(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **3 082 424**

(21) Número de solicitud: 202530287

(51) Int. CI.:

C12N 9/16 (2006.01)
C12N 15/09 (2006.01)

(12) SOLICITUD DE PATENTE A1

(22) Fecha de presentación:

07.04.2025

(43) Fecha de publicación de la solicitud:

22.10.2026

(71) Solicitantes:

UNIVERSIDAD DE ALICANTE (100,00%)
CARRETERA SAN VICENTE DEL RASPEIG, S/N
03690 SAN VICENTE DEL RASPEIG (Alicante) ES

(72) Inventor/es:

BAQUEDANO MOZOS, Ignacio;
ESPINOSA MANZANO, Javier;
ROSSELLI, Riccardo;
MARCO GUZMÁN, Noemí y
MARTÍNEZ MOJICA, Francisco Juan

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 ENDONUCLEASA Cas12a Y SISTEMA CRISPR-Cas12a ASOCIADO**

(57) Resumen:

La presente invención se refiere a una proteína endonucleasa Casl2a ('AlCas12a') que comprende una secuencia aminoacídica según SEQ ID NO: 2, a la secuencia nucleotídica que codifica dicha proteína, al vector de expresión que comprende dicha secuencia, a la célula que comprende dicha proteína, secuencia, y/o vector, al sistema CRISPR-Cas12a ('CRISPR-AlCas12a') que comprende dicha proteína, al método para editar un genoma usando dicha proteína, y a los usos asociados de la misma para ingeniería genética, modificación epigenética, y producción de antibacterianos. El sistema CRISPR-AlCas12a es funcional en presencia de Mg^{2+}, llevando a cabo interferencia guiada por crRNA dependiente de PAM (consenso 5'-TTN-3') contra moléculas diana de ADN bicatenario (ADNbc) en un amplio rango de temperaturas (de 20 ºC a 45 ºC). Además, la proteína AlCas12a puede realizar cortes tanto en trans como independientes de crRNA en ADN monocatenario (ADNcs) no diana, mientras que no muestra actividad ARNasa, siendo incapaz de procesar precursores de los crRNAs. La herramienta CRISPR-AlCas12a se aplicó con éxito a la selección de mutantes de *E. coli* en experimentos de recombinación genética. AlCas12a expresada de forma heteróloga en células de *E. coli*, disminuye su susceptibilidad a infección por virus de ADNcs o ADNbc en ausencia de un crRNA guía.



FIG. 1

ES 3 082 424 A1

DESCRIPCIÓN

PROTEÍNA ENDONUCLEASA Cas12a Y SISTEMA CRISPR-Cas12a ASOCIADO

CAMPO DE LA INVENCIÓN

La presente invención se enmarca en el campo de la ingeniería genética. Más concretamente, el objeto de la invención se refiere a una nueva proteína endonucleasa Cas12a y a un sistema CRISPR-Cas que comprende dicha proteína para la edición genética en células, la degradación inespecífica de secuencias de ADN de cadena sencilla y la producción de antibacterianos.

ANTECEDENTES DE LA INVENCIÓN

Los sistemas CRISPR-Cas se han identificado en la mayoría de las arqueas y aproximadamente la mitad de los genomas de bacterias como mecanismos de defensa contra la infección por ácidos nucleicos exógenos; esto es, plásmidos o virus. Cada sistema está constituido por una o varias agrupaciones de repeticiones de secuencias de ADN denominadas CRISPR (del inglés *Clustered Regularly Interspaced Short Palindromic Repeats*) y un conjunto de genes que codifican las proteínas Cas (del inglés *CRISPR associated*), cuya presencia dará lugar a las endonucleasas encargadas de cortar y degradar el ácido nucleico exógeno. Dichas repeticiones se encuentran regularmente espaciadas dentro de cada agrupación por secuencias no reiteradas denominadas ‘espaciadores’, al menos algunas de las cuales derivan de fragmentos genéticos de origen extracromosómico. Adyacente a cada agrupación de repeticiones-espaciadores hay una secuencia denominada ‘líder’ donde se localiza el promotor responsable de la transcripción de dicha agrupación en un ARN precursor (pre-crRNA) que abarca la totalidad de la misma. Este pre-crRNA es procesado mediante una ribonucleasa (RNAsa) dando lugar a crRNA maduros, cada uno de los cuales contiene un único espaciador. Cada uno de estos crRNA, comúnmente denominado ‘ARN guía’, aparea con su secuencia complementaria en la molécula de ácido nucleico exógeno, activando con ello a una proteína Cas específica que produce la degradación de dicho ácido nucleico, protegiendo así a la célula huésped de la infección.

Más allá de su función como sistema natural de inmunidad adaptativa en organismos procariotas, los sistemas CRISPR-Cas se han convertido en unas de las herramientas de

edición genética más poderosas en los campos de la biología, la biomedicina y la biotecnología. En concreto, permiten el silenciamiento o la eliminación de genes, mutagénesis, y correcciones de secuencias específicas del genoma de cualquier célula de una manera fácil, rápida, y altamente precisa. Entre sus numerosas aplicaciones destacan el diagnóstico y tratamiento de enfermedades y la producción de antimicrobianos específicos de secuencia.

Si bien los sistemas CRISPR-Cas9 son los más populares, existen otros que están despertando un gran interés por su aplicación en edición génica y diagnóstico de enfermedades. Es el caso de los sistemas CRISPR-Cas de Clase 2 - Tipo V - subtipo A (de ahora en adelante, sistemas CRISPR-Cas12a). Por ejemplo, la patente estadounidense US 10648020 B2 divulga el uso de sistemas CRISPR-Cas12a procedentes de *Acidaminococcus* sp., *Lachnospiracaea bacterium*, o *Francisella novicida* en la identificación de secuencias de ADN. Por su parte, la solicitud de patente CA 3149273 A1 divulga una variedad de sistemas CRISPR-Cas12a artificiales basados en los procedentes de *Acidaminococcus* sp., *Lachnospiracaea bacterium*, *Francisella novicida* y *Moraxella bovoculi,* así como su uso en edición genética de plantas.

De acuerdo con las caracterizaciones bioquímicas realizadas hasta la fecha, los sistemas CRISPR-Cas12a están constituidos por una agrupación de un número variable de repeticiones CRISPR-espaciadores y un grupo de genes *cas* que codifican las proteínas Cas1, Cas2 y Cas4, las cuales participan en la adquisición de nuevos espaciadores, así como Cas12a, cuyo gen es empleado como marcador genético distintivo de este subtipo de sistemas CRISPR-Cas. Los ARN guía (crRNA) son generados a partir del procesamiento de los pre-cRNAs por la actividad ARNasa de Cas12a. Cada uno de estos crRNA, que permanece unido a una molécula Cas12a formando el complejo de vigilancia, aparea con su secuencia complementaria en la molécula diana de ADN bicatenario exógeno a degradar. Este apareamiento requiere la unión previa del complejo de vigilancia a través de Cas12a a unas determinadas secuencias adyacentes a los proto-espaciadores diana, conocidas como PAM (*Protospacer Adjacent Motif*), que difieren de la secuencia adyacente a los espaciadores en la posición correspondiente dentro de la agrupación CRISPR, evitando así la autoinmunidad. La hibridación crRNA:diana provoca la activación del sitio desoxirribonucleasa dependiente de metales en el domino RuvC de Cas12a. Este sitio cataliza una rotura en cada una de las hebras de ADN, generando un producto de digestión con extremos sobresalientes de cadena sencilla (“corte en cis”). Algunas proteínas Cas12a mantienen la actividad nucleasa después del corte en cis, conduciendo

a la degradación colateral, de forma indiscriminada, de ADN no diana tanto bicatenario (ADNbc) como de cadena sencilla (ADNcs) (“corte en trans”). Diversos ortólogos de Cas12a se han empleado con éxito para la edición de genomas basándose en su actividad de corte en cis [SENTHILNATHAN, R., et al. *An update on CRISPR-Cas12 as a versatile tool in genome editing*. Mol. Biol. Rep., 2023, vol. 50, no 3, p. 2865-2881], y la actividad de corte en trans de ADNcs se ha explotado para diagnóstico molecular [LEUNG, R. K-K., et al. *CRISPR-Cas12-based nucleic acids detection systems*. Methods, 2022, vol. 203, p. 276-281].

Además de los cortes en cis y trans dependientes de guía, ensayos de digestión *in vitro* han demostrado que algunas proteínas Cas12a catalizan la degradación metal-dependiente de ADN en ausencia de crRNA [SUNDARESAN, R., et al. *RNA-independent DNA cleavage activities of Cas9 and Cas12a*. Cell Rep., 2017, vol. 21, no 13, p. 3728-3739]. Este corte independiente de crRNA sugiere la posibilidad de que la activación de estos ortólogos de Cas12a en la célula podría resultar en una inmunidad inespecífica, debido a la degradación de cualquier ADN invasor, sin importar su secuencia. De forma similar al corte en trans del propio genoma, esta actividad podría desencadenar una infección abortiva. Sin embargo, la degradación *in vivo* de ADN independiente de crRNA, y la contribución de esta actividad a la inmunidad, no se ha comprobado hasta ahora para ninguna proteína de la familia Cas12a.

Entre las limitaciones de la utilización de las herramientas CRISPR-Cas12a disponibles destacan el requerimiento de un motivo PAM, que restringe el número de dianas elegibles, y el gran tamaño de la proteína (entre 1206 y 1373 aminoácidos), el cual dificulta tanto su administración directa a células como la introducción del ADN codificante. Se requiere, por tanto, la identificación y caracterización bioquímica y funcional de proteínas Cas12a alternativas de menor tamaño y con motivos PAM poco restrictivos.

DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

La presente invención está orientada a resolver las limitaciones expuestas anteriormente mediante una nueva proteína endonucleasa Cas12a con un tamaño inferior al promedio de la familia (1226 aminoácidos) y un motivo PAM con una extensión mínima (2nts, TTN), llamada de aquí en adelante AlCas12a, o proteína de la invención. Esta proteína es apta para su uso en diversas herramientas de biología molecular, tanto para realizar cortes específicos de secuencia dependientes de guía, equivalentes a las implementadas con

otras endonucleasas Cas12a, como para la degradación inespecífica (no guiada) de moléculas de ADN.

La presente invención soluciona el problema del estado de la técnica expuesto en la sección anterior al proporcionar una proteína endonucleasa Cas12a con un motivo PAM poco restrictivo, y con un tamaño tal que facilita su administración tanto a células procariotas como eucariotas mediante vectores comúnmente empleados en biotecnología y biomedicina; esto es, plásmidos o bacteriófagos, en el caso de las bacterias, y lentivirus, virus adenoasociados (AAV, por sus siglas en inglés), o similar, para células de mamíferos.

Así pues, en un primer aspecto, la presente invención se refiere a una proteína endonucleasa Cas12a que comprende una secuencia aminoacídica con al menos un 70% de identidad de secuencia con SEQ ID NO: 2. En modos de realización, la proteína endonucleasa AlCas12a comprende una secuencia aminoacídica con al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, o al menos un 99% de identidad de secuencia con SEQ ID NO: 2. En un modo particular de realización, la proteína Cas12a comprende una secuencia aminoacídica con un 100% de identidad de secuencia con SEQ ID NO: 2.

En el ámbito de interpretación de la presente invención, el término “identidad de secuencia” se entenderá como el grado de similitud entre dos secuencias nucleotídicas o aminoacídicas, expresado a modo de porcentaje, que se obtiene al alinear dichas secuencias. Este dependerá del número de nucleótidos o residuos comunes entre las secuencias alineadas. Se determina mediante programas bioinformáticos bien establecidos en el estado de la técnica, tales como BLAST (del inglés, *Basic Local Alignment Search Tool*) o FASTA (EMBL-EBI). Se considerarán también dentro de la presente invención aquellas secuencias análogas, derivadas o equivalentes a la secuencia aminoacídica de la invención (por ejemplo, SEQ ID NO: 2) que comprendan al menos un residuo de aminoácido alterado por una inserción, sustitución, deleción, o modificación química de un aminoácido respecto a la secuencia aminoacídica de la proteína de la presente invención.

En una realización particular, la proteína de la presente invención comprende una secuencia aminoacídica según SEQ ID NO: 5. En otra realización particular, la secuencia aminoacídica comprende una inserción de al menos un aminoácido. En el ámbito de

interpretación de la presente invención, se entenderá por “inserción” cualquier tipo de mutación en la secuencia aminoacídica de la proteína de la presente invención que implique la adición de uno o más aminoácidos. La secuencia aminoacídica identificada como SEQ ID NO: 5 comprende una inserción de 19 aminoácidos tras el primer aminoácido de la secuencia identificada como SEQ ID NO: 2, donde dicha inserción incluye una cola de 6 histidinas para facilitar la purificación de la proteína de la presente invención (AlCas12a-His; SEQ ID NO: 5). En el contexto de la invención, el experto en la materia entiende que se podría usar igualmente cualquier etiqueta peptídica sobradamente conocida que no interfiera con la funcionalidad de la proteína. Ejemplos de etiquetas peptídicas incluyen:

- Avitag™ (GLNDIFEAQKIEWHE, SEQ ID NO: 13)
- Etiqueta Calmodulina (KRRWKKNFIAVSAANRFKKKISSSGAL, SEQ ID NO: 14)
- Etiqueta FLAG (DYKDDDDK, SEQ ID NO: 15)
- Etiqueta HA (YPYDVPDYA, SEQ ID NO: 16)
- Etiqueta Myc (EQKLISEDL, SEQ ID NO: 17)
- Etiqueta S (KETAAAAKFERQHMDS, SEQ ID NO: 18)
- Etiqueta SBP (MDEKTTGWRGHVVEGLAGELEQLRARLEHHPQGQREP, SEQ ID NO: 19)
- Softag 1 (SLAELLNAGLGGS, SEQ ID NO: 20)
- Softag 3 (TQDPSRVG, SEQ ID NO: 21)
- Etiqueta V5 (GKPIPNPLLGLDST, SEQ ID NO: 22)
- Strep-tag II (WSHPQFEK, SEQ ID NO: 23)

En modos de realización de la invención, la secuencia aminoacídica comprende una sustitución de al menos un aminoácido. En el ámbito de interpretación de la presente invención, se entenderá por “sustitución” cualquier tipo de mutación en la secuencia aminoacídica de la proteína de la presente invención que implique el reemplazo de uno o más aminoácidos. En modos particulares de realización, esta sustitución de al menos un aminoácido puede afectar, de manera controlada, a la funcionalidad de la proteína Cas12a de la presente invención. Así, en un modo de realización de la invención, la proteína comprende:

- una mutación D831A, preferiblemente donde la proteína comprende una secuencia aminoacídica según SEQ ID NO: 8, y/o
- una mutación E921A, preferiblemente donde la proteína comprende una secuencia aminoacídica según SEQ ID NO: 10, y/o
- una mutación D1183A, preferiblemente donde la proteína comprende una secuencia aminoacídica según SEQ ID NO: 12.

De acuerdo con modos de realización de la invención, la proteína Cas12a puede contener, de entre estas tres mutaciones: solo la mutación D831A; solo la mutación E921A; solo la mutación D1183A; las mutaciones D831A y E921A, pero no la D1183A; las mutaciones D831A y D1183A, pero no la E921A; las mutaciones y E921A y D1183A, pero no la D831A; las tres mutaciones D831A, E921A y D1183A. Estas mutaciones pueden estar incorporadas en una secuencia de aminoácidos con una identidad de secuencia de al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, o al menos un 99% de identidad de secuencia con SEQ ID NO: 2. En un modo de realización particular, la proteína de la presente invención comprende una secuencia aminoacídica según SEQ ID NO: 8. La secuencia aminoacídica identificada como SEQ ID NO: 8 (dAlCas12a-D831A) se corresponde con la secuencia identificada como SEQ ID NO: 2 pero incorporando la mutación D831A. En un modo de realización particular, la proteína de la presente invención comprende una secuencia aminoacídica según SEQ ID NO: 10. La secuencia aminoacídica identificada como SEQ ID NO: 10 (dAlCas12a-E921A) se corresponde con la secuencia identificada como SEQ ID NO: 2 pero incorporando la mutación E921A. En un modo de realización particular, la proteína de la presente invención comprende una secuencia aminoacídica según SEQ ID NO: 12. La secuencia aminoacídica identificada como SEQ ID NO: 12 (dAlCas12a-D1183A) se corresponde con la secuencia identificada como SEQ ID NO: 2 pero incorporando la mutación D1183A.

El interés en este tipo de mutaciones se deriva de que dichas modificaciones producen una proteína Cas12a con una pérdida total o parcial de actividad ADNasa (dCas12a). Por lo tanto, estos mutantes de AlCas12a sin actividad de corte de ADNbc (dAlCas12a-D831A, dAlCas12a-E921A y dAlCas12a-D1183A) se podrían fusionar con distintos dominios funcionales para implementar herramientas CRISPR-dAlCas12a capaces de controlar la transcripción génica y modificar o marcar ADN en lugares específicos, sin cortar su secuencia, tanto *in vivo* como *in vitro*.

En un segundo aspecto, la presente invención se refiere a una secuencia nucleotídica que codifica una proteína según el primer aspecto de la presente invención (de aquí en adelante, “secuencia nucleotídica de la presente invención”). En el ámbito de interpretación de la presente invención, se entenderá por “secuencia nucleotídica que codifica la proteína de la presente invención” cualquier secuencia de nucleótidos que, bajo un control de expresión adecuado, sea capaz de transcribir y traducir la secuencia aminoacídica de la proteína de la presente invención. El término "ácido nucleico", tal como se utiliza en el presente documento, se refiere a un polímero que tiene dos o más moléculas de desoxirribonucleótido, ribonucleótido o análogos de nucleótidos, así como moléculas que son estructuralmente similares a un ácido nucleico nativo, pero difieren del ácido nucleico nativo (por ejemplo, a través de la modificación química) en una o más de la columna vertebral del ácido nucleico (por ejemplo, fosfato en ácidos nucleicos nativos), azúcar de ácido nucleico (por ejemplo, desoxirribosa en el ADN nativo y ribosa en el ARN nativo) y base de ácido nucleico (por ejemplo, adenosina, citosina, guanina o timidina en ácidos nucleicos nativos).

En modos particulares de realización de la invención, la secuencia nucleotídica comprende una secuencia nucleotídica con al menos un 70% de identidad de secuencia con SEQ ID NO: 1, preferiblemente que comprende una secuencia nucleotídica con al menos un 70% de identidad de secuencia con SEQ ID NO: 3, o más preferiblemente que comprende una secuencia nucleotídica según SEQ ID NO: 4. En modos de realización, la secuencia nucleotídica comprende una secuencia nucleotídica con al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, o al menos un 99 de identidad de secuencia con SEQ ID NO: 1.

En una realización preferente, la secuencia nucleotídica de la presente invención comprende o es la secuencia nucleotídica identificada como SEQ ID NO: 1 (*alcas12a*). En otra realización preferente, la secuencia nucleotídica de la presente invención comprende o es la secuencia nucleotídica identificada como SEQ ID NO: 3 (*alcas12a-ecoli*). En otra realización preferente, la secuencia nucleotídica de la presente invención comprende o es la secuencia nucleotídica identificada como SEQ ID NO: 4 (*alcas12a-his*). En otra realización preferente, la secuencia nucleotídica de la presente invención comprende o es la secuencia nucleotídica identificada como SEQ ID NO: 7 (*dalcas12a*-D831A). En otra realización preferente, la secuencia nucleotídica de la presente invención comprende o es

la secuencia nucleotídica identificada como SEQ ID NO: 9 (*dalcas12a*-E921A). En otra realización preferente, la secuencia nucleotídica de la presente invención comprende o es la secuencia nucleotídica identificada como SEQ ID NO: 11 (*dalcas12a*-D1183A).

En un tercer aspecto, la presente invención se refiere a un vector de expresión que comprende la secuencia nucleotídica de la presente invención (de aquí en adelante, “vector de expresión de la presente invención”). En el ámbito de interpretación de la presente invención, se entenderá por “vector de expresión” cualquier molécula de ADN que pueda utilizarse como vehículo para transportar la secuencia nucleotídica de la presente invención al interior de una célula huésped. El vector de expresión de la presente invención puede comprender una molécula de ácido nucleico monocatenaria, bicatenaria o parcialmente bicatenaria; una molécula de ADN, ARN, o híbrida ADN:ARN. Ejemplos de vectores de expresión son los plásmidos y los bacteriófagos o fagos.

En una realización preferente, el vector de expresión de la presente invención comprende una secuencia nucleotídica según SEQ ID NO: 1. En otra realización preferente, el vector de expresión de la presente invención comprende una secuencia nucleotídica según SEQ ID NO: 3. En otra realización preferente, el vector de expresión de la presente invención comprende una secuencia nucleotídica según SEQ ID NO: 4. En otra realización preferente, el vector de expresión de la presente invención comprende una secuencia nucleotídica según SEQ ID NO: 7. En otra realización preferente, el vector de expresión de la presente invención comprende una secuencia nucleotídica según SEQ ID NO: 9. En otra realización preferente, el vector de expresión de la presente invención comprende una secuencia nucleotídica según SEQ ID NO: 11.

En un cuarto aspecto, la presente invención se refiere a una célula que comprende una proteína de acuerdo con la presente invención, y/o una secuencia nucleotídica de acuerdo con la presente invención, y/o un vector de expresión de acuerdo con la presente invención (de aquí en adelante, “célula de la presente invención”). En el ámbito de interpretación de la presente invención, se entenderá por “célula” cualquier unidad básica, estructural y funcional de un ser vivo susceptible a una o más de las siguientes alteraciones genéticas: transformación (absorción directa, incorporación y expresión de la secuencia nucleotídica de la presente invención), transfección o transducción (introducción de material genético externo mediante el vector de expresión de la invención), y translocación (introducción de la proteína de la presente invención en el interior del ribosoma). El término “célula”, o “célula huésped”, tal como se utiliza en el presente documento, se refiere a una célula en

la que se ha introducido un ácido nucleico de la invención, como un polinucleótido o un vector según la invención. Debe entenderse que tales términos se refieren no sólo a la célula en cuestión en particular, sino también a la progenie o progenie potencial de dicha célula. Debido a que ciertas modificaciones pueden ocurrir en generaciones sucesivas debido a mutaciones o influencias ambientales, dicha progenie puede, de hecho, no ser exactamente idéntica a la célula madre, pero aun así se incluyen dentro del alcance del término tal como se usa en este documento.

En un quinto aspecto, la presente invención se refiere a un sistema CRISPR-Cas que comprende un ARN guía y una proteína de acuerdo con el primer aspecto de la presente invención (de aquí en adelante, “sistema CRISPR-Cas de la presente invención”).

En el ámbito de interpretación de la presente invención, se entenderá por “sistema CRISPR-Cas” cualquier sistema que comprenda los elementos implicados en la expresión y/o actividad de los genes asociados a dicho sistema, incluyendo tanto la/s secuencia/s nucleotídica/s que se transcribe/n para generar el ARN guía como la secuencia nucleotídica que codifica la proteína de la presente invención.

Asimismo, se entenderá por “ARN guía”, cualquier construcción de ARN monocatenaria, bicatenaria o parcialmente bicatenaria que se asocie con la proteína de la presente invención y que comprenda una secuencia ribonucleotídica complementaria a una secuencia de ADN específica de una célula (“secuencia diana”); esto es, que forma enlaces de hidrógeno con las bases nitrogenadas de los nucleótidos de la secuencia diana. Preferiblemente, el ARN guía es un ARN monocatenario. En modos particulares de realización de la presente invención, el ARN guía es un ARN CRISPR (crRNA), o un pre-crRNA de un tamaño próximo a un crRNA. Típicamente, un crRNA posee, en su extremo 5’ (correspondiente con el *5’-handle*), una región invariable correspondiente con una repetición palindrómica corta derivada de un locus CRISPR (por ejemplo, del locus CRISPR-Cas12a) seguida, en el extremo 3’ del crRNA, de una secuencia del espaciador, que es complementaria de la secuencia diana. En modos particulares de realización, la longitud del extremo 5’ (región invariable) puede ser de entre 15-35 nucleótidos, de entre 16-30 nucleótidos, de entre 17-25 nucleótidos, o de entre 18-22 nucleótidos. En modos particulares de la invención, la longitud del extremo 5’ (región invariable) puede ser de 17, 18, 19, 20, 21 o 22 nucleótidos. En un modo preferido de realización, la longitud del extremo 5’ (región invariable) es de 19 nucleótidos. En modos particulares de realización, la longitud del extremo 3’ (secuencia del espaciador) puede ser de entre 15-35 nucleótidos,

de entre 18-30 nucleótidos, o de entre 20-25 nucleótidos. En modos particulares de la invención, la longitud del extremo 5' (región invariable) puede ser de 20, 21, 22, 23, 24, 25, o 26 nucleótidos. En un modo preferido de realización, la longitud del extremo 3' (secuencia del espaciador) es de 24 nucleótidos. Así, la longitud total del crRNA puede ser de entre 30-70 nucleótidos.

En modos de realización, la secuencia concreta del extremo 5' (región invariable) puede ser cualquier secuencia que permita la formación de una estructura en horquilla que permita la unión entre la proteína Cas12a y el crRNA (o RNA guía). En un modo de realización, el sistema CRISPR-Cas de la presente invención comprende un ARN guía con un extremo 5' con una secuencia nucleotídica según SEQ ID NO: 6.

Se incluye, en el ámbito de la presente invención, cualquier modificación en las moléculas de ARN que sirven como guía que puedan mejorar la eficacia y la especificidad de corte de ADN bicatenario (ADNbc) por proteínas Cas12a. Estas modificaciones son conocidas para un experto en la materia [JEDRZEJCZYK *et al.*, *CRISPR-Cas12a nucleases function with structurally engineered crRNAs: SynThetic trAcrRNA*, Sci Rep. 2022 Jul 16;12(1):12193].

En un sexto aspecto, la invención se refiere a un método para editar un genoma que incluye una secuencia nucleotídica diana, donde dicho método comprende una etapa de poner en contacto una secuencia nucleotídica diana con un sistema CRISPR-Cas según el quinto aspecto de la presente invención. Preferentemente, para la edición de secuencias nucleotídicas concretas de células eucariotas. La edición de genomas se basa en la actividad de corte en cis de la proteína [SENTHILNATHAN R, et al., *An update on CRISPR-Cas12 as a versatile tool in genome editing*, Mol Biol Rep 50: 2865–2881 (2023)].

En un séptimo aspecto, la invención se refiere a un uso de una proteína según el primer aspecto de la presente invención, y/o una secuencia nucleotídica según el segundo aspecto de la presente invención, y/o un vector de expresión según el tercer aspecto de la presente invención, y/o una célula según el cuarto aspecto de la presente invención, y/o un sistema CRISPR-Cas según el quinto aspecto de la presente invención para:

- la edición y/o modificación genética, y/o
- la degradación de moléculas de ADNcs en soluciones de ácidos nucleicos; y/o
- la producción de un agente antimicrobiano específico de secuencia.

Preferentemente, para la degradación de secuencias nucleotídicas concretas en células procariotas. En un modo particular de este séptimo aspecto de la invención, el agente antimicrobiano específico de secuencia es un agente antibacteriano, preferiblemente un agente antibacteriano frente a *Escherichia coli*. Esta aplicación particular se basa en la observación que muestra que la herramienta CRISPR-AlCas12a provoca muerte celular cuando se dirige al cromosoma de *E. coli*, teniendo por lo tanto aplicación como antimicrobiano específico de secuencia, al menos para esta especie bacteriana. Está dentro de las capacidades de un experto en la materia desarrollar herramientas antibacterianas equivalentes para especies bacterianas distintas de *E. coli*.

En la degradación de moléculas de ADNcs en soluciones de ácidos nucleicos, este aspecto aprovecha el uso de la proteína AlCas12 para la degradación de cualquier molécula de ADNcs como consecuencia de su capacidad para degradar ADNcs inespecíficos, aprovechando su actividad de corte en trans (desatada por la actividad guiada) o independiente de guía. Así, la actividad endonucleasa de ADNcs dependiente de $MgCl_2$ e independiente de crRNA se podría emplear para la degradación de moléculas de ADNcs en soluciones de ácidos nucleicos o para la detección de regiones desapareadas en moléculas de ADNbc.

En un octavo aspecto, la invención se refiere a un uso de una proteína según el primer aspecto de la presente invención, y/o una secuencia nucleotídica según el segundo aspecto de la presente invención, y/o un vector de expresión según el tercer aspecto de la presente invención, y/o una célula según el cuarto aspecto de la presente invención, y/o un sistema CRISPR-Cas según el quinto aspecto de la presente invención, donde la proteína (o proteína resultante) comprende al menos una mutación de entre D831A, E921A y D1183A, para:

- la regulación de expresión génica; y/o
- la visualización y/o marcaje de secuencias nucleotídicas concretas; y/o
- la edición y/o modificación genética.

Este aspecto de la invención aprovecha los mutantes de AlCas12a sin actividad de corte de ADNbc (dAlCas12a-D831A, dAlCas12a-E921A y dAlCas12a-D1183A), donde estos mutantes se pueden fusionar con distintos dominios funcionales para implementar herramientas CRISPR-dAlCas12a capaces de controlar la transcripción génica y modificar

o marcar ADN en lugares específicos, sin cortar su secuencia, tanto *in vivo* como *in vitro* [KHAN, S. and SALLARD, E., *Current and Prospective Applications of CRISPR-Cas12a in Pluricellular Organisms*. Mol Biotechnol 65: 196–205 (2023)].

En un noveno aspecto, la invención se refiere a un método para el diagnóstico *in vitro* de una enfermedad caracterizada por un marcador molecular en un sujeto, donde el método comprende una etapa de poner en contacto una muestra biológica del sujeto con una proteína según el primer aspecto de la presente invención, y/o una secuencia nucleotídica según el segundo aspecto de la presente invención, y/o un vector de expresión según el tercer aspecto de la presente invención, y/o una célula según el cuarto aspecto de la presente invención, y/o un sistema CRISPR-Cas según el quinto aspecto de la presente invención. En el ámbito de interpretación de la presente invención, se entenderá por “muestra biológica” cualquier material biológico aislado de un tejido o un fluido biológico de un sujeto. Por "muestra biológica" aislada de un tejido se entenderá cualquier material biológico procedente de una biopsia. Por "muestra biológica" aislada de un fluido biológico se entenderá cualquier material biológico procedente de sangre, plasma, suero, orina, líquido cefalorraquídeo, líquido pleural, líquido peritoneal y ascítico, líquido pericárdico, líquido amniótico, exudados, heces, esputo, saliva, o semen.

En este aspecto de la invención cabe destacar que el corte en trans de otras proteínas Cas12a se ha aprovechado para implementar métodos de diagnóstico molecular *in vitro*, para los que también se podría utilizar AlCas12a. Estos métodos de diagnóstico típicamente incluyen el uso de un reportero que se activa tras el corte en trans. Estos reporteros pueden tener una estructura de marcador e inhibidor, unidos por un espaciador de ADN monocatenario (ADNcs) de tal manera que cuando se produce el corte en trans, el inhibidor se separa del marcador (por ejemplo, un marcador fluorescente) y se activa la señal de dicho marcador. Sistemas reporteros son conocidos para el experto en la materia [ZEFENG MAO et al., *CRISPR/Cas12a-based technology: A powerful tool for biosensing in food safety*, Trends in Food Science & Technology, Volume 122, April 2022, Pages 211-222], [ROSS KA-KIT LEUNG et al., *CRISPR-Cas12-based nucleic acids detection systems*, Methods, Volume 203, July 2022, Pages 276-281].

En un décimo aspecto, la invención se refiere a una proteína según el primer aspecto de la presente invención, y/o una secuencia nucleotídica según el segundo aspecto de la presente invención, y/o un vector de expresión según el tercer aspecto de la presente

invención, y/o una célula según el cuarto aspecto de la presente invención, y/o un sistema CRISPR-Cas según el quinto aspecto de la presente invención para su uso en medicina.

En un undécimo aspecto, la invención se refiere a una proteína según el primer aspecto de la presente invención, y/o una secuencia nucleotídica según el segundo aspecto de la presente invención, y/o un vector de expresión según el tercer aspecto de la presente invención, y/o una célula según el cuarto aspecto de la presente invención, y/o un sistema CRISPR-Cas según el quinto aspecto de la presente invención para su uso en el tratamiento de una infección, preferiblemente de una infección bacteriana, más preferiblemente de una infección bacteriana donde la bacteria presenta resistencia a al menos un antibiótico. De manera general, se conoce el uso de sistemas CRISPR-Cas en el tratamiento de infecciones microbianas con especies que presentan resistencia a agentes antimicrobianos, particularmente el caso de bacterias que presentan resistencia a antibióticos [AHMED S.A. et al., *New frontiers in CRISPR: Addressing antimicrobial resistance with Cas9, Cas12, Cas13, and Cas14*, Heliyon, Vol. 11, Is. 2 e42013, 2025] [YIZHENG QIAN et al., *Application of CRISPR-Cas system in the diagnosis and therapy of ESKAPE infections*, Front. Cell. Infect. Microbiol., Volume 13 - 2023].

En un duodécimo aspecto, la invención se refiere a un método para reducir la susceptibilidad de una bacteria a infección por virus de ADN monocatenario (ADNcs) o bicatenario (ADNbc), donde el método comprende:

- proporcionar una bacteria que comprende una proteína según el primer aspecto de la presente invención, y/o una secuencia nucleotídica según el segundo aspecto de la presente invención; o
- transformar una bacteria con un vector de expresión según el tercer aspecto de la presente invención.

En este sentido, los métodos de transformación de una bacteria con material exógeno son sobradamente conocidos para el experto en la materia.

Todos los términos y modos de realización descritos anteriormente son aplicables a cualquier aspecto y modo de realización de la invención. De acuerdo con la presente invención, el término en singular “el”, “la”, “un”, “uno”, “una”, se refiere igualmente a su correspondiente en plural “los”, “las”, “unos”, “unas”, salvo que se desprenda del contexto que claramente el término se refiere a una especie en el singular. El término “comprende”

o “que comprende”, tal y como se usa en el presente documento, también describe “consiste en” o “que consiste en” de acuerdo con la práctica de patentes generalmente aceptada.

EJEMPLOS

La siguiente invención se describe por medio de los siguientes ejemplos, que deben interpretarse como meramente ilustrativos y no limitativos del alcance de la invención.

La presente invención consiste en el diseño de la herramienta CRISPR-AlCas12a que comprende un ARN guía (AlcrRNA) y una proteína, AlCas12a, derivada de un sistema CRISPR-Cas12a anotado en un metagenoma público, pero previamente inexplorado, y su uso para la digestión de secuencias específicas (digestión guiada por crRNA) de ADNbc con la finalidad de llevar a cabo edición genética o inactivar bacterias, entre otras aplicaciones.

Ejemplo 1: Predicción de la identidad y origen de la proteína de la invención

Mediante el análisis con el programa informático CRISPRCasFinder [COUVIN, D., et al. *CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins*. Nucleic acids research, 2018, vol. 46, no W1, p. W246-W251] se predijo, en un cóntigo de ~10Kb (véase Figura 1) de un metagenoma de las comunidades bacterianas de aguas residuales de la depuradora de Alicante, España (Replicate C1 metagenome dataset; IMG ID 3300027924; scaffold Ga0268731_1002260; MAESTRE-CARBALLA, L., et al. *Insights into the antibiotic resistance dissemination in a wastewater effluent microbiome: bacteria, viruses and vesicles matter*. Environ. Microbiol., 2019, vol. 21, no 12, p. 4582-4596), un locus CRISPR-Cas12a formado por una agrupación CRISPR con nueve repeticiones de 35 pb intercaladas con espaciadores de entre 25 pb y 28 pb, y ORFs de proteínas predichas como pertenecientes a las familias Cas1, Cas2, Cas4 y Cas12a (*alcas12a*: SEQ ID NO: 1; AlCas12a: SEQ ID NO: 2).

Esta última proteína se identifica en el gen (Ga0268731_10022604; IMG ID 3300027924) como una proteína hipotética de 1226 aminoácidos. En la base de datos IMG/MER, se indica que la proteína con mayor porcentaje de identidad (31,29 %) es el producto de un gen (Gene ID 2526218409) del genoma de la bacteria *Prevotella albensis* DSM 11370,

identificado por similitud de secuencia como *CRISPR-associated protein Cpf1 subtype PREFRAN*, sin que esta asignación a la familia de proteínas Cpf1 (actualmente Cas12a) haya sido validada experimentalmente. El análisis bioinformático realizado por los autores de esta invención reveló que la proteína funcionalmente caracterizada con el mayor porcentaje de identidad (44%) con la Cpf1 de *Prevotella albensis* DSM 11370 (1253 aminoácidos) es FnCas12a (Genbank protein ID WP_003040289.1; 1300 aminoácidos) de *Francisella tularensis subsp. novicida* U112 (Zetsche et al., 2015). Este porcentaje de identidad es insuficiente *per se* para establecer una asignación fehaciente a la familia Cas12a, la proteína efectora (responsable del corte de ADN diana) de los sistemas CRISPR-Cas12a, por lo que se requiere validación experimental.

La dirección de transcripción de la agrupación CRISPR de AlCas12a se dedujo a partir del alineamiento con Clustal Omega (EMBL-EBI) de la secuencia de sus repeticiones con las de proteínas Cas12a previamente caracterizadas; en concreto, FnCas12a, AsCas12a, LbCas12a, Lb2Cas12a, PdCas12a, MbCas12a y PbCas12a (véase Figura 2) [ZETSCHE, B., et al. *A survey of genome editing activity for 16 Cas12a orthologs*. The Keio journal of medicine, 2020, vol. 69, no 3, p. 59-65].

Por otra parte, la búsqueda mediante la herramienta informática Blastp de proteínas depositadas en la base de datos *nr* (del inglés “non-redundant protein sequences”) del NCBI con secuencias similares a la de AlCas12a, revelaron porcentajes de identidad muy bajos: el valor más alto fue de 32,37% para una posible proteína Cas12a de una bacteria no cultivada perteneciente a la especie *Alistipes* sp., mientras que la búsqueda con la misma herramienta en la base de datos de secuencias patentadas del NCBI (“pataa”) reveló porcentajes de identidad inferiores al 32%.

Las secuencias de aminoácidos de 188 proteínas Cas12a depositadas en la base de datos Uniprot (www.uniprot.org) se alinearon con la de AlCas12a utilizando MAFFT v.7.505 (tras ser recortadas para enriquecer las regiones conservadas utilizando BMGE v1.12, con las opciones -m BLOSUM30, -h 0.55). Los alineamientos de estas secuencias recortadas se utilizaron para calcular la topología filogenética general de Cas12a utilizando IQ-TREE v1.6.12, estableciendo las opciones -m LG -nt AUTO -bb 1000 -alrt 1000. Los resultados se visualizaron utilizando la herramienta en línea para mostrar árboles filogenéticos iTOL v5. La evaluación de las identidades recíprocas de secuencias se realizó mediante alineamientos por pares todo-vs-todo entre las proteínas Cas12a utilizando Diamond v2.0.13.151 blastp, con las opciones --matrix BLOSUM45 para las proteínas divergentes

y fijando el umbral de e-value en 0,001. Los valores de identidad se ordenaron según la ubicación de cada Cas12a en el árbol filogenético relativo y se visualizaron mediante el paquete R heatmap3. Los resultados confirmaron que AlCas12a está más estrechamente relacionada con los ortólogos de Cas12a codificados por microorganismos no cultivados pertenecientes al grupo Patescibacteria (véase Figura 3).

Para tratar de identificar el organismo portador del cóntigo del sistema CRISPR-AlCas12a, se compararon las secuencias de las proteínas codificadas por los siete genes aparentemente completos detectados en dicho fragmento (incluyendo las cuatro secuencias Cas) con la base de datos *nr* del NCBI. Con excepción de AlCas12a (28,9%), los mayores porcentajes de identidad (entre el 53% y el 85,1%) se encontraron con proteínas codificadas en el genoma de *Candidatus Saccharibacteria bacterium* del grupo Patescibacteria. Estos resultados sugieren una relación lejana entre el hospedador de CRISPR-AlCas12a, probablemente perteneciente al grupo de las Patescibaterias, y los organismos conocidos.

1.1 Caracterización in silico de la proteína AlCas12a

Análisis previos mediante cristalografía de rayos X y criomicroscopía electrónica de la estructura de proteínas Cas12a han revelado que adoptan una arquitectura bilobulada consistente en un lóbulo de reconocimiento alfa-helicoidal (Rec), formado por los dominios Rec1 y Rec2, y un lóbulo Nuc constituido por dominios Nuc, RuvC (responsable del corte de ADN), PI (interacción con PAM), BH (Bridge Helix) y WED (Wedge). Los dominos WED y RuvC están a su vez conformados por tres subdominios cada uno [JIANWEI, L., et al. *Structures of apo Cas12a and its complex with crRNA and DNA reveal the dynamics of ternary complex formation and target DNA cleavage*. Plos Biology, 2023, vol. 21, no 3, p. e3002023].

Nuestro análisis de la secuencia de AlCas12a con la herramienta InterPro [BLUM M, et al., *InterPro: the protein sequence classification resource in 2025*, Nucleic Acids Research. 2024] (https://www.ebi.ac.uk/interpro/) la clasificó tentativamente dentro de la familia Cas12a, prediciendo subdominios RuvCII y RuvCIII, así como dominios PI, Rec1 y Nuc, característicos de la familia Cas12a. Esté análisis localizó en la secuencia residuos equivalentes a los previamente identificados en proteínas de la familia Cas12a como esenciales para la actividad nucleasa de RuvC (D831 y E921) y alguno de los aminoácidos implicados en el procesamiento del pre-crRNA (K782), la unión a la secuencia PAM, al

proto-espaciador y al crRNA. Nuestro alineamiento mediante Clustal Omega de la secuencia de aminoácidos de AlCas12a con la de proteínas Cas12a con estructuras conocidas nos permitió predecir la localización aproximada en AlCas12 de los diferentes dominios de la familia, y el alineamiento con proteínas Cas12a funcionalmente caracterizadas nos permitió evaluar la conservación de residuos implicados en las distintas actividades que pueden llevar a cabo (véase Figura 4).

La baja identidad de la secuencia de AlCas12a con las de Cas12a validadas y la ausencia de algunos de los residuos identificados en ortólogos de la familia Cas12a como importantes para su actividad, en especial de aminoácidos esenciales para el procesamiento del crRNA (H843 y K852 en FnCas12a) y que están muy conservados en la familia [SWARTS, D.C., et al. *Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a*. Molecular cell, 2017, vol. 66, no 2, p. 221-233. e4; FONFARA, I., et al. *The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA*. Nature, 2016, vol. 532, no 7600, p. 517-521], no permiten establecer una asignación inequívoca de AlCas12a a la misma, requiriendo por lo tanto una validación experimental.

La familia Cas12a2 es un ejemplo ilustrativo de esta incertidumbre [BRAVO, J.P.K, et al. *RNA targeting unleashes indiscriminate nuclease activity of CRISPR–Cas12a2*. Nature, 2023, vol. 613, no 7944, p. 582-587]. Aunque en su estructura se aprecian dominios Rec1, Rec2, WED, PI y RuvC con dos residuos catalíticos esenciales para su actividad ADNasa, Cas12a2 corta de forma específica (corte en cis) dianas de ARN en lugar de ADNbc, y tras el corte en cis cataliza cortes inespecíficos (en trans) en ARNcs, ADNcs y ADNbc, características que la distinguen de la familia Cas12a. Además, son numerosas las proteínas (e.g., AbCas12a, C6Cas12a, LpCas12a, SaCas12a, PsCas12a, PbCas12a, PrCas12a, CsbCas12a, BhCas12; SsCas12a, Lb3Cas12a, BpCas12a, PdCas12a) asignadas tentativamente a la familia Cas12a basándose en similitud de secuencia y predicción de dominios característicos de la familia que no han mostrado actividad en ensayos de digestión de ADN diana [CHEN, P., et al. *A Cas12a ortholog with stringent PAM recognition followed by low off-target editing rates for genome editing*. Genome Biology, 2020, vol. 21, p. 1-13].

1.2 Caracterización in silico de los espaciadores CRISPR-AlCas12a

Las secuencias coincidentes con espaciadores CRISPR del sistema CRISPR-AlCas12a (es decir, proto-espaciadores), y que por tanto podrían ejercer como dianas de la interferencia, se exploraron con la aplicación informática CRISPRTarget (http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html), y mediante búsquedas Blastn en la base de datos nr/nt del NCBI, así como en metagenomas disponibles en la página web *Integrated Microbial Genomes & Microbiomes* (https://img.jgi.doe.gov/). Mediante estos análisis se detectaron posibles proto-espaciadores en genomas virales (véase Tabla 1), apoyando la posibilidad de que CRISPR-AlCas12a ejerza como sistema inmune adaptativo.

Tabla 1. Proto-espaciadores del sistema CRISPR-AlCas12a identificados en bases de datos de secuencias.

Secuencia del espaciador (5' -> 3')	**Secuencia PAM_proto-espaciador (5' -> 3')[a]**	**Base de datos**	**Origen**
TGGTACTTGGCAATATACTCCAGTAT (SEQ ID NO: 24)	CTTA_TGGTACTTGGCAATgTACTCCAGTAT (SEQ ID NO: 25)	IMGVR	Comunidad vírica
	CTTA_TGGTACTTGGCAATgTACTCCAGTAT (SEQ ID NO: 25)	IMGVR	Comunidad microbiana
	CTTA_TGGTACTTGGCAATgTACTCCAGTAT (SEQ ID NO: 25)	ENV/env_nt	Metagenoma intestinal
	CTTA_TGGTACTTGGCAATgTACTCCAGTAT (SEQ ID NO: 25)	HuVirDB	Metagenoma intestinal
TCAAGGCATACCCTGATAACTTTAGAG (SEQ ID NO: 26)	TTTA_TCAAGGCATACCCTGATAACTTTAGAG (SEQ ID NO: 27)	IMGVR_UViGs	Comunidad vírica

[a] Los nucleótidos que difieren de los de la secuencia espaciadora se indican con minúsculas.

Ejemplo 2: Comprobación funcional y caracterización bioquímica de CRISPR-AlCas12a

Dada la imposibilidad de identificar la bacteria portadora de CRISPR-AlCas12a, la evaluación *in vivo* de su actividad de interferencia se realizó en *E. coli*.

Para investigar la capacidad de realizar cortes específicos en secuencias diana de ADNbc por AlCas12a e identificar al mismo tiempo las secuencias PAM necesarias, generamos los plásmidos pMML30 (véase Figura 5), portador del gen *alcas12a* y de una agrupación CRISPR, y pMML31 con la misma agrupación, pero sin el gen:

- El plásmido pMML31 se generó insertando en pBAD33 (resistente a cloranfenicol) una agrupación CRISPR flanqueada por un promotor constitutivo (Part:BBa_J23101; colección BioBricks) y la secuencia terminadora BBa_B1006 (adquirida como bloque G de NZYTech), que había sido sintetizada y clonada en un plásmido derivado de pUC57 (pUC57-array) por NZYTech. La agrupación CRISPR estaba formada por dos repeticiones AlCas12a, de 35 pb de longitud cada una, separadas por un espaciador de 28 pb (5'-CTGTATATATCGTGCGAAAAAGGATGGATA-3', SEQ ID NO: 28) dirigido al plásmido pSEVA431.
- Para generar el plásmido pMML30, el gen *alcas12a* optimizado en codones de *E. coli* fue sintetizado químicamente y clonado por NZYtech en el vector pUC57 (pUC57-Alcas12a). Este gen fue amplificado por PCR e insertado en pMML31 entre el promotor de arabinosa y el terminador rrnB.

2.1 Cribado de PAMs

Células electrocompetentes de *E. coli* BW27783 portadoras de pMML30 o pMML31 fueron transformadas con una genoteca de plásmidos derivados de pSEVA431 (resistencia a espectinomicina) conteniendo una misma secuencia coincidente con la del espaciador pero adyacente a distintas secuencias en la región PAM (cribado de PAMs). En estos experimentos de cribado de PAMs se probaron cuatro posiciones (numeradas de -4 a -1, siendo -1 la contigua a la secuencia diana) en la región PAM canónica de Cas12a (5’ de la secuencia correspondiente a la de espaciador en el crRNA), incluyendo todas las combinaciones posibles de nucleótidos. Los amplicones obtenidos por PCR de la región PAM a partir de un conjunto de colonias transformantes resistentes a espectinomicina se secuenciaron mediante secuenciación masiva (HTS) y se comparó la incidencia de cada tetranucleótido en las posiciones -4 a -1 en presencia de AlCas12a con la estimada en ausencia de la proteína.

Se observó una infrarrepresentación de lecturas con secuencias particulares en la región PAM cuando se expresaba AlCas12a con respecto al control sin la proteína, lo que

demuestra tanto su actividad nucleasa sobre ADNbc, específica de secuencia y dependiente de PAM, como la generación de un ARN guía funcional en la célula.

Las secuencias PAM necesarias para un corte eficaz, deducidas a partir de las lecturas con secuencias en la región PAM subrepresentadas, se ajustaban al consenso 5'-BTTV-3' (B = C o G o T; V = A o C o G) (véase Figura 6A). La funcionalidad de este motivo PAM también está respaldada por la observación de que en la región PAM de los proto-espaciadores víricos correspondientes a los dos espaciadores de la agrupación asociada a AlCas12a para los que se detectaron proto-espaciadores, se encuentran las secuencias 5'-CTTA-3' y 5'-TTTA-3' (véase la Tabla 1).

A continuación, validamos *in vivo* un subconjunto de secuencias PAM. Para ello, células de *E. coli* BW27783 portadoras de pMML30 o pMML31, fueron transformadas con derivados de pSEVA431 (plásmidos pMML33 a pMML38) con una secuencia diana adyacente a las secuencias 5'-CTTG-3', 5'-ATTG-3', 5'-ATTA-3', 5'-TTTG-3', 5'-CTTC-3', 5'-TTTT-3' o 5'-AATC-3' en la región PAM. En consonancia con la secuencia consenso inferida a partir del cribado de PAMs, cuando el nucleótido timina estaba presente en las posiciones -2 y -3 la eficiencia de transformación en presencia de AlCas12a disminuyó varios órdenes de magnitud en comparación con experimentos equivalentes sin *alcas12a* (véase Figura 6B). Además, no se observó interferencia contra el plásmido diana en ausencia de timina en la posición -3 (véase la eficiencia de transformación con 5'-AATC-3' en la Figura 6B). Sin embargo, en contraste con los resultados del cribado de PAMs, donde la frecuencia de transformación de plásmidos diana con timina en la posición PAM -1 o adenina en la posición PAM -4 no se había visto disminuida en presencia de AlCas12, tanto 5'-TTTT-3' como 5'-ATTG-3' y 5'-ATTA-3' permitieron la escisión de la diana.

En conjunto, nuestros resultados indican que mientras que la timina en las posiciones -2 y -3 es requerida por AlCas12a para la escisión eficiente de la diana, cualquier nucleótido en -1 y -4 es tolerado, apoyando una PAM consenso 5'-TTN-3'. Este motivo es de los menos restrictivos identificados hasta la fecha para proteínas Cas12a, solo superada en flexibilidad por HkCas12a, RbCas12a y Cas12aM29-1 entre las funcionalmente caracterizadas, que además de timina toleran citosina en las posiciones -2 y -3 (consenso 5’-YYN-3’) (véase Tabla 2).

Tabla 2. Origen, tamaño y secuencia PAM consenso de proteínas Cas12a que han sido caracterizadas funcionalmente.

Nombre	**Origen**	**Aminoácidos**	**PAM[a]**
AlCas12a	Metagenoma	1226	TTN
AsCas12a	*Acidaminococcus* sp. BV3L6	1307	TTTV
LbCas12a	*Lachnospiraceae bacterium* ND2006	1228	TTTV
FnCas12a	*Francisella novicida* U112	1300	TTTV
MbCas12a	*Moraxella bovoculi* 237	1373	TTTV
ArCas12a	*Agathobacter rectalis* 2789STDY5834884	1262	TTN
ErCas12a	*Eubacterium rectale*	1263	YTTN
LpCas12a	*Lachnospira pectinoschiza* 2789STDY5834886	1306	TTTN
Mb2Cas12a	*Moraxella bovoculi* AAX08_00205	1251	TTV
Mb3Cas12a	*Moraxella bovoculi* AAX11_00205	1261	TTV
PrCas12a	*Pseudobutyrivibrio ruminis* CF1b	1213	TTN
PxCas12a	*Pseudobutyrivibrio xylanivorans* DSM 10317	1215	TTTN
TsCas12a	*Thiomicrospira sp.* Xs5	1298	TTV
PiCas12a	*Prevotella disiens*	1323	KKYV
BsCas12a	*Butyrivibrio* sp. NC3005	1206	TTN
HkCas12a	*Helcococcus kunzii* ATCC 51366	1310	YYN
BoCas12a	Bacteroidetes oral taxon 274 str. F0058	1261	TTN
Pb2Cas12a	*Prevotella bryantii* B14	1247	TTTV
MlCas12a	*Moraxella lacunata*	1251	BTTV
MaCas12a	*Methanomethyloplilus alvi*	1227	TTV
BfCas12a	*Butyrivibrio fibrisolvens* MD2001	1231	YYYV
Lb2Cas12a	*Lachnospiraceae bacterium* MA2020	1206	TTTN
EeCas12a	*Lachnospira eligens* ATCC 27750	1282	TTTN
RbCas12a	*Ruminococcus bromii* sp. Butyrivibrio	1247	YYN
CeCas12a	*Coprococcus eutactus* sp.	1229	TTTV
Cas12aM13-1	Metagenoma	1223	TTTN
Cas12aM26-1	Metagenoma	1309	YYN
Cas12aM28-1	Metagenoma	1293	TTYN
Cas12aM29-1	Metagenoma	1280	YYN
Cas12aM31-1	Metagenoma	1307	YTTN
Cas12aM32-1	Metagenoma	1294	YTTN

PdCas12a	*Prevotella disiens*	1323	TTTV
Fn3Cas12a	*Francisella novicida*	1307	YTV
Adurb336Cas12a	Archaeaon ADurb.Bin336	1299	TTTV
Adurb193Cas12a	Firmicutes bacterium ADurb.Bin193	1356	TTTV
CmeCas12a	Metagenoma	1296	TTTV
YmeCas12a	Metagenoma	1371	TTTV

a. Y = C o T; B = C, G o T; V = A, C o T; K = G o T; N = A, C, G o T.

2.2 Actividad endorribonucleasa

Las proteínas Cas12a se caracterizan generalmente por poseer actividad endorribonucleasa que les permite procesar el pre-crRNA (ARN CRISPR inmaduro).

En el subdominio WED III de ortólogos Cas12a se han identificado cuatro residuos clave implicados en dicho procesamiento (por ejemplo, H843, K852, K869 y F873 en FnCas12a). Sin embargo, el alineamiento de AlCas12a con la secuencia de FnCas12a (véase Figura 4) reveló la ausencia de dos de ellos (H843 y K852), cuestionando su capacidad para llevar a cabo dicha actividad. Para evaluar el procesamiento del precrRNA por AlCas12a, se realizaron ensayos de digestión de dos moléculas diferentes de precrRNA (precrRNA1 y precrRNA2):

- El precrRNA1 está constituido por dos repeticiones completas (de 35 nt cada una) del sistema CRISPR-AlCas12a flanqueando una región espaciadora de 28 nt con la misma secuencia que el espaciador de la agrupación CRISPR del plásmido pUC57-array. Para la síntesis *in vitro* del precrRNA1, el ADNbc producto de la hibridación de los oligonucleótidos 169 y 170 se utilizó como molde en ensayos de transcripción empleando el HiScribe T7 *In Vitro* Transcription Kit (NEB) siguiendo las instrucciones del fabricante.

- El gen de precrRNA2, clonado por la empresa GenScript en el plásmido pUC57-precrRNA4, está constituido por cuatro repeticiones idénticas a las del precrRNA1 intercaladas por espaciadores de 28 pb. Uno de estos espaciadores coincide con el de precrRNA1 y el otro (5'-ACCGAAACGGAGGAATGGGAACGGCGCG-3', SEQ ID NO: 29) también está dirigido contra una secuencia localizada en el plásmido pSEVA431. Para la síntesis del precrRNA2, el plásmido pUC57-precrRNA4 linealizado por digestión con la enzima BamHI se empleó como molde en ensayos de transcripción *in vitro* equivalentes a los utilizados para la síntesis del precrRNA1.

Mezclas de reacción, conteniendo 0,5 µM de AlCas12a, 1 µM de sustrato precrRNA, 50 mM de Tris-HCl (pH 7,5 a 37°C), 100 mM de NaCl, 0,1 mg/ml de BSA, y 10 mM de sales de distintos cationes divalentes (Mg^{+2}, Ca^{+2}, Mn^{+2}, Co^{+2}, o Zn^{+2}), se incubaron durante 30 min a 37°C. Las reacciones se detuvieron añadiendo EDTA y proteinasa K a concentraciones finales de 40 mM y 1 mg/ml, respectivamente, incubándose durante 30 min más a 37°C. Posteriormente, se añadió un mismo volumen de tampón de carga de ARN 2x, desnaturalizando las muestras mediante tratamiento a 95°C durante 10 min, tras lo cual se sometieron a electroforesis en gel desnaturalizante (8M Urea) de poliacrilamida al 15%, no observándose en ningún caso procesamiento de los precrRNA (Figura 7).

Estos resultados *in vitro* sugieren que, a diferencia de las otras proteínas de la familia Cas12a que han sido evaluadas para dicha actividad, AlCas12a no tiene actividad ARNasa y no es capaz de procesar el precrRNA.

Esta diferencia es importante ya que es posible mejorar la eficacia y especificidad del corte de ADNbc diana por proteínas Cas12a introduciendo diversas modificaciones en las moléculas de ARN que sirven como guía [JEDRZEJCZYK *et al.*, *CRISPR-Cas12a nucleases function with structurally engineered crRNAs: SynThetic trAcrRNA*, Sci Rep. 2022 Jul 16;12(1):12193]. Sin embargo, la actividad ARNasa de la proteína efectora podría eliminar algunas de estas modificaciones. En este contexto, la incapacidad de AlCas12a para procesar el ARN guía representa una ventaja, ya que permite probar una mayor diversidad de modificaciones sin el riesgo de que estas sean alteradas por la proteína efectora.

Ejemplo 3: Diseño y funcionalidad de la herramienta CRISPR-AlCas12a

La herramienta CRISPR-AlCas12a comprende la proteína AlCas12a y un ARN guía (Al crRNA) diseñado a partir de la secuencia de la repetición CRISPR asociada al sistema CRISPR-AlCas12a, así como los vectores construidos para la administración de estos elementos a bacterias y su producción.

3.1 Proteína AlCas12a

Para la purificación de AlCas12a, la secuencia codificante con codones optimizados para *E. coli* (SEQ ID NO: 3) se insertó en el vector pHTP1 (resistente a la kanamicina),

generando el plásmido pMML32 (véase Figura 8), de forma que la transcripción del gen *alcas12a* a partir del promotor T7 controlado por el operador *lac* (inducible por IPTG) del vector diera lugar a una proteína AlCas12a fusionada a una etiqueta hexahistidina N-terminal (*alCas12a-his*: SEQ ID NO:4; AlCas12a-His: SEQ ID NO: 5).

Células de *E. coli* BL21(DE3) electrocompetentes se transformaron con pMML32 y se crecieron a 37ºC en caldo LB suplementado con kanamicina. Cuando el cultivo alcanzó una DO_{600} = 0,5, se indujo la expresión de la proteína añadiendo 1mM IPTG y tras 16 h de incubación a 16ºC se recogieron las células por centrifugación (5000 x g durante 15 min a 4ºC) y se resuspendieron en tampón de unión compuesto por 50 mM de tampón fosfato pH 7,6, 500 mM de NaCl, 10 mM de imidazol, 5% de glicerol, 10 mM de β-mercaptoetanol y 1mM de fluoruro de fenilmetilsulfonilo (PMSF). Las células se lisaron por sonicación. La fracción soluble de proteínas de *E. coli* se sometió a cromatografía de afinidad, y aquellas fracciones en las que se observó la presencia de AlCas12a-His tras electroforesis y tinción con Coomassie-Blue, se sometieron a cromatografía de exclusión molecular para obtener una preparación de AlCas12a-His de alta pureza.

3.2 ARN guía (AlcrRNA)

Dada la alta conservación tanto de la longitud del espaciador como de la secuencia de la repetición (excepto los nucleótidos del bucle de la región que forma la horquilla) en los crRNA de subtipo V-A que han sido validados funcionalmente [TENG, F., et al. *Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds*. Genome biology, 2019, vol. 20, p. 1-6] (véase Figura 2), resulta probable que el crRNA maduro de AlCas12a sea similar a los previamente descritos para estos sistemas.

Bajo esta premisa, diseñamos un crRNA para AlCas12a (AlcrRNA) que abarcaba en el extremo 5' (es decir, el *5'-handle*; SEQ ID NO: 6) una región invariable de 19 nt coincidente con la región 3' de la secuencia de la repetición, seguida de una región espaciadora de 24 nt que coincidía con una secuencia del plásmido pSEVA431. Fragmentos de ADNbc con la secuencia del AlcrRNA bajo un promotor T7 se obtuvieron por hibridación (incubación a 100°C durante 10 minutos seguida de 3 horas a temperatura ambiente) de mezclas de los oligonucleótidos complementarios id. 163 y 164, a una concentración equimolar, y se transcribieron *in vitro* con HiScribe T7 *In Vitro* Transcription Kit siguiendo las instrucciones del fabricante, incluido el tratamiento con DNasa.

3.3 Actividad de corte en cis

Para probar la funcionalidad del AlcrRNA diseñado, moléculas generadas por transcripción *in vitro* se incubaron con proteína AlCas12a-His purificada junto con un fragmento de ADNbc lineal de 840 pb de longitud que contenía una secuencia coincidente con el espaciador flanqueada en el extremo 5' (es decir, la región PAM) por 5'-CTTG-3' (compatible con la escisión), utilizando una proporción molar guía:proteína:diana de 40:20:1. Para comprobar si precrRNAs no procesados pueden ser utilizados como guía por AlCas12a, se realizaron experimentos similares con precrRNA1 y precrRNA2 (portadores del mismo espaciador que el AlcrRNA) en lugar del AlcrRNA. Las reacciones se incubaron a 37ºC durante 30 min en 10 mM $MgCl_2$ y los productos se analizaron mediante electroforesis en gel de agarosa (véase Figura 9A), observándose tan solo en el caso del AlcrRNA digestión del 100% sustrato, con la aparición de dos bandas de aproximadamente 520 pb y 320 pb, compatibles con la generación de una rotura de doble cadena dentro de la diana. Mientras que con el precrRNA2 no se observó escisión del ADN, con precrRNA1 se detectó una digestión parcial de la diana.

Estos resultados confirman la funcionalidad del AlcrRNA diseñado y corroboran la actividad específica de corte de la diana de ADNbc por parte de AlCas12a, al tiempo que demuestran que precrRNAs con un tamaño próximo al del AlcrRNA pueden actuar como guía, mientras que guías de mayor tamaño no son funcionales.

A continuación, evaluamos diversas condiciones de reacción para caracterizar con detalle y optimizar los ensayos de digestión *in vitro* mediada por AlCas12a-His guiada por AlcrRNA:

- En primer lugar, llevamos a cabo ensayos de digestión con mezclas de reacción conteniendo AlCas12a-His, $MgCl_2$, AlcrRNA y ADNbc diana con una PAM compatible (5'-CTTC-3') en las mismas concentraciones que los experimentos preliminares, así como reacciones en las que faltaba alguno de estos componentes además de una reacción con todos los componentes, pero con un sustrato diana sin una PAM compatible (5'-AATC-3' en lugar de 5'-CTTC-3'; control negativo). Los resultados obtenidos confirmaron el requerimiento de Mg^{+2} (véase Figura 9B).
- Seguidamente probamos concentraciones de proteína entre 10 nM y 1 µM en las mismas condiciones y concentraciones de los demás componentes utilizados en los otros ensayos (véase Figura 9C). La máxima proporción de escisión de sustrato se

observó con concentraciones de proteína superiores a 0,25 µM. En base a este resultado, los ensayos de digestión posteriores se realizaron con 0,5 µM de AlCas12a, manteniendo la proporción molar 40:20:1 de guía:proteína:diana.

- El análisis mediante electroforesis en geles de agarosa de muestras tomadas tras varios tiempos de incubación de mezclas de reacción a 37ºC (véase Figura 9D), mostró que aproximadamente el 44 % del sustrato se escindió tras 5 min, y que el 94 % ya había sido escindido tras 15 min, mientras que el mayor porcentaje de producto cortado (99 %) se observó a partir de los 30 min. Estos porcentajes representan la relación porcentual entre las intensidades de las bandas correspondientes a los productos escindidos y la suma de las intensidades de todas las bandas presentes en la misma calle del gel, calculadas mediante el software Image Lab (BIO-RAD).

- Por último, se determinó el intervalo de temperatura para la digestión en las condiciones y concentraciones establecidas hasta el momento para el resto de los parámetros. Para ello se incubaron a temperaturas de entre 15ºC y 65ºC durante 60 min mezclas de reacción obtenidas tras combinar soluciones de ADNbc diana con soluciones incluyendo todos los componentes de reacción excepto la diana preincubadas a la temperatura correspondiente durante 3 min. AlCas12a escindió la diana entre 20ºC y 45ºC, con un óptimo dentro del rango de 35ºC a 40ºC (véase Figura 9E), lo cual permite su aplicación en una gran diversidad de ambientes y organismos, incluyendo plantas y animales tanto ectotermos como endotermos.

Para identificar residuos de AlCas12a que participan en la escisión de ADNbc diana, se generaron mutantes puntuales de AlCas12a-His reemplazando por alanina los aminoácidos D831 (*alcas12a*-D831A: SEQ ID NO: 7; AlCas12a-D831A: SEQ ID NO: 8), E921 (*alcas12a*-E921A: SEQ ID NO: 9; AlCas12a-E921A: SEQ ID NO: 10) o D1183 (*alcas12a*-E921A: SEQ ID NO: 11; AlCas12a- D1183A: SEQ ID NO: 12), equivalentes a D917, E1006 y D1255, respectivamente, del dominio RuvC de FnCas12a cuya mutación se ha demostrado que resulta en la pérdida total (D917, E1006) o parcial (D1255) de la actividad ADNasa. El gen *alcas12a-his* del plásmido pMML32 se modificó empleando el sistema de mutagénesis QuickChange con cebadores diseñados para dirigirse a los tripletes de ADN que codifican cada uno de estos aminoácidos. Los plásmidos, pMML42, pMML43 y pMML44 (expresando los mutantes dAlCas12a-D831A - SED dAlCas12a-E921A y dAlCas12a-D1183A, respectivamente), obtenidos tras amplificación por PCR con polimerasa Phusion® High-Fidelity DNA Polymerase (NEB) se digirieron con DpnI y se transformaron en *E. coli* NZYstar. Las sustituciones se confirmaron mediante

secuenciación de los plásmidos purificados a partir de colonias de transformantes y las proteínas mutantes se purificaron siguiendo el procedimiento descrito anteriormente.

En ensayos de digestión *in vitro* de moléculas diana con cada una de las proteínas AlCas12a-His mutantes (dAlCas12a-D831A, dAlCas12a-E921A o dAlCas12a-D1183A) en presencia de un AlcrRNA guía bajo las mismas condiciones de reacción optimizadas para la proteína nativa (véase Figura 9F), no se detectó escisión de ADN diana tras 30 min de incubación, demostrando la implicación de cada uno los tres aminoácidos en dicha actividad.

3.4 Actividad de corte en trans

Para investigar la actividad de corte en trans de AlCas12a, se preincubaron a 37ºC durante 15 min la proteína AlCas12a-His en tampón 10 mM $MgCl_2$ y un AlcrRNA generado por transcripción *in vitro*, con un potencial activador consistente en un oligonucleótido de cadena sencilla complementario a la región de espaciador del crRNA sin una secuencia PAM adyacente, en una proporción molar de 16:8:1 (ARNcr:proteína:activador). Tras la preincubación se añadió a la mezcla ADNcs circular M13mp18, en cuya molécula no se encuentra ninguna secuencia similar a la del espaciador del crRNA, en una proporción molar de 1:40 con respecto a la proteína, y la mezcla se incubó a la misma temperatura durante 30 min. Como control previamente validado de escisión colateral inducida por activador, se empleó LbCas12a (EnGen® Lba Cas12a de NEB) y un crRNA (LbcrRNA) con un *5'-handle* compatible con la enzima y el mismo espaciador que el del AlcrRNA. También se realizaron ensayos omitiendo $MgCl_2$, crRNA o activador en las mezclas de reacción (véase Figura 10A). El sustrato M13mp18 fue degradado por ambas proteínas en presencia de todos los componentes. Sin embargo, en ausencia de activador no se detectó degradación de M13mp18 por la proteína LbCas12, mientras que AlCas12a-His lo degradó completamente tras 30 min de incubación tanto en ausencia del activador como del AlcrRNA, pero no cuando faltaba $MgCl_2$.

Experimentos equivalentes llevados a cabo con las proteínas dAlCas12a-D831A, dAlCas12a-E921A o dAlCas12a-D1183A (0,5 µM de proteína en tampón O 1X de Thermo Scientific y 13 nM de ADN M13mp18), en comparación con los realizados con AlCas12a-His en las mismas condiciones, mostraron una disminución en la velocidad de degradación del ADN de M13mp18 en ausencia tanto del activador como del AlcrRNA (véase Figura 10B), confirmando la implicación de estos aminoácidos en el corte inespecífico de ADNcs.

3.5 Actividad de corte independiente de guía

Para corroborar la capacidad de AlCas12a para cortar de manera inespecífica moléculas de ADNcs, se realizaron en las mismas condiciones incubaciones de AlCas12a-His con oligodesoxinucleótidos no diana de 67 nt y 100 nt en lugar de M13mp18, observándose degradación de ambos en ausencia de activador o AlcrRNA (véase Figura 10C). En conjunto, estos resultados muestran que AlCas12a presenta una actividad endonucleasa de ADNcs independiente de crRNA guía y dependiente de $MgCl_2$, pero no permiten discernir si la degradación de ADNcs no diana observada en presencia de activador y crRNA es debida exclusivamente a la actividad de corte independiente de guía o también participa una actividad de corte en trans.

Para comprobar si AlCas12a es capaz de llevar a cabo este tipo de actividad, comparamos la velocidad de degradación de un oligonucleótido de ADNcs no diana de 100 nt en una mezcla de reacción conteniendo tan solo $MgCl_2$ y la proteína AlCas12a-His, bien preincubada con activador y ADNbc diana (proteína activada) o sin preincubar (véase Figura 10D). El corte del oligonucleótido fue más rápido con AlCas12a-His activada que con la proteína sin activar, confirmando su capacidad de corte de ADNcs en trans.

El corte en trans de otras proteínas Cas12a se ha aprovechado para implementar métodos de diagnóstico molecular *in vitro*, para los que también se podría utilizar AlCas12a. Por otra parte, la actividad endonucleasa de ADNcs dependiente de $MgCl_2$ e independiente de crRNA se podría emplear para la degradación de moléculas de ADNcs en soluciones de ácidos nucleicos o para la detección de regiones desapareadas en moléculas de ADNbc, aplicaciones no implementadas para otras proteínas de la familia Cas12a.

Ejemplo 4: Aplicación de la proteína de la invención para la selección positiva de mutantes en *E. coli*

Para evaluar la aplicación de CRISPR-AlCas12a como herramienta para la selección positiva de mutantes de *E. coli*, se construyó a partir de pMML30 el plásmido pMML40 que expresa AlCas12a bajo el control de un promotor de arabinosa y un AlcrRNA dirigido al gen cromosómico *pyrF*, sustituyendo la región entre el promotor y el terminador de la agrupación CRISPR por una secuencia codificante del crRNA que abarca un fragmento de 19 nt de la repetición del sistema CRISPR-AlCas12a y una región espaciadora de 24

nt que coincide con una secuencia en el gen cromosómico *pyrF* con 5'-GTTC-3' en la región PAM, bajo el control del promotor constitutivo (Part:BBa_J23101, colección BioBricks). Siguiendo el mismo procedimiento, pero utilizando pMML31 en lugar de pMML30, se generó el plásmido pMML41 que carece del gen *alcas12a*. La secuenciación Sanger de las colonias obtenidas tras la transformación con pMML40 reveló en todos los casos un mismo plásmido mutante (pMML40bis) (véase Figura 11), con la región codificante de AlcrRNA duplicada en tándem. Por otra parte, se generó mediante ensamblaje Gibson un molde de recombinación de ADNbc lineal constituido por un fragmento de 145 pb, que coincide con una secuencia localizada en la región intergénica aguas arriba de *pyrF*, y un fragmento de 163 pb que coincide con una secuencia aguas abajo del mismo gen, de manera que la recombinación del ADN molde con las secuencias homólogas en el cromosoma daría lugar a una deleción de 607 pb.

Células electrocompetentes de *E. coli* BW 27783 fueron transformadas con el plásmido termosensible pKD46 (resistencia a la ampicilina), que codifica el sistema de recombinación Lambda Red (Datsenko and Wanner, 2000). Los transformantes se crecieron en placas de agar LB con ampicilina a 30ºC en lugar de a 37ºC para evitar la pérdida del plásmido. Colonias portadoras de pKD46 se transfirieron a caldo LB suplementado con ampicilina y se incubaron a 30ºC hasta una DO_{600} = 0,2. A continuación, se transfirieron a caldo LB suplementado con ampicilina y se incubaron a 30ºC hasta alcanzar una DO_{600} = 0,2, momento en el que se añadió L-arabinosa (concentración final 0,2%) para inducir la expresión de las proteínas Lamba Red; cuando el cultivo alcanzó una DO_{600} = 0,5 se prepararon células electrocompetentes. Tres alícuotas de células electrocompetentes fueron co-transformadas de forma independiente con 150 ng de ADN molde y 50 ng de pMML40bis o pMML41. Las colonias transformantes se crecieron a 37ºC, impidiendo así la replicación de pKD46, en agar LB suplementado con cloranfenicol (para la selección de plásmidos pMML40bis y pMML41) y 0,2% de L-arabinosa (para la inducción de la transcripción de *alcas12a*).

La amplificación por PCR de la región *pyrF* de los genomas de 100 colonias seleccionadas al azar de células que expresaban AlCas12a procedentes de los tres experimentos, reveló bandas con el tamaño de las esperadas tras la deleción en 90 de las colonias analizadas (frecuencia de detección de mutantes del 90%). En cambio, con las 35 colonias analizadas de la misma manera procedentes de células sin *alcas12a* tan solo se obtuvieron productos de amplificación cuyo tamaño coincidía con el del gen nativo (véase Figura 12).

Ejemplo 5: Reducción de la susceptibilidad de *E. coli* a infección por virus de ADNcs y ADNbc como consecuencia de la producción de la proteína AlCas12a.

Para evaluar la susceptibilidad de *E. coli* a la infección por virus no diana de ADNcs (M13mp18) o ADNbc (P1-vir) en presencia de AlCas12a, se emplearon células de la cepa de *E. coli* NZYstar (resistente a tetraciclina) transformadas con pMML30 o pMML39. Ambos plásmidos confieren resistencia a cloranfenicol y codifican AlCas12a bajo el control de un promotor inducible por arabinosa. El plásmido pMML30 contiene además una agrupación CRISPR con 2 repeticiones que flanquean un espaciador dirigido contra el plásmido pSEVA431, con expresión constitutiva. Como control sin AlCas12a ni ARN guía se utilizaron células portadoras del plásmido pBAD33 (resistencia a cloranfenicol).

Células electrocompetentes de *E. coli* NZYstar fueron transformadas con los respectivos plásmidos y los transformantes se seleccionaron en placas de LB suplementadas con tetraciclina y cloranfenicol. A continuación, las colonias resistentes se transfirieron a caldo LB suplementado con los mismos antibióticos y 0,2% D-glucosa (para reprimir la expresión de AlCas12a), incubando el cultivo a 37ºC en agitación hasta alcanzar la fase estacionaria de crecimiento. Posteriormente, las células de cada cultivo se lavaron en caldo LB suplementado con tetraciclina, cloranfenicol y 0,2% de L-arabinosa (para inducir la expresión de AlCas12a) y se diluyeron usando el mismo medio hasta una DO_{600} = 0,05, incubando de nuevo los cultivos en las mismas condiciones. Cuando los cultivos alcanzaron una DO_{600} ~ 0,5 se tomaron 0,5 mL de cada uno y se mezclaron con 4,5 mL de una solución a 50°C de LB con agar al 0,6% (agar de cobertura) fundido, depositando a continuación las mezclas sobre placas de medio sólido LB suplementadas con los antibióticos, 0,1mM IPTG y 0,2 mg/mL X-gal. Una vez solidificada la capa de agar de cobertura, se depositaron gotas de 4 µL de diluciones decimales, preparadas a partir de una suspensión con ~10^{12} ufp/mL de virus ADNcs M13mp18 y otra con ~10^{8} ufp/mL de virus ADNbc P1-vir, sobre la superficie del agar de cobertura, incubando las placas a 37ºC durante 16 h. Tras la incubación las placas fueron fotografiadas.

M13mp18 es una variante de M13 que codifica la enzima β-galactosidasa y que no presenta ninguna secuencia que coincida con la del espaciador de la agrupación CRISPR de pMML30. La hidrólisis de X-gal por la β-galactosidasa genera un compuesto de color azul (5,5'-dibromo-4,4'-dicloro-índigo), facilitando la visualización de la multiplicación del virus en células de *E. coli* crecidas en medio suplementado con este compuesto e IPTG (inductor). P1-vir es una variante virulenta del fago bacteriano P1, incapaz de establecer

lisogenia, que tampoco contiene ninguna secuencia que coincida con el espaciador de la agrupación CRISPR de pMML30. Tanto en las cepas portadoras del plásmido pMML30 como en las portadoras de pMML39 se observó una reducción en el tamaño y transparencia de los halos de inhibición de crecimiento producidos por M13mp18 y P1-vir, en comparación con los obtenidos en la cepa control (portadora del plásmido pBAD33) para una misma concentración de virus. En el caso de los halos generados por M13mp18, la expresión de AlCas12 resulta en halos con una menor intensidad del color azul, especialmente en los bordes (véase Figura 13).

Esta menor susceptibilidad a infección por parte de las células que expresan AlCas12a sugiere que su actividad desoxirribonucleasa indiscriminada interfiere con la propagación de virus tanto ADNcs como ADNbc. Sin embargo, este efecto inhibidor fue más pronunciado en la cepa que expresaba únicamente AlCas12a (portadora de pMML39) que para la que además expresaba la agrupación CRISPR (portadora de pMML30), indicando que el AlcrRNA reduce dicha actividad de corte inespecífico. Esta consecuencia biológica de la actividad no guiada de AlCas12a, que no ha sido documentada para otras proteínas Cas12a, puede utilizarse para proteger a las células que la expresan frente a virus con ADNcs, independientemente de su secuencia.

Ejemplo 6. Materiales y métodos

Los cultivos de la bacteria *Escherichia coli* se realizaron en caldo Luria-Bertani (LB) con agitación rotatoria a 180 rpm o en agar LB sólido, y se incubaron a 37ºC a menos que se indique lo contrario. Para la selección de células portadoras de plásmidos, los medios de cultivo se suplementaron con cloranfenicol (25 µg/ml), espectinomicina (50 µg/ml), ampicilina (100 µg/ml) o kanamicina (50 µg/ml), según correspondiera. El ADN plasmídico se aisló de *E. coli* con PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen) o PureLink™ HiPure Plasmid Miniprep Kit (Invitrogen). Los productos de PCR y los fragmentos de ADN se purificaron con GFX™ PCR DNA and Gel Band Purification Kit (Cytiva). Las clonaciones se llevaron a cabo mediante ensamblaje Gibson empleando el Gibson Assembly® Cloning Kit (NEB). Las concentraciones de ácidos nucleicos se midieron con un NanoDrop ND-1000 (Thermo Scientific). Los geles de agarosa utilizados en las electroforesis de ácidos nucleicos se tiñeron con GreenSafe premium (NZYTech) y los de acrilamida con GelRed Nucleic Acid Gel Stain (Thermo Fisher Scientific), siendo visualizados con un sistema ChemiDoc XRS+ Gel Imaging System (BIORAD).

Las cepas bacterianas, los plásmidos, y los oligonucleótidos utilizados en la presente invención se relacionan en las Tablas 3, 4 y 5, respectivamente.

Tabla 3. Cepas de *E. coli* utilizadas

Cepa	Genotipo Relevante	Uso	Fuente / Referencia
NZYstar	*end*A1 *hsd*R17(r_k-, m_k+) *sup*E44 thi -1 *rec*A1 *gyr*A96 *rel*A1 *lac*[F´ *pro*A^+B^+ *lac*I^q ZDM15 :Tn10(Tc^R)]	Clonación de plásmidos, generación de la genoteca PAM e infección con M13	NZYTech
BW 27783	*lacIq rrnB3* Δ*lacZ*4787 Δ*(araBAD)*567 Δ(*rhaBAD*)568 *hsdR*514 Δ(K) Φ(*ΔaraEp* PCP8−*araE*)	Escrutinio y validación de ensayos PAM *in vivo*. Ensayos de edición genética en *E. coli*	[1]
BL21(DE3)	*F- ompT gal dcm lon* HsdSB (rB-mB-) λ(DE3 [*lacI lacUV5-T7* gene 1 *ind1 sam7 nin5])*	Sobreexpresión y purificación de AlCas12a	[2]

[1] Khlebnikov,A., Datsenko,K.A., Skaug,T., Wanner,B.L. and Keasling,J.D. (2001) Homogeneous expression of the P BAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology (Reading), 147(Pt 12), 3241–3247.

[2] Rosenberg,A.H., Lade,B.N., Dao-shan,C., Lin,S.-W., Dunn,J.J. and Studier,F.W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene, 56, 125–135.

Tabla 4. Plásmidos empleados.

Plásmido	Descripción	Referencia
pBAD33	Vector de expresión bacteriano. Resistencia a cloranfenicol y promotor inducible por arabinosa.	[1]
pSEVA431	Vector de clonación. Resistencia a estreptomicina/ espectinomicina.	[2]
pMML33	Derivado de pSEVA431 con secuencia diana flanqueada por 5'-ATTA-3'.	Esta invención
pMML34	Derivado de pSEVA431 con secuencia diana flanqueada por 5'-ATTG-3'.	Esta invención
pMML35	Derivado de pSEVA431 con secuencia diana flanqueada por 5'-CTTC-3'.	Esta invención

pMML36	Derivado de pSEVA431 con secuencia diana flanqueada por 5'-CTTG-3'.	Esta invención
pMML37	Derivado de pSEVA431 con secuencia diana flanqueada por 5'-TTTG-3'.	Esta invención
pMML38	Derivado de pSEVA431 con secuencia diana flanqueada por 5'-CTTT-3'.	Esta invención
pHTP1	Vector de expresión bacteriano. Fusión a etiqueta hexahistidina y promotor dependiente de la polimerasa de T7. Resistencia a kanamicina.	NZYTech
pKD46	Plásmido con replicación sensible a temperatura que codifica el sistema de recombinación Lambda Red. Resistencia a ampicilina.	[3]
pMML30	Derivado del plásmido pBAD33 que contiene el gen que codifica la proteína AlCas12a-His, con codones optimizados para *E. coli*, bajo el control del promotor *BAD* inducible por arabinosa, y una agrupación CRISPR de 2 repeticiones que flanquean el espaciador 5'-CTGTATATATCGTGCGAAAAAGGATGGATA-3', SEQ ID NO: 28.	Esta invención
pMML31	Similar al plásmido pMML30, excepto por carecer del gen *alcas12a*.	Esta invención
pMML39	Similar al plásmido pMML30, excepto por carecer de la agrupación CRISPR.	Esta invención
pMML32	Derivado del vector pHTP1, portador del gen *alcas12a* con expresión dependiente del promotor T7 y controlado por un operador lac (inducible por IPTG) que genera una proteína AlCas12a fusionada a una etiqueta hexahistidina N-terminal (AlCas12a-His).	Esta invención
pMML42	Derivado del plásmido pMML32 con una mutación en el gen *alcas12a* que provoca la sustitución Asp831Ala (D831A) en la proteína AlCas12a.	Esta invención
pMML43	Derivado del plásmido pMML32 con una mutación en el gen *alcas12a* que provoca la sustitución Glu921Ala (E921A) en la proteína AlCas12a.	Esta invención
pMML44	Derivado del plásmido pMML32 con una mutación en el gen *alcas12a* que provoca la sustitución Asp1183Ala (D1183A) en la proteína AlCas12a.	Esta invención
pMML40	Derivado del plásmido pMML30 por reemplazamiento del espaciador y la repetición corriente abajo de la agrupación CRISPR por una región guía de 24 nt que coincide con una	Esta invención

	secuencia en el gen cromosómico de *E. coli pyrF* con 5'-GTTC-3' en la región PAM.	
pMML41	Derivado del plásmido pMML31 por reemplazamiento del espaciador y la repetición corriente abajo de la agrupación CRISPR por una región guía de 24 nt que coincide con una secuencia en el gen cromosómico de *E. coli pyrF* con 5'-GTTC-3' en la región PAM.	Esta invención
pMML40bis	Derivado del plásmido pMML40 con la región codificante del AlcrRNA duplicada.	Esta invención
pUC57-array	Derivado del plásmido pUC57 que contiene dos repeticiones de la agrupación CRISPR de CRISPR-AlCas12a de 35 pb de longitud cada una, separadas por un espaciador de 28 pb dirigido al plásmido pSEVA431.	Esta invención
pUC57-AlCas12a	Derivado del plásmido pUC57 que contiene el gen *alcas12a* cuyo uso de codones está optimizado para *E. coli*	Esta invención
pUC57-precrRNA4	Derivado del plásmido pUC57 que contiene repeticiones de la agrupación CRISPR de CRISPR-AlCas12a de 35 pb de longitud cada una que flanquean 2 espaciadores distintos que dispuestos de manera alterna constituyen una agrupación de 4 espaciadores, todos de 28 pb y dirigidos al plásmido pSEVA431.	Esta invención

[1] Guzman,L.-M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P BAD Promoter. J. Bacteriol. 177:4121–4130.

[2] Silva-Rocha,R., Martínez-García,E., Calles,B., Chavarría,M., Arce-Rodríguez,A., de Las Heras,A., Páez-Espino,A.D., Durante-Rodríguez,G., Kim,J., Nikel,P.I., et al. (2013) The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res. 41: D666-D675.

[3] Datsenko,K.A. and Wanner,B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. U.S.A. 2000; 97:6640–6645.

Tabla 5. Oligonucleótidos utilizados.

# Id	**Uso**	**Diana**	**Secuencia (5'→3')**
1	Cebador de PCR para construcción de pMML30	pMML31	aactggtgacgggcgaataatctagagtcgacctgcaggc (SEQ ID NO: 30)

2	Cebador de PCR para construcción de pMML30	pMML31	ttggtaaaattgctgtacatatgtatatctccttcttaaaggtaccgag (SEQ ID NO: 31)
3	Cebador de PCR para construcción de pMML30	pUC57-AlCas12a	tttaagaaggagatatacatatgtacagcaattttaccaacatctattccct (SEQ ID NO: 32)
4	Cebador de PCR para construcción de pMML30	pUC57-AlCas12a	gcctgcaggtcgactctagattattcgcccgtcaccagttga (SEQ ID NO: 33)
8	Cebador de PCR para construcción de pMML31	pUC57-array	gaagcatttatcagggttatcgcactcgagtttacagctagc (SEQ ID NO: 34)
9	Cebador de PCR para construcción de pMML31	pUC57-array	atgtatccgctcatgagacatgcgctcgagaaaaaaaaaccccg (SEQ ID NO: 35)
10	Cebador de PCR para construcción de pMML31	pBAD33	tagctgtaaactcgagtgcgataaccctgataaatgcttcaataatattgaaaaaggaag (SEQ ID NO: 36)
11	Cebador de PCR para construcción de pMML31	pBAD33	ggtttttttctcgagcgcatgtctcatgagcggatacatatttgaatgt (SEQ ID NO: 37)
109	Cebador de PCR para construcción de genoteca PAM	pSEVA431	ccttttcgcacgatatacagnnnnttgccaaagggttcgt (SEQ ID NO: 38)
110	Cebador de PCR para construcción de genoteca PAM	pSEVA431	acgaacccttggcaannnnctgtatatcgtgcgaaaaagg (SEQ ID NO: 39)
160	Cebador de PCR para construcción de pMML33-38	pSEVA431	atggatataccgaaaaaatcgctataatgac (SEQ ID NO: 40)
148	Cebador de PCR para construcción de pMML33	pSEVA431	ccttttcgcacgatatacagtaatttgccaaagggttcgt (SEQ ID NO: 41)
149	Cebador de PCR para construcción de pMML34	pSEVA431	ccttttcgcacgatatacagcaatttgccaaagggttcgt (SEQ ID NO: 42)
141	Cebador de PCR para construcción de pMML35	pSEVA431	ccttttcgcacgatatacaggaagttgccaaagggttcgt (SEQ ID NO: 43)
140	Cebador de PCR para construcción de pMML36	pSEVA431	ccttttcgcacgatatacagcaagttgccaaagggttcgt (SEQ ID NO: 44)
139	Cebador de PCR para construcción de pMML37	pSEVA431	ccttttcgcacgatatacagcaaattgccaaagggttcgt (SEQ ID NO: 45)
154	Cebador de PCR para construcción de pMML38	pSEVA431	ccttttcgcacgatatacagaaaattgccaaagggttcgt (SEQ ID NO: 46)
70	Secuenciación de regiones PAM (escrutinio PAM)	pSEVA431	gcgaataagggacagtgaa (SEQ ID NO: 47)
60	Secuenciación de regiones PAM (escrutinio PAM)	pSEVA431	ggcccagtatcagcccgtc (SEQ ID NO: 48)
163	Generación de AlcrRNA	Oligonucleótido 164	aagctaatacgactcactatagggaatttctactaacgtagatctgtatatcgtgcgaaaaaggatg (SEQ ID NO: 49)

164	Generación de AlcrRNA	Oligonucleotido 163	catccttttcgcacgatatacagatctacgttagtagaaattccctatagtgagtcgtattagctt (SEQ ID NO: 50)
167	Generación de LbcrRNA	Oligonucleótido 168	aagctaatacgactcactatagggtaatttctactgttgtagatctgtatatcgtgcgaaaaag (SEQ ID NO: 51)
168	Generación de LbcrRNA	Oligonucleótido 167	ctttttcgcacgatatacagatctacaacagtagaaattaccctatagtgagtcgtattagctt (SEQ ID NO: 52)
169	Generación de precrRNA1	Oligonucleótido 170	aagctaatacgactcactatagggtctaattcgtccatcaatttctactaacgtagatctgtatatcgtgcgaaaaaggatggatagtctaattcgtccatcaatttctactaacgtagat (SEQ ID NO: 53)
170	Generación de precrRNA1	Oligonucleótido 169	atctacgttagtagaaattgatggacgaattagactatccatccttttcgcacgatatacagatctacgttagtagaaattgatggacgaattagaccctatagtgagtcgtattagctt (SEQ ID NO: 54)
181	Activador de AlCas12	No aplicable	accctgcttcggggtcattatagcgatttttcggtatatccatccttttcgcacgatatacaggattttgccaaagggttcgtgtaga (SEQ ID NO: 55)
115	Cebador de PCR para generación de fragmento de ADNbc para ensayos de corte *in vitro*	pSEVA431	attcttgcaggtatcttcgagcc (SEQ ID NO: 56)
15	Cebador de PCR para generación de fragmento de ADNbc para ensayos de corte *in vitro*	pSEVA431	ccagccttgaccgaaacg (SEQ ID NO: 57)
263	Cebador de PCR para construcción de pMML40-41	pMML30-31 (respectivamente)	tggggatatcgtggaatttcaggtatctacgttagtagaaattgctagcataatacctag (SEQ ID NO: 58)
265	Cebador de PCR para construcción de pMML40-41	pMML30-31 (respectivamente)	ctgaaattccacgatatccccacatgtctcatgagcggatacatat (SEQ ID NO: 59)
188	Cebador de PCR para generación de ADNbc molde de recombinación	*E. coli* BW 27783	tccattgtgctccggcaac (SEQ ID NO: 60)

189	Cebador de PCR para generación de ADNbc molde de recombinación	*E. coli* BW 27783	ctgttctggcgtcataatgcgaccagaccttcttgatgatgg (SEQ ID NO: 61)
190	Cebador de PCR para generación de ADNbc molde de recombinación	*E. coli* BW 27783	atcatcaagaaggtctggtcgcattatgacgccagaacaa (SEQ ID NO: 62)
191	Cebador de PCR para generación de ADNbc molde de recombinación	*E. coli* BW 27783	tgagtagaccagacggctg (SEQ ID NO: 63)
234	Cebador de PCR para amplificación de región pyrF	Colonias de transformantes *E. coli*	tactggcattgtccgtcttgtc (SEQ ID NO: 64)
235	Cebador de PCR para amplificación de región pyrF	Colonias de transformantes *E. coli*	gcaccacaccgtcgcct (SEQ ID NO: 65)
236	Cebador de PCR para mutagénesis D831A. Construcción de pMML42	pMML32	tggcattgctcgcggtgagaaggaactgggctacatc (SEQ ID NO: 66)
237	Cebador de PCR para mutagénesis D831A. Construcción de pMML42	pMML32	ccgcgagcaatgccaataacggttttgataccaagag (SEQ ID NO: 67)
238	Cebador de PCR para mutagénesis E921A. Construcción de pMML43	pMML32	attgtgctggcgaatctgaatgttggctttaaacaagg (SEQ ID NO: 68)
239	Cebador de PCR para mutagénesis E921A. Construcción de pMML43	pMML32	agattcgccagcacaatgatcgctttgtgcttgatcgc (SEQ ID NO: 69)
240	Cebador de PCR para mutagénesis D1183A. Construcción de pMML44	pMML32	ggaacggtgccgcgaatggcgcctaccacattgg (SEQ ID NO: 70)
241	Cebador de PCR para mutagénesis D1183A. Construcción de pMML44	pMML32	attcgcggcaccgttccacgcatagccctggacc (SEQ ID NO: 71)

DESCRIPCIÓN DE LAS FIGURAS

La Figura 1 muestra una representación esquemática del cóntigo donde se localiza el locus CRISPR-*alcas12a*. Los siete genes predichos (*alcas12a*, *cas4*, *cas1*, *cas2* y tres *orfs* no relacionados) se representan como flechas que indican su dirección de transcripción. La

agrupación CRISPR se representa como una serie de rectángulos, precedida por una flecha que señala la dirección de transcripción predicha.

La Figura 2 muestra el alineamiento de la repetición CRISPR de AlCas12a con las de sistemas CRISPR-Cas12a validados funcionalmente. Las posiciones subrayadas corresponden a la repetición invertida interna que forma el tallo de la estructura en horquilla y con asteriscos se indican las posiciones conservadas.

La Figura 3 muestra las relaciones filogenéticas entre proteínas Cas12 mediante un árbol filogenético de AlCas12a, proteínas asignadas a la familia Cas12a, incluyendo ortólogos de Patescibacteria (sombreado), miembros del grupo Cas12a2 (sombreado) y ortólogos de la familia Cas12b1 (sombreado). Se indica la localización en el árbol de AlCas12a (en negrita) y de las proteínas Cas12a con estructura conocida (AsCas12a, LbCas12a, Lb2Cas12a y FnCas12a), así como la proteína AbCas12a de la Patescibacteria Aliarcobacter *butzleri* L348 y SuCas12a2 de *Sulfuricurvum* sp. PC08-66.

La Figura 4 muestra el alineamiento mediante Clustal Omega de las secuencias de la proteína AlCas12a con las de las proteínas Cas12a con estructuras conocidas (AsCas12a, FnCas12a, LbCas12a, Lb2Cas12a). En cada una de ellas se identifican los dominios Rec1, Rec2, Nuc, RuvC (compuesto por los subdominios RuvC-I, RuvC-II y RuvC-III), PI, BH y WED (compuesto por los subdominios WED-I, WED-II y WED-III), delimitados por líneas verticales y sombreados en gris o no sombreados alternativamente. Las posiciones de los residuos que se ha comprobado experimentalmente que están implicados en el procesamiento del precrRNA (H843, K852, K869 y F873, en el subdominio WED-III) y en el corte de ADNbc diana (D917, E1006 y D1255, en los subdominios RuvC) están marcadas con flechas e identificadas por su abreviatura del aminoácido y la posición que ocupan en FnCas12a. Los aminoácidos conservados en las cinco proteínas se destacan sombreados en negro.

La Figura 5 muestra el mapa del plásmido pMML30, que incluye un gen que confiere resistencia a cloranfenicol (CmR), el gen *alcas12a* modificado para adecuar su secuencia al uso de codones de *E. coli*, y una agrupación CRISPR constituida por dos repeticiones del sistema CRISPR-AlCas12a separadas por un espaciador de 28 pb cuya secuencia coincide con una región del plásmido pSEVA431. La transcripción del gen *alcas12a* está bajo el control del promotor BAD inducible por arabinosa, y finaliza en la región terminadora rrnB T1 T2. La transcripción de la agrupación CRISPR está controlada por un promotor

constitutivo (Part:BBa_J23101) y finaliza en un terminador artificial (Part:BBa_B1006). El plásmido incluye el gen *araC* bajo el control de su propio promotor, que codifica la proteína AraC (regulador del promotor BAD).

La Figura 6 muestra los resultados del cribado y validación de PAM. (A) Logotipo de secuencia de la región PAM preferida por la proteína de la presente invención para la escisión de la diana, según lo determinado por el cribado *in vivo* de una genoteca PAM. Se indican las posiciones de los nucleótidos desde el extremo 5' de la secuencia diana (cadena coincidente con el espaciador). Se ensayaron los nucleótidos de las posiciones -4 a -1. (B) Validación de PAMs. Se representa la eficiencia de transformación (UFC por µg de ADN) de células *E. coli* que expresan (+ AlCas12a) o no (- AlCas12a) AlCas12a en experimentos de transformación con plásmidos diana que contienen secuencias específicas en las posiciones -4 a -1 de la región PAM. Los datos representados son la media de tres experimentos biológicos independientes (las barras de error corresponden a la desviación estándar).

La Figura 7 muestra la evaluación *in vitro* del procesamiento del precrRNA1 y precrRNA2 por AlCas12a. Se incluye una muestra del AlcrRNA como referencia de procesamiento. Se indica el tiempo transcurrido en minutos tras mezclar los componentes de reacción con (+) o sin (-) AlCas12a.

La Figura 8 muestra el mapa del plásmido pMML32, que incluye un gen que confiere resistencia a kanamicina (Km^R), el gen *alcas12a-his*, que codifica la proteína AlCas12a fusionada a una etiqueta hexahistidina N-terminal (AlCas12a-His), modificado para adecuar su secuencia al uso de codones de *E. coli*, y el gen *lacI* (codifica la proteína LacI) bajo el control de su propio promotor. La transcripción del gen *alcas12a-his* a partir del promotor T7 está controlada por un operador *lac* (reprimido por LacI e inducible por IPTG).

La Figura 9 muestra varios ejemplos de los resultados de electroforesis en gel de agarosa de muestras de ensayos *in vitro* de escisión de ADNbc por AlCas12a en diferentes condiciones de reacción. Excepto cuando se especifica algún cambio, 500 nM de AlCas12a, 1 mM de ARN guía (precrRNA o AlcrRNA), 25 nM de ADNbc diana (fragmento lineal de 840 pb con una secuencia diana y una PAM compatible, 5'-CTTG-3') en una solución 10 mM de MgCl2 se incubaron durante 30 min a 37ºC (condiciones estándar). Se indica el tamaño de las bandas más relevantes del marcador de peso molecular – M – 1 Kb Plus DNA Ladder (Invitrogen) y el de las correspondientes al tamaño del sustrato de

ADN no cortado y cortado. (A) Productos de reacciones de digestión en presencia de precrRNA1, precrRNA2 o crRNA. (B) Productos de reacciones de digestión obtenidos a partir de mezclas de reacción con distinta composición. Todas las muestras incluyen un sustrato de ADNbc con secuencia diana, bien con una secuencia en la región PAM compatible (5'-CTTG-3') con el corte por la proteína (+) o con una secuencia incompatible con la actividad ADNasa guiada de la proteína (5'-AATC-3') en dicha región (-). (C) Muestras correspondientes a reacciones de digestión en condiciones estándar, salvo por la concentración de AlCas12a. (D) Muestras correspondientes a digestiones en condiciones estándar salvo por los tiempos de incubación. (E) Muestras correspondientes a reacciones de digestión en condiciones estándar salvo por la temperatura (entre 15ºC y 50ºC) y el tiempo (1 hora) de incubación. (F) Muestras de reacciones en condiciones estándar con la proteína AlCas12a (WT) y los mutantes dAlCas12a-D831A, dAlCas12a-E921A o dAlCas12a-D1183A, con (+) o sin (-) AlcrRNA.

La Figura 10 muestra ejemplos de electroforesis en gel de agarosa de los productos de reacciones de digestión de ADNcs catalizadas por AlCas12a. Las reacciones se llevaron a cabo en condiciones estándar excepto cuando se especifica alguna diferencia. (A) Muestras tomadas a diferentes tiempos de incubación a partir de reacciones de digestión de M13mp18 no diana, catalizadas por LbCas12a y AlCas12a en presencia (+) o ausencia (-) de activador, su correspondiente crRNA (LbcrRNA o AlcrRNA, respectivamente) o MgCl2. Se indica el tamaño de las bandas más relevantes del marcador de peso molecular – M – 1 Kb Plus DNA Ladder (Invitrogen). (B) Productos de reacción obtenidos tras distintos tiempos de incubación de M13mp18 con la proteína AlCas12a (WT) y los mutantes dAlCas12a-D831A, dAlCas12a-E921A o dAlCas12a-D1183A. (C) Productos de reacción obtenidos tras distintos tiempos de incubación de AlCas12a con oligonucleótidos no diana de 67 nt (Oligo67) o de 100 nt (Oligo100). (D) Productos de reacción obtenidos tras distintos tiempos de incubación con un oligonucleótido no diana de 100 nt (Oligo100), de AlCas12a o AlCas12a activada tras la escisión guiada por AlcrRNA de ADNbc diana.

La Figura 11 muestra el mapa del plásmido pMML40bis, que incluye un gen que confiere resistencia a cloranfenicol (CmR), el gen *alcas12a-his*, que codifica la proteína AlCas12a fusionada a una etiqueta hexahistidina N-terminal (AlCas12a-His), modificado para adecuar su secuencia al uso de codones de *E. coli*, y una agrupación CRISPR generada por duplicación espontánea de la región que codifica un AlcrRNA dirigido al gen cromosómico *pyrF*. La transcripción del gen *alcas12a-his* está bajo el control del promotor BAD inducible por arabinosa, y finaliza una la región terminadora rrnB T1 T2. La

transcripción de la agrupación CRISPR está controlada por un promotor constitutivo (Part:BBa_J23101). El plásmido incluye además el gen *araC* bajo el control de su propio promotor, que codifica la proteína AraC (regulador del promotor BAD).

La Figura 12 muestra un ejemplo de resultados de electroforesis en gel de agarosa de los productos de PCR del gen *pyrF* tras la transformación de células *E. coli* que expresan el sistema de recombinación Lambda Red con un molde de ADNbc y un plásmido (pMML40bis) que codifica AlCas12a-His y un crRNA con una repetición y espaciador (que coincide con una secuencia *pyrF* con una PAM compatible) duplicados (+ AlCas12a), o un plásmido (pMML41) con la región que codifica el AlcrRNA con una repetición y el mismo espaciador sin *alcas12a* (- AlCas12a). Se analizaron colonias transformantes seleccionadas al azar y crecidas en placas con inductor del gen *alcas12a*. Se indica el tamaño de las bandas relevantes del marcador de peso molecular de ADNbc lineal GeneRuler DNA Ladder Mix (Thermo Scientific) (M). Los productos de PCR con una longitud de aprox. de 1,0 kb corresponden al *pyrF* de tipo silvestre (WT) y los amplicones de aprox. 0,5 kb se corresponden con el tamaño del amplificado del gen delecionado (Mutante).

La Figura 13 muestra los halos de inhibición del crecimiento de cultivos de *E. coli* incubados durante 16 h a 37ºC en placas tras depositar en su superficie gotas de 4 µL de diluciones decimales (10^{-1} a 10^{-5}) de una suspensión del fago de ADNcs M13mp18 (M13) o del fago de ADNbc P1-*vir* (P1). Se incluyen cultivos de células de *E. coli* portadoras de un plásmido que codifica la proteína AlCas12a (pMML39), AlCas12a y un AlcrRNA (pMML30), o ninguno de los dos elementos CRISPR (plásmido control pBAD33), crecidas en medio con IPTG, X-gal y arabinosa.

REIVINDICACIONES

1. Una proteína endonucleasa Cas12a que comprende una secuencia aminoacídica con al menos un 70% de identidad de secuencia con SEQ ID NO: 2.

2. La proteína según la reivindicación 1, donde la proteína comprende una secuencia aminoacídica según SEQ ID NO: 2, preferiblemente donde la proteína comprende una secuencia aminoacídica según SEQ ID NO: 5.

3. La proteína según una cualquiera de las reivindicaciones 1 o 2, donde la proteína comprende:
   - una mutación D831A, preferiblemente donde la proteína comprende una secuencia aminoacídica según SEQ ID NO: 8, y/o
   - una mutación E921A, preferiblemente donde la proteína comprende una secuencia aminoacídica según SEQ ID NO: 10, y/o
   - una mutación D1183A, preferiblemente donde la proteína comprende una secuencia aminoacídica según SEQ ID NO: 12.

4. Una secuencia nucleotídica que codifica una proteína según una cualquiera de las reivindicaciones 1 o 2.

5. La secuencia nucleotídica según la reivindicación 4, que comprende una secuencia nucleotídica con al menos un 70% de identidad de secuencia con SEQ ID NO: 1, preferiblemente que comprende una secuencia nucleotídica con al menos un 70% de identidad de secuencia con SEQ ID NO: 3, o más preferiblemente que comprende una secuencia nucleotídica según SEQ ID NO: 4.

6. Una secuencia nucleotídica que codifica una proteína según la reivindicación 3.

7. La secuencia nucleotídica según la reivindicación 6, que comprende:
   - una secuencia nucleotídica según SEQ ID NO: 7;
   - una secuencia nucleotídica según SEQ ID NO: 9; o
   - una secuencia nucleotídica según SEQ ID NO: 11.

8. Un vector de expresión que comprende una secuencia nucleotídica según una cualquiera de las reivindicaciones 4 a 7.

9. Una célula que comprende una proteína según una cualquiera de las reivindicaciones 1 a 3, y/o una secuencia nucleotídica según una cualquiera de las reivindicaciones 4 a 7, y/o un vector según la reivindicación 8.

10. Un sistema CRISPR-Cas que comprende un ARN guía y una proteína según cualquiera de las reivindicaciones 1 a 3.

11. Un sistema CRISPR-Cas según la reivindicación 10, que comprende un ARN guía con un extremo 5’ con una secuencia nucleotídica según SEQ ID NO: 6.

12. Un método para editar un genoma que incluye una secuencia nucleotídica diana, donde dicho método comprende una etapa de:

    - poner en contacto una secuencia nucleotídica diana con un sistema CRISPR-Cas según una cualquiera de las reivindicaciones 10 u 11.

13. Un uso de una proteína según una cualquiera de las reivindicaciones 1 o 2, y/o una secuencia nucleotídica según una cualquiera de las reivindicaciones 4 o 5, y/o un vector de expresión según la reivindicación 8, y/o una célula según la reivindicación 9, y/o un sistema CRISPR-Cas según una cualquiera de las reivindicaciones 10 u 11 para:

    - la edición y/o modificación genética; y/o
    - la degradación de moléculas de ADNcs en soluciones de ácidos nucleicos; y/o
    - la producción de un agente antimicrobiano específico de secuencia.

14. El uso según la reivindicación 13, donde el agente antimicrobiano específico de secuencia es un agente antibacteriano, preferiblemente un agente antibacteriano frente a *Escherichia coli*.

15. Un uso de una proteína según la reivindicación 3, y/o una secuencia nucleotídica según una cualquiera de las reivindicaciones 6 o 7, y/o un vector de expresión según la reivindicación 8, y/o una célula según la reivindicación 9, y/o un sistema CRISPR-Cas según una cualquiera de las reivindicaciones 10 u 11 para:

- la regulación de expresión génica; y/o
- la visualización y/o marcaje de secuencias nucleotídicas concretas; y/o
- la edición y/o modificación genética.

16. Un método para el diagnóstico *in vitro* de una enfermedad caracterizada por un marcador molecular en un sujeto, donde el método comprende una etapa de:
    - poner en contacto una muestra biológica del sujeto con una proteína según una cualquiera de las reivindicaciones 1 a 3, y/o una secuencia nucleotídica según una cualquiera de las reivindicaciones 4 a 7, y/o un vector de expresión según la reivindicación 8, y/o una célula según la reivindicación 9, y/o un sistema CRISPR-Cas según una cualquiera de las reivindicaciones 10 u 11.

17. Una proteína según una cualquiera de las reivindicaciones 1 a 3, y/o una secuencia nucleotídica según una cualquiera de las reivindicaciones 4 a 7, y/o un vector de expresión según la reivindicación 8, y/o una célula según la reivindicación 9, y/o un sistema CRISPR-Cas según una cualquiera de las reivindicaciones 10 u 11 para su uso en medicina.

18. Una proteína según una cualquiera de las reivindicaciones 1 a 3, y/o una secuencia nucleotídica según una cualquiera de las reivindicaciones 4 a 7, y/o un vector de expresión según la reivindicación 8, y/o una célula según la reivindicación 9, y/o un sistema CRISPR-Cas según una cualquiera de las reivindicaciones 10 u 11 para su uso en el tratamiento de una infección, preferiblemente de una infección bacteriana, más preferiblemente de una infección bacteriana donde la bacteria presenta resistencia a al menos un antibiótico.

19. Un método para reducir la susceptibilidad de una bacteria a infección por virus de ADN monocatenario o bicatenario, donde el método comprende:
    - proporcionar una bacteria que comprende una proteína según una cualquiera de las reivindicaciones 1 a 3, y/o una secuencia nucleotídica según una cualquiera de las reivindicaciones 4 a 7; o
    - transformar una bacteria con un vector de expresión según la reivindicación 8.



FIG. 1

```
AlCas12a   GTCTAATTCGTCCAT-CAATTTCTACTAA-CGTAGAT
FnCas12a   GTCTAAGAACTTTAAATAATTTCTACTG-TTGTAGAT
LbCas12a   G-TTTCAAAGATTAAATAATTTCTACTAAGTGTAGAT
Lb2Cas12a  GCTTAGAACATTTAAAGAATTTCTACTA-TTGTAGAT
PdCas12a   GTCAATAAGACTCATTTAATTTCTACTTC-GGTAGAT
MbCas12a   GTCTAACGACCTT-TTAAATTTCTACTGTTTGTAGAT
AsCas12a   GTCAAAAGACCTT-TTTAATTTCTACTC-TTGTAGAT
           *                **********    ******
```

SEQ ID NO: 72
SEQ ID NO: 73
SEQ ID NO: 74
SEQ ID NO: 75
SEQ ID NO: 76
SEQ ID NO: 77
SEQ ID NO: 78

FIG. 2

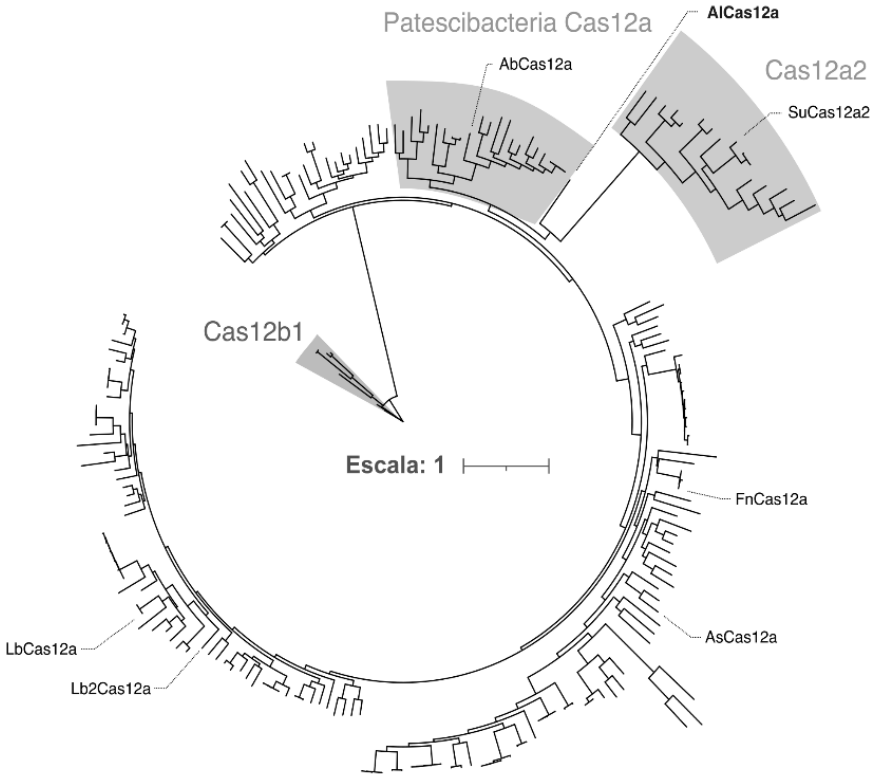


FIG. 3

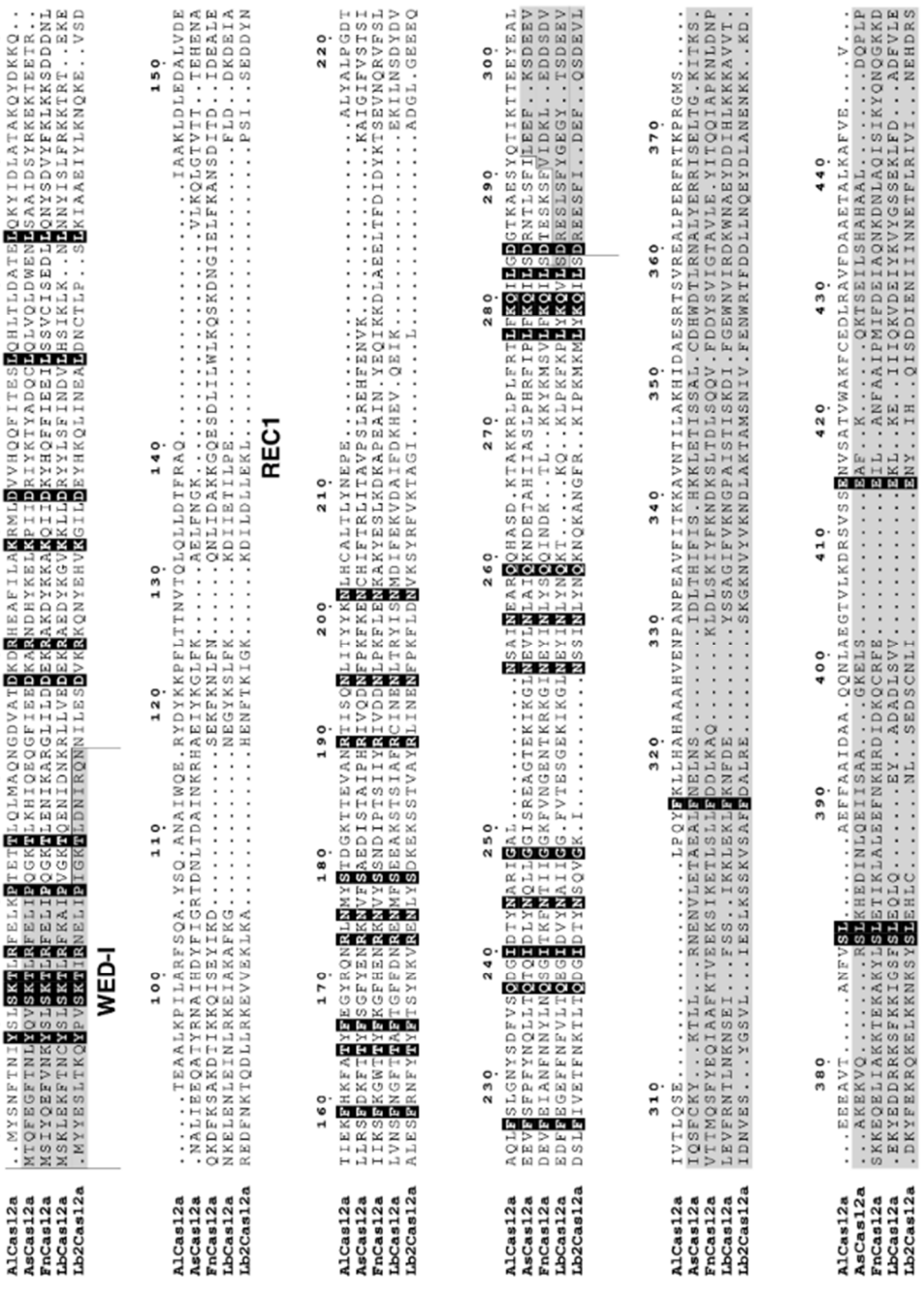

WED-I
REC1
REC2
AlCas12a
AsCas12a
FnCas12a
LbCas12a
Lb2Cas12a

FIG. 4

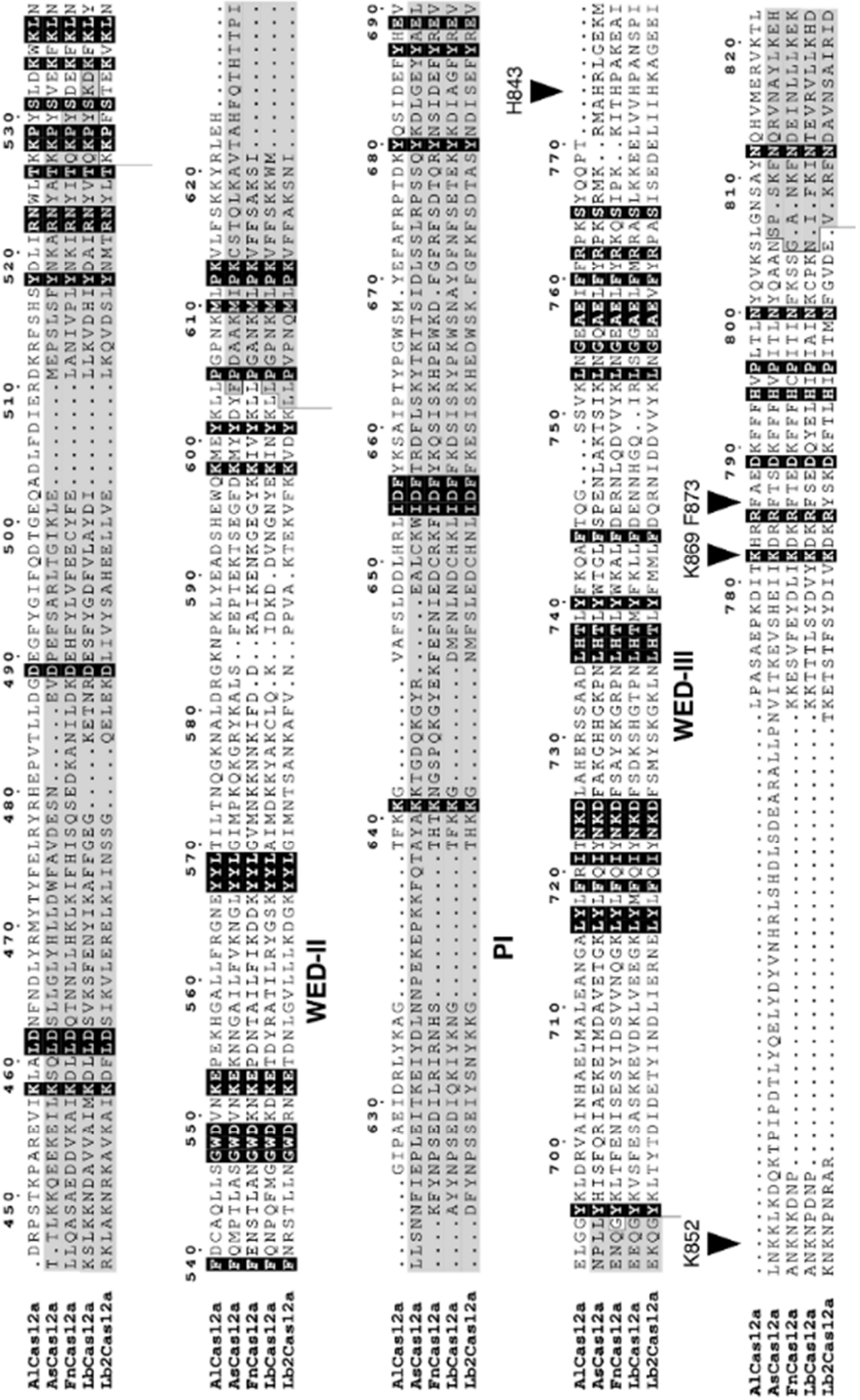


FIG. 4 (Cont.)

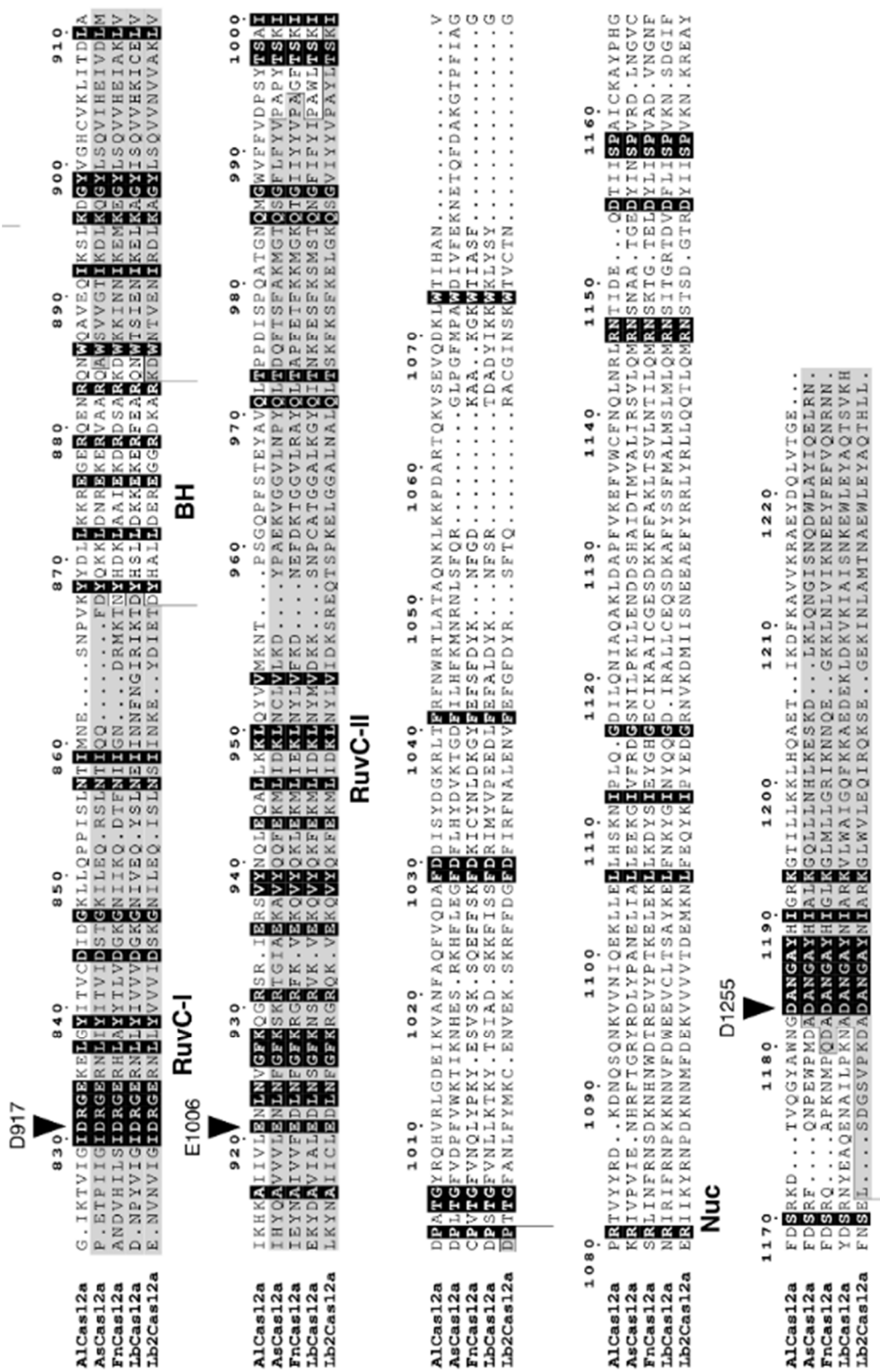

D917
E1006
D1255
RuvC-I
BH
RuvC-II
Nuc
RuvC-III
AlCas12a
AsCas12a
FnCas12a
IbCas12a
Ib2Cas12a

FIG. 4 (Cont.)

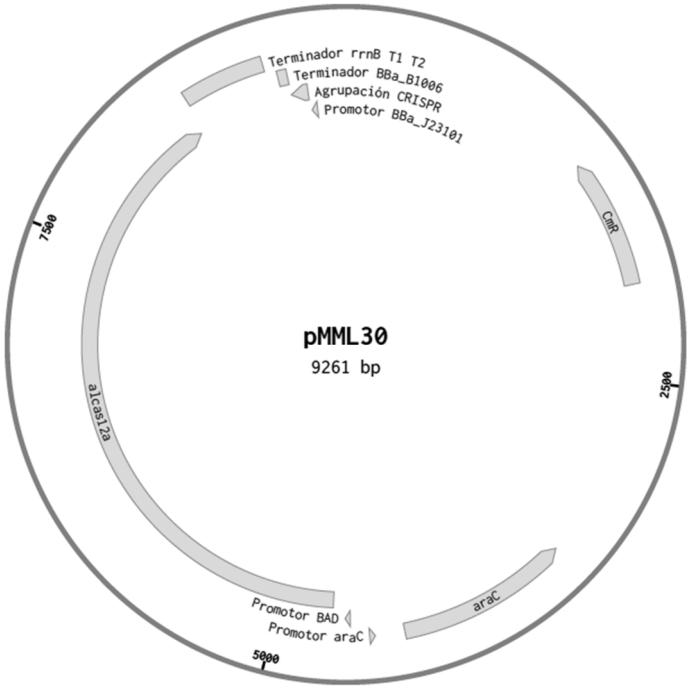

Terminador rrnB T1 T2
Terminador BBa_B1006
Agrupación CRISPR
Promotor BBa_J23101
CmR
pMML30
9261 bp
alcas12a
Promotor BAD
Promotor araC
araC
7500
2500
5000

FIG. 5

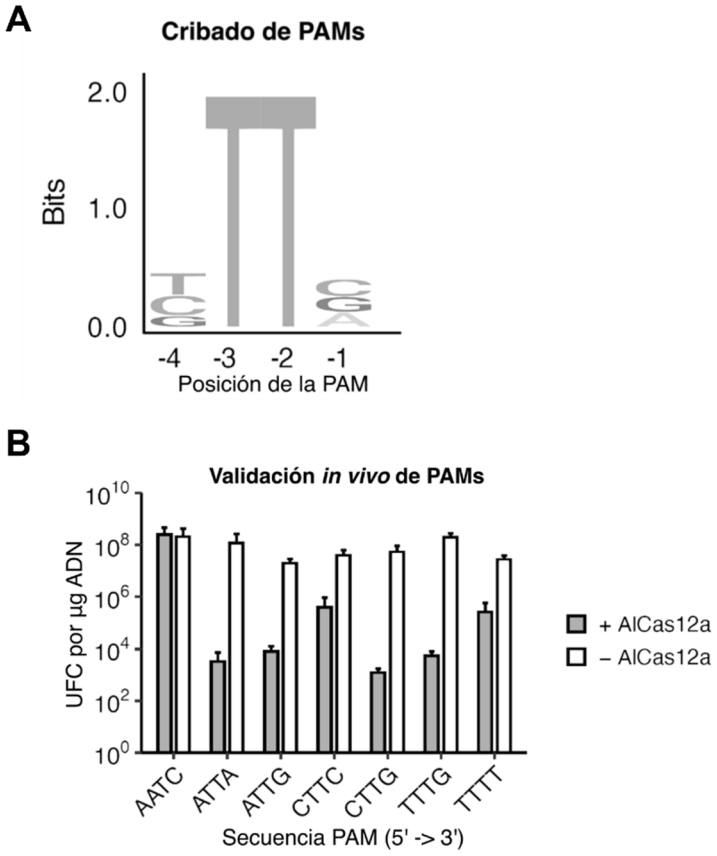

A
Cribado de PAMs
Bits
2.0
1.0
0.0
-4
-3
-2
-1
Posición de la PAM
B
Validación *in vivo* de PAMs
UFC por µg ADN
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
10^{0}
AATC
ATTA
ATTG
CTTC
CTTG
TTTG
TTTT
Secuencia PAM (5' -> 3')
+ AlCas12a
– AlCas12a

FIG. 6

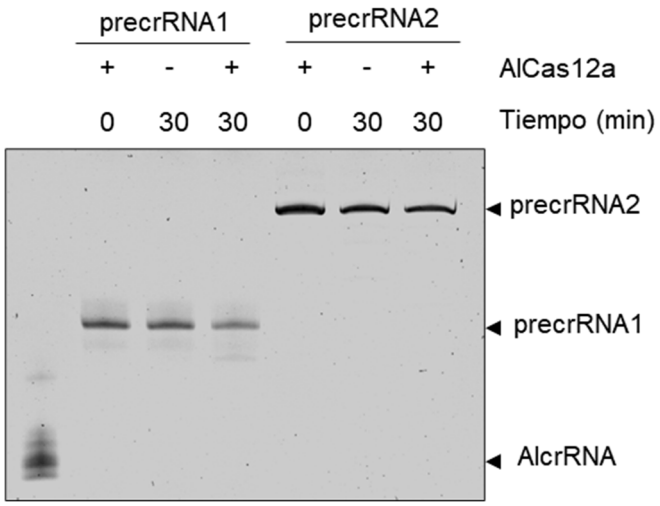

precrRNA1
precrRNA2
+ - + + - +
AlCas12a
0 30 30 0 30 30
Tiempo (min)
precrRNA2
precrRNA1
AlcrRNA

FIG. 7

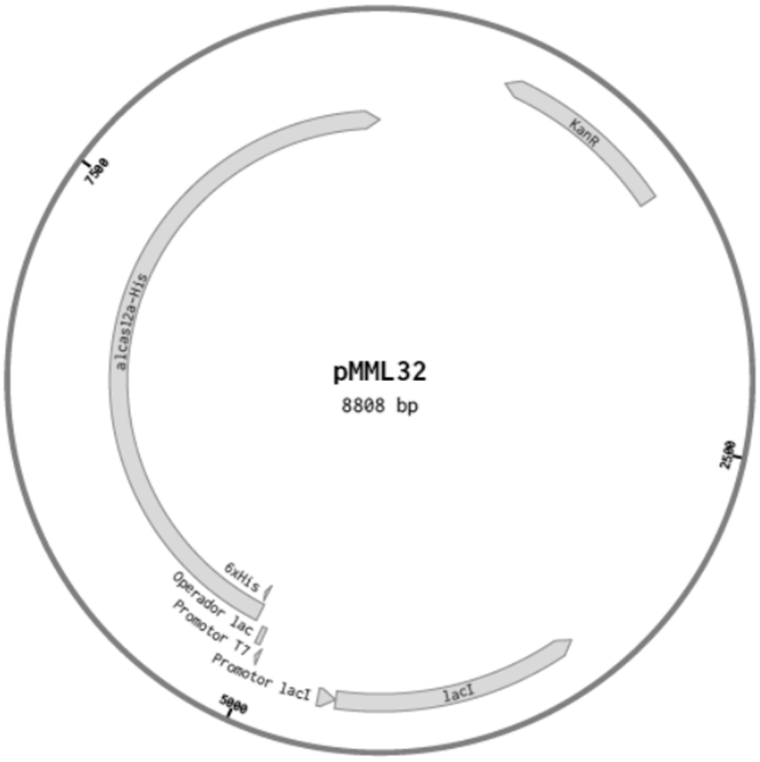

pMML32
8808 bp
KanR
alcas12a-His
6xHis
Operador lac
Promotor T7
Promotor lacI
lacI
7500
2500
5000

FIG. 8

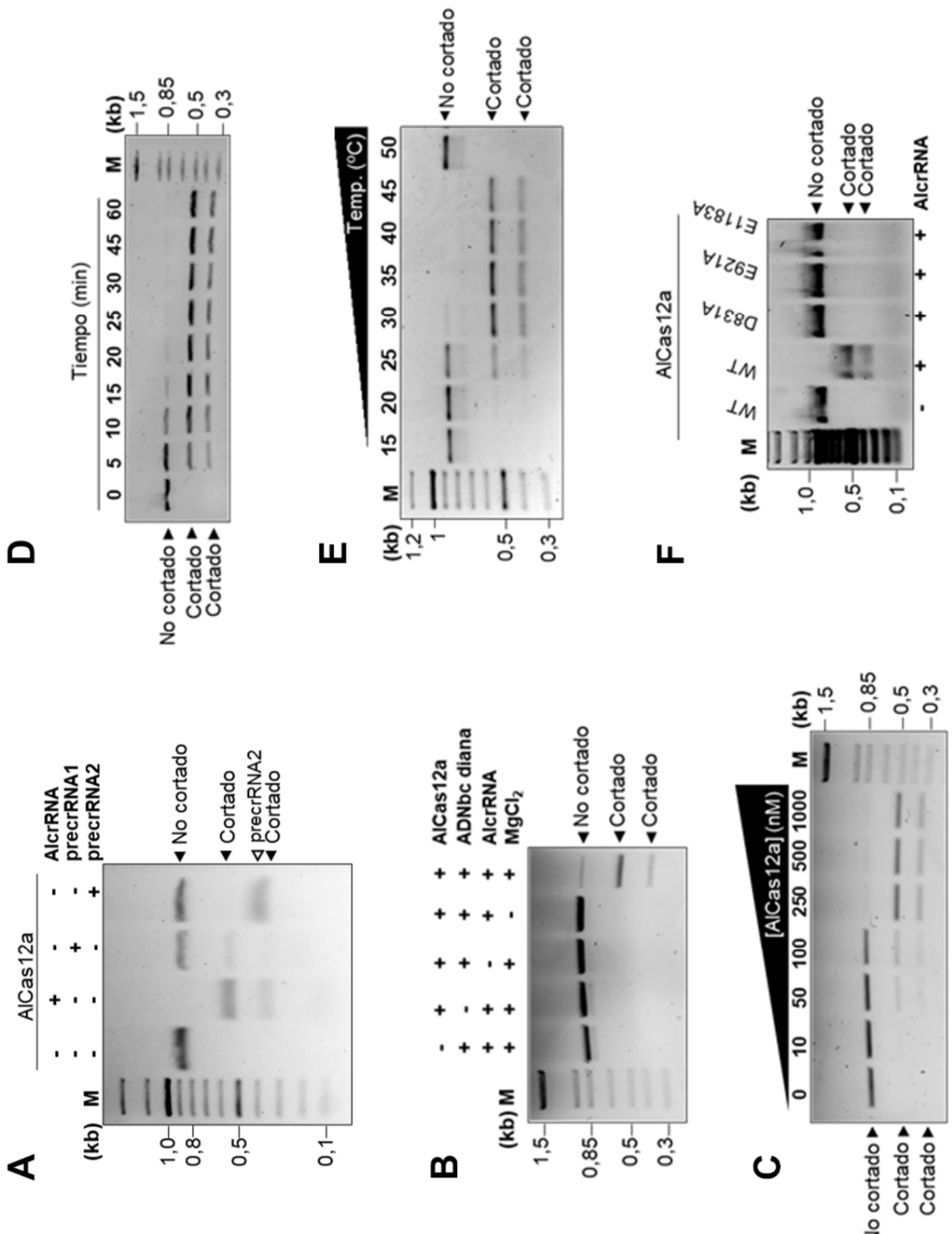

A
AlCas12a
AlcrRNA
precrRNA1
precrRNA2
(kb) M
1,0
0,8
0,5
0,1
No cortado
Cortado
precrRNA2
Cortado
B
AlCas12a
ADNbc diana
AlcrRNA
MgCl2
(kb) M
1,5
0,85
0,5
0,3
No cortado
Cortado
Cortado
C
[AlCas12a] (nM)
0 10 50 100 250 500 1000
M
(kb)
1,5
0,85
0,5
0,3
No cortado
Cortado
Cortado
D
Tiempo (min)
0 5 10 15 20 25 30 45 60
M
(kb)
1,5
0,85
0,5
0,3
No cortado
Cortado
Cortado
E
Temp. (ºC)
M 15 20 25 30 35 40 45 50
(kb)
1,2
1
0,5
0,3
No cortado
Cortado
Cortado
F
AlCas12a
M
WT
WT
D831A
E921A
E1183A
(kb)
1,0
0,5
0,1
No cortado
Cortado
Cortado
AlcrRNA

FIG. 9

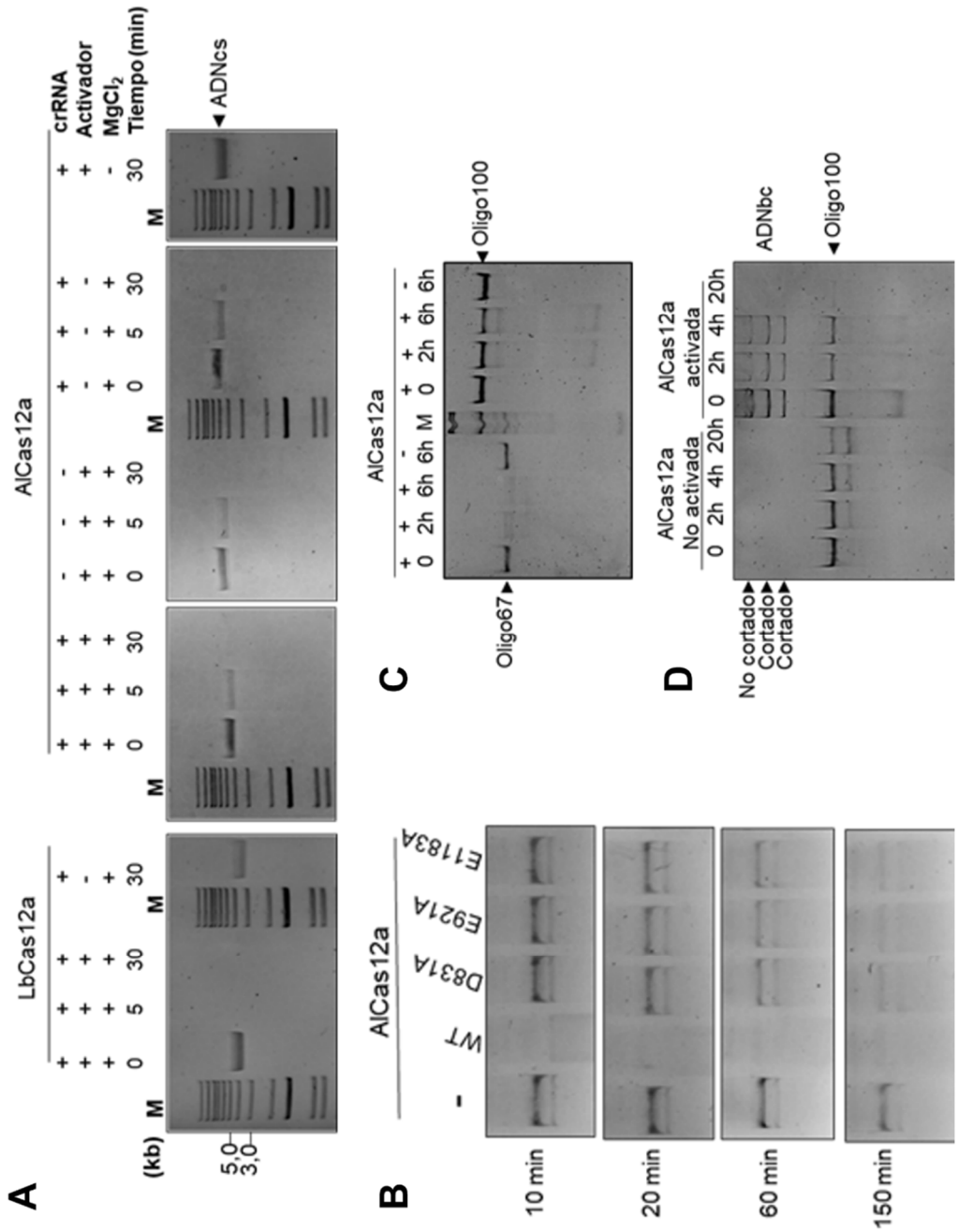

A
LbCas12a
AlCas12a
crRNA
Activador
MgCl2
Tiempo (min)
M
(kb)
5,0
3,0
ADNcs
B
AlCas12a
-
WT
D831A
E921A
E1183A
10 min
20 min
60 min
150 min
C
AlCas12a
Oligo100
Oligo67
D
AlCas12a
No activada
AlCas12a
activada
ADNbc
Oligo100
No cortado
Cortado
Cortado

FIG. 10

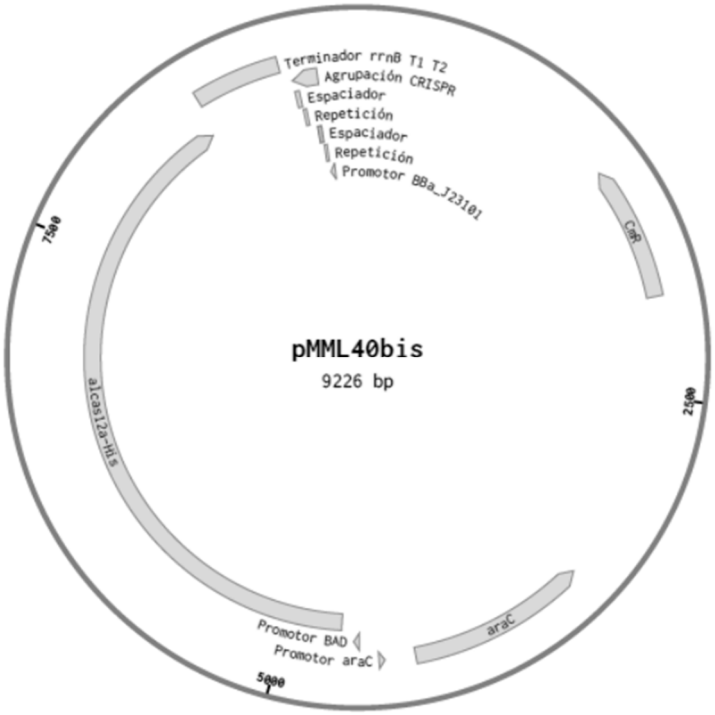


FIG. 11

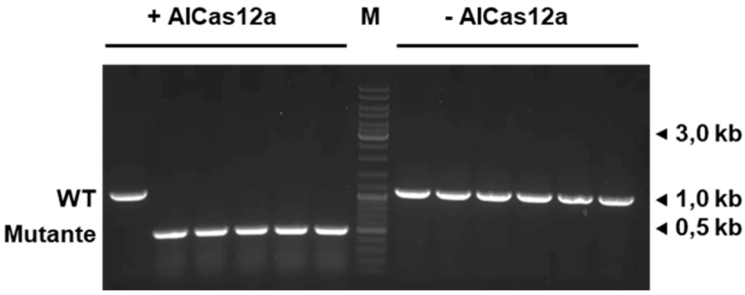


FIG. 12

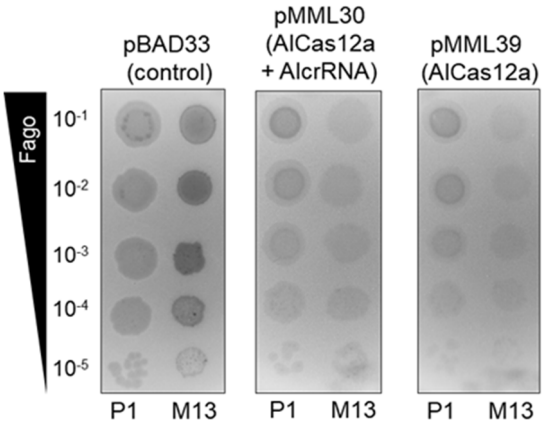


FIG. 13

OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

(21) N.º solicitud: 202530287

(22) Fecha de presentación de la solicitud: 07.04.2025

(32) Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

(51) Int. CI.: ***C12N9/16*** (2006.01)
C12N15/09 (2006.01)

DOCUMENTOS RELEVANTES

Categoría	(56) Documentos citados	Reivindicaciones afectadas
X	TENG, FEI et al. Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds. *Genome Biology*. 05/02/2019, Vol. 20, Nº 1, artículo nº 15, <DOI: 10.1186/s13059-019-1620-8>. Página 2, columna 1, líneas 11 y siguientes; material suplementario 2, tabla S5, crRNA4n78-Nrl; y material suplementario 5: AbCas12a, BsCas12a, OsCas12a, PrCas12a, PsCas12a, PxCas12a, SaCas12a.	1-19
A	US 2023113805 A1 (BRAIN BIOTECH AG) 13/04/2023	1-19
A	BURSTEIN, DAVID et al. New CRISPR–Cas systems from uncultivated microbes. *Nature*. 22/12/2016 (en línea), 09/02/2017 (impreso), Vol. 542, Nº 7640, Páginas 237-241, <DOI: 10.1038/nature21059>	1-19
A	US 2020149021 A1 (INST ZOOLOGY CAS) 14/05/2020	1-19

Categoría de los documentos citados
X: de particular relevancia
Y: de particular relevancia combinado con otro/s de la misma categoría
A: refleja el estado de la técnica
O: referido a divulgación no escrita
P: publicado entre la fecha de prioridad y la de presentación de la solicitud
E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones ☐ para las reivindicaciones nº:

Fecha de realización del informe	Examinador	Página
09.02.2026	Á. Gálvez Flores	1/2

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

C12N

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

AbS colección patentes, AbS colección literatura no patente, INVENES/LATIPAT, Google Patents, Espacenet.