

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 897 017**

21 Número de solicitud: 202030890

51 Int. Cl.:

**C07K 14/245** (2006.01)

12

PATENTE DE INVENCION CON EXAMEN

B2

22 Fecha de presentación:

**28.08.2020**

43 Fecha de publicación de la solicitud:

**28.02.2022**

Fecha de modificación de las reivindicaciones:

**27.05.2022**

Fecha de concesión:

**23.06.2022**

45 Fecha de publicación de la concesión:

**30.06.2022**

73 Titular/es:

**UNIVERSIDAD DE CANTABRIA (72.0%)  
Pabellón de Gobierno. Avda. los Castros s/n  
39005 Santander (Cantabria) ES;  
CONSEJO SUPERIOR DE INVESTIGACIONES  
CIENTÍFICAS (8.0%) y  
INSTITUTO PASTEUR (20.0%)**

72 Inventor/es:

**LLOSA BLAS, Matxalen;  
GUZMÁN HERRADOR, Dolores Lucía y  
BIKARD, David**

74 Agente/Representante:

**PONS ARIÑO, Ángel**

Observaciones:

La lista de secuencias está accesible al público en la página web de la OEPM para su descarga en formato electrónico.

54 Título: **PROTEÍNA QUIMÉRICA RELAXASA-Cas**

57 Resumen:

La presente invención se refiere a la proteína de fusión Relaxasa-Cas12a, al sistema CRISPR/Cas que comprende dicha proteína de fusión y al uso de la proteína de fusión Relaxasa-Cas12a y/o del sistema CRISPR/Cas para la translocación de endonucleasas y/o endonucleasas unidas a moléculas de ADN, a células diana a través del sistema de secreción bacteriano tipo IV, y para la modificación genética de las células diana.

ES 2 897 017 B2

Aviso: Se puede realizar consulta prevista por el art. 41 LP 24/2015. Dentro de los seis meses siguientes a la publicación de la concesión en el Boletín Oficial de la Propiedad Industrial cualquier persona podrá oponerse a la concesión. La oposición deberá dirigirse a la OEPM en escrito motivado y previo pago de la tasa correspondiente (art. 43 LP 24/2015).

**DESCRIPCIÓN****PROTEÍNA QUIMÉRICA RELAXASA–Cas****Campo de la invención**

La presente invención se encuadra en el campo general de la biotecnología y en particular se refiere a una proteína de fusión relaxasa-proteína Cas y al uso de la misma para la edición  
5 génica de diferentes células.

**Estado de la técnica**

Los sistemas CRISPR-Cas (del inglés, *Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated proteins*) constituyen un sistema inmune adaptativo en la mayoría de procariontes que confiere resistencia a elementos genéticos externos, y han dado  
10 lugar al desarrollo de una revolucionaria herramienta de edición genómica <sup>1</sup>. De manera general, estos sistemas se basan en una serie de secuencias de ADN CRISPR y proteínas Cas. Los clústers de secuencias CRISPR están formados por secuencias espaciadoras, separadas por secuencias repetidas. Estos espaciadores codifican unos ARNs guía (gARNs), que tienen como diana una secuencia de ADN exógeno. Las proteínas Cas procesan esos  
15 ARNs, que llevan hasta la secuencia diana a la endonucleasa Cas, que realiza un corte de doble cadena en dicha secuencia de ADN exógeno. Su principal función biológica es por tanto defender a la bacteria hospedadora de invasores externos (elementos genéticos móviles, como fagos o plásmidos).

El uso de los sistemas CRISPR-Cas como herramienta biotecnológica se basa en la  
20 introducción de un corte de doble cadena en cualquier secuencia diana de interés. Este corte, considerado por la célula como una señal de peligro, sirve de punto de reclutamiento de los sistemas de reparación celular, que pueden reparar esa herida introduciendo pequeñas mutaciones, o incorporando moléculas de ADN exógeno que introduzcan la información deseada en el genoma receptor. Existían sistemas de nucleasas sitio-específicas, como las  
25 Zinc o las TALEN, que podían ser diseñadas para actuar sobre secuencias específicas, pero ello conllevaba un laborioso e incierto proceso de diseño y/o evolución molecular. Lo que ha convertido a los sistemas CRISPR-Cas en una revolucionaria técnica de edición genética es su sencillez, puesto que basta con proporcionar el gARN adecuado a la endonucleasa Cas, de modo que le guíe a la secuencia diana, donde cortará la doble hebra del ADN <sup>1</sup>. De este  
30 modo, solo se necesita diseñar un gARN específico para un gen en concreto, seleccionando una secuencia que cuente con la PAM (motivo adyacente de protoespaciador) específica para la unión de la endonucleasa Cas.

Actualmente todos los sistemas CRISPR-Cas están englobados en dos clases (Clase I y Clase II)<sup>2</sup>, siendo los segundos los más adecuados para su uso como sistemas de modificación genética, debido a que codifican en una sola proteína (la endonucleasa Cas) todas las funciones necesarias (reconocimiento y unión al gARN, reconocimiento y unión a la secuencia diana del ADN, y realización de corte de doble cadena). A pesar de que Cas9<sup>3</sup> es la endonucleasa más utilizada, cada vez más se estudian otros sistemas CRISPR por las ventajas y diferencias que pueden tener sobre Cas9, siendo Cas12a<sup>4</sup> (también conocida como Cpf1) una de las más estudiadas.

Aunque en un inicio las aplicaciones de los sistemas CRISPR-Cas se centraron en la edición genética de organismos eucariotas (para generación de modelos de laboratorio, o para modificación genética con fines terapéuticos), han ido creciendo las posibilidades que ofrecen. Actualmente, las aplicaciones de los sistemas CRISPR Cas son prácticamente ilimitadas debido a que podemos utilizar como diana ADN codificante (permitiendo, por ejemplo, una modificación génica), ARN (control del "splicing") o promotores (permitiendo una regulación génica). Esta versatilidad abre un enorme campo en la aplicación de estos sistemas<sup>5</sup>. La posibilidad de guiar la endonucleasa a un sitio específico del genoma puede servir para su localización, mutación, modulación epigenética, herramientas de diagnóstico, estudios de interacción de la cromatina, etc.

Su uso en procariontes cada vez tiene un mayor impacto en diferentes campos, como la ingeniería metabólica, la biología sintética, o la medicina<sup>6</sup>. El uso de bacterias para un fin biotecnológico como "fábricas microbianas", cada vez más extendido, requiere de su manipulación genética para mejorar y potenciar el rendimiento. El desarrollo de la tecnología CRISPR-Cas ha supuesto un gran avance en este campo ya que ha facilitado enormemente la manipulación tanto de bacterias modelo como no modelo. Algunos ejemplos son la modificación de *Escherichia coli* (organismo modelo), especies del género *Clostridium* (producción de acetona, butanol o etanol), *Streptomyces* (producción de productos naturales con actividad antimicrobiana) o Bacterias del Ácido Láctico (uso como probióticos). También es muy importante la aplicación de estos sistemas en el campo de la investigación biomédica, para modificar genéticamente o estudiar posibles genes esenciales en patógenos como *Mycobacterium tuberculosis*, *Yersinia pestis* o *Klebsiella pneumoniae*. Otra aplicación de gran importancia es su uso como antimicrobianos de nueva generación, gracias a la elevada especificidad en su secuencia diana, permitiendo la eliminación de una población bacteriana específica o la eliminación de un plásmido en concreto dentro de una población<sup>7</sup>. Esta

capacidad de eliminar plásmidos de manera específica le otorga un papel sumamente importante en su uso para combatir resistencia a antibióticos<sup>8</sup>.

A pesar de su alto grado de eficacia y el enorme impacto que ha supuesto, la tecnología CRISPR-Cas no carece de limitaciones que hay que solventar para un correcto uso. La constante mejora se centra en dos aspectos fundamentales: aumentar la especificidad de la endonucleasa por su diana (disminución del efecto off-target), y mejorar el sistema de envío a la célula receptora *in vivo*, teniendo en cuenta que tanto el gARN como la endonucleasa tienen que ser enviados a la célula que se quiere modificar.

Por una parte, existe el problema de la actividad de corte de la endonucleasa en secuencias distinta de su diana (efecto off-target), lo que puede acarrear la introducción de mutaciones no deseadas. Aunque este problema puede verse reducido dependiendo del método de envío de la proteína, se han buscado otro tipo de soluciones que lo reduzcan directamente. Para ello, se han centrado todos los esfuerzos en dos direcciones principales. Una consiste en desarrollar métodos de detección de eventos off-target, mediante el desarrollo de programas informáticos como Cas-OFFinder, CCTop o DISCOVER. La segunda consistiría en generar sistemas CRISPR-Cas modificados de alta fidelidad<sup>5</sup>. Para ello, se han usado dos abordajes principales: modificar la endonucleasa efectora, o modificar el gARN. Actualmente existen diferentes tipos de Cas9 que han sido diseñadas introduciendo ciertas modificaciones estructurales que le otorgan una mayor especificidad, como es el caso de SpCas9, HiFiCas9 y HypaCas9. Además, se han encontrado en librerías de mutantes de Cas9, endonucleasas con una mayor especificidad que la nativa, como SnieperCas9. Por otro lado, se han desarrollado modificaciones en el gARN como la adición de nucleótidos o el truncamiento de los nucleótidos. Otras estrategias son la adición de enlaces fosfodiéster o la introducción de desoxinucleótidos para generar guías híbridos ADN-ARN<sup>9</sup>.

Por otro lado, está el problema del envío del sistema. Existen tres posibles formatos de enviar la endonucleasa a la célula: enviando la propia proteína, enviando el ARN mensajero (ARNm), o enviando el gen que la codifica. El envío de ADN se caracteriza por ser un método barato, simple y estable. Una limitación importante es el tamaño; la mayoría de vectores, o bien no admiten grandes cantidades de ADN (como los vectores virales), o baja mucho su eficiencia de entrada a la célula al aumentar de tamaño (caso de los plásmidos), y sólo el gen *cas9* ya mide 4.5 Kb aproximadamente. En eucariotas, existe el requerimiento adicional de que el ADN exógeno entre en el núcleo para su transcripción, lo que disminuye la eficiencia de edición genómica. Además, la introducción de ADN exógeno a la célula genera otros problemas como

la probabilidad de integración del ADN introducido en el genoma diana, o la producción de proteína durante más tiempo del necesario, lo que puede dar lugar efectos tóxicos, incremento de la actividad off-target o importantes respuestas inmunes <sup>10</sup>.

5 El ARNm es fácil de obtener por transcripción *in vitro*. Es un método más directo que el ADN, ya que no hace falta que ocurra la transcripción, y, además, la expresión es más transitoria, lo que disminuye los efectos de toxicidad y off-target. Sin embargo, el ARNm es muy inestable, y junto con su expresión transitoria, resulta en una disminución de la efectividad.

10 Por último, podemos introducir la proteína Cas directamente. En principio es el método más eficiente ya que no hay que transcribir ni traducir el ADN. Además, la expresión es también transitoria lo que disminuye toxicidad y efectos off target. Sin embargo, la purificación de proteínas es un proceso costoso y complejo. Además, el gran tamaño de las proteínas Cas, que además tienen que ser introducidas junto con su ARN guía, dificulta el transporte efectivo de complejos ribonucleoprotéicos (RNPs).

15 Tan relevante como el formato de envío es el método utilizado para introducirlo en la célula diana (el vector). En eucariotas, los más extendidos son los vectores virales. Consisten en utilizar un virus para introducir el ADN que codifica para la endonucleasa y el gARN mediante transducción. Los más usados son los Virus Adeno-Asociados (AAV) y lentivirus (LV) <sup>11</sup>. A pesar de que es uno de los sistemas más utilizado por su facilidad y su posibilidad de utilizarse en ensayos *in vivo*, tienen las limitaciones ya explicadas por el uso de ADN. En procariontes se ha explorado también el uso de la transformación y la conjugación bacteriana como medio de introducción del ADN en bacterias. Las ventajas de este último son que no hay limitación de tamaño y permite hacer un envío *in vivo* a un amplio número de huéspedes (dependiendo del sistema conjugativo)<sup>12</sup>, aunque mantiene los problemas derivados del envío de ADN.

25 Otras formas de envío se basan en métodos físicos o químicos <sup>13</sup>. Dentro de los mecanismos físicos se engloban técnicas como electroporación, transformación o inyección hidrodinámica. Estas técnicas permiten el envío de ADN, ARN o proteína, lo cual les da una gran versatilidad. Han sido muy utilizados en ensayos *in vitro* o *ex vivo*; sin embargo, su difícil traslación a ensayos *in vivo* es una importante limitación en sus usos. Por otro lado, el transporte químico se basa en la modificación o adición de moléculas complementarias al sistema CRISPR-Cas para que éste sea capaz de traspasar las barreras celulares o para protegerlo de la degradación. Algunas formas son la encapsulación lipídica, la unión a nanopartículas como oro o sílice, o la modificación química como la adición de péptidos de penetración celular (CPP). La utilización de estos sistemas está cada vez más extendida en eucariotas, teniendo

una gran aplicación en modelos *in vivo*. Estos mecanismos permiten una protección de los complejos evitando su degradación y permite introducir la proteína Cas directamente. Sin embargo, también tienen una serie de limitaciones como es la obtención y formación de los complejos, la dificultad, en ocasiones, de formar complejos entre el gARN y la proteína Cas, la estabilidad de estos complejos en condiciones fisiológicas, o la dificultad de acumularlos en la célula receptora. En procariotas este tipo de mecanismo no está estudiado prácticamente.

Existe pues la necesidad de proporcionar una herramienta de edición génica basado en el sistema CRISPR/Cas, que solucione todos los problemas descritos anteriormente.

Los Sistemas de Secreción Tipo IV (T4SS, por sus siglas en inglés) son complejos multiproteicos situados en las membranas bacterianas, involucrados en diferentes procesos como la transferencia de proteínas efectoras a una célula eucariota durante una infección <sup>14</sup> o de complejos nucleoproteicos de una bacteria donadora a otra bacteria receptora (conjugación) <sup>15</sup>. La conjugación bacteriana es un mecanismo muy eficiente de transferencia horizontal de ADN desde una bacteria donadora a una receptora a través de un contacto físico entre ambas, confiriéndoles una elevada plasticidad genómica <sup>15</sup>. Durante la conjugación, la relaxasa, que es la proteína encargada de iniciar y terminar el proceso, se une a su diana (el *oriT*), corta la cadena de ADN a transferir en una zona en concreto conocida como sitio *nic* y se une de forma covalente a él formando el complejo relaxasa-ADNmc (ADN monocatenario), que va a ser reclutado por el T4SS y translocado a la célula receptora, donde esta relaxasa catalizará la recirculación de la cadena de ADN transferido <sup>16</sup>.

Existen antecedentes del uso de sistemas de secreción bacterianos con fines terapéuticos. Hay estudios preclínicos en los que utilizan sistemas de secreción bacterianos tipo 3 (T3SS) como sistemas de vacunación o inmunoterapia, o trabajos de reemplazamientos de proteínas defectuosas (también con T3SS) <sup>17</sup>. El uso de T4SS abre un importante campo ya que no sólo permite la translocación de proteínas, sino también de ADN. En particular, existen diferentes trabajos en los que se ha intentado unir las ventajas de los sistemas de secreción con tecnologías de modificación genética (CRISPR-Cas o similares): utilizando el T4SS de *Agrobacterium tumefaciens* para introducir un plásmido en la célula receptora codificando Cas9 y el gARN <sup>18</sup>, o para introducir nucleasas con dedos de zinc (ZFN). En otro caso, han fusionado una señal reconocida por un T3SS con una nucleasa TALEN <sup>19</sup>, y han sido capaces de translocarlo a células eucariotas mediante infección bacteriana. Por último, se ha intentado generar una fusión entre Cas9 y una señal reconocida por el T3SS de *Pseudomonas*

*aeruginosa*, aunque el resultado ha sido negativo debido a la toxicidad de la fusión generada en la bacteria <sup>20</sup>.

### **Breve descripción de la invención**

La presente invención soluciona los problemas descritos en el estado de la técnica puesto  
5 que proporciona una proteína de fusión entre una proteína Cas y una relaxasa conjugativa,  
que es reconocida por el T4SS, de tal modo que la proteína Cas es translocada a través del  
T4SS a la bacteria receptora. De este modo, los T4SS bacterianos actúan como sistemas de  
envío in vivo de proteínas Cas. Además, la proteína de fusión, debido a su actividad relaxasa,  
10 puede unirse covalentemente a una molécula de ADN de elección (la cual debe contener una  
secuencia *oriT* reconocida por la relaxasa de la fusión), que puede codificar tanto el gARN  
como el ADN para editar el genoma receptor.

Así pues, en un primer aspecto, la presente invención se refiere a una proteína de fusión (de  
aquí en adelante, proteína de fusión de la presente invención que comprende una proteína  
relaxasa unida a una proteína Cas).

15 En la presente invención los términos “proteína de fusión”, “péptido de fusión” o “polipéptido  
de fusión” se consideran equivalentes y se pueden usar indistintamente, y se refiere a una  
proteína, péptido o polipéptido que comprende la secuencia aminoacídica de una relaxasa o  
una variante de la misma y de una proteína Cas o una variante de la misma, entendiéndose por  
variante de la proteína relaxasa, cualquier proteína con función relaxasa, y por variante de la  
20 proteína CRISPR/Cas, cualquier proteína Cas con función endonucleasa .

En una realización particular de la presente invención, la proteína relaxasa es la relaxasa  
TrwC. Mas en particular, la relaxasa es la relaxasa TrwC del plásmido conjugativo R388.

En otra realización particular de la presente invención, la proteína Cas, es la proteína Cas12a,  
(también conocida como Cpf1), más en particular es la variante AsCas12a originaria de la  
25 bacteria *Acidaminococcus* sp. *BV3L6*.

En la presente invención el término AsCas12a, Cpf1 y Cas12a se utilizarán indistintamente.

En otra realización particular de la presente invención, la endonucleasa Cas12a se une al  
extremo C-terminal de la relaxasa TrwC.

En otra realización particular, la proteína de fusión de la presente invención comprende la  
30 secuencia aminoacídica SEQ ID NO: 1 o una secuencia con al menos un 70% de identidad  
con la secuencia aminoacídica SEQ ID NO: 1.

En la presente invención los términos "identidad de secuencia", "identidad", "similitud" e "idéntico", se consideran equivalentes y se pueden usar indistintamente. Se entiende que el término "identidad de secuencia" significa el grado de similitud entre dos secuencias de aminoácidos obtenidas alineando las dos secuencias. El grado de identidad se expresa como porcentaje y depende del número de residuos comunes entre las secuencias alineadas. El grado de identidad entre secuencias de aminoácidos se determina mediante métodos bien establecidos en el estado de la técnica, tales como BLAST.

Las proteínas con al menos un 70% de identidad con la secuencia aminoacídica SEQ ID NO: 1 se consideran variantes de la secuencia SEQ ID NO: 1 y también se consideran dentro de la presente invención. Concretamente una secuencia con al menos un 70, 75, 80, 85, 90, 95, 100 % de identidad con la SEQ ID NO: 1. Se consideran también dentro la presente invención aquellas secuencias análogas, derivadas o equivalentes a la secuencia SEQ ID NO: 1, y que comprenden al menos un residuo de aminoácido alterado por una, adición, sustitución, delección o modificación química de un aminoácido, respecto a la proteína de fusión de la presente invención.

En un segundo aspecto, la presente invención se refiere a una secuencia nucleotídica (de aquí en adelante secuencia nucleotídica de la presente invención), que codifica la proteína de fusión de la presente invención.

En la presente invención el término "secuencia nucleotídica que codifica para la proteína de fusión" se refiere a una secuencia de nucleótidos, polinucleótidos, o ácidos nucleicos, que, bajo un control de expresión adecuado, es capaz de transcribir y traducir la secuencia aminoacídica de la proteína de la presente invención.

En una realización más en particular de la presente invención, la secuencia nucleotídica que codifica para la proteína de fusión comprende una secuencia nucleotídica según la SEQ ID NO: 2.

En otro aspecto, la presente invención se refiere a un vector de expresión (de aquí en adelante vector de expresión de la presente invención) que comprende la secuencia nucleotídica de la presente invención.

En la presente invención el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido, preferentemente el ácido nucleico de la presente invención que codifica para la proteína de fusión de la presente invención y, está unido operativamente a nucleótidos adicionales tales como secuencias control, que

proporcionan su expresión. El vector de la presente invención, puede comprender moléculas de ácido nucleico monocatenarias, bicatenarias o parcialmente bicatenarias; moléculas de ácido nucleico que comprenden ADN, ARN o ambos; y otras variedades de polinucleótidos conocidos en el estado de la técnica. Un tipo de vector es un plásmido, que se refiere a un bucle circular de ADN bicatenario en donde se pueden insertar segmentos de ADN adicionales. Otro tipo de vector es un vector viral, que comprenden polinucleótidos para ser transfectados en una célula huésped.

En la presente invención por “unido operativamente” se refiere a una disposición de dos o más componentes, en donde dichos componentes están en una relación que les permite funcionar de manera coordinada.

En otro aspecto, la presente invención se refiere a una célula (de aquí en adelante célula de la presente invención) que comprende la proteína de fusión de la presente invención y/o el ácido nucleico de la presente invención, y/o el vector de la presente invención. La célula de la presente invención se refiere a cualquier tipo de célula que es susceptible de transformación, transfección, transducción, conjugación y translocación con el ácido nucleico de la presente invención o con el vector de expresión que comprende dicho ácido nucleico, o con la proteína de fusión de la presente invención.

En otro aspecto, la presente invención se refiere al sistema CRISPR/Cas (de aquí en adelante, sistema CRISPR/Cas de la presente invención) que comprende la proteína de fusión de la presente invención.

En una realización más en particular, el sistema CRISPR/Cas de la presente invención comprende

- una secuencia nucleotídica que codifica para el ARN guía,
- la proteína de fusión relaxasa-Cas de la presente invención,

donde la proteína relaxasa-Cas se transloca a una célula receptora a través de un sistema de secreción bacteriano tipo IV.

En la presente invención por “sistema CRISPR/Cas” se refiere al sistema que comprende los elementos implicados en la expresión y/o actividad de los genes asociados a CRISPR/Cas, incluyendo las secuencias nucleotídicas que codifican para el ARN guía, secuencias nucleotídicas que codifican la proteína de fusión relaxasa-Cas de la presente invención.

En la presente invención por “ARN guía”, se refiere a una secuencia polinucleotídica que contiene una secuencia repetida, específica de cada sistema CRISPR y una secuencia espaciadora con complementariedad con la secuencia diana de tal forma que es capaz de hibridar con la misma.

- 5 En la presente invención por “secuencia diana” se refiere a cualquier secuencia que se encuentra en la célula receptora, que tiene un motivo adyacente de protoespaciador (PAM, de sus siglas en inglés) correcto y que es capaz de hibridar con la secuencia guía.

En la presente invención por “motivo adyacente de protoespaciador (PAM)” se refiere a la secuencia nucleotídica concreta, situada al lado de la secuencia diana, reconocida por la  
10 endonucleasa Cas.

En la presente invención por “complementariedad” se refiere a la capacidad de un ácido nucleico para formar enlaces de hidrógeno con otra secuencia de ácido nucleico de forma tradicional, según Watson and Crick, o de alguna otra manera. El porcentaje de complementariedad indica el porcentaje de residuos en una molécula de ácido nucleico que  
15 puede formar enlaces de hidrógeno con otra molécula de ácido nucleico. “suficientemente complementario” significa que el grado de complementariedad es de al menos 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 99%, 100%.

En la presente invención por “hibridación” se refiere a una reacción en la que uno o más polinucleótidos reaccionan para formar un complejo que se estabiliza a través de la formación  
20 de enlaces de hidrógeno entre las bases de los restos de nucleótidos. La formación de enlaces de hidrógeno se produce mediante el emparejamiento de bases de según Watson and Crick, o de cualquier otra manera específica de secuencias, conocido por un experto en la materia.

En una realización más en particular de la presente invención, el sistema de secreción bacteriano tipo IV es el del plásmido conjugativo R388

- 25 En una realización más en particular, el sistema CRISPR/Cas de la presente invención, comprende:

- una secuencia nucleotídica que codifica para el ARN guía,
- la proteína de fusión relaxasa- Cas de la presente invención

donde la proteína relaxasa-Cas es traslocada a la célula receptora a través del T4SS y, en la célula receptora se une a la secuencia nucleotídica que codifica para el ARN guía e hibrida con el ADN diana.

5 En una realización más en particular, el sistema CRISPR/Cas de la presente invención, comprende:

- una secuencia nucleotídica que codifica para el ARN guía,
- la proteína de fusión relaxasa- Cas de la presente invención,

10 donde la proteína relaxasa-Cas reconoce una secuencia *oriT* de vector movilizable que contiene la secuencia nucleotídica que codifica para el ARN guía y transloca dicha molécula de ADN unida a la proteína relaxasa-Cas a través del T4SS de la presente invención a la célula diana donde el ARN guía hibrida con el ADN diana.

15 En la presente invención, por “vector movilizable” se refiere a cualquier vector el cual contenga un origen de transferencia (*oriT*). Se entiende por *oriT* a una secuencia nucleotídica concreta, reconocida por la relaxasa del sistema conjugativo y necesaria para la movilización del vector mediante el proceso conjugativo.

En una realización más en particular de la presente invención, la proteína de fusión relaxasa-Cas del sistema CRISPR/Cas de la presente invención, comprende la relaxasa TrwC unida a la endonucleasa Cas12a. Más en particular, comprende la secuencia aminoacídica SEQ ID NO: 1.

20 En una realización particular de la presente invención, la célula diana es una célula eucariota.

En otra realización particular de la presente invención, la célula diana es una célula procarionta.

En otro aspecto, la presente invención se refiere a la proteína de fusión de la presente invención, o al sistema CRISPR/Cas de la presente invención para la translocación in vivo, de ADN y/o endonucleasas a células diana a través del T4SS de la presente invención.

25 En otro aspecto, la presente invención se refiere al uso de la proteína de fusión de la presente invención y/o al ácido nucleico de la presente invención, y/o al vector de la presente invención, para la edición génica en células. En una realización particular, para la edición génica en células eucariotas. En una realización particular, para la edición génica en células procariontas.

30

### Descripción de las figuras

La figura 1, muestra los resultados de estabilidad de la proteína de fusión, mediante Western Blot utilizando un anticuerpo anti-TrwC. La proteína TrwC purificada se utilizó como control positivo. El lisado celular de HEK293T se usó como control negativo (-). Cas12a y TrwC:Cas12a corresponden con los lisados celulares tras transfectar los plásmidos pY010 y pLG08 respectivamente en HEK293T. Los tamaños esperados se muestran en la parte derecha de la figura. La flecha superior corresponde a 250 kDa y la inferior a 100 kDa.

La figura 2, muestra la actividad de Cas12a.

a) TrwC:Cas12a muestra actividad endonucleasa en eucariotas. Gel de agarosa tras tratar las muestras con el Surveyor Mutation Detection kit. La actividad de Cas12a se mide por la generación de diferentes bandas de digestión producidas por la surveyor endonucleasa.

b) TrwC:Cas12a muestra actividad endonucleasa en procariotas. Número de transformantes tras coelectroporar el plásmido pLG24 (pTET:trwC:cas12a) con diferentes gARNs en condiciones de inducción (+) y de no inducción. Los datos corresponden a la media 9 experimentos independientes. (\*\*\*\*,  $P < 0.0001$ ).

La figura 3 muestra la actividad de la proteína de fusión en la bacteria receptora tras su translocación por el T4SS mediante conjugación. TrwC:Cas12a produce un incremento de la señal SOS en la célula receptora en presencia de un gARN con diana en el cromosoma bacteriano. a) En la imagen se observan los filtros de conjugación tras 3 horas de conjugación en condiciones de inducción. En el filtro de la parte derecha de la imagen (con diana para el gARN) se observa un incremento en la fluorescencia con respecto al de la parte izquierda (sin diana para el gARN). b) En la parte derecha se muestran los valores de GFP/OD obtenidos con el TECAN Infinite M200 Pro tras detener la conjugación. (\*\*\*,  $P < 0.0005$ )

La figura 4 muestra el alineamiento de la secuencia del gen sacB de los diferentes mutantes. En la imagen se muestra una zona del alineamiento del gen sacB, en el que se comparan 6 transconjugantes mutantes resistentes a sacarosa (EcLG1-6) con la cepa MG1655:sacB. En la parte inferior del alineamiento se marca la PAM (subrayada) y el espaciador utilizados para generar el gARN.

### Descripción detallada de la invención

La presente invención proporciona una proteína de fusión TrwC-Cas12a estable y funcional, que es reconocida a través de un Sistema de Secreción Tipo IV bacteriano (T4SS) y translocada a una célula receptora, manteniendo su actividad.

#### 5 *Ejemplo 1: Diseño y construcción de la proteína de fusión de la presente invención*

La proteína Cas12a (también conocida como Cpf1), en concreto la variante AsCas12a tiene actividad tanto en células procariontas como eucariotas<sup>4</sup>. Esta proteína tiene una serie de diferencias importantes con respecto a Cas9 (la proteína modelo por excelencia), como la PAM que reconoce (5'-TTTN-3' en lugar de 5'-NGG-3') o el tipo de doble corte que produce  
 10 en la hebra de ADN (extremos cohesivos en lugar de romos). Además, Cas12a tiene un peso molecular de 127kDa, mientras que Cas9 de 163 kDa. Esta diferencia es importante, ya que a la hora de ser translocadas a través del T4SS, va a ser más sencillo hacerlo con una proteína de menor tamaño.

Con respecto al sistema conjugativo, se utilizó el del plásmido R388, por tratarse de un  
 15 sistema bien caracterizado y de muy amplio rango de huéspedes<sup>21</sup>. La relaxasa conjugativa de este sistema, la proteína TrwC, se trata de una proteína que permite la fusión de dominios en su extremo C terminal sin afectar a su función en la conjugación<sup>22</sup>.

Para generar la proteína de fusión entre la relaxasa TrwC y la endonucleasa Cas12a, se construyó el gen adecuado mediante la técnica de isothermal assembly<sup>23</sup>, eliminando el codón  
 20 de STOP del gen *trwC* (amplificado a partir del plásmido pAA12<sup>22</sup>) y colocando a continuación el codón de inicio de *cas12a* amplificado a partir del plásmido pY010 (addgene número 69982). Hay que hacer notar que ese gen ha sido optimizado para su expresión en mamíferos, y codifica para una proteína que contiene una señal nuclear y una cola de histidinas C-terminal.

#### 25 *Ejemplo 2: Estabilidad de la proteína de fusión de la presente invención*

En primer lugar, se comprobó si la proteína de fusión era estable mediante un Western Blot a partir de lisados celulares de HEK293T. Para ello, se emplazó el gen de fusión en un plásmido (pLG08) bajo el promotor CMV, que permite la expresión en eucariotas.

El plásmido pLG08, se construyó mediante el método de clonación Isothermal Assembly  
 30 (Gibson et al, 2009). Para ello, se ensambló el gen *trwC* y el vector pY010. Se amplificó por PCR la secuencia de *trwC* del plásmido pAA12 (en él, *trwC* carece de codón de STOP). El

vector pY010 se amplificó por PCR. El fragmento de *trwC* fue ensamblado delante del codón de inicio de Cas12a en el pY010, bajo el control del promotor CMV. Se transfectaron los plásmidos pY010 (Cas12a) y pLG08 (TrwC:Cas12a), y se realizó un lisado celular. A continuación, se realizó el Western Blot utilizando un anticuerpo anti TrwC (Figura 1). Los resultados mostraron como la mayoría de la proteína se encontraba en el tamaño esperado para la fusión, indicando que se producía la proteína esperada, y era estable.

*Ejemplo 3: Actividad de la proteína de fusión de la presente invención*

A continuación, se realizó una validación funcional de las diferentes partes de la fusión, para comprobar si mantenían las funciones de las proteínas parentales. En primer lugar, se realizaron conjugaciones bacterianas, en las que se complementó un plásmido derivado de R388 carente de una proteína TrwC funcional, con TrwC o con TrwC:Cas12a. Las frecuencias de conjugación obtenidas en ambos casos fueron muy similares (Tabla 1). Estos datos no solo muestran que TrwC:Cas12a mantiene la actividad de TrwC, sino que, además, la proteína de fusión es reconocida y translocada a través del T4SS a la célula receptora, puesto que se sabe que la relaxasa se requiere en la bacteria receptora para completar el proceso conjugativo<sup>24</sup>.

Plásmido con T4SS	Complementación	Relaxasa	Frecuencia de conjugación
pSU1445 <sup>25</sup> (R388::Tn5 <i>tac1</i> en <i>trwC</i> )	pHP159 <sup>26</sup> pLG24	TrwC TrwC:Cas12a	3.8x10 <sup>-1</sup> ±0.24 4.1x10 <sup>-1</sup> ±0.26

Tabla 1. Frecuencias de conjugación.

En todos los casos, la bacteria donadora corresponde a la cepa de *Escherichia coli* D1210 y la receptora a DH5α. Los datos corresponden a la media de 4 experimentos independientes.

Por otra parte, se validó la actividad endonucleasa de Cas12a en la proteína de fusión, tanto en células eucariotas como en células procariontas (Figura 2).

Para la validación en células eucariotas se utilizó de nuevo el plásmido pLG08. El gARN se construyó en un casete de expresión, bajo el promotor U6, utilizando como diana el gen *dnmt1*. Los diferentes elementos se introdujeron por transfección a células HEK293T (ATCC® CRL-3216™) utilizando el kit Lipofectamine® 3000 Reagent de Invitrogen. Las transfecciones realizadas se muestran en la Tabla 2.

Transfección	Plásmido	Proteína Cas codificada	U6:: <i>dnmt1<sub>gRNA</sub></i> )
1	pY010	Cas12a	-
2	-	-	+
3	pY010	Cas12a	+
4	pLG08	TrwC:Cas12a	+

Tabla 2. Transfecciones realizadas para determinar la actividad de Cas12a en eucariotas.

Tras tres días, se extrajo el ADN genómico con el kit GenElute™ Mammalian Genomic DNA Miniprep Kits (Sigma Aldrich). Se realizó una PCR amplificando la zona diana y se trató con el kit Surveyor® Mutation Detection Kits (IDT), el cual detecta la formación de indels, pequeñas inserciones-delecciones producidas por los sistemas de reparación no homólogos de la célula eucariota en respuesta a los cortes generados por Cas12a.

Los resultados mostraron que TrwC:Cas12a era capaz de generar estos indels, al igual que Cas12a. Con esto, demostramos que Cas12a en la proteína de fusión mantiene su actividad (Figura 2a).

Por otro lado, se realizó la comprobación de la actividad endonucleasa de TrwC-Cas12a en células procariontas. La mayoría de las bacterias no contienen un sistema de reparación del ADN basado en la recombinación no homóloga (NHEJ en inglés) por lo que si Cas12a corta el ADN de forma eficaz, producirá cortes en la doble cadena de ADN y la bacteria morirá, pudiendo medir la actividad de Cas12a en un decrecimiento de nuestra población bacteriana <sup>27</sup>. Para estos experimentos se generó el plásmido pLG24 que expresa el gen de fusión *trwC:cas12a* bajo el control del promotor pTet, el cual, en ausencia del inductor anhidrotetraciclina (aTc) está muy reprimido, mientras que en presencia de aTc es muy activo, permitiendo una regulación muy estricta de la cantidad de TrwC:Cas12a.

El plásmido pLG14 es el plásmido en el que se basa el pLG24. Se construyó ensamblando la secuencia de *Cas12a-NLS-3Ha* (amplificada del plásmido pY010) con el vector pBBR6, amplificado por PCR, dejando el gen *Cas12a* en el pBBR6, aguas abajo del promotor lactosa.

El plásmido pLG24 se construyó ensamblando cuatro amplicones: el promotor pTET se amplificó del plásmido pOSIP-CO-RBS-library-dCas9. La secuencia de *trwC* se amplificó del pAA12. El pLG14 se amplificó en dos fragmentos, eliminando el promotor lactosa. El fragmento de *trwC* se ensambló delante del codón de inicio de *cas12a* y la secuencia del pTET se colocó aguas arriba de ambos fragmentos.

Para estos experimentos también se diseñaron dos gARNs que se clonaron bajo el control del promotor plac (inducible por isopropil- $\beta$ -D-1-tiogalactopiranosido, IPTG), cuyas dianas son los genes *lacZ* (pLG15) o *sacB* (pLG19). En la tabla 5 se muestran los plásmidos utilizados en la presente invención. En la tabla 6 se muestra la construcción detallada de los plásmidos y en la tabla 7 se muestra el listado de oligonucleótidos utilizados en la presente invención.

Los experimentos se realizaron coelectroporando en la cepa de *Escherichia coli* D1210 el plásmido pTet:TrwC:Cas12a (pLG24) con un gARN contra el gen *lacZ* (presente en la cepa diana) o contra el gen *sacB* (no presente en la cepa diana). Los transformantes se seleccionaron en condiciones de inducción (aTc 1 $\mu$ g/ml e IPTG 0.5mM) y de no inducción (Figura 2b). El número de transformantes cayó drásticamente solamente en los casos en los que se coelectroporaron los plásmidos con el gARN cuya diana se encontraba presente en la bacteria (*lacZ*) y se seleccionó en condiciones de inducción, demostrando la actividad endonucleasa de TrwC:Cas12a también en procariontes.

#### *Ejemplo 4: Actividad de TrwC:Cas12a tras ser translocada por un T4SS*

Uno de los problemas planteados en el estado de la técnica y que se han superado en la presente invención, es la recuperación de la actividad de la proteína de fusión en la célula receptora tras ser translocada a través del T4SS. La translocación de sustratos a través de sistemas de secreción conlleva un desplegamiento total o parcial de los mismos<sup>28,29</sup>. La actividad de las proteínas tras haber perdido parte de su configuración tridimensional es muy variable e imposible de saber a priori. Se sabe que las relaxasas recuperan su funcionalidad en la célula receptora, pero no se podía prever qué ocurriría con la actividad endonucleasa de la fusión TrwC:Cas12a.

Para detectar la actividad Cas12a en las bacterias receptoras, se realizaron dos ensayos distintos: incremento de la señal SOS y selección de mutaciones en *sacB*.

Para el primer ensayo, se ha adaptado el ensayo de la inducción de la señal SOS propuesto por Lun Cui y David Bikard<sup>27</sup>. En este trabajo, demuestran como los cortes de doble cadena producidos por Cas9 en el cromosoma bacteriano generan una activación de la señal SOS en cepas RecA+. Para demostrarlo, generan el plásmido pZA31-*sulA*-GFP (número en addgene: 78493), que tiene el gen que codifica la proteína fluorescente verde (GFP) bajo el control del promotor del gen *sulA*, involucrado en la respuesta SOS; por tanto, una inducción de la misma debida a los cortes producidos por Cas se traduce en producción de GFP. Se ha adaptado este ensayo para la detección de la actividad de TrwC:Cas12a en los transconjugantes. Las

bacterias utilizadas en este ensayo se muestran en la Tabla 3. Básicamente, la idea es realizar conjugaciones a cepas que tengan un gARN con (*lacZ*) o sin (*sacB*) diana en el cromosoma de la bacteria receptora. La translocación de TrwC:Cas12a desde la bacteria donadora daría lugar a su acción endonucleasa en el cromosoma en el caso de la receptora con el gARN contra *lacZ*.

Plásmidos en la bacteria donadora <sup>1</sup>	Descripción de plásmidos	Plásmidos en la receptora <sup>2</sup>	Descripción de plásmidos	Gen diana
pSU1445 <sup>25</sup> + pLG24	R388::Tn5 <i>lac1</i> en <i>trwC</i> + pTET: <i>trwC:cas12a</i>	pZA31-sulA:gfp <sup>27</sup> + pLG15	psulA:gfp + plac: <i>lacZ</i> <sub>gRNA</sub>	<i>lacZ</i>
		pZA31-sulA:gfp <sup>27</sup> + pLG19	psulA:gfp + plac: <i>sacB</i> <sub>gRNA</sub>	<i>sacB</i>

Tabla 3. Plásmidos utilizados para la detección de la señal SOS en transconjugantes

<sup>1</sup>Bacteria donadora: *E. coli* D1210

<sup>2</sup>Bacteria receptora: *E. coli* MG1655 (RecA+)

Para las conjugaciones, los cultivos bacterianos crecidos en fase estacionaria se diluyeron 1:20 y se indujeron, las donadoras con aTc 1µg/ml y las receptoras con IPTG 0.5 mM, durante 3 horas. A continuación, se realizaron las conjugaciones en filtro, suplementando las placas de conjugación con aTc e IPTG. La detección de GFP se realizó directamente tras 3 horas, sobre la placa de conjugación con el lector de placas Azure Biosystems c400 (Figura 3a). A continuación, se detuvo la conjugación y se midieron los valores de fluorescencia de GFP (los valores de excitación y emisión fueron 515nm y 475nm respectivamente) y de OD (600nm) utilizando el TECAN infinite M200 Pro (Figura 3b).

En los experimentos realizados, se observó un ruido de fondo en la detección de GFP en ausencia de corte por Cas12a, debido a que la propia conjugación induce la señal SOS<sup>30</sup>. Por encima de ese ruido, se observó un incremento de GFP específicamente cuando en la cepa receptora estaba presente el gARN y su diana genómica (*lacZ*), confirmando la actividad Cas12a procedente de la proteína TrwC:Cas12a translocada desde las donadoras.

Para el segundo ensayo de actividad de Cas12a se seleccionaron cambios en el fenotipo de los transconjugantes. El gen elegido como diana fue *sacB*. La expresión de *sacB* en presencia de sacarosa hace que la bacteria que lo esté expresando muera<sup>31</sup>. De hecho, es una contraselección tan efectiva, que las bacterias que expresan *sacB* solo sobreviven en un medio suplementado con sacarosa si pierden la actividad SacB por mutación. No existe

bibliografía acerca de los reordenamientos que se puedan producir como consecuencia del corte de Cas12a en el cromosoma bacteriano, sin embargo, está descrito que cortes por Cas9 en el cromosoma, pueden dar lugar a deleciones producidas por eventos de recombinación entre secuencias homólogas o con microhomologías que se encuentren distanciadas en el cromosoma, de una manera *recA* dependiente <sup>27</sup>. De este modo, si se usa como diana de Cas12a en la célula receptora el gen *sacB*, y se siembra en medio suplementado con sacarosa los transconjugantes resultantes de la conjugación, si Cas12a ha introducido cortes en el gen *sacB*, se encontrará un mayor número de mutantes resistentes a sacarosa debido a las deleciones producidas en el gen por el corte de Cas12a y posteriores recombinaciones.

10 Las donadoras y receptoras utilizadas en los experimentos se muestran en la Tabla 4.

Donadora (D1210)			Receptor a*	Frec. conjugación	Trs. sacR**
Plásmidos	Descripción	Relaxasa			
pSU1445+pHP 159	R388:: <i>Tn5tac1</i> en TrwC + oritTrwABC	TrwC	-sacB	$7 \times 10^{-1}$	>>
			+sacB	$1.23 \times 10^{-3}$	0
pSU1445+pLG 24	R388:: <i>Tn5tac1</i> en TrwC + pTET:TrwC:Cas12a	TrwC:Cas12a	-sacB	$1.3 \times 10^{-3}$	>>
			+sacB	$1.27 \times 10^{-1}$	11

**Tabla 4. Selección de transconjugantes resistentes a sacarosa.** Los datos corresponden a un experimento.

\*Las cepas receptoras utilizadas fueron MG1655 (-sacB) y MG1655::*sacB* (+sacB). Ambas cepas contienen los plásmidos pZA31-*sulA:gfp* <sup>27</sup> y pLG19 (*plac:sacB*gRNA)

15 \*\*Números de transconjugantes en la dilución 0 en placas suplementadas con sacarosa 1%. >> significa incontables (crecimiento continuo).

En este caso, se compararon los datos ensayando de forma paralela TrwC con TrwC:Cas12a. De esta manera, se puede comparar el número de mutantes resistentes a sacarosa cuando se transloque a la célula receptora (que contiene el gARN contra *sacB*) la relaxasa sola, o la proteína de fusión. Tras realizar los ensayos de conjugación en las condiciones de inducción (iguales que las explicadas anteriormente) se seleccionaron transconjugantes en medio suplementado con sacarosa al 1% y sin sacarosa. En el caso de las conjugaciones en las que se usó MG1655 como receptora, en todos los casos crecieron receptoras resistentes a sacarosa 1%. Por el contrario, cuando la cepa utilizada como receptora fue MG1655::*sacB*, únicamente en el caso de TrwC:Cas12a se obtuvieron mutantes resistentes a sacarosa. Estos mutantes fueron analizados amplificando por PCR la región de *sacB* con los oligonucleótidos

5'-CTACCGCACTGCTGGCAG-3' (SEQ ID NO: 25) y 5'-GATGCTGTCTTTGACAACAG-3' (SEQ ID NO: 26). En todos los casos se obtuvo la amplificación del producto del tamaño esperado, que fue utilizado para determinar la secuencia de ADN. En los 6 transconjugantes analizados se observaron deleciones en la zona donde el gARN va dirigido (Figura 4). Este resultado mostró de manera inequívoca la actividad de Cas12a en la bacteria receptora, demostrando que la proteína es activa tras ser translocada a través de un T4SS. Por lo tanto, se confirma que, Cas12a se puede translocar a través de un T4SS bacteriano uniéndola covalentemente a una relaxasa conjugativa, y que una vez que llega a la bacteria receptora, es capaz de recuperar su actividad de endonucleasa sitio-específica guiada por el gARN adecuado.

Los ensayos mostraron que la proteína de fusión de la presente invención, es una proteína de fusión funcional TrwC:Cas12a, que permitió la translocación de la endonucleasa sitio-específica Cas12a a través de un sistema de secreción bacteriano tipo IV. Con este sistema, se lleva directamente la endonucleasa Cas a la célula receptora, con los beneficios que esto aporta, como son una mayor eficiencia y rapidez en el sistema, sin necesidad de purificación de la misma. Además, al trabajar con proteínas y no con ADN o mRNA, se disminuye el riesgo de actividad off-target y toxicidad, derivadas de la expresión sostenida en el tiempo de la endonucleasa. También se evita el riesgo de integración del ADN en el genoma diana.

La proteína de fusión de la presente invención puede ser utilizada como herramienta de modificación genética de cualquier bacteria que actúe como receptora conjugativa del plásmido R388, cuyo rango de huéspedes abarca una gran variedad de bacterias Gram-negativas como *Escherichia coli* o del género *Pseudomonas*, *Salmonella*, *Stenotrophomonas* o *Shigella*<sup>32</sup>, y probablemente vaya mucho más allá, puesto que también se ha demostrado su conjugación a cianobacterias<sup>33</sup>.

La proteína de fusión de la presente invención, proporciona un vehículo de introducción de las herramientas CRISPR/Cas en bacterias no modelo, que representan el grueso de nuestra microbiota, cuyo papel en nuestra salud se está mostrando cada vez más relevante, y para las que las herramientas genéticas convencionales generalmente no son válidas<sup>34</sup>. Los ensayos demuestran que la proteína de fusión de la presente invención mantiene su actividad de endonucleasa sitio-específica en células humanas y, habida cuenta que los T4SS también median la translocación de proteínas directamente a células de mamífero, puede tener una posible aplicación en la modificación genética de mamíferos.

Plásmido	Descripción	Referencia
pAA12	pHP159:: <i>trwC-RalF</i> <i>TS</i>	22
pBBR6	Vector de clonación derivado del pBBR1-MCS	35
pHP159	pBBR6:: <i>oriT</i> <i>trwAB*C::pCMV</i> <i>eGFP</i>	26
pLG08	pY010:: <i>trwC-Cas12a</i>	Este trabajo
pLG14	pBBR6:: <i>Cas12a</i>	Este trabajo
pLG15	pUC8:: <i>plac lacZ<sub>gRNA</sub></i>	Este trabajo
pLG19	pUC18:: <i>plac</i> <i>sacB<sub>gRNA</sub></i>	Este trabajo
pLG24	pBBR6:: <i>pTET::trwC-Cas12a Δplac</i>	Este trabajo
pOSIP-CO-RBS-library-dCas9	Contiene pTET	36
pSU1445	R388:: <i>Tn5tac1</i> en <i>trwC</i>	37
pUC8, pUC18	Vectores de clonación	38,39
pY010	pcDNA3.1:: <i>hAsCas12a</i>	4
pZA31-sulA-GFP	pZA31:: <i>luc:GFP mut2</i>	27

Tabla 5: Plásmidos utilizados en la presente invención

Oligonucleótido	Secuencia (5' a 3')	USO
U6_crRNA_F	GAGGGCCTATTTCCCATGATTCCT (SEQ ID NO: 21)	Casete de expresión U6::gRNA <sub>DNMT1</sub> en eucariotas
DNMT1_crRNA3_R	GAGTAACAGACATGGACCATCAGATC TACAAGAGTAGAAATTACGGTGTTC GTCCTTTCCACAAG (SEQ ID NO: 22)	
DNMT1_F	CTGGGACTCAGGCGGGTCCAC (SEQ ID NO: 23)	Amplificación región <i>dnmt1</i> para el Surveyor Assay
DNMT1_R	CCTCACACAACAGCTTCATGTCAGC (SEQ ID NO: 24)	
SacB_F	CTACCGCACTGCTGGCAG (SEQ ID NO: 25)	Análisis de transconjugantes resistentes a sacarosa
SacB_R	GATGCTGTCTTTGACAACAG (SEQ ID NO: 26)	

Tabla 6: Oligonucleótidos usados

Método de clonación	Plásmido	Descripción	Oligonucleótido	Secuencia de oligonucleótidos (5' a 3') <sup>2</sup>	Inserto (Vector de procedencia)	Vector
IA	pLG08	pY010:: <i>trwC-Cas12a</i>	pY010_TrwC_F Cpf1_TrwC_R	TTAAGCTTGGTACCGCCACCCATGGCTCAGTCACAT <b>GGTATTG</b> (SEQ ID NO:3) AAGCCCTCGAAGCTGTGTGCAT <b>CCTTCCGGCCCTCCAT</b> (SEQ ID NO: 4)	trwC (pAA12)	
IA	pLG14	pBBR6:Cpf1	Cpf1_1_F pY010_929_R	<b>GGTGGGGTACCAAG</b> (SEQ ID NO: 5) <b>ATGACACAGTTCGAGGG</b> (SEQ ID NO: 6)		pY010
IA	pLG14	pBBR6:Cpf1	Cpf1_pBBR6_F pLG15_Cpf1_R	TGAGCGGATAACAAATTCACACAGGAAACAGCTAT <b>GACACAGTTCGAGGG</b> (SEQ ID NO: 7) TTTAAACGGGAAATTTAACAAAAATTTAACGGGGCAT <b>AGTCGGGGACA</b> (SEQ ID NO: 8)	Cas12a (pY010)	
ER	pLG15	pUC8:lacZ <sub>gR</sub> NA	pBBR6_1440_F pBBR6_1074_R <u>lacZ-gRNA</u>	GCGTTAATATTTTGTAAATTCG (SEQ ID NO: 9) AGCTGTTCCCTGTGTGAAA (SEQ ID NO: 10) <u>GAATTCGTCAAAAGACCTTTTTTAATTTCTACTCTTG</u> TAGATCCGACCGCACGCCGCATCCAGCGCTGTCA AAAGACCTTTTTAAATTTCTACTCTTGTAGATAAGC T (SEQ ID NO: 11)		pBBR6
ER	pLG19	pUC18:sacB <sub>g</sub> RNA	<u>sacB-gRNA</u>	<u>GAATTCGTCAAAAGACCTTTTTTAATTTCTACTCTTG</u> TAGATGGACAGCTGGCCATTACAAAACGGTCAAAA GACCTTTTTAAATTTCTACTCTTGTAGATAAGC T (SEQ ID NO: 12)	lacZ <sub>gRNA</sub> (gblock)	pUC8
IA	pLG24	pBBR6:pTET :: <i>trwC-Cas12a</i> $\Delta$ plac	tetR_F ptet_R  ptet_TrwC_F TrwC_Cpf1_R	<b>TTAAGACCCACTTTCACATTTAAG</b> (SEQ ID NO: 13) <b>TTTTGCCTCCTAACTAGGTCAT</b> (SEQ ID NO: 14)  TATCCGGAGGCATATCAAATGACCCTAGTTAGGAGG CAAAA <b>ATGCTCAGTCACATGGTATTGA</b> (SEQ ID NO: 15) TCACCTGATACAGGTTGGTAAAGCCCTCGAACTGT GTCAT <b>TGACCTCCGGCCCTCCAT</b> (SEQ ID NO: 16)	pTET (pOSIP- KH- pTET_RBS2- dCas9)	

			<p>Cpf1_F                  pLG14_R                  pLG14_F                  pLG14_2_R</p>	<p><b>ATGACACAGTTCGAGGGGCT</b> (SEQ ID NO: 17)  <b>GCGTAGCACACCAGGCGT</b> (SEQ ID NO: 18)  <b>ACCAATAGGCCGACTGCGAT</b> (SEQ ID NO: 19)                  CGGATTAGAAAAACAACCTTAAATGTGAAAAGTGGGT  <b>CTTAATTAGGTGGCGGTACTTGGGTCG</b> (SEQ ID                  NO: 20)</p>		pLG14
--	--	--	--	--	--	-------

Tabla 7: Construcción detallada de los plásmidos

## REFERENCIAS

1. Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R. & van der Oost, J. Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends Biotechnol* **34**, 575–587 (2016).
- 5 2. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* vol. 18 67–83 (2020).
3. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-. )*. **337**, 816–821 (2012).
4. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
- 10 5. Moon, S. Bin, Kim, D. Y., Ko, J. H. & Kim, Y. S. Recent advances in the CRISPR genome editing tool set. *Experimental and Molecular Medicine* vol. 51 1–11 (2019).
6. Yao, R. *et al.* CRISPR-Cas9/Cas12a biotechnology and application in bacteria. *Synthetic and Systems Biotechnology* vol. 3 135–149 (2018).
- 15 7. Luo, M. L., Leenay, R. T. & Beisel, C. L. Current and future prospects for CRISPR-based tools in bacteria. *Biotechnol. Bioeng.* **113**, 930–943 (2016).
8. Shabbir, M. A. B. *et al.* CRISPR-cas system: Biological function in microbes and its use to treat antimicrobial resistant pathogens. *Annals of Clinical Microbiology and Antimicrobials* vol. 18 21 (2019).
- 20 9. Moon, S. Bin, Kim, D. Y., Ko, J. H., Kim, J. S. & Kim, Y. S. Improving CRISPR Genome Editing by Engineering Guide RNAs. *Trends in Biotechnology* vol. 37 870–881 (2019).
10. Hughes, T. S. *et al.* Immunogenicity of intrathecal plasmid gene delivery: Cytokine release and effects on transgene expression. *J. Gene Med.* **11**, 782–790 (2009).
11. Xu, C. L., Ruan, M. Z. C., Mahajan, V. B. & Tsang, S. H. Viral Delivery Systems for CRISPR. *Viruses* **11**, 28 (2019).
- 25 12. Hamilton, T. A. *et al.* Efficient inter-species conjugative transfer of a CRISPR nuclease for targeted bacterial killing. *Nat. Commun.* **10**, 1–9 (2019).
13. Glass, Z., Lee, M., Li, Y. & Xu, Q. Engineering the Delivery System for CRISPR-Based Genome Editing. *Trends in Biotechnology* vol. 36 173–185 (2018).
- 30 14. Llosa, M., Schröder, G. & Dehio, C. New perspectives into bacterial DNA transfer to human cells. *Trends in Microbiology* vol. 20 355–359 (2012).
15. de la Cruz, F. & Davies, J. Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol* **8**, 128–133 (2000).
16. Guzmán-Herrador, D. L. & Llosa, M. The secret life of conjugative relaxases. *Plasmid* vol. 104 102415 (2019).
- 35 17. Walker, B. J., Stan, G.-B. V & Polizzi, K. M. Intracellular delivery of biologic therapeutics by bacterial secretion systems. (2017) doi:10.1017/erm.2017.7.
18. Jiang, W. *et al.* Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. doi:10.1093/nar/gkt780.
- 40 19. Jia, J. *et al.* Efficient Gene Editing in Pluripotent Stem Cells by Bacterial Injection of Transcription Activator-Like Effector Nuclease Proteins. *Stem Cells Transl. Med.* **4**, 913–926 (2015).
20. Bai, F., Li, Z., Umezawa, A., Terada, N. & Jin, S. Bacterial type III secretion system as a protein delivery tool for a broad range of biomedical applications. *Biotechnology Advances* vol. 36 482–493 (2018).
- 45 21. Datta, N. & Hedges, R. W. Trimethoprim resistance conferred by W plasmids in Enterobacteriaceae. *J Gen Microbiol* **72**, 349–355 (1972).
22. Alperi, A. *et al.* A translocation motif in relaxase trwc specifically affects recruitment by its conjugative type IV Secretion System. *J. Bacteriol.* **195**, 4999–5006 (2013).
- 50 23. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred

- kilobases. *Nat Methods* **6**, 343–345 (2009).
24. Draper, O., César, C. E., Machón, C., De La Cruz, F. & Llosa, M. Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16385–16390 (2005).
- 5 25. Llosa, M., Bolland, S. & de la Cruz, F. Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *J Mol Biol* **235**, 448–464.
26. Fernández-González, E. *et al.* Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. *J. Bacteriol.* **193**, 6257–6265 (2011).
- 10 27. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Res* **44**, 4243–4251 (2016).
28. Trokter, M. & Waksman, G. Translocation through the Conjugative Type IV Secretion System Requires Unfolding of Its Protein Substrate. (2018) doi:10.1128/JB.00615-17.
29. Akeda, Y. & Galán, J. E. Chaperone release and unfolding of substrates in type III secretion. *Nature* **437**, 911–915 (2005).
- 15 30. Baharoglu, Z., Bikard, D. & Mazel, D. Conjugative DNA Transfer Induces the Bacterial SOS Response and Promotes Antibiotic Resistance Development through Integron Activation. *PLoS Genet.* **6**, e1001165 (2010).
31. Reytrat, J. M., Pelicic, V., Gicquel, B. & Rappuoli, R. Counterselectable markers: Untapped tools for bacterial genetics and pathogenesis. *Infection and Immunity* vol. 66 4011–4017 (1998).
- 20 32. Fernández-López, R. *et al.* Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiology Reviews* vol. 30 942–966 (2006).
- 25 33. Encinas, D. *et al.* Plasmid Conjugation from Proteobacteria as Evidence for the Origin of Xenologous Genes in Cyanobacteria. (2014) doi:10.1128/JB.01464-13.
34. Waller, M. C., Bober, J. R., Nair, N. U. & Beisel, C. L. Toward a genetic tool development pipeline for host-associated bacteria. *Current Opinion in Microbiology* vol. 38 156–164 (2017).
- 30 35. Vergunst, A. C. *et al.* VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. *Science (80-. )*. **290**, 979–982 (2000).
36. Cui, L. *et al.* A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nat. Commun.* **9**, 1–10 (2018).
37. Llosa, M., Bolland, S., Grandoso, G. & de la Cruz, F. Conjugation-independent, site-specific recombination at the oriT of the IncW plasmid R388 mediated by TrwC. *J Bacteriol* **176**, 3210–3217 (1994).
- 35 38. Vieira, J. & Messing, J. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259–268 (1982).
- 40 39. Norrander, J., Kempe, T. & Messing, J. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**, 101–106 (1983).

**REIVINDICACIONES**

1. Proteína de fusión que comprende una proteína relaxasa unida a una proteína CRISPR-Cas, donde la proteína relaxasa es la relaxasa TrwC y la proteína CRISPR-Cas es la proteína Cas12a, y donde la endonucleasa Cas12a se une al extremo C-terminal de la relaxasa TrwC.
- 5
2. Proteína de fusión según la reivindicación 1 que comprende la secuencia aminoacídica SEQ ID NO: 1.
- 10
3. Secuencia nucleotídica que codifica la proteína de fusión según cualquiera de las reivindicaciones 1 o 2.
4. Secuencia nucleotídica según la reivindicación 3, que comprende una secuencia nucleotídica según la SEQ ID NO: 2.
- 15
5. Vector de expresión que comprende la secuencia nucleotídica según cualquiera de las reivindicaciones 3 o 4.
- 20
6. Célula que comprende una proteína según cualquiera de las reivindicaciones 1 o 2 y/o una secuencia nucleotídica según cualquiera de las reivindicaciones 3 o 4, y/o un vector según la reivindicación 5.
7. Sistema CRISPR/Cas que comprende la proteína de fusión según cualquiera de las reivindicaciones 1 o 2, donde dicho sistema comprende:
- 25
- una secuencia nucleotídica que codifica para el ARN guía,
  - una proteína de fusión relaxasa-Cas según cualquiera de las reivindicaciones 1 o 2, donde la proteína relaxasa-Cas se transloca a una célula receptora a través de un sistema de secreción bacteriano tipo IV, donde el sistema de secreción bacteriano tipo
- 30
- IV es el plásmido conjugativo R388.
8. Sistema CRISPR/Cas según la reivindicación 7, donde la proteína relaxasa-Cas en la célula receptora se une a la secuencia nucleotídica que codifica para el ARN guía donde hibrida con el ADN diana.

35

9. Sistema CRISPR/Cas según cualquiera de las reivindicaciones 7 u 8, donde la proteína relaxasa-Cas reconoce una secuencia oriT de un vector movilizable que comprende la secuencia nucleotídica que codifica para el ARN guía y transloca dicha molécula de ADN unida a la proteína relaxasa-Cas a través del sistema de secreción bacteriano tipo IV a la célula diana donde el ARN guía hibrida con el ADN diana.
- 5
10. Sistema CRISPR/Cas según cualquiera de las reivindicaciones 7 a 9, que comprende la secuencia aminoacídica SEQ ID NO: 1.
- 10 11. Sistema CRISPR/Cas según cualquiera de las reivindicaciones 7 a 10, donde la célula diana es una célula procarionota.
12. Uso de una proteína de fusión según cualquiera de las reivindicaciones 1 o 2, o un sistema CRISPR/Cas según cualquiera de las reivindicaciones 7 a 11 para la translocación de endonucleasas y/o endonucleasas unidas a moléculas de ADN, a células diana a través del sistema de secreción bacteriano tipo IV.
- 15
13. Uso de una proteína de fusión según cualquiera de las reivindicaciones 1 o 2, y/o una secuencia nucleotídica según cualquiera de las reivindicaciones 3 o 4, y/o un vector según la reivindicación 5, y/o un sistema CRISPR/Cas según cualquiera de las reivindicaciones 7 a 11 para la modificación genética de las células diana.
- 20

FIG. 1

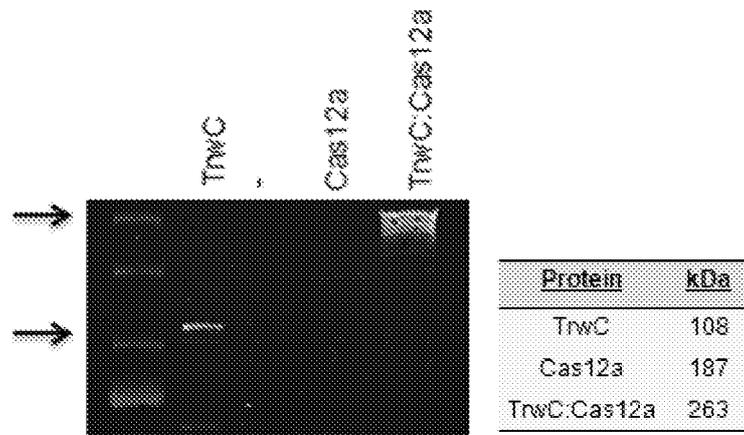


FIG. 2

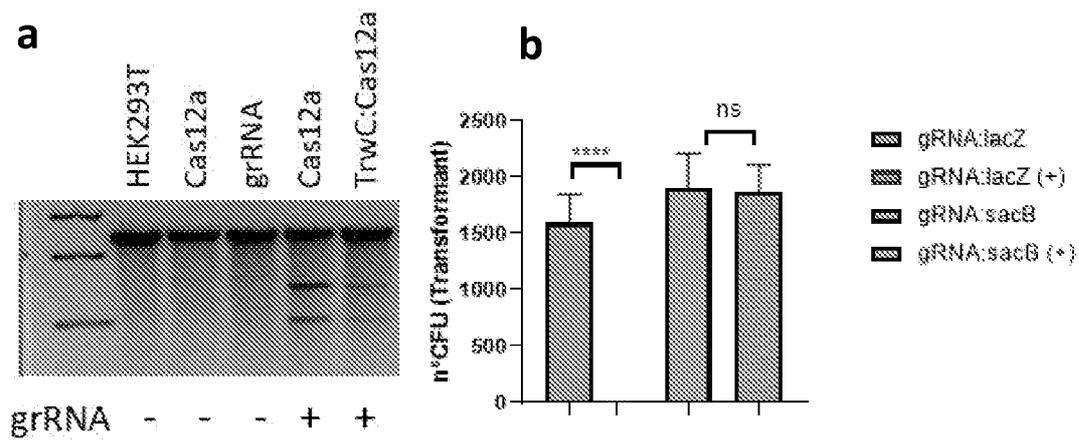


FIG. 3

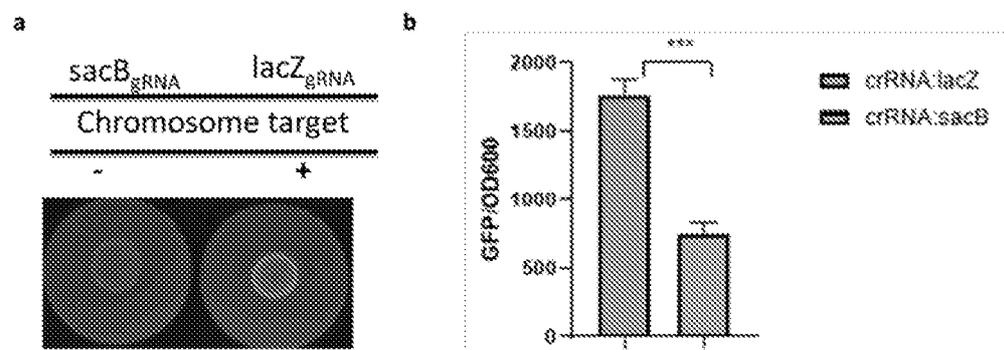


FIG.4

