420					425					430		
Pro	Ser	Lys	Lys	Glu	Gln	Xaa	Leu	Ile	Ala	Lys	Lys	Thr	Glu	Lys	Ala
		435					440					445			
Lys	Tyr	Leu	Ser	Leu	Glu	Thr	Ile	Lys	Leu	Ala	Leu	Glu	Glu	Phe	Asn
	450					455					460				
Lys	His	Arg	Asp	Ile	Asp	Lys	Gln	Cys	Arg	Phe	Glu	Glu	Ile	Leu	Ala
465					470					475					480
Asn	Phe	Ala	Ala	Ile	Pro	Met	Ile	Phe	Asp	Glu	Ile	Ala	Gln	Asn	Lys
				485					490					495	
Asp	Asn	Leu	Ala	Gln	Ile	Ser	Xaa	Lys	Tyr	Gln	Asn	Gln	Gly	Lys	Lys
			500					505					510		

ES 2 768 976 T3

Asp	Leu	Leu	Gln	Ala	Ser	Ala	Glu	Xaa	Asp	Val	Lys	Ala	Ile	Lys	Asp
		515					520					525			
Leu	Leu	Asp	Gln	Thr	Asn	Asn	Leu	Leu	His	Xaa	Leu	Lys	Ile	Phe	His
	530					535					540				
Ile	Ser	Gln	Ser	Glu	Asp	Lys	Ala	Asn	Ile	Leu	Asp	Lys	Asp	Glu	His
545					550					555					560
Phe	Tyr	Leu	Val	Phe	Glu	Glu	Cys	Tyr	Phe	Glu	Leu	Ala	Asn	Ile	Val
				565					570					575	
Pro	Leu	Tyr	Asn	Lys	Ile	Arg	Asn	Tyr	Ile	Thr	Gln	Lys	Pro	Tyr	Ser
			580					585					590		
Asp	Glu	Lys	Phe	Lys	Leu	Asn	Phe	Glu	Asn	Ser	Thr	Leu	Ala	Asn	Gly
		595					600					605			
Trp	Asp	Lys	Asn	Lys	Glu	Pro	Asp	Asn	Thr	Ala	Ile	Leu	Phe	Ile	Lys
	610					615					620				
Asp	Asp	Lys	Tyr	Tyr	Leu	Gly	Val	Met	Asn	Lys	Lys	Asn	Asn	Lys	Ile
625					630					635					640
Phe	Asp	Asp	Lys	Ala	Ile	Lys	Glu	Asn	Lys	Gly	Glu	Gly	Tyr	Lys	Lys
				645					650					655	
Ile	Val	Tyr	Lys	Leu	Leu	Pro	Gly	Ala	Asn	Lys	Met	Leu	Pro	Lys	Val
			660					665					670		
Phe	Phe	Ser	Ala	Lys	Ser	Ile	Lys	Phe	Tyr	Asn	Pro	Ser	Glu	Asp	Ile
		675					680					685			
Leu	Arg	Ile	Arg	Asn	His	Ser	Thr	His	Thr	Lys	Asn	Gly	Asn	Pro	Gln
	690					695					700				
Lys	Gly	Tyr	Glu	Lys	Phe	Glu	Phe	Asn	Ile	Glu	Asp	Cys	Arg	Lys	Phe
705					710					715					720
Ile	Asp	Phe	Tyr	Lys	Glu	Ser	Ile	Ser	Lys	His	Pro	Glu	Trp	Lys	Asp
				725					730					735	

ES 2 768 976 T3

Phe	Gly	Phe	Arg	Phe	Ser	Asp	Thr	Gln	Arg	Tyr	Asn	Ser	Ile	Asp	Glu	
			740					745					750			
Phe	Tyr	Arg	Glu	Val	Glu	Asn	Gln	Gly	Tyr	Lys	Leu	Thr	Phe	Glu	Asn	
		755					760					765				
Ile	Ser	Glu	Ser	Tyr	Ile	Asp	Ser	Val	Val	Asn	Gln	Gly	Lys	Leu	Tyr	
	770					775					780					
Leu	Phe	Gln	Ile	Tyr	Asn	Lys	Asp	Phe	Ser	Ala	Tyr	Ser	Lys	Gly	Xaa	
785					790					795					800	
Pro	Asn	Leu	His	Thr	Leu	Tyr	Trp	Lys	Ala	Leu	Phe	Asp	Glu	Arg	Asn	
				805					810					815		
Leu	Gln	Asp	Val	Val	Tyr	Lys	Leu	Asn	Gly	Glu	Ala	Glu	Leu	Phe	Tyr	
			820					825					830			
Arg	Lys	Gln	Ser	Ile	Pro	Lys	Lys	Ile	Thr	His	Pro	Ala	Lys	Glu	Ala	
		835					840					845				
Ile	Ala	Asn	Lys	Asn	Lys	Asp	Asn	Pro	Lys	Lys	Glu	Ser	Val	Phe	Glu	
	850					855					860					
Tyr	Asp	Leu	Ile	Lys	Asp	Lys	Arg	Phe	Thr	Glu	Asp	Lys	Phe	Phe	Phe	
865					870					875					880	
His	Cys	Pro	Ile	Thr	Ile	Asn	Phe	Lys	Ser	Ser	Gly	Ala	Asn	Lys	Phe	
				885					890					895		
Asn	Asp	Glu	Ile	Asn	Leu	Leu	Leu	Lys	Glu	Lys	Ala	Asn	Asp	Val	His	
			900					905					910			
Ile	Leu	Ser	Ile	Asp	Arg	Gly	Glu	Arg	His	Leu	Ala	Tyr	Tyr	Thr	Leu	
		915					920					925				
Val	Asp	Gly	Lys	Gly	Asn	Ile	Ile	Lys	Gln	Asp	Thr	Phe	Asn	Ile	Ile	
	930					935					940					
Gly	Asn	Asp	Arg	Met	Lys	Thr	Asn	Tyr	His	Asp	Lys	Leu	Ala	Ala	Ile	
945					950					955					960	
Glu	Lys	Asp	Arg	Asp	Ser	Ala	Arg	Lys	Asp	Trp	Lys	Lys	Ile	Asn	Asn	
				965					970					975		

ES 2 768 976 T3

Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile	980	985	990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu	995	1000	1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val	1010	1015	1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu	1025	1030	1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg	1040	1045	1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly	1055	1060	1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser	1070	1075	1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys	1085	1090	1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp	1100	1105	1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe	1115	1120	1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr	1130	1135	1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp	1145	1150	1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu	1160	1165	1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly	1175	1180	1185

ES 2 768 976 T3

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Asp Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290

Phe Val Gln Asn Arg Asn Asn Ser Ser Lys Ile
1295 1300

<210> 2

<211> 5

5 <212> PRT

<213> Secuencia artificial

<220>

10 <223> Residuos 786-791 de SEQ ID NO:1

<400> 2

Phe Gln Ile Tyr Asn

1 5

<210> 3

15 <211> 6

<212> PRT

<213> Secuencia artificial

<220>

20 <223> Residuos 786-791 de SEQ ID NO:1

<400> 3

Phe Gln Ile Tyr Asn Lys

1 5

25 <210> 4

<211> 7

<212> PRT

<213> Secuencia artificial

30 <220>

<223> Residuos 786 - 792 de SEQ ID NO: 1

<400> 4

Phe Gln Ile Tyr Asn Lys Asp

1 5

35

<210> 5
 <211> 14
 <212> PRT
 <213> Secuencia artificial
 5
 <220>
 <223> Residuos 781 - 794 de SEQ ID NO:1
 <220>
 10 <221> misc_feature
 <222> (1)..(1)
 <223> Xaa puede ser Gly o Lys
 <220>
 15 <221> misc_feature
 <222> (2)..(2)
 <223> Xaa puede ser Lys, Ser o Asp
 <220>
 20 <221> misc_feature
 <222> (3)..(3)
 <223> Xaa puede ser Leu o Ile
 <220>
 25 <221> misc_feature
 <222> (4)..(4)
 <223> Xaa puede ser Tyr o Phe
 <220>
 30 <221> misc_feature
 <222> (5)..(5)
 <223> Xaa puede ser Leu o Met
 <220>
 35 <221> misc_feature
 <222> (13)..(13)
 <223> Xaa puede ser Phe o Tyr
 <220>
 40 <221> misc_feature
 <222> (14)..(14)
 <223> Xaa puede ser Ser, Ala o Val
 <400> 5
 Xaa Xaa Xaa Xaa Xaa Phe Gln Ile Tyr Asn Lys Asp Xaa Xaa
 45 1 5 10
 <210> 6
 <211> 14
 <212> PRT
 50 <213> Secuencia artificial
 <220>
 <223> Residuos 781 - 794 de SEQ ID NO: 1
 <400> 6
 Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser
 1 5 10
 <210> 7
 <211> 7
 60 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Dominio RuvC de Cpf1 catalíticamente inactivo

<220>
 <221> misc_feature
 <222> (1)..(1)
 5 <223> Xaa puede ser Gly o Ser

 <400> 7
 Xaa Ile Asp Arg Gly Glu Arg
 1 5
 10 <210> 8
 <211> 6
 <212> PRT
 <213> Secuencia artificial
 15 <220>
 <223> Dominio RuvC de Cpf1 catalíticamente inactivo

 <400> 8
 Asp Ala Asn Gly Ala Tyr
 1 5
 20 <210> 9
 <211> 4
 <212> PRT
 <213> Secuencia artificial
 25 <220>
 <223> Dominio RuvC de Cpf1 catalíticamente inactivo

 <220>
 30 <221> misc_feature
 <222> (2)..(2)
 <223> Xaa puede ser Asp, Asn o Glu

 <400> 9
 Glu Xaa Leu Asn
 35 1
 <210> 10
 <211> 4
 <212> PRT
 40 <213> Secuencia artificial

 <220>
 <223> Dominio RuvC de Cpf1 catalíticamente inactivo
 45 <400> 10
 Glu Asp Leu Asn
 1
 <210> 11
 <211> 7
 50 <212> PRT
 <213> Secuencia artificial

 <220>
 <223> dominio RuvC
 55 <400> 11
 Ser Ile Asp Arg Gly Glu Arg
 1 5
 <210> 12
 <211> 7
 60 <212> PRT
 <213> Secuencia artificial

<220>
 <223> dominio RuvC

 <400> 12
 Gly Ile Asp Arg Gly Glu Arg
 5 1 5

 <210> 13
 <211> 4
 <212> PRT
 10 <213> Secuencia artificial

 <220>
 <223>
 Dominio RuvC
 15
 <400> 13
 Glu Asn Leu Asn
 1

 <210> 14
 20 <211> 4
 <212> PRT
 <213> Secuencia artificial

 <220>
 25 <223> dominio RuvC

 <400> 14
 Glu Glu Leu Asn
 1

 30 <210> 15
 <211> 19
 <212> PRT
 <213> Secuencia artificial

 35 <220>
 <223> Motivo rico en arginina

 <220>
 <221> misc_feature
 40 <222> (1)..(1)
 <223> Xaa puede ser Asp o Asn

 <220>
 <221> misc_feature
 45 <222> (3)..(3)
 <223> Xaa puede ser Arg, Gln o His

 <220>
 <221> misc_feature
 50 <222> (4)..(4)
 <223> Xaa puede ser Lys, Glu, Ser o Asp

 <220>
 <221> misc_feature
 55 <222> (5)..(5)
 <223> Xaa puede ser Ala, Lys o Leu

 <220>
 <221> misc_feature
 60 <222> (7)..(7)
 <223> Xaa puede ser Asp, Asn o Ala

 <220>

```

<221> misc_feature
<222> (8)..(8)
<223> Xaa puede ser Val, Asn, Gln, Lys o Ala

5    <220>
      <221> misc_feature
      <222> (9)..(9)
      <223> Xaa puede ser Arg, Lys o Ile

10   <220>
      <221> misc_feature
      <222> (11)..(11)
      <223> Xaa puede ser Tyr, Lys o Ile

15   <220>
      <221> misc_feature
      <222> (12)..(12)
      <223> Xaa puede ser Asp o Glu

20   <220>
      <221> misc_feature
      <222> (13)..(13)
      <223> Xaa puede ser Asn, Arg o Met

25   <220>
      <221> misc_feature
      <222> (14)..(14)
      <223> Xaa puede ser Lys, Val, Phe o Asp

30   <220>
      <221> misc_feature
      <222> (15)..(15)
      <223> Xaa puede ser Glu, Ala, Asp o Ser

35   <220>
      <221> misc_feature
      <222> (18)..(18)
      <223> Xaa puede ser Arg, Gln o Lys

40   <220>
      <221> misc_feature
      <222> (19)..(19)
      <223> Xaa puede ser Asn, Ala, Ser o Asp

45   <400> 15
      Xaa Tyr Xaa Xaa Xaa Leu Xaa Xaa Xaa Glu Xaa Xaa Xaa Xaa Xaa Ala
      1          5          10          15

      Arg Xaa Xaa

50   <210> 16
      <211> 19
      <212> PRT
      <213> Secuencia artificial

55   <220>
      <223> Motivo rico en arginina

      <400> 16
      Asp Tyr Arg Lys Ala Leu Asp Val Arg Glu Tyr Asp Asn Lys Glu Ala
      1          5          10          15

      Arg Arg Asn

60   <210> 17

```

ES 2 768 976 T3

<211> 19
<212> PRT
<213> Secuencia artificial

5 <220>
<223> Motivo rico en arginina

<400> 17
Asp Tyr Gln Lys Lys Leu Asp Asn Arg Glu Lys Glu Arg Val Ala Ala
1 5 10 15

Arg Gln Ala

10

<210> 18
<211> 19
<212> PRT
<213> Secuencia artificial

15
 <220>
 <223> Motivo rico en arginina

<400> 18
Asp Tyr Arg Glu Lys Leu Asn Gln Arg Glu Ile Glu Met Lys Asp Ala
1 5 10 15

20 Arg Gln Ser

25 <210> 19
 <211> 19
 <212> PRT
 <213> Secuencia artificial

<220>
<223> Motivo rico en arginina

30 <400> 19
Asp Tyr His Ser Leu Leu Asp Lys Lys Glu Lys Glu Arg Phe Glu Ala
 5 10 15

Arg Gln Asn

35 <210> 20
 <211> 19
 <212> PRT
 <213> Secuencia artificial

<220>
<223> Motivo rico en arginina

40 <400> 20
Asn Tyr His Asp Lys Leu Ala Ala Ile Glu Lys Asp Arg Asp Ser Ala
1 5 10 15

Arg Lys Asp

45 <210> 21
 <211> 29
 <212> ADN
 <213> Francisella novicida

```
<400> 21
agattaaaag gtaattctat cttgttgag
29
```

<210> 22
 <211> 45
 <212> ADN
 <213> Francisella novicida
 5
 <400> 22
 ataatttaag attaaaaggt aattctatatt tgttgagatc tgagc
 45
 <210> 23
 10 <211> 26
 <212> ADN
 <213> Francisella novicida
 <400> 23
 tagcgattta tgaaggtcat tttttt
 15 26
 <210> 24
 <211> 42
 <212> ADN
 20 <213> Francisella novicida
 <400> 24
 ctaaattata gcgatttatg aaggtcattt ttttaaaaag tt
 42
 25 <210> 25
 <211> 29
 <212> ADN
 <213> Francisella novicida
 30 <400> 25
 atggattatt acttaactgg agtgtttac
 29
 <210> 26
 <211> 45
 35 <212> ADN
 <213> Francisella novicida
 <400> 26
 aatgttcaat ggattattac ttaattggag tgtctacgtc gatgg
 45
 40 <210> 27
 <211> 28
 <212> ADN
 <213> Francisella novicida
 45 <400> 27
 gccacaaata ctacaaaaaa taacttaa
 28
 <210> 28
 50 <211> 44
 <212> ADN
 <213> Francisella novicida
 <400> 28
 attttttggc tccaaatact acaaaaaata acttaaacctt tgaa
 55 44
 <210> 29
 <211> 32
 <212> ADN
 60 <213> Francisella novicida

ES 2 768 976 T3

<400> 29
attgtcaaaa cataagcagc tgcttcaaat at
32

5 <210> 30
<211> 48
<212> ADN
<213> Francisella novicida

10 <400> 30
ggtcttttac tggtattaca taagcagccg cttcaaatat cttagcaa
48

<210> 31
<211> 10
15 <212> PRT
<213> Secuencia artificial

<220>
<223> Un polipéptido aislado o fragmento del mismo que tiene una actividad de nucleasa.

20 <400> 31
Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe
1 5 10

25 <210> 32
<211> 1184
<212> PRT
<213> Methanomethylophilus alvus

<400> 32
Met Asp Ala Lys Glu Phe Thr Gly Gln Tyr Pro Leu Ser Lys Thr Leu
1 5 10 15

Arg Phe Glu Leu Arg Pro Ile Gly Arg Thr Trp Asp Asn Leu Glu Ala
20 25 30

Ser Gly Tyr Leu Ala Glu Asp Arg His Arg Ala Glu Cys Tyr Pro Arg
35 40 45

30 Ala Lys Glu Leu Leu Asp Asp Asn His Arg Ala Phe Leu Asn Arg Val
50 55 60

ES 2 768 976 T3

Leu Pro Gln Ile Gln Asp Ala Asp Gly Tyr Lys Gly Leu Phe Ala Lys
65 70 75 80

Pro Ala Leu Asp Glu Ala Met Lys Ile Ala Lys Glu Asn Gly Asn Glu
85 90 95

Ser Asp Ile Glu Val Leu Glu Ala Phe Asn Gly Phe Ser Val Tyr Phe
100 105 110

Thr Gly Tyr His Glu Ser Arg Glu Asn Ile Tyr Ser Asp Glu Asp Met
115 120 125

Val Ser Val Ala Tyr Arg Ile Thr Glu Asp Asn Phe Pro Arg Phe Val
130 135 140

Ser Asn Ala Leu Ile Phe Asp Lys Leu Asn Glu Ser His Pro Asp Ile
145 150 155 160

Ile Ser Glu Val Ser Gly Asn Leu Gly Val Asp Asp Ile Gly Lys Tyr
165 170 175

Phe Asp Val Ser Asn Tyr Asn Asn Phe Leu Ser Gln Ala Gly Ile Asp
180 185 190

Asp Tyr Asn His Ile Ile Gly Gly His Thr Thr Glu Asp Gly Leu Ile
195 200 205

Gln Ala Phe Asn Val Val Leu Asn Leu Arg His Gln Lys Asp Pro Gly
210 215 220

Phe Glu Lys Ile Gln Phe Lys Gln Leu Tyr Lys Gln Ile Leu Ser Val
225 230 235 240

Arg Thr Ser Lys Ser Tyr Ile Pro Lys Gln Phe Asp Asn Ser Lys Glu
245 250 255

Met Val Asp Cys Ile Cys Asp Tyr Val Ser Lys Ile Glu Lys Ser Glu
260 265 270

Thr Val Glu Arg Ala Leu Lys Leu Val Arg Asn Ile Ser Ser Phe Asp
275 280 285

Leu Arg Gly Ile Phe Val Asn Lys Lys Asn Leu Arg Ile Leu Ser Asn

ES 2 768 976 T3

290		295		300
Lys Leu Ile Gly Asp Trp Asp Ala Ile Glu Thr Ala Leu Met His Ser				
305		310		315 320
Ser Ser Ser Glu Asn Asp Lys Lys Ser Val Tyr Asp Ser Ala Glu Ala				
	325		330	335
Phe Thr Leu Asp Asp Ile Phe Ser Ser Val Lys Lys Phe Ser Asp Ala				
	340		345	350
Ser Ala Glu Asp Ile Gly Asn Arg Ala Glu Asp Ile Cys Arg Val Ile				
	355		360	365
Ser Glu Thr Ala Pro Phe Ile Asn Asp Leu Arg Ala Val Asp Leu Asp				
	370		375	380
Ser Leu Asn Asp Asp Gly Tyr Glu Ala Ala Val Ser Lys Ile Arg Glu				
385		390		395 400
Ser Leu Glu Pro Tyr Met Asp Leu Phe His Glu Leu Glu Ile Phe Ser				
	405		410	415
Val Gly Asp Glu Phe Pro Lys Cys Ala Ala Phe Tyr Ser Glu Leu Glu				
	420		425	430
Glu Val Ser Glu Gln Leu Ile Glu Ile Ile Pro Leu Phe Asn Lys Ala				
	435		440	445
Arg Ser Phe Cys Thr Arg Lys Arg Tyr Ser Thr Asp Lys Ile Lys Val				
	450		455	460
Asn Leu Lys Phe Pro Thr Leu Ala Asp Gly Trp Asp Leu Asn Lys Glu				
465		470		475 480
Arg Asp Asn Lys Ala Ala Ile Leu Arg Lys Asp Gly Lys Tyr Tyr Leu				
	485		490	495
Ala Ile Leu Asp Met Lys Lys Asp Leu Ser Ser Ile Arg Thr Ser Asp				
	500		505	510
Glu Asp Glu Ser Ser Phe Glu Lys Met Glu Tyr Lys Leu Leu Pro Ser				
	515		520	525

ES 2 768 976 T3

Pro Val Lys Met Leu Pro Lys Ile Phe Val Lys Ser Lys Ala Ala Lys
530 535 540

Glu Lys Tyr Gly Leu Thr Asp Arg Met Leu Glu Cys Tyr Asp Lys Gly
545 550 555 560

Met His Lys Ser Gly Ser Ala Phe Asp Leu Gly Phe Cys His Glu Leu
565 570 575

Ile Asp Tyr Tyr Lys Arg Cys Ile Ala Glu Tyr Pro Gly Trp Asp Val
580 585 590

Phe Asp Phe Lys Phe Arg Glu Thr Ser Asp Tyr Gly Ser Met Lys Glu
595 600 605

Phe Asn Glu Asp Val Ala Gly Ala Gly Tyr Tyr Met Ser Leu Arg Lys
610 615 620

Ile Pro Cys Ser Glu Val Tyr Arg Leu Leu Asp Glu Lys Ser Ile Tyr
625 630 635 640

Leu Phe Gln Ile Tyr Asn Lys Asp Tyr Ser Glu Asn Ala His Gly Asn
645 650 655

Lys Asn Met His Thr Met Tyr Trp Glu Gly Leu Phe Ser Pro Gln Asn
660 665 670

Leu Glu Ser Pro Val Phe Lys Leu Ser Gly Gly Ala Glu Leu Phe Phe
675 680 685

Arg Lys Ser Ser Ile Pro Asn Asp Ala Lys Thr Val His Pro Lys Gly
690 695 700

Ser Val Leu Val Pro Arg Asn Asp Val Asn Gly Arg Arg Ile Pro Asp
705 710 715 720

Ser Ile Tyr Arg Glu Leu Thr Arg Tyr Phe Asn Arg Gly Asp Cys Arg
725 730 735

Ile Ser Asp Glu Ala Lys Ser Tyr Leu Asp Lys Val Lys Thr Lys Lys
740 745 750

Ala Asp His Asp Ile Val Lys Asp Arg Arg Phe Thr Val Asp Lys Met

ES 2 768 976 T3

755								760								765
Met	Phe	His	Val	Pro	Ile	Ala	Met	Asn	Phe	Lys	Ala	Ile	Ser	Lys	Pro	
770						775					780					
Asn	Leu	Asn	Lys	Lys	Val	Ile	Asp	Gly	Ile	Ile	Asp	Asp	Gln	Asp	Leu	
785					790					795					800	
Lys	Ile	Ile	Gly	Ile	Asp	Arg	Gly	Glu	Arg	Asn	Leu	Ile	Tyr	Val	Thr	
				805					810					815		
Met	Val	Asp	Arg	Lys	Gly	Asn	Ile	Leu	Tyr	Gln	Asp	Ser	Leu	Asn	Ile	
			820					825					830			
Leu	Asn	Gly	Tyr	Asp	Tyr	Arg	Lys	Ala	Leu	Asp	Val	Arg	Glu	Tyr	Asp	
		835					840					845				
Asn	Lys	Glu	Ala	Arg	Arg	Asn	Trp	Thr	Lys	Val	Glu	Gly	Ile	Arg	Lys	
	850					855					860					
Met	Lys	Glu	Gly	Tyr	Leu	Ser	Leu	Ala	Val	Ser	Lys	Leu	Ala	Asp	Met	
865					870					875					880	
Ile	Ile	Glu	Asn	Asn	Ala	Ile	Ile	Val	Met	Glu	Asp	Leu	Asn	His	Gly	
				885					890					895		
Phe	Lys	Ala	Gly	Arg	Ser	Lys	Ile	Glu	Lys	Gln	Val	Tyr	Gln	Lys	Phe	
			900					905					910			
Glu	Ser	Met	Leu	Ile	Asn	Lys	Leu	Gly	Tyr	Met	Val	Leu	Lys	Asp	Lys	
		915					920					925				
Ser	Ile	Asp	Gln	Ser	Gly	Gly	Ala	Leu	His	Gly	Tyr	Gln	Leu	Ala	Asn	
	930					935					940					
His	Val	Thr	Thr	Leu	Ala	Ser	Val	Gly	Lys	Gln	Cys	Gly	Val	Ile	Phe	
945					950					955					960	
Tyr	Ile	Pro	Ala	Ala	Phe	Thr	Ser	Lys	Ile	Asp	Pro	Thr	Thr	Gly	Phe	
				965					970					975		
Ala	Asp	Leu	Phe	Ala	Leu	Ser	Asn	Val	Lys	Asn	Val	Ala	Ser	Met	Arg	
		980						985					990			

ES 2 768 976 T3

Glu Phe Phe Ser Lys Met Lys Ser Val Ile Tyr Asp Lys Ala Glu Gly
995 1000 1005

Lys Phe Ala Phe Thr Phe Asp Tyr Leu Asp Tyr Asn Val Lys Ser
1010 1015 1020

Glu Cys Gly Arg Thr Leu Trp Thr Val Tyr Thr Val Gly Glu Arg
1025 1030 1035

Phe Thr Tyr Ser Arg Val Asn Arg Glu Tyr Val Arg Lys Val Pro
1040 1045 1050

Thr Asp Ile Ile Tyr Asp Ala Leu Gln Lys Ala Gly Ile Ser Val
1055 1060 1065

Glu Gly Asp Leu Arg Asp Arg Ile Ala Glu Ser Asp Gly Asp Thr
1070 1075 1080

Leu Lys Ser Ile Phe Tyr Ala Phe Lys Tyr Ala Leu Asp Met Arg
1085 1090 1095

Val Glu Asn Arg Glu Glu Asp Tyr Ile Gln Ser Pro Val Lys Asn
1100 1105 1110

Ala Ser Gly Glu Phe Phe Cys Ser Lys Asn Ala Gly Lys Ser Leu
1115 1120 1125

Pro Gln Asp Ser Asp Ala Asn Gly Ala Tyr Asn Ile Ala Leu Lys
1130 1135 1140

Gly Ile Leu Gln Leu Arg Met Leu Ser Glu Gln Tyr Asp Pro Asn
1145 1150 1155

Ala Glu Ser Ile Arg Leu Pro Leu Ile Thr Asn Lys Ala Trp Leu
1160 1165 1170

Thr Phe Met Gln Ser Gly Met Lys Thr Trp Lys
1175 1180

<210> 33

<211> 1265

5 <212> PRT

<213> Acidaminococcus sp. BV3L6 8

<400> 33

ES 2 768 976 T3

Met	Thr	Gln	Phe	Glu	Gly	Phe	Thr	Asn	Leu	Tyr	Gln	Val	Ser	Lys	Thr
1				5					10					15	
Leu	Arg	Phe	Glu	Leu	Ile	Pro	Gln	Gly	Lys	Thr	Leu	Lys	His	Ile	Gln
			20					25					30		
Glu	Gln	Gly	Phe	Ile	Glu	Glu	Asp	Lys	Ala	Arg	Asn	Asp	His	Tyr	Lys
		35					40					45			
Glu	Leu	Lys	Pro	Ile	Ile	Asp	Arg	Ile	Tyr	Lys	Thr	Tyr	Ala	Asp	Gln
	50					55					60				
Cys	Leu	Gln	Leu	Val	Gly	Arg	Thr	Asp	Asn	Leu	Thr	Asp	Ala	Ile	Asn
65					70					75					80
Lys	Arg	His	Ala	Glu	Ile	Tyr	Lys	Gly	Leu	Phe	Lys	Ala	Glu	Leu	Phe
				85					90					95	
Asn	Gly	Lys	Val	Leu	Lys	Gln	Leu	Gly	Thr	Val	Thr	Thr	Thr	Glu	His
			100					105						110	
Glu	Asn	Ala	Leu	Leu	Arg	Ser	Phe	Asp	Lys	Phe	Thr	Thr	Tyr	Phe	Ser
		115					120					125			
Gly	Phe	Tyr	Glu	Asn	Arg	Lys	Asn	Val	Phe	Ser	Ala	Glu	Asp	Ile	Ser
	130					135					140				
Thr	Ala	Ile	Pro	His	Arg	Ile	Val	Gln	Asp	Asn	Phe	Pro	Lys	Phe	Lys
145					150					155					160
Glu	Asn	Cys	His	Ile	Phe	Thr	Arg	Leu	Ile	Thr	Ala	Val	Pro	Ser	Leu
				165				170						175	
Arg	Glu	His	Phe	Glu	Asn	Val	Lys	Lys	Ala	Ile	Gly	Ile	Phe	Val	Ser
			180					185					190		
Thr	Ser	Ile	Glu	Glu	Val	Phe	Ser	Phe	Pro	Phe	Tyr	Asn	Gln	Leu	Leu
		195					200					205			
Thr	Gln	Thr	Gln	Ile	Asp	Leu	Tyr	Asn	Gln	Leu	Leu	Gly	Gly	Ile	Ser
	210					215					220				

ES 2 768 976 T3

Arg Glu Ala Gly Thr Glu Lys Ile Lys Gly Leu Asn Glu Val Leu Asn
 225 230 235 240
 Leu Ala Ile Gln Lys Asn Asp Glu Thr Ala His Ile Ile Ala Ser Leu
 245 250 255
 Pro His Arg Phe Ile Pro Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn
 260 265 270
 Thr Leu Ser Phe Ile Leu Glu Glu Phe Lys Ser Asp Glu Glu Val Ile
 275 280 285
 Gln Ser Phe Cys Lys Tyr Lys Thr Leu Leu Arg Asn Glu Asn Val Leu
 290 295 300
 Glu Thr Ala Glu Ala Leu Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr
 305 310 315 320
 His Ile Phe Ile Ser His Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu
 325 330 335
 Cys Asp His Trp Asp Thr Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile
 340 345 350
 Ser Glu Leu Thr Gly Lys Ile Thr Lys Ser Ala Lys Glu Lys Val Gln
 355 360 365
 Arg Ser Leu Lys His Glu Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala
 370 375 380
 Ala Gly Lys Glu Leu Ser Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile
 385 390 395 400
 Leu Ser His Ala His Ala Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu
 405 410 415
 Lys Lys Gln Glu Glu Lys Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu
 420 425 430
 Leu Gly Leu Tyr His Leu Leu Asp Trp Phe Ala Val Asp Glu Ser Asn
 435 440 445
 Glu Val Asp Pro Glu Phe Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu
 450 455 460

ES 2 768 976 T3

Met	Glu	Pro	Ser	Leu	Ser	Phe	Tyr	Asn	Lys	Ala	Arg	Asn	Tyr	Ala	Thr
465					470					475					480
Lys	Lys	Pro	Tyr	Ser	Val	Glu	Lys	Phe	Lys	Leu	Asn	Phe	Gln	Met	Pro
				485					490					495	
Thr	Leu	Ala	Ser	Gly	Trp	Asp	Val	Asn	Lys	Glu	Lys	Asn	Asn	Gly	Ala
			500					505					510		
Ile	Leu	Phe	Val	Lys	Asn	Gly	Leu	Tyr	Tyr	Leu	Gly	Ile	Met	Pro	Lys
		515					520					525			
Gln	Lys	Gly	Arg	Tyr	Lys	Ala	Leu	Ser	Phe	Glu	Pro	Thr	Glu	Lys	Thr
	530					535					540				
Ser	Glu	Gly	Phe	Asp	Lys	Met	Tyr	Tyr	Asp	Tyr	Phe	Pro	Asp	Ala	Ala
545					550					555					560
Lys	Met	Ile	Pro	Lys	Cys	Ser	Thr	Gln	Leu	Lys	Ala	Val	Thr	Ala	His
				565					570					575	
Phe	Gln	Thr	His	Thr	Thr	Pro	Ile	Leu	Leu	Ser	Asn	Asn	Phe	Ile	Glu
			580					585					590		
Pro	Leu	Glu	Ile	Thr	Lys	Glu	Ile	Tyr	Asp	Leu	Asn	Asn	Pro	Glu	Lys
		595					600					605			
Glu	Pro	Lys	Lys	Phe	Gln	Thr	Ala	Tyr	Ala	Lys	Lys	Thr	Gly	Asp	Gln
	610					615					620				
Lys	Gly	Tyr	Arg	Glu	Ala	Leu	Cys	Lys	Trp	Ile	Asp	Phe	Thr	Arg	Asp
625					630					635					640
Phe	Leu	Ser	Lys	Tyr	Thr	Lys	Thr	Thr	Ser	Ile	Asp	Leu	Ser	Ser	Leu
				645					650					655	
Arg	Pro	Ser	Ser	Gln	Tyr	Lys	Asp	Leu	Gly	Glu	Tyr	Tyr	Ala	Glu	Leu
			660					665					670		
Asn	Pro	Leu	Leu	Tyr	His	Ile	Ser	Phe	Gln	Arg	Ile	Ala	Glu	Lys	Glu
		675					680					685			

ES 2 768 976 T3

Ile Met Asp Ala Val Glu Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr
690 695 700

Asn Lys Asp Phe Ala Lys Gly His His Gly Lys Pro Asn Leu His Thr
705 710 715 720

Leu Tyr Trp Thr Gly Leu Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser
725 730 735

Ile Lys Leu Asn Gly Gln Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg
740 745 750

Met Lys Arg Met Ala His Arg Leu Gly Glu Lys Met Leu Asn Lys Lys
755 760 765

Leu Lys Asp Gln Lys Thr Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu
770 775 780

Tyr Asp Tyr Val Asn His Arg Leu Ser His Asp Leu Ser Asp Glu Ala
785 790 795 800

Arg Ala Leu Leu Pro Asn Val Ile Thr Lys Glu Val Ser His Glu Ile
805 810 815

Ile Lys Asp Arg Arg Phe Thr Ser Asp Lys Phe Phe Phe His Val Pro
820 825 830

Ile Thr Leu Asn Tyr Gln Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln
835 840 845

Arg Val Asn Ala Tyr Leu Lys Glu His Pro Glu Thr Pro Ile Ile Gly
850 855 860

Ile Asp Arg Gly Glu Arg Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser
865 870 875 880

Thr Gly Lys Ile Leu Glu Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe
885 890 895

Asp Tyr Gln Lys Lys Leu Asp Asn Arg Glu Lys Glu Arg Val Ala Ala
900 905 910

Arg Gln Ala Trp Ser Val Val Gly Thr Ile Lys Asp Leu Lys Gln Gly
915 920 925

ES 2 768 976 T3

Tyr Leu Ser Gln Val Ile His Glu Ile Val Asp Leu Met Ile His Tyr
 930 935 940

Gln Ala Val Val Val Leu Glu Asn Leu Asn Phe Gly Phe Lys Ser Lys
 945 950 955 960

Arg Thr Gly Ile Ala Glu Lys Ala Val Tyr Gln Gln Phe Glu Lys Met
 965 970 975

Leu Ile Asp Lys Leu Asn Cys Leu Val Leu Lys Asp Tyr Pro Ala Glu
 980 985 990

Lys Val Gly Gly Val Leu Asn Pro Tyr Gln Leu Thr Asp Gln Phe Thr
 995 1000 1005

Ser Phe Ala Lys Met Gly Thr Gln Ser Gly Phe Leu Phe Tyr Val
 1010 1015 1020

Pro Ala Pro Tyr Thr Ser Lys Ile Asp Pro Leu Thr Gly Phe Val
 1025 1030 1035

Asp Pro Phe Val Trp Lys Thr Ile Lys Asn His Glu Ser Arg Lys
 1040 1045 1050

His Phe Leu Glu Gly Phe Asp Phe Leu His Tyr Asp Val Lys Thr
 1055 1060 1065

Gly Asp Phe Ile Leu His Phe Lys Met Asn Arg Asn Leu Ser Phe
 1070 1075 1080

Gln Arg Gly Leu Pro Gly Phe Met Pro Ala Trp Asp Ile Val Phe
 1085 1090 1095

Glu Lys Asn Glu Thr Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile
 1100 1105 1110

Ala Gly Lys Arg Ile Val Pro Val Ile Glu Asn His Arg Phe Thr
 1115 1120 1125

Gly Arg Tyr Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu
 1130 1135 1140

ES 2 768 976 T3

Leu Glu Glu Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu
1145 1150 1155

Pro Lys Leu Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met
1160 1165 1170

Val Ala Leu Ile Arg Ser Val Leu Gln Met Arg Asn Ser Asn Ala
1175 1180 1185

Ala Thr Gly Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn
1190 1195 1200

Gly Val Cys Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met
1205 1210 1215

Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln
1220 1225 1230

Leu Leu Leu Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln
1235 1240 1245

Asn Gly Ile Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu
1250 1255 1260

Arg Asn
1265

<210> 34

<211> 1187

5 <212> PRT

<213> Lachnospiraceae

<400> 34

Met Ser Lys Leu Glu Lys Phe Thr Asn Cys Tyr Ser Leu Ser Lys Thr
1 5 10 15

Leu Arg Phe Lys Ala Ile Pro Val Gly Lys Thr Gln Glu Asn Ile Asp
20 25 30

Asn Lys Arg Leu Leu Val Glu Asp Glu Lys Arg Ala Glu Asp Tyr Lys
35 40 45

Gly Val Lys Lys Leu Leu Asp Arg Tyr Tyr Leu Ser Phe Ile Asn Asp
50 55 60

10

ES 2 768 976 T3

Val Leu His Ser Ile Gly Asn Glu Gly Tyr Lys Ser Leu Phe Lys Lys
65 70 75 80

Asp Ile Ile Glu Thr Ile Leu Pro Glu Phe Leu Asp Asp Lys Asp Glu
85 90 95

Ile Ala Leu Val Asn Ser Phe Asn Gly Phe Thr Thr Ala Phe Thr Gly
100 105 110

Phe Phe Asp Asn Arg Glu Asn Met Phe Ser Glu Glu Ala Lys Ser Thr
115 120 125

Ser Ile Ala Phe Arg Cys Ile Asn Glu Asn Leu Thr Arg Tyr Ile Ser
130 135 140

Asn Met Asp Ile Phe Glu Lys Val Asp Ala Ile Phe Asp Lys His Glu
145 150 155 160

Val Gln Glu Ile Lys Glu Lys Ile Leu Asn Ser Asp Tyr Asp Val Glu
165 170 175

Asp Phe Phe Glu Gly Glu Phe Phe Asn Phe Val Leu Thr Gln Glu Gly
180 185 190

Ile Asp Val Tyr Asn Ala Ile Ile Gly Gly Phe Val Thr Glu Ser Gly
195 200 205

Glu Lys Ile Lys Gly Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Lys
210 215 220

Thr Lys Gln Lys Leu Pro Lys Phe Lys Pro Leu Tyr Lys Gln Val Leu
225 230 235 240

Ser Asp Arg Glu Ser Leu Ser Phe Tyr Gly Glu Gly Tyr Thr Ser Asp
245 250 255

Glu Glu Val Leu Glu Val Phe Arg Asn Thr Leu Asn Lys Asn Ser Glu
260 265 270

Ile Phe Ser Ser Ile Lys Lys Leu Glu Lys Leu Phe Lys Asn Phe Asp
275 280 285

Glu Tyr Ser Ser Ala Gly Ile Phe Val Lys Asn Gly Pro Ala Ile Ser

ES 2 768 976 T3

290		295		300
Thr Ile Ser Lys Asp Ile Phe Gly Glu Trp Asn Val Ile Arg Asp Lys				
305		310		315 320
Trp Asn Ala Glu Tyr Asp Asp Ile His Leu Lys Lys Lys Ala Val Val				
	325		330	335
Thr Glu Lys Tyr Glu Asp Asp Arg Arg Lys Ser Phe Lys Lys Ile Gly				
	340		345	350
Ser Phe Ser Leu Glu Gln Leu Gln Glu Tyr Ala Asp Ala Asp Leu Ser				
	355		360	365
Val Val Glu Lys Leu Lys Glu Ile Ile Ile Gln Lys Val Asp Glu Ile				
	370		375	380
Tyr Lys Val Tyr Gly Ser Ser Glu Lys Leu Phe Asp Ala Asp Phe Val				
385		390		395 400
Leu Glu Lys Ser Leu Lys Lys Asn Asp Ala Val Val Ala Ile Met Lys				
	405		410	415
Asp Leu Leu Asp Ser Val Lys Ser Phe Glu Asn Tyr Ile Lys Ala Phe				
	420		425	430
Phe Gly Glu Gly Lys Glu Thr Asn Arg Asp Glu Ser Phe Tyr Gly Asp				
	435		440	445
Phe Val Leu Ala Tyr Asp Ile Leu Leu Lys Val Asp His Ile Tyr Asp				
	450		455	460
Ala Ile Arg Asn Tyr Val Thr Gln Lys Pro Tyr Ser Lys Asp Lys Phe				
465		470		475 480
Lys Leu Tyr Phe Gln Asn Pro Gln Phe Met Gly Gly Trp Asp Lys Asp				
	485		490	495
Lys Glu Thr Asp Tyr Arg Ala Thr Ile Leu Arg Tyr Gly Ser Lys Tyr				
	500		505	510
Tyr Leu Ala Ile Met Asp Lys Lys Tyr Ala Lys Cys Leu Gln Lys Ile				
	515		520	525

ES 2 768 976 T3

Asp	Lys	Asp	Asp	Val	Asn	Gly	Asn	Tyr	Glu	Lys	Ile	Asn	Tyr	Lys	Leu
530						535					540				
Leu	Pro	Gly	Pro	Asn	Lys	Met	Leu	Pro	Lys	Val	Phe	Phe	Ser	Lys	Lys
545					550					555					560
Trp	Met	Ala	Tyr	Tyr	Asn	Pro	Ser	Glu	Asp	Ile	Gln	Lys	Ile	Tyr	Lys
				565					570					575	
Asn	Gly	Thr	Phe	Lys	Lys	Gly	Asp	Met	Phe	Asn	Leu	Asn	Asp	Cys	His
			580					585					590		
Lys	Leu	Ile	Asp	Phe	Phe	Lys	Asp	Ser	Ile	Ser	Arg	Tyr	Pro	Lys	Trp
		595					600					605			
Ser	Asn	Ala	Tyr	Asp	Phe	Asn	Phe	Ser	Glu	Thr	Glu	Lys	Tyr	Lys	Asp
	610					615					620				
Ile	Ala	Gly	Phe	Tyr	Arg	Glu	Val	Glu	Glu	Gln	Gly	Tyr	Lys	Val	Ser
625					630					635					640
Phe	Glu	Ser	Ala	Ser	Lys	Lys	Glu	Val	Asp	Lys	Leu	Val	Glu	Glu	Gly
				645					650					655	
Lys	Leu	Tyr	Met	Phe	Gln	Ile	Tyr	Asn	Lys	Asp	Phe	Ser	Asp	Lys	Ser
			660					665					670		
His	Gly	Thr	Pro	Asn	Leu	His	Thr	Met	Tyr	Phe	Lys	Leu	Leu	Phe	Asp
		675					680					685			
Glu	Asn	Asn	His	Gly	Gln	Ile	Arg	Leu	Ser	Gly	Gly	Ala	Glu	Leu	Phe
	690					695					700				
Met	Arg	Arg	Ala	Ser	Leu	Lys	Lys	Glu	Glu	Leu	Val	Val	His	Pro	Ala
705					710					715					720
Asn	Ser	Pro	Ile	Ala	Asn	Lys	Asn	Pro	Asp	Asn	Pro	Lys	Lys	Thr	Thr
				725					730					735	
Thr	Leu	Ser	Tyr	Asp	Val	Tyr	Lys	Asp	Lys	Arg	Phe	Ser	Glu	Asp	Gln
			740					745					750		
Tyr	Glu	Leu	His	Ile	Pro	Ile	Ala	Ile	Asn	Lys	Cys	Pro	Lys	Asn	Ile

ES 2 768 976 T3

755								760								765
Phe	Lys	Ile	Asn	Thr	Glu	Val	Arg	Val	Leu	Leu	Lys	His	Asp	Asp	Asn	
770						775					780					
Pro	Tyr	Val	Ile	Gly	Ile	Asp	Arg	Gly	Glu	Arg	Asn	Leu	Leu	Tyr	Ile	
785					790					795					800	
Val	Val	Val	Asp	Gly	Lys	Gly	Asn	Ile	Val	Glu	Gln	Tyr	Ser	Leu	Asn	
				805					810					815		
Glu	Ile	Ile	Asn	Asn	Phe	Asn	Gly	Ile	Arg	Ile	Lys	Thr	Asp	Tyr	His	
			820					825					830			
Ser	Leu	Leu	Asp	Lys	Lys	Glu	Lys	Glu	Arg	Phe	Glu	Ala	Arg	Gln	Asn	
	835						840					845				
Trp	Thr	Ser	Ile	Glu	Asn	Ile	Lys	Glu	Leu	Lys	Ala	Gly	Tyr	Ile	Ser	
	850					855					860					
Gln	Val	Val	His	Lys	Ile	Cys	Glu	Leu	Val	Glu	Lys	Tyr	Asp	Ala	Val	
865					870					875					880	
Ile	Ala	Leu	Glu	Asp	Leu	Asn	Ser	Gly	Phe	Lys	Asn	Ser	Arg	Val	Lys	
				885					890					895		
Val	Glu	Lys	Gln	Val	Tyr	Gln	Lys	Phe	Glu	Lys	Met	Leu	Ile	Asp	Lys	
			900					905					910			
Leu	Asn	Tyr	Met	Val	Asp	Lys	Lys	Ser	Asn	Pro	Cys	Ala	Thr	Gly	Gly	
	915						920					925				
Ala	Leu	Lys	Gly	Tyr	Gln	Ile	Thr	Asn	Lys	Phe	Glu	Ser	Phe	Lys	Ser	
	930					935					940					
Met	Ser	Thr	Gln	Asn	Gly	Phe	Ile	Phe	Tyr	Ile	Pro	Ala	Trp	Leu	Thr	
945					950					955					960	
Ser	Lys	Ile	Asp	Pro	Ser	Thr	Gly	Phe	Val	Asn	Leu	Leu	Lys	Thr	Lys	
				965					970					975		
Tyr	Thr	Ser	Ile	Ala	Asp	Ser	Lys	Lys	Phe	Ile	Ser	Ser	Phe	Asp	Arg	
			980					985					990			

ES 2 768 976 T3

Ile Met Tyr Val Pro Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr
 995 1000 1005

Lys Asn Phe Ser Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys
 1010 1015 1020

Leu Tyr Ser Tyr Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys
 1025 1030 1035

Lys Asn Asn Val Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala
 1040 1045 1050

Tyr Lys Glu Leu Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly
 1055 1060 1065

Asp Ile Arg Ala Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr
 1070 1075 1080

Ser Ser Phe Met Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn
 1085 1090 1095

Ser Ile Thr Gly Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val
 1100 1105 1110

Lys Asn Ser Asp Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala
 1115 1120 1125

Gln Glu Asn Ala Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala
 1130 1135 1140

Tyr Asn Ile Ala Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys
 1145 1150 1155

Lys Ala Glu Asp Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser
 1160 1165 1170

Asn Lys Glu Trp Leu Glu Tyr Ala Gln Thr Ser Val Lys His
 1175 1180 1185

<210> 35

<211> 1255

<212> PRT

<213> Francisella tularensis U112 1

<400> 35

ES 2 768 976 T3

Met	Ser	Ile	Tyr	Gln	Glu	Phe	Val	Asn	Lys	Tyr	Ser	Leu	Ser	Lys	Thr	1	5	10	15
Leu	Arg	Phe	Glu	Leu	Ile	Pro	Gln	Gly	Lys	Thr	Leu	Glu	Asn	Ile	Lys	20	25	30	
Ala	Arg	Gly	Leu	Ile	Leu	Asp	Asp	Glu	Lys	Arg	Ala	Lys	Asp	Tyr	Lys	35	40	45	
Lys	Ala	Lys	Gln	Ile	Ile	Asp	Lys	Tyr	His	Gln	Phe	Phe	Ile	Glu	Glu	50	55	60	
Ile	Leu	Ser	Ser	Val	Asp	Ser	Glu	Lys	Phe	Lys	Asn	Leu	Phe	Asn	Gln	65	70	75	80
Asn	Leu	Ile	Asp	Ala	Lys	Lys	Gly	Gln	Glu	Ser	Asp	Leu	Ile	Leu	Trp	85	90	95	
Leu	Lys	Gln	Ser	Lys	Asp	Asn	Gly	Ile	Glu	Leu	Phe	Lys	Ala	Asn	Ser	100	105	110	
Asp	Ile	Thr	Asp	Ile	Asp	Glu	Ala	Leu	Glu	Ile	Ile	Lys	Ser	Phe	Lys	115	120	125	
Gly	Trp	Thr	Thr	Tyr	Phe	Lys	Gly	Phe	His	Glu	Asn	Arg	Lys	Asn	Val	130	135	140	
Tyr	Ser	Ser	Asn	Asp	Ile	Pro	Thr	Ser	Ile	Ile	Tyr	Arg	Ile	Val	Asp	145	150	155	160
Asp	Asn	Leu	Pro	Lys	Phe	Leu	Glu	Asn	Lys	Ala	Lys	Tyr	Glu	Ser	Leu	165	170	175	
Lys	Asp	Lys	Ala	Pro	Glu	Ala	Ile	Asn	Tyr	Glu	Gln	Ile	Lys	Lys	Asp	180	185	190	
Leu	Ala	Glu	Glu	Leu	Thr	Phe	Asp	Ile	Asp	Tyr	Lys	Thr	Ser	Glu	Val	195	200	205	
Asn	Gln	Arg	Val	Phe	Ser	Leu	Asp	Glu	Val	Phe	Glu	Ile	Ala	Asn	Phe	210	215	220	

ES 2 768 976 T3

Asn	Asn	Tyr	Leu	Asn	Gln	Ser	Gly	Ile	Thr	Lys	Phe	Asn	Thr	Ile	Ile
225					230					235					240
Gly	Gly	Lys	Phe	Val	Asn	Gly	Glu	Asn	Thr	Lys	Arg	Lys	Gly	Ile	Asn
				245					250					255	
Glu	Tyr	Ile	Asn	Leu	Tyr	Ser	Gln	Gln	Ile	Asn	Asp	Lys	Thr	Leu	Lys
			260					265					270		
Lys	Tyr	Lys	Met	Ser	Val	Leu	Phe	Lys	Gln	Ile	Leu	Ser	Asp	Thr	Glu
		275					280					285			
Ser	Lys	Ser	Phe	Val	Ile	Asp	Lys	Leu	Glu	Asp	Asp	Ser	Asp	Val	Val
	290					295					300				
Thr	Thr	Met	Gln	Ser	Phe	Tyr	Glu	Gln	Ile	Ala	Ala	Phe	Lys	Thr	Val
305					310					315					320
Glu	Glu	Lys	Ser	Ile	Lys	Glu	Thr	Leu	Ser	Leu	Leu	Phe	Asp	Asp	Leu
				325					330					335	
Lys	Ala	Gln	Lys	Leu	Asp	Leu	Ser	Lys	Ile	Tyr	Phe	Lys	Asn	Asp	Lys
			340					345					350		
Ser	Leu	Thr	Asp	Leu	Ser	Gln	Gln	Val	Phe	Asp	Asp	Tyr	Ser	Val	Ile
		355					360					365			
Gly	Thr	Ala	Val	Leu	Glu	Tyr	Ile	Thr	Gln	Gln	Ile	Ala	Pro	Lys	Asn
	370					375					380				
Leu	Asp	Asn	Pro	Ser	Lys	Lys	Glu	Gln	Glu	Leu	Ile	Ala	Lys	Lys	Thr
385					390					395					400
Glu	Lys	Ala	Lys	Tyr	Leu	Ser	Leu	Glu	Thr	Ile	Lys	Leu	Ala	Leu	Glu
				405					410					415	
Glu	Phe	Asn	Lys	His	Arg	Asp	Ile	Asp	Lys	Gln	Cys	Arg	Phe	Glu	Glu
			420					425					430		
Ile	Leu	Ala	Asn	Phe	Ala	Ala	Ile	Pro	Met	Ile	Phe	Asp	Glu	Ile	Ala
		435					440					445			
Gln	Asn	Lys	Asp	Asn	Leu	Ala	Gln	Ile	Ser	Ile	Lys	Tyr	Gln	Asn	Gln
	450					455					460				

ES 2 768 976 T3

Gly	Lys	Lys	Asp	Leu	Leu	Gln	Ala	Ser	Ala	Glu	Asp	Asp	Val	Lys	Ala
465					470					475					480
Ile	Lys	Asp	Leu	Leu	Asp	Gln	Thr	Asn	Asn	Leu	Leu	His	Lys	Leu	Lys
				485					490					495	
Ile	Phe	His	Ile	Ser	Gln	Ser	Glu	Asp	Lys	Ala	Asn	Ile	Leu	Asp	Lys
			500					505					510		
Asp	Glu	His	Phe	Tyr	Leu	Val	Phe	Glu	Glu	Cys	Tyr	Phe	Glu	Leu	Ala
		515					520					525			
Asn	Ile	Val	Pro	Leu	Tyr	Asn	Lys	Ile	Arg	Asn	Tyr	Ile	Thr	Gln	Lys
	530					535					540				
Pro	Tyr	Ser	Asp	Glu	Lys	Phe	Lys	Leu	Asn	Phe	Glu	Asn	Ser	Thr	Leu
545					550					555					560
Ala	Asn	Gly	Trp	Asp	Lys	Asn	Lys	Glu	Pro	Asp	Asn	Thr	Ala	Ile	Leu
				565					570					575	
Phe	Ile	Lys	Asp	Asp	Lys	Tyr	Tyr	Leu	Gly	Val	Met	Asn	Lys	Lys	Asn
			580					585					590		
Asn	Lys	Ile	Phe	Asp	Asp	Lys	Ala	Ile	Lys	Glu	Asn	Lys	Gly	Glu	Gly
		595					600					605			
Tyr	Lys	Lys	Ile	Val	Tyr	Lys	Leu	Leu	Pro	Gly	Ala	Asn	Lys	Met	Leu
	610					615					620				
Pro	Lys	Val	Phe	Phe	Ser	Ala	Lys	Ser	Ile	Lys	Phe	Tyr	Asn	Pro	Ser
625					630					635					640
Glu	Asp	Ile	Leu	Arg	Ile	Arg	Asn	His	Ser	Thr	His	Thr	Lys	Asn	Gly
				645					650					655	
Ser	Pro	Gln	Lys	Gly	Tyr	Glu	Lys	Phe	Glu	Phe	Asn	Ile	Glu	Asp	Cys
			660					665					670		
Arg	Lys	Phe	Ile	Asp	Phe	Tyr	Lys	Gln	Ser	Ile	Ser	Lys	His	Pro	Glu
		675					680					685			

ES 2 768 976 T3

Trp	Lys	Asp	Phe	Gly	Phe	Arg	Phe	Ser	Asp	Thr	Gln	Arg	Tyr	Asn	Ser
690						695					700				
Ile	Asp	Glu	Phe	Tyr	Arg	Glu	Val	Glu	Asn	Gln	Gly	Tyr	Lys	Leu	Thr
705					710					715					720
Phe	Glu	Asn	Ile	Ser	Glu	Ser	Tyr	Ile	Asp	Ser	Val	Val	Asn	Gln	Gly
				725					730					735	
Lys	Leu	Tyr	Leu	Phe	Gln	Ile	Tyr	Asn	Lys	Asp	Phe	Ser	Ala	Tyr	Ser
			740					745					750		
Lys	Gly	Arg	Pro	Asn	Leu	His	Thr	Leu	Tyr	Trp	Lys	Ala	Leu	Phe	Asp
		755					760					765			
Glu	Arg	Asn	Leu	Gln	Asp	Val	Val	Tyr	Lys	Leu	Asn	Gly	Glu	Ala	Glu
	770					775					780				
Leu	Phe	Tyr	Arg	Lys	Gln	Ser	Ile	Pro	Lys	Lys	Ile	Thr	His	Pro	Ala
785					790					795					800
Lys	Glu	Ala	Ile	Ala	Asn	Lys	Asn	Lys	Asp	Asn	Pro	Lys	Lys	Glu	Ser
				805					810					815	
Val	Phe	Glu	Tyr	Asp	Leu	Ile	Lys	Asp	Lys	Arg	Phe	Thr	Glu	Asp	Lys
			820					825					830		
Phe	Phe	Phe	His	Cys	Pro	Ile	Thr	Ile	Asn	Phe	Lys	Ser	Ser	Gly	Ala
		835					840					845			
Asn	Lys	Phe	Asn	Asp	Glu	Ile	Asn	Leu	Leu	Leu	Lys	Glu	Lys	Ala	Asn
	850					855					860				
Asp	Val	His	Ile	Leu	Ser	Ile	Asp	Arg	Gly	Glu	Arg	His	Leu	Ala	Tyr
865					870					875					880
Tyr	Thr	Leu	Val	Asp	Gly	Lys	Gly	Asn	Ile	Ile	Lys	Gln	Asp	Thr	Phe
				885					890					895	
Asn	Ile	Ile	Gly	Asn	Asp	Arg	Met	Lys	Thr	Asn	Tyr	His	Asp	Lys	Leu
			900					905					910		
Ala	Ala	Ile	Glu	Lys	Asp	Arg	Asp	Ser	Ala	Arg	Lys	Asp	Trp	Lys	Lys
		915					920					925			

ES 2 768 976 T3

Ile Asn Asn Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val	930	935	940
His Glu Ile Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe	945	950	955 960
Glu Asp Leu Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys	965	970	975
Gln Val Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr	980	985	990
Leu Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg	995	1000	1005
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly	1010	1015	1020
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser	1025	1030	1035
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys	1040	1045	1050
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp	1055	1060	1065
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe	1070	1075	1080
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr	1085	1090	1095
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp	1100	1105	1110
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu	1115	1120	1125
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly	1130	1135	1140

ES 2 768 976 T3

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1145 1150 1155

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1160 1165 1170

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1175 1180 1185

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1190 1195 1200

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1205 1210 1215

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1220 1225 1230

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1235 1240 1245

Phe Val Gln Asn Arg Asn Asn
1250 1255

<210> 36

<211> 1196

<212> PRT

<213> Candidatus Methanoplasma termitum 10

<400> 36

Met Asn Asn Tyr Asp Glu Phe Thr Lys Leu Tyr Pro Ile Gln Lys Thr
1 5 10 15

Ile Arg Phe Glu Leu Lys Pro Gln Gly Arg Thr Met Glu His Leu Glu
20 25 30

Thr Phe Asn Phe Phe Glu Glu Asp Arg Asp Arg Ala Glu Lys Tyr Lys
35 40 45

Ile Leu Lys Glu Ala Ile Asp Glu Tyr His Lys Lys Phe Ile Asp Glu
50 55 60

His Leu Thr Asn Met Lys Asp Asp Arg Phe Lys Asp Leu Phe Ser Lys
65 70 75 80

5

10

ES 2 768 976 T3

Lys	Leu	Phe	Ser	Glu	Leu	Leu	Lys	Glu	Glu	Ile	Tyr	Lys	Lys	Gly	Asn		
				85					90					95			
His	Gln	Glu	Ile	Asp	Ala	Leu	Lys	Ser	Phe	Asp	Lys	Phe	Ser	Gly	Tyr		
			100					105					110				
Phe	Ile	Gly	Leu	His	Glu	Asn	Arg	Lys	Asn	Met	Tyr	Ser	Asp	Gly	Asp		
		115					120					125					
Glu	Ile	Thr	Ala	Ile	Ser	Asn	Arg	Ile	Val	Asn	Glu	Asn	Phe	Pro	Lys		
		130					135				140						
Phe	Leu	Asp	Asn	Leu	Gln	Lys	Tyr	Gln	Glu	Ala	Arg	Lys	Lys	Tyr	Pro		
145					150					155					160		
Glu	Trp	Ile	Ile	Lys	Ala	Glu	Ser	Ala	Leu	Val	Ala	His	Asn	Ile	Lys		
				165					170					175			
Met	Asp	Glu	Val	Phe	Ser	Leu	Glu	Tyr	Phe	Asn	Lys	Val	Leu	Asn	Gln		
			180					185					190				
Glu	Gly	Ile	Gln	Arg	Tyr	Asn	Leu	Ala	Leu	Gly	Gly	Tyr	Val	Thr	Lys		
		195					200					205					
Ser	Gly	Glu	Lys	Met	Met	Gly	Leu	Asn	Asp	Ala	Leu	Asn	Leu	Ala	His		
	210					215					220						
Gln	Ser	Glu	Lys	Ser	Ser	Lys	Gly	Arg	Ile	His	Met	Thr	Pro	Leu	Phe		
225					230					235					240		
Lys	Gln	Ile	Leu	Ser	Glu	Lys	Glu	Ser	Phe	Ser	Tyr	Ile	Pro	Asp	Val		
				245					250					255			
Phe	Thr	Glu	Asp	Ser	Gln	Leu	Leu	Pro	Ser	Ile	Gly	Gly	Phe	Phe	Ala		
			260					265					270				
Gln	Ile	Glu	Asn	Asp	Lys	Asp	Gly	Asn	Ile	Phe	Asp	Arg	Ala	Leu	Glu		
		275					280					285					
Leu	Ile	Ser	Ser	Tyr	Ala	Glu	Tyr	Asp	Thr	Glu	Arg	Ile	Tyr	Ile	Arg		
	290					295					300						
Gln	Ala	Asp	Ile	Asn	Arg	Val	Ser	Asn	Val	Ile	Phe	Gly	Glu	Trp	Gly		

ES 2 768 976 T3

305		310		315		320									
Thr	Leu	Gly	Gly	Leu	Met	Arg	Glu	Tyr	Lys	Ala	Asp	Ser	Ile	Asn	Asp
				325					330					335	
Ile	Asn	Leu	Glu	Arg	Thr	Cys	Lys	Lys	Val	Asp	Lys	Trp	Leu	Asp	Ser
			340					345					350		
Lys	Glu	Phe	Ala	Leu	Ser	Asp	Val	Leu	Glu	Ala	Ile	Lys	Arg	Thr	Gly
		355					360					365			
Asn	Asn	Asp	Ala	Phe	Asn	Glu	Tyr	Ile	Ser	Lys	Met	Arg	Thr	Ala	Arg
	370					375					380				
Glu	Lys	Ile	Asp	Ala	Ala	Arg	Lys	Glu	Met	Lys	Phe	Ile	Ser	Glu	Lys
385					390					395					400
Ile	Ser	Gly	Asp	Glu	Glu	Ser	Ile	His	Ile	Ile	Lys	Thr	Leu	Leu	Asp
				405					410					415	
Ser	Val	Gln	Gln	Phe	Leu	His	Phe	Phe	Asn	Leu	Phe	Lys	Ala	Arg	Gln
			420					425					430		
Asp	Ile	Pro	Leu	Asp	Gly	Ala	Phe	Tyr	Ala	Glu	Phe	Asp	Glu	Val	His
		435					440					445			
Ser	Lys	Leu	Phe	Ala	Ile	Val	Pro	Leu	Tyr	Asn	Lys	Val	Arg	Asn	Tyr
	450					455					460				
Leu	Thr	Lys	Asn	Asn	Leu	Asn	Thr	Lys	Lys	Ile	Lys	Leu	Asn	Phe	Lys
465					470					475					480
Asn	Pro	Thr	Leu	Ala	Asn	Gly	Trp	Asp	Gln	Asn	Lys	Val	Tyr	Asp	Tyr
				485					490					495	
Ala	Ser	Leu	Ile	Phe	Leu	Arg	Asp	Gly	Asn	Tyr	Tyr	Leu	Gly	Ile	Ile
			500					505					510		
Asn	Pro	Lys	Arg	Lys	Lys	Asn	Ile	Lys	Phe	Glu	Gln	Gly	Ser	Gly	Asn
		515					520					525			
Gly	Pro	Phe	Tyr	Arg	Lys	Met	Val	Tyr	Lys	Gln	Ile	Pro	Gly	Pro	Asn
	530					535					540				

ES 2 768 976 T3

Lys Asn Leu Pro Arg Val Phe Leu Thr Ser Thr Lys Gly Lys Lys Glu
545 550 555 560

Tyr Lys Pro Ser Lys Glu Ile Ile Glu Gly Tyr Glu Ala Asp Lys His
565 570 575

Ile Arg Gly Asp Lys Phe Asp Leu Asp Phe Cys His Lys Leu Ile Asp
580 585 590

Phe Phe Lys Glu Ser Ile Glu Lys His Lys Asp Trp Ser Lys Phe Asn
595 600 605

Phe Tyr Phe Ser Pro Thr Glu Ser Tyr Gly Asp Ile Ser Glu Phe Tyr
610 615 620

Leu Asp Val Glu Lys Gln Gly Tyr Arg Met His Phe Glu Asn Ile Ser
625 630 635 640

Ala Glu Thr Ile Asp Glu Tyr Val Glu Lys Gly Asp Leu Phe Leu Phe
645 650 655

Gln Ile Tyr Asn Lys Asp Phe Val Lys Ala Ala Thr Gly Lys Lys Asp
660 665 670

Met His Thr Ile Tyr Trp Asn Ala Ala Phe Ser Pro Glu Asn Leu Gln
675 680 685

Asp Val Val Val Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr Arg Asp
690 695 700

Lys Ser Asp Ile Lys Glu Ile Val His Arg Glu Gly Glu Ile Leu Val
705 710 715 720

Asn Arg Thr Tyr Asn Gly Arg Thr Pro Val Pro Asp Lys Ile His Lys
725 730 735

Lys Leu Thr Asp Tyr His Asn Gly Arg Thr Lys Asp Leu Gly Glu Ala
740 745 750

Lys Glu Tyr Leu Asp Lys Val Arg Tyr Phe Lys Ala His Tyr Asp Ile
755 760 765

Thr Lys Asp Arg Arg Tyr Leu Asn Asp Lys Ile Tyr Phe His Val Pro

ES 2 768 976 T3

770		775		780
Leu Thr Leu Asn Phe Lys Ala Asn Gly Lys Lys Asn Leu Asn Lys Met				
785		790		795
				800
Val Ile Glu Lys Phe Leu Ser Asp Glu Lys Ala His Ile Ile Gly Ile				
		805		810
				815
Asp Arg Gly Glu Arg Asn Leu Leu Tyr Tyr Ser Ile Ile Asp Arg Ser				
		820		825
				830
Gly Lys Ile Ile Asp Gln Gln Ser Leu Asn Val Ile Asp Gly Phe Asp				
		835		840
				845
Tyr Arg Glu Lys Leu Asn Gln Arg Glu Ile Glu Met Lys Asp Ala Arg				
		850		855
				860
Gln Ser Trp Asn Ala Ile Gly Lys Ile Lys Asp Leu Lys Glu Gly Tyr				
		865		870
				875
				880
Leu Ser Lys Ala Val His Glu Ile Thr Lys Met Ala Ile Gln Tyr Asn				
		885		890
				895
Ala Ile Val Val Met Glu Glu Leu Asn Tyr Gly Phe Lys Arg Gly Arg				
		900		905
				910
Phe Lys Val Glu Lys Gln Ile Tyr Gln Lys Phe Glu Asn Met Leu Ile				
		915		920
				925
Asp Lys Met Asn Tyr Leu Val Phe Lys Asp Ala Pro Asp Glu Ser Pro				
		930		935
				940
Gly Gly Val Leu Asn Ala Tyr Gln Leu Thr Asn Pro Leu Glu Ser Phe				
		945		950
				955
				960
Ala Lys Leu Gly Lys Gln Thr Gly Ile Leu Phe Tyr Val Pro Ala Ala				
		965		970
				975
Tyr Thr Ser Lys Ile Asp Pro Thr Thr Gly Phe Val Asn Leu Phe Asn				
		980		985
				990
Thr Ser Ser Lys Thr Asn Ala Gln Glu Arg Lys Glu Phe Leu Gln Lys				
		995		1000
				1005

ES 2 768 976 T3

Phe	Glu	Ser	Ile	Ser	Tyr	Ser	Ala	Lys	Asp	Gly	Gly	Ile	Phe	Ala
1010						1015					1020			
Phe	Ala	Phe	Asp	Tyr	Arg	Lys	Phe	Gly	Thr	Ser	Lys	Thr	Asp	His
1025						1030					1035			
Lys	Asn	Val	Trp	Thr	Ala	Tyr	Thr	Asn	Gly	Glu	Arg	Met	Arg	Tyr
1040						1045					1050			
Ile	Lys	Glu	Lys	Lys	Arg	Asn	Glu	Leu	Phe	Asp	Pro	Ser	Lys	Glu
1055						1060					1065			
Ile	Lys	Glu	Ala	Leu	Thr	Ser	Ser	Gly	Ile	Lys	Tyr	Asp	Gly	Gly
1070						1075					1080			
Gln	Asn	Ile	Leu	Pro	Asp	Ile	Leu	Arg	Ser	Asn	Asn	Asn	Gly	Leu
1085						1090					1095			
Ile	Tyr	Thr	Met	Tyr	Ser	Ser	Phe	Ile	Ala	Ala	Ile	Gln	Met	Arg
1100						1105					1110			
Val	Tyr	Asp	Gly	Lys	Glu	Asp	Tyr	Ile	Ile	Ser	Pro	Ile	Lys	Asn
1115						1120					1125			
Ser	Lys	Gly	Glu	Phe	Phe	Arg	Thr	Asp	Pro	Lys	Arg	Arg	Glu	Leu
1130						1135					1140			
Pro	Ile	Asp	Ala	Asp	Ala	Asn	Gly	Ala	Tyr	Asn	Ile	Ala	Leu	Arg
1145						1150					1155			
Gly	Glu	Leu	Thr	Met	Arg	Ala	Ile	Ala	Glu	Lys	Phe	Asp	Pro	Asp
1160						1165					1170			
Ser	Glu	Lys	Met	Ala	Lys	Leu	Glu	Leu	Lys	His	Lys	Asp	Trp	Phe
1175						1180					1185			
Glu	Phe	Met	Gln	Thr	Arg	Gly	Asp							
1190						1195								

REIVINDICACIONES

1. Un vector de expresión que comprende una secuencia de nucleótidos que codifica un polipéptido Cpf1 que comprende la secuencia de aminoácidos YLFQIYNKDF (residuos de aminoácidos 784-793 de la SEQ ID NO: 1).
2. El vector de expresión de la reivindicación 1, en el que dicho polipéptido Cpf1 tiene al menos el 60% de identidad con la SEQ ID NO: 1 y tiene actividad de nucleasa.
3. El vector de expresión de la reivindicación 1, en el que el polipéptido Cpf1 comprende la secuencia de aminoácidos GKLYLFQIYNKDFS (residuos de aminoácidos 781-794 de la SEQ ID NO: 1).
4. El vector de expresión de la reivindicación 1, en el que el polipéptido Cpf1 es una nickasa.
5. El vector de expresión de cualquiera de las reivindicaciones 1 a 4, en el que el polipéptido Cpf1 está fusionado en su extremo N-terminal o C-terminal a una secuencia de aminoácidos adicional.
6. El vector de expresión de la reivindicación 5, en el que la secuencia de aminoácidos adicional es una proteína seleccionada de entre una helicasa, una nucleasa, una nucleasa-helicasa, una ADN metiltransferasa, una ADN desmetilasa, una histona metiltransferasa, una histona desmetilasa, una acetilasa, una desacetilasa, una fosfatasa, una quinasa, un (co)activador de la transcripción, una subunidad de ARN polimerasa, un represor de la transcripción, una proteína de unión a ADN, una proteína de estructuración de ADN, una proteína marcadora, una proteína reportera, una proteína fluorescente, una proteína de unión a ligando, un péptido señal, una secuencia de localización subcelular o un epítipo de anticuerpo.
7. El vector de expresión de la reivindicación 6, en el que la secuencia de aminoácidos adicional es una secuencia de localización nuclear.
8. El vector de expresión de la reivindicación 6, en el que la secuencia de aminoácidos adicional es un dominio FokI.
9. El vector de expresión de cualquiera de las reivindicaciones anteriores, que además comprende una secuencia de nucleótidos que codifica un ARN guía que tiene una complementariedad sustancial con una secuencia deseada presente en una cadena de ácido nucleico diana.
10. Un sistema vector que comprende (a) un vector de expresión de cualquiera de las reivindicaciones 1 a 8 y (b) un vector que comprende una secuencia de nucleótidos que codifica un ARN guía que tiene una complementariedad sustancial con una secuencia deseada presente en una cadena de ácido nucleico diana.
11. El vector o el sistema de expresión de cualquiera de las reivindicaciones anteriores, en el que dichos vectores son vectores víricos.
12. El vector o el sistema de expresión de la reivindicación 11, en el que dichos vectores víricos son vectores víricos adenoasociados.
13. Una célula huésped que comprende una secuencia de nucleótidos que codifica un polipéptido Cpf1 que comprende la secuencia de aminoácidos YLFQIYNKDF (residuos de aminoácidos 784-793 de la SEQ ID NO: 1), no siendo la célula huésped parte de un cuerpo humano en ninguna etapa de formación o de desarrollo.
14. El vector de expresión de la reivindicación 1, en el que dicho polipéptido Cpf1 tiene al menos el 85% de identidad con la SEQ ID NO: 1.
15. El vector de expresión de la reivindicación 1, en el que dicho polipéptido Cpf1 tiene al menos el 95% de identidad con la SEQ ID NO: 1.

Figura 1

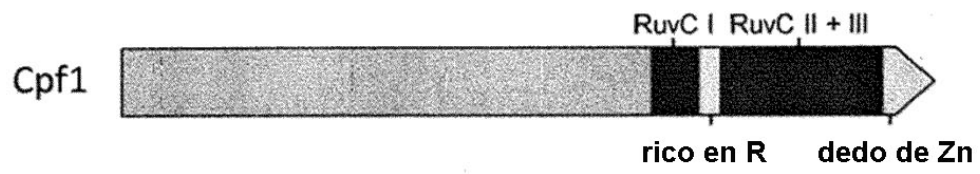
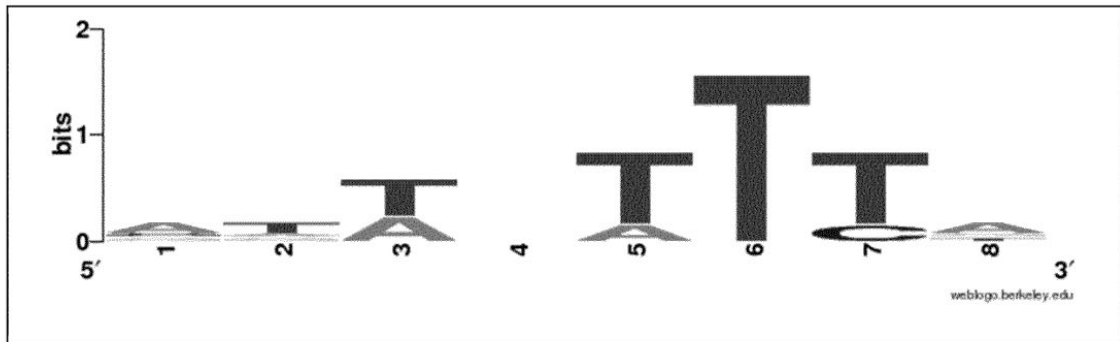


Figura 2

A.



B.

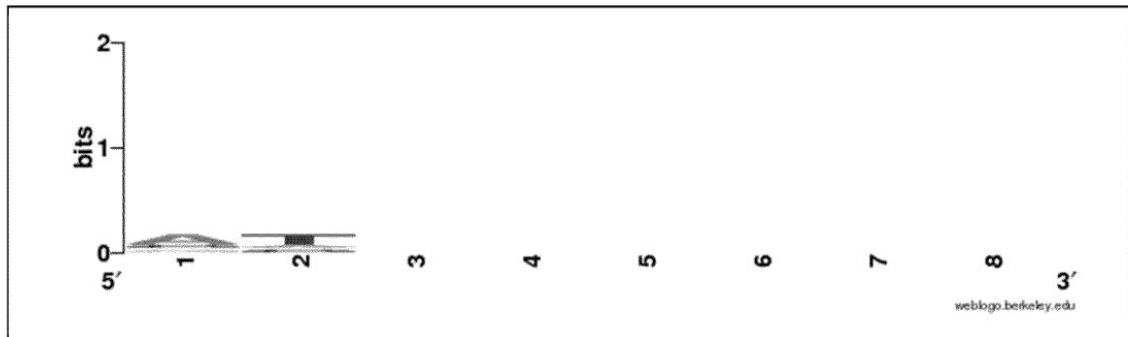


Figura 3A

505317677	Methanomethylophilus alvus	505317677	Methanomethylophilus alvus
545612232	Acidaminococcus sp BV3L6 8	545612232	Acidaminococcus sp BV3L6 8
851218172	Candidatus Methanoplasma termitum 10	851218172	Candidatus Methanoplasma termitum 10
737666241	Lachnospiraceae	737666241	Lachnospiraceae
489130501	Francisella tularensis U112 1	489130501	Francisella tularensis U112 1
505317677	Methanomethylophilus alvus	505317677	Methanomethylophilus alvus
545612232	Acidaminococcus sp BV3L6 8	545612232	Acidaminococcus sp BV3L6 8
851218172	Candidatus Methanoplasma termitum 10	851218172	Candidatus Methanoplasma termitum 10
737666241	Lachnospiraceae bacterium ND2006 14	737666241	Lachnospiraceae bacterium ND2006 14
489130501	Francisella tularensis U112 1	489130501	Francisella tularensis U112 1
505317677	Methanomethylophilus alvus	505317677	Methanomethylophilus alvus
545612232	Acidaminococcus sp BV3L6 8	545612232	Acidaminococcus sp BV3L6 8
851218172	Candidatus Methanoplasma termitum 10	851218172	Candidatus Methanoplasma termitum 10
737666241	Lachnospiraceae bacterium ND2006 14	737666241	Lachnospiraceae bacterium ND2006 14
489130501	Francisella tularensis U112 1	489130501	Francisella tularensis U112 1
505317677	Methanomethylophilus alvus	505317677	Methanomethylophilus alvus
545612232	Acidaminococcus sp BV3L6 8	545612232	Acidaminococcus sp BV3L6 8
851218172	Candidatus Methanoplasma termitum 10	851218172	Candidatus Methanoplasma termitum 10
737666241	Lachnospiraceae bacterium ND2006 14	737666241	Lachnospiraceae bacterium ND2006 14
489130501	Francisella tularensis U112 1	489130501	Francisella tularensis U112 1
505317677	Methanomethylophilus alvus	505317677	Methanomethylophilus alvus
545612232	Acidaminococcus sp BV3L6 8	545612232	Acidaminococcus sp BV3L6 8
851218172	Candidatus Methanoplasma termitum 10	851218172	Candidatus Methanoplasma termitum 10
737666241	Lachnospiraceae bacterium ND2006 14	737666241	Lachnospiraceae bacterium ND2006 14
489130501	Francisella tularensis U112 1	489130501	Francisella tularensis U112 1

.....MDAKFTQYPLSKTIRFELPQ.....GRWNLEA.....S.GYLAEDHRAE...CY.PRAKELLDDNHRAFINRVLPQI
.....MTQEGFTNIYQVSKTIRFELPQ.....GKTLKHQE.....Q.GFTEEDKARD..HY.KELKPIIDRIYKTYADQCLQIV
.....MNVDEFTKLYPQKTRFELPQ.....GRTMHLET.....F.NFTEEDKRAE..KY.KIKKALIDEYHKKFTDEHITNM
.....MSKLEKFTNCYSLSKTIRFELPQ.....GKTCNIN.....K.RLIVEDEKRAE..DY.KGVKKILDRYVLSFINDVLHSI
.....MSIQEFTVNVKYSLSKTIRFELPQ.....GKTLNKA.....R.GLILDEKRAK..DY.KKAKQIIDRYHQFFTEELISSV
.....QADGVKGLFAKPALE.....AMTKAKENGESDIEV.....LEAF.....NGF.SVYFTCY
.....GRDNLIDAINKRAHEIYKGLFAKLENGK.....VLKQIGTVTTEHNAI.....IRSF....DKF.TTYFSGF
.....K.....DDR...FKDLFSKLFSEL.....IKEEIYKKGNHQIDA.....IKSF....DKF..SGYFGL
.....G.....NEG...YKSLFKKQIDB.T.....ILPEFLDDKQIDAL.....VNSF....NGF..TTAFTGF
.....D.....SEK...FKNLFNQNLIDAKK.....GOESDILILWLKQSKONGIELELFKANSDITDIDEALBI.....IKSF....KGW.TTYFKGF
HESRENIY.SDE.....DM.VSAVYRTIEDNFRFVSNALIFDKINESHPD.....I.....ISEVSGNLGVDD.....
YENRKNVF.SAE.....DISTAIPHRIVQDNFPKFKENCHIETRLITAVPS.....LREHPENVKKAIGIEVSTS.....
HENRKNWY.SDG.....DEITAINRIVNENFPKFLDNLOKYOEARKKYPE.....WTIKAESALVAHNK.....
FDNRENWE.SEE.....AKSTISAFRCINENLTRYISNMIDIEKV..DAIF.....DKHEVQEIKKIILNSDYD.....
HENRKNVY.SSN.....DIPSTIIIRIVDDNLKPLENKAKAYESL.KOKAP.....EAINVEQIKKDLAEELTFDIDYKTSFVNQRFVS
IGKVPDVSNNFLSQAGIDD.....YNHIIGG...HTTEDGLIOAFNVIM.LRHQKDP.....GFE....KI.....QFKQLYKQILSVRTSK
IEEVFSFPFNQLLTQIDIL.....YNOLLGGIS.REAGTEKIKGNEVLN.LAIQKND.....ET...A.HIASLPHRFTPLFKQILSDRNTL
MDEVFSLEYENKVLNOEGIOR.....YNALGGYV.TKSQE.KWMLNDALN.LAHQSEK.....SSK..G.RI.....HMTPLFKQILSEKESF
VEDEFEGEFNFVLTQEGIDV.....YNALIGGEV.TESGE.KIKGNEYN.LYNQKTEK.....Q.KL.....PKKPLFKQVLSDSRESL
LDEVFEIANFNNNVLOSQITK.....ENTIIIGGF.VNGENTRRKGINEYN.LYSQOIN.....DKT..L.KA.....YKMSVLFKQILSDTESK
SYIPKQFDSKE.....MVCICDY..VSKIEKSETVER.....ALKVRNI.....S..SEDLRGIFVNKK.NUR
SFILIEEKSDEE.....VIQSFCKY..KTLLENVLET.....AEALFNEI.....N..SIDLTHIFTSK.KLE
SVIPDVFTDSQ.....LLPSIGCF..FAQIEN..DKDG.....NIFDRALELISSY.....A..EYDTERIYIROA.DIN
SFYEGEGTSDEE.....VLEVRNT..LNKNSEIFESSIK.....KLEKLEKPF.....D.EYSSAGIFVKNGPAIS
SFVIDKLEDDSD.....VVTIMOSF..YEQIAAFKIVEE....KSIKETUSLULFDDL.....KAKQUDLSKIYFKNDKSLT
TILSNKL.....IGDWDIAIETALMHSSES.....NDKKSVDSEAFITLDDIFSSVKKF.....
TISSAL.....CDHWTIRNALYERRISE.....ITGKITK.SAKEKVQKSLK.HEDINLOEIIISAAG.....
RVSNVI.....FGEMTIGGLMREYKADS.....INDINLE.RTCKKVDKWL.D.SKEFALSVDLEAKRT.....GN
TISKDI.....FGEMNVRHDKWNAEYDDI.....HLKKKAVTE.KVEDDRKSKFKIGTSLEQLQOYADAD.....
DLSQOV.....FDDYSVIGTAVLEIYITQO.....IAPKNLONPSK.KEQELIAKKTEKAKY.SLETIKIALEEF...NKHRDIDK

Figura 3B

505317677 Methanomethylophilus alvus	SDA....SAEDI.....GNRAEDICRVISETAPEINDRA.....VLDLSL.....NDDCYEAADVSKIRSELEPYWDLF
545612232 Acidaminococcus sp BV3L6 8	HLLDMFAVDSENE.....VDEEF....SARLTGKTLEM.....E.PSISFVKARNAVTKRPS....VEKFL.NFOMPTLAS.GW.....DVNKE
851218172 Candidatus Methanoplasma termitum 10	NDA....FNEVI.....SKM.....RIARCKIDAAARKMKFI.....SKKI.....SCDESHIKITLLDSVOOFL
737666241 Lachnospiraceae bacterium ND2006 14	LSV....VEKIK.....EII.....IOKVEIVKYVGSSEKLPADF.....VLESKL.....KKNDAAVATMKDILDSVKSFE
89130501 Francisella tularensis U112 1	OCR....FEEIL...ANF.....AAI.....PMIFDEIAQNKDNIAGISIKYONQ.....CKKDL.....QASAEEDVKAIKDLLOQTNLL
505317677 Methanomethylophilus alvus	HELEIFSVQDE.....FPKCAAF....YSELEEVSQL.....I.EIIPLENKARSCFCKRYS....TDKIKV.NIKFPTLIAD..GW.....DENKE
545612232 Acidaminococcus sp BV3L6 8	HFNENLF.KARQD.....TLPDCAF....YAEDEVHSKL.....F.AIVPLYNKVRNYLTKNIN..TKAIKL.NFKNPTIAN..GW.....DONKV
851218172 Candidatus Methanoplasma termitum 10	NYTKAF.FEGEKE....TNRDESE....YGDVIAIDIL.....L.KVDHIYDAIRNYVTKRPS....KDAFKL.YFONQPMG..GW.....DKDKE
737666241 Lachnospiraceae bacterium ND2006 14	HKIKFHSQSED.KANILDKDEHF....YIVFECCYFEL....A.NIVPLYNKIRNYITQKPS....DEKFKL.NFNENSTIAN..GW.....DNKKE
89130501 Francisella tularensis U112 1	RDKKAILRK.DGK....YITAILDM.KK....DIS.....SIRTSDE..DE.....SSF.EKWEYKLIPSPV...KMLPKI.....
505317677 Methanomethylophilus alvus	KNGCALIVK.NGI.....YIIGIMP.K.OK....CRYKALISEPTEKTS.....SGF.DKMYDYFFDAA...KMLPCSTQLKAVTAHFQT
545612232 Acidaminococcus sp BV3L6 8	TDYRATILRY.GSK.....YIITAIMDK.KY....AKGIQIKDDVN.....GNY.EKINYKLLPGN...KMLPRV.....
851218172 Candidatus Methanoplasma termitum 10	PDNTAILFIK.DDK....YILGVNWK.KN..NKIED.DKAIKENKG.....EGY.KTIVYKLLPGAN...KMLPRV.....
737666241 Lachnospiraceae bacterium ND2006 14	FVK.....SKAAKCK.....YGL.....T.....DMLECYDK.....GMH.....KSGS....
89130501 Francisella tularensis U112 1	HTTPILL.....SNFIEP.....LEI.....T.....KEIYDIANNPEKP.....KKSQTAYAKA...KTGDDQK....
505317677 Methanomethylophilus alvus	FLT.....STAGKE.....YKP.....S.....KEITEGVEA.....DKH.....IRGD....
545612232 Acidaminococcus sp BV3L6 8	FF.....SKKWAY.....YNP.....S.....EDIOKIYKN.....GTF.....KKGD....
851218172 Candidatus Methanoplasma termitum 10	FF.....SAKSIF.....YNP.....S.....EDILIRNH.....STHT.....KNGSPQKG
737666241 Lachnospiraceae bacterium ND2006 14	AFD...IGFC.....HELID.....KWD.....
89130501 Francisella tularensis U112 1	GVR...EALC.....KWD.....HKLID.....
505317677 Methanomethylophilus alvus	MEN...LINC.....HKLID.....
545612232 Acidaminococcus sp BV3L6 8	YKX.....EFN...IEDC.....RKFD.....
851218172 Candidatus Methanoplasma termitum 10	YKRC.CIAEY.....PGWD...VF.DFK.F.....RET...SDYG.SMKEENEDVAGA..GYMSL.RKIPCEV
737666241 Lachnospiraceae bacterium ND2006 14	FTRD.FLSKY.....TKYT...SL.DLSL.....RPS..SQYK.DLGEYVAENPL..LYHISE.QRIAEKEI
89130501 Francisella tularensis U112 1	FFKE.SIEKH.....KWS...KF.NFY.F.....SPT...ESYG.DISEFYLDVKEQ..GYRWHF.ENISAEI
505317677 Methanomethylophilus alvus	FTRD.SISRY.....PWSN..AY.DFN.F.....SET..EKYK.DIAGFYREVEEQ..GYKVSF.ESASKEV
545612232 Acidaminococcus sp BV3L6 8	FYKQ.SISKH.....PEWK...DP.GFR.F.....SDT...QRVN.SIDEFYREVENQ..GYKLTF.ENISSEYI
851218172 Candidatus Methanoplasma termitum 10	YRLLD.EXS.IYFOIYNKDYFS.....ENAHGNKNMHTYWEGL...FSPQN.IESP.....VFKUSGGAELFFRKSSIPNDAKTWHPKGVLVP
737666241 Lachnospiraceae bacterium ND2006 14	MDAVE.TBK.IYFOIYNKDEFA.....KGHCKPMLHTLYTGL...FSPEN.IAKT.....STKLNGQAEIFYRQKSRMKR..MAHRLGEMLN
89130501 Francisella tularensis U112 1	DEYVE.KGD.IYFOIYNKDFV.....KAATCKKDMHTIYVNA...FSPEN.IQDV.....VVKLNGEALFYRQKSDIKE..IVHREGEILVN
505317677 Methanomethylophilus alvus	DKLIVE.EGK.IYFOIYNKDFS.....DKSHGTPLNHTMYFKLL...FDEAN.HGOI.....RLSCGAELFMRASLKEBELVWHFANSPIAN
545612232 Acidaminococcus sp BV3L6 8	DSVWN.QCK.IYFOIYNKDES.....AYSGRPNLHTLYWKAL...FDEAN.IQDV.....VVKLNGEALFYRQKSDIKE..ITHPAKEAIAN
851218172 Candidatus Methanoplasma termitum 10	RN.DVNGRRIP....DSIVRELTRYFNRGDCRISDEAK.....SYLDKVTKKADH...DIVKDRRTVDKMMHFVPIAMN.....
737666241 Lachnospiraceae bacterium ND2006 14	KKIKDQKTPIP...DTLQDELYDVNHR.....LSHDSIDEARALLPNVITKEVSHIELIKDRETSOKFFHFVPIITN.....
89130501 Francisella tularensis U112 1	RT.YNGRTPVP....DKIHKK.TDYHNGRTKDLGEAKELYDKVY.F.....KAHY.....DITKDRYLNKDIYFHFVPIITN.....
505317677 Methanomethylophilus alvus	KN.PD..NPXK....TIT.....L.....SY.....DVKDKRSEDOYELHIPITAN.....
545612232 Acidaminococcus sp BV3L6 8	KN.KD..NPXK....ESV.....F.....EV.....DLIKDKRETDKFFHFHCPTIN.....
851218172 Candidatus Methanoplasma termitum 10	FKAISKP..NUNKKVIDGIID.DQB....LXICIDRGERNLIYVTWVD..RKNILIYQD..SLNILL.....NG.....Y.....
737666241 Lachnospiraceae bacterium ND2006 14	YQAANSP..SKFNQVNAVILKE.HPE....TPHICIDRGERNLIYITVID.STGKILEQR..SIANTI...QQ.....F.....
89130501 Francisella tularensis U112 1	FRANKCK..NUNKWTEKFLS.DEK...AHICIDRGERNLIYVSIID.RSGKIIDQO..SIANVI...DG.....F.....
505317677 Methanomethylophilus alvus	KCPKNIF..KINTEVAVILKH.DDN...PVVICIDRGERNLIYITVVD.GKGNIVEOY..SIMEI...IN.NENGI.RIKT.....
545612232 Acidaminococcus sp BV3L6 8	FKSSCAN..KENDEINLLIKERAND...VHIIS.DRGERHILAYITLVD.GKGNILIKQD..TFNII...GN.....D.RUKT.....

RawC-1

Figura 3C

505317677	Methanomethylophilus alvus	
545612232	Acidaminococcus sp BV3L6 8	
851218172	Candidatus Methanoplasma termitum 10	
737666241	Lachnospiraceae bacterium ND2006 14	
489130501	Francisella tularensis U112 1	
505317677	Methanomethylophilus alvus	
545612232	Acidaminococcus sp BV3L6 8	
851218172	Candidatus Methanoplasma termitum 10	
737666241	Lachnospiraceae bacterium ND2006 14	
489130501	Francisella tularensis U112 1	
505317677	Methanomethylophilus alvus	
545612232	Acidaminococcus sp BV3L6 8	
851218172	Candidatus Methanoplasma termitum 10	
737666241	Lachnospiraceae bacterium ND2006 14	
489130501	Francisella tularensis U112 1	
505317677	Methanomethylophilus alvus	
545612232	Acidaminococcus sp BV3L6 8	
851218172	Candidatus Methanoplasma termitum 10	
737666241	Lachnospiraceae bacterium ND2006 14	
489130501	Francisella tularensis U112 1	
505317677	Methanomethylophilus alvus	
545612232	Acidaminococcus sp BV3L6 8	
851218172	Candidatus Methanoplasma termitum 10	
737666241	Lachnospiraceae bacterium ND2006 14	
489130501	Francisella tularensis U112 1	
505317677	Methanomethylophilus alvus	
545612232	Acidaminococcus sp BV3L6 8	
851218172	Candidatus Methanoplasma termitum 10	
737666241	Lachnospiraceae bacterium ND2006 14	
489130501	Francisella tularensis U112 1	

Hélice rica en R

RuvC-II

RuvC-III