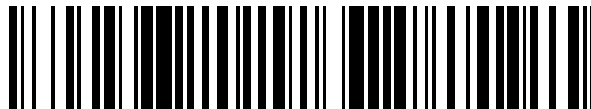


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 522 765**

21 Número de solicitud: 201300203

51 Int. Cl.:

**C12Q 1/68** (2006.01)

**C12N 15/74** (2006.01)

12

SOLICITUD DE PATENTE

A1

22 Fecha de presentación:

**22.02.2013**

43 Fecha de publicación de la solicitud:

**17.11.2014**

71 Solicitantes:

**UNIVERSIDAD DE ALICANTE (100.0%)**

**Carrera San Vicente del Raspeig s/n**

**03690 San Vicente del Raspeig (Alicante) ES**

72 Inventor/es:

**DÍEZ VILLASEÑOR, César;**

**MARTÍNEZ MOJICA, Francisco J.;**

**MARCO GUZMÁN, Noemi;**

**ALMENDROS ROMERO, Cristobal y**

**GARCÍA MARTÍNEZ, Jesús**

54 Título: **Método para detectar inserciones de espaciadores en estructuras CRISPR**

57 Resumen:

La presente invención se refiere a un método para detectar inserciones de espaciadores mediante selección independiente de la interferencia a partir de estructuras artificiales basadas en repeticiones palindrómicas cortas agrupadas regularmente espaciadas (CRISPR) que comprende la inserción de al menos una unidad CRISPR-espaciador, en la estructura artificial dentro de la secuencia que codifica un gen testigo, que restaura la pauta de lectura de traducción, donde la expresión del gen testigo es indicativo de la inserción del espaciador. También se refiere a la estructura artificial utilizada.

ES 2 522 765 A1

## DESCRIPCIÓN

### MÉTODO PARA DETECTAR INSERCIONES DE ESPACIADORES EN ESTRUCTURAS CRISPR

#### Campo de la invención

La presente invención se encuadra en general en el campo de la genética y en particular se refiere a un método para detectar inserciones de espaciadores mediante selección independiente de la interferencia a partir de estructuras artificiales basadas en CRISPR.

#### Estado de la técnica

Los sistemas CRISPR-Cas se han identificado en la mayoría de las arqueas y aproximadamente la mitad de los genomas de bacterias. Cada sistema está constituido por una o varias agrupaciones de repeticiones de ADN denominadas CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) y un conjunto de genes que codifican las proteínas Cas (CRISPR associated) funcionalmente relacionados con la actividad de las CRISPR. 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 localizados en elementos genéticos móviles, tales como plásmidos y virus. 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 RNA primario (pre-crRNA) que abarca la totalidad de la misma.

La identidad de los genes *cas* asociados a una agrupación CRISPR puede variar de manera considerable entre sistemas, en base a la cual los CRISPR-Cas se clasifican en 3 tipos principales, cada uno de los cuales incluye varios subtipos. Tan solo *cas1* y *cas2* están presentes en todos los subtipos identificados hasta la fecha.

Los sistemas CRISPR-Cas producen la degradación (interferencia) de ADN exógeno conteniendo secuencias complementarias a los espaciadores. Un sistema CRISPR-Cas activo puede insertar nuevos espaciadores (proceso conocido como adquisición y catalizado por Cas1 y Cas2) entre el líder y la repetición contigua. La inserción conlleva la duplicación de la repetición adyacente al sitio de inserción. La adquisición de un nuevo espaciador proporciona inmunidad frente a moléculas conteniendo la secuencia de la cual deriva. Para ello, en un primer paso del proceso de interferencia los pre-crRNAs son procesados dando lugar a moléculas crRNA conteniendo un único

5 espaciador. Cada uno de estos crRNAs aparea con su secuencia complementaria en una molécula de ADN reclutando a continuación proteínas Cas específicas que producen su degradación. Además de una función como sistema inmune, se ha propuesto una implicación de los CRISPR-Cas en regulación génica y reparación de ácidos nucleicos.

10 La inserción de nuevas unidades CRISPR-espaciador es un evento poco frecuente y cuya detección requiere realizar un escrutinio al azar de las agrupaciones CRISPR de un gran número de clones (por ejemplo mediante amplificación por PCR). Tan solo cuando dicha adquisición produzca un nuevo patrón de interferencia se podrían seleccionar las células adaptadas, requiriendo para ello condiciones experimentales que permitan dicha interferencia y excluyendo la detección de inserciones "silenciosas".

15 La disponibilidad de una herramienta para detectar inserciones de espaciadores mediante una selección independiente de la interferencia resulta por tanto ventajosa respecto a sistemas no seleccionables.

### **Descripción de la invención**

20 Así pues la presente invención en un primer aspecto se refiere a un método para detectar inserciones de espaciadores, de ahora en adelante método de la presente invención, mediante selección independiente de la interferencia a partir de estructuras artificiales basadas en repeticiones palindrómicas cortas agrupadas regularmente espaciadas (CRISPR) que comprende la inserción de al menos una unidad CRISPR-espaciador, en la estructura artificial dentro de la secuencia que codifica un gen testigo, que restaura la pauta de lectura de traducción, donde la expresión del gen testigo es indicativo de la inserción del espaciador.

25 En una realización más en particular la unidad CRISPR-espaciador insertada comprende un número de nucleótidos no múltiplo de 3.

En una realización más en particular, el espaciador de la unidad CRISPR-espaciador del método de la presente invención comprende un codón de terminación. Más en particular el codón de terminación es críptico.

30 En una realización más en particular, la estructura artificial del método de la presente invención comprende la secuencia identificada como SEQ ID NO: 103, en una realización más en particular, la estructura artificial del método de la presente

invención se corresponde con la fusión traduccional de un *cassette* CRISPR-espaciador-CRISPR-líder y la secuencia codificante del gen de la cloranfenicol acetil-transferasa (*cat*), donde los nucleótidos del 96-124 y 156-165 se corresponden con las unidades CRISPR y la región líder (nucleótidos 165-234) del *cassette* insertado 5' de la región codificante de la proteína CAT (nucleótidos 275-953) que se encuentra fuera de fase respecto al inicio de la traducción.

En una realización más en particular de la presente invención, el espaciador es seleccionado de entre las secuencias identificadas como SEQ ID NO: 1- SEQ ID NO: 89.

En un segundo aspecto, la presente invención se refiere a una estructura artificial, de ahora en adelante, estructura artificial de la presente invención) identificada caracterizada por que comprende la secuencia identificada como SEQ ID NO: 103.

En un tercer aspecto, la presente invención se refiere al uso de la estructura artificial de la presente invención para la detección de inserciones de espaciadores.

#### 15 Descripción de las figuras

Figura 1: Representación esquemática de la integración en los *cassettes* testigo clonados en los plásmidos pCSIR y la estrategia usada para detección específica de integración. A. *Cassette* de integración en pCSIR-A o pCSIR-T consta de un líder y dos duplicones CRISPR (CR) interespaciados por un espaciador (Sp-1) fusionado aguas arriba a un fragmento del gen *lacZ- $\alpha$*  (*lacZ $\alpha$ '*) y, aguas abajo la secuencia completa del gen *cat* fuera de fase (*cat* +2). Se indican el punto de unión al ribosoma (rbs) y el codón de inicio de la traducción *lacZ- $\alpha$*  (punta de flecha blanca). La traducción de los transcritos generados desde el promotor *lac* ( $P_{lac}$ , representado con una flecha) termina en el líder (punta de flecha negra). Aguas abajo del gen *cat* se localiza un promotor de T7 ( $P_{T7}$ ), que inicia la transcripción en sentido opuesto (la flecha apunta en el sentido de la transcripción). Se muestran las secuencias de los fragmentos en la región líder-*cat* de pCSIR-T y pCSIR-A. Los nucleótidos idénticos a la variante de líder T o A correspondiente están sombreados, los codones de terminación y la caja -10 subrayados y etiquetados (el punto de inicio se indica con una flecha que apunta en la dirección de transcripción). Para evitar codones de terminación, se reemplazaron las timinas en las posiciones 42 y 70 del líder de pCSIR-T por citosina y adenina respectivamente (ver el cebador TLR en la Tabla 2). La traducción del

transcrito generado de *P<sub>lac</sub>* acaba en el líder en caso de delección (*cat* seguiría fuera de fase, *cat*+1) (B), en el espaciador duplicado (Sp-1) en caso de duplicación CRISPR-espaciador (C), o en el codón de terminación de la secuencia codificante de *cat* en el caso de la inserción (D) de un espaciador nuevo (Sp-2).

5 Figura 2: Distribución, en pCas1-2[K], de las secuencias correspondientes a espaciadores muestreados en los ensayos de adquisición. Los espaciadores se muestran como flechas negras apuntando al líder tal y como se insertaron en el *cassette* de integración. Se muestra el origen de replicación (CDF ori), los genes codificando el represor del operón *lac* (*LacI*), la proteína de resistencia a estreptomicina (*SmR*), y el promotor T7-*lac* (*P<sub>T7</sub>*) que promueve la transcripción de la  
10 variante K de los genes *cas1-2* (*cas1-K* y *cas2-K*).

Figura 3: WebLogos generados por el alineamiento de las regiones de protoespaciadores de CRISPR2 de *E. coli*. Las secuencias conteniendo protoespaciadores (posiciones -1 a -33) y los dos nucleótidos adyacentes de la región  
15 PAM (posiciones 0 y 1) estaban orientados del mismo modo que los espaciadores correspondientes en la agrupación CRISPR (el extremo 3' hacia el líder). Los logos se obtuvieron a partir de secuencias no CRISPR con una identidad mayor al 90% a espaciadores de genomas de *E. coli*, asociados con el líder-T (A) o el líder-A (B), así como protoespaciadores correspondientes a espaciadores encontrados en los  
20 experimentos de adquisición realizados en cepas alojando los genes Cas-EK (C) o Cas-EO (D).

Figura 4: Mecanismo hipotético de inserción de espaciadores. Se distinguen los tres pasos principales en el proceso de integración de espaciadores: (i) corte en el punto de inserción, (ii) integración del espaciador y (iii) duplicación CRISPR. A. Las  
25 inserciones aberrantes (secuencias CRISPR-espaciador atípicas) obtenidas en ensayos de adquisición con pCSIR-A y el plásmido pCas1-2[K] pueden explicarse con un nick inicial (triángulo negro; nuestros datos no permiten precisar en qué cadena) entre las posiciones primera y segunda del líder (cursiva) y un corte secundario (triángulo blanco) en la hebra complementaria a una distancia fija (28 nt) hacia el  
30 duplicón CRISPR adyacente (posiciones sombreadas). Después se produce la ligación del espaciador entrante de 33 pb (indicado con letras N) a los extremos libres de los puntos de inserción, los huecos se rellenan por polimerización del DNA (negrita), generándose los duplicones CRISPR de 26 pb (recortados en 2 pb) manteniéndose la

periodicidad típica de 61 pb. B. En un proceso normal de inserción, el primer punto de corte se produce en la unión entre la CRISPR y el líder, y el segundo a una distancia fija de 28 nt (es decir, en el límite entre el espaciador y la CRISPR), lo que conduce a la duplicación de una unidad CRISPR completa y mantiene la periodicidad de 61 pb.

5 Se indican las longitudes de relevancia.

Figura 5: muestra agrupaciones CRISPR2.1 de *E.coli*.

### Descripción detallada de la invención

#### *Ejemplo 1: Ensayo de adquisición de espaciadores*

Las cepas de *E. coli* utilizadas como hospedadores en los ensayos de integración  
 10 fueron K12 str. MG1655, O157:H7 str. EDL931 (Colección Española de Cultivos Tipo, CECT 4267) y BL21-AI (Novagen). Las tres cepas contienen los *loci* CRISPR2.2 y CRISPR2.3-líder. Las cepas de K12 contienen además una agrupación CRISPR2.1 compuesta de duplicones CRISPR-T asociados con una líder de tipo T y todo el conjunto de genes *cas-EK*. El *locus* CRISPR2.1 de *E. coli* O157:H7 se compone de  
 15 repeticiones y líder tipo A, y adjunta el conjunto completo de genes *cas-EO*. El *locus* CRISPR2.1 de cepas derivadas de BL21 consta de duplicones CRISPR-A, y carece tanto de líder como de genes *cas*. El genoma de BL21-AI codifica la RNA polimerasa de T7 inducible por arabinosa, mientras que las otras cepas carecen de genes homólogos a ella.

20 El *locus* CRISPR2.1 de *E. coli* ECOR69, usado como molde para la amplificación por PCR del *cassette* testigo de integración clonado en pCSIR-A, contiene CRISPRs y líder tipo A (Diez-Villasenor C, Almendros C, Garcia-Martinez J, Mojica FJ. Diversity of CRISPR loci in Escherichia coli. Microbiology 2010).

Los plásmidos desarrollados en este estudio y los detalles de su construcción están  
 25 indicados en la Tabla 1 y los cebadores respectivos utilizados en la Tabla 2.

El plásmido de 9Kb pWUR339 (proveído por el laboratorio de John van der Oost) se deriva del vector de bajo número de copias pCDF-1b (Novagen) y contiene el operón Cascada-*cas1-cas2* de *E. coli* K12 bajo el promotor T7-*lac*. Las construcciones se realizaron clonando fragmentos de DNA obtenidos por PCR con cebadores  
 30 conteniendo puntos de corte de enzimas de restricción (ver Tablas 1 y 2) o mediante el método de extremos con 3'-T sobresalientes.

Tabla 1. Plásmidos construidos durante el estudio.

Nombre	Tamaño (bp)	CTT/Kb	CAT/Kb	Vector	Sitio(s) de Clonación	Inserto(s)	Cebador	Patrón (Template)
pCR2.1-Cm	4,592	35.9	27.9	pCR2.1	XhoI+XbaI	caf ORF	CATF1/CATR1	pKK232-8
pCSIR-A	4,662	35.8	27.9	pCR2.1-Cm	HindIII+XhoI	cassete CRISPR tipo A	AS2F /ALR	ECOR69
pCSIR-T	4,770	36.1	27.7	pCR2.1	EcoRI XhoI+XbaI	cassete CRISPR tipo T caf ORF	TS2F/ TLR	K12
pCas1-2[O]	4,777	26.8	28.7	pCDF-1b	NcoI+NofI	cas1- cas2 variante E-O	CAS1EOF /CAS2EOR	O157:H7
pCas1-2[K]	4,711	29.9	30.1	pCDF-1b	NcoI+NofI	cas1- cas2 variante E-K	CAS1EKF/ BG2375 <sup>30</sup>	pWUR399

**Tabla 2.** Cebadores usados en la presente invención.

Nombre	Secuencia(5'-3') <sup>c</sup>	Características
CATF1	SEQ ID NO:90	<i>sitio XhoI</i> incluye el codón de inicio <i>cat</i> mutado (CTC)
CATR1	SEQ ID NO:91	el codón de stop <i>cat</i> (TAA) precede al <i>sitio XbaI</i>
CAS1EKF	SEQ ID NO:92	el <i>sitio NcoI</i> incluye el condón de inicio (ATG)
BG2375	SEQ ID NO:93	el codón de stop <i>cas2</i> (TGA) precede al <i>sitio NotI</i>
CAS1EOF	SEQ ID NO:94	el <i>sitio NcoI</i> incluye el codón de inicio (ATG)
CAS2EOR	SEQ ID NO:95	el codón de stop <i>cas2</i> stop TAG) precede al <i>sitio NotI</i>
TS2F	SEQ ID NO:96	
TLR	SEQ ID NO:97	Incluye mutaciones para evitar codones de stop
AS2F	SEQ ID NO:98	<i>sitio HindIII</i> site
ALR	SEQ ID NO:99	<i>sitio XhoI</i>
T7t	SEQ ID NO:100	
T7	SEQ ID NO:101	
M13R	SEQ ID NO:102	

- El plásmido pCR2.1-Cm se obtuvo clonando el gen de la cloranfenicol acetiltransferasa (*cat*) de pKK232-8 en el vector de alto número de copias pCR2.1 (Invitrogen; origen de replicación pUC). pCSIR-A deriva de pCR2.1-Cm mediante la inserción de una
- 5 secuencia que contiene 43 pb del líder2.1 y el fragmento CRISPR-espaciador-CRISPR adyacente de la cepa ECOR69. Para construir pCSIR-T se clonó en el vector pCR®2.1 por el método de 3'-T sobresalientes (TA Cloning® Kit de Invitrogen) un segmento de 69 pb del líder2.1 y el fragmento CRISPR-espaciador-CRISPR adyacente amplificados
- 10 de *E. coli* K12 str. MG1655, y la subsecuente inserción de un amplicón conteniendo el gen *cat* del mismo modo que para la construcción de pCR2.1-Cm. pCas1-2[K] y pCas1-2[O] se obtuvieron mediante inserción en el vector pCDF-1b de amplicones conteniendo los genes *cas1* y *cas2* de pWUR399 y los de *E. coli* O157:H7 str. EDL931 respectivamente.
- 15 Para clonar en los extremos 3'-T sobresalientes del vector pCR®2.1 linearizado, los productos de PCR se amplificaron con *Taq* polimerasa recombinante (Invitrogen). Para el resto de los casos (Tabla S4) se obtuvieron con Expand High Fidelity PCR System (Roche). Para realizar las PCRs se utilizó un termociclador Mastercycler Gradient (Eppendorf). Los productos de PCR se purificaron y digirieron posteriormente con



enzimas de Thermo Scientific según las recomendaciones de la herramienta web Double Digest. Las ligaciones se realizaron con DNA ligase de T4 (New England Biolabs). Tras la ligación, se transformó la cepa TOP10 (Invitrogen) o NovaBlue (Novagen) con los plásmidos derivados de pCR2.1 o pCDF-1b respectivamente, y se verificaron las construcciones mediante secuenciación con los cebadores T7 (pCR2.1 y pCDF-1b) y M13R (pCR2.1) o T7t (pCDF-1b).

Los líderes de las cepas ECOR se determinaron mediante secuenciación de productos de PCR como se describe en Diez-Villasenor C et al. Diversity of CRISPR loci in *Escherichia coli*. Microbiology 2010. Las CRISPR y los líderes de genomas de *E. coli* disponibles se identificaron con el programa CRISPRFinder.

Para la identificación de los motivos PAM de las CRISPR2 asociados con las variantes CRISPR-líder A o T, las regiones de los protoespaciadores con una identidad de al menos 30 nt con espaciadores de los sistemas CRISPR-Cas I-E de cepas de *E. coli* se buscaron con el programa BLASTn contra la base de datos Nucleotide Collection (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) y alineadas usando el extremo proximal al líder como referencia. De modo similar, los motivos PAMs experimentales se determinaron mediante alineamiento de secuencias con un 100% de similitud a espaciadores adquiridos en los ensayos de integración y detectados en los replicones del hospedador correspondiente usando BLASTn o Geneious R6 creado por Biomatters Ltd. Las hebras de DNA portadoras de la secuencia complementaria a la del espaciador correspondiente en el crRNA se alinearon usando WebLogo para obtener logos de secuencia como se describe en Mojica FJM, et al. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 2009; 155:733-40.

El DNA genómico y plasmídico se purificó de células crecidas en medio LB (triptona 10 g/l, extracto de levadura 5 g/l, y NaCl 10 g/l) usando Wizard® Genomic DNA Purification Kit (Promega) y High Pure Plasmid Isolation Kit (Roche) respectivamente. Los productos de PCR y fragmentos de restricción se purificaron con GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare).

Las secuenciaciones se realizaron con el Big Dye Terminator Cycle Sequencing Kit en un secuenciador ABI PRISM 310 DNA Sequencer (Applied Biosystems) siguiendo las instrucciones del fabricante (Servicios Técnicos de Investigación, Universidad de Alicante, Spain).

Las transformaciones se efectuaron por electroporación (2.45 KV, 25  $\mu$ F, 200 $\Omega$ ) en un Electroporator 2510 (Eppendorf). Las células electrocompetentes se prepararon siguiendo el procedimiento descrito por Shi et al. Enhancing *Escherichia coli* electrotransformation competency by invoking physiological adaptations to stress and modifying membrane integrity. Anal Biochem 2003; 320:152-5. Las colonias transformadas se seleccionaron en placas de LB con agar conteniendo los antibióticos apropiados.

Como paso previo a los ensayos de adquisición, para confirmar la sensibilidad a cloranfenicol, se sembraron placas Petri de LB suplementadas con 25  $\mu$ g/ml de cloranfenicol (Sigma) con colonias de *E. coli* llevando el plásmido testigo de integración y otro con gen(es) *cas*. Para cada ensayo de adquisición se utilizó un clon sensible distinto, que se creció 12 h a 37°C con agitación (150 rpm) en LB líquido suplementado con los antibióticos necesarios para la selección de plásmidos (ampicilina 100  $\mu$ g/ml - Sigma – o estreptomycin 20  $\mu$ g/ml - Sigma - para pCSIR y los plásmidos conteniendo genes *cas* respectivamente). Si fue necesario se realizaron ciclos adicionales (máximo de seis) de 12h de crecimiento (diluciones 1:300 del cultivo en medio fresco e incubación en las mismas condiciones), hasta que aparecieran colonias resistentes a cloranfenicol (en expectativa con una unidad espaciador-CRISPR extra) tras sembrar muestras en medio LB sólido conteniendo cloranfenicol 25  $\mu$ g/ml. Para evitar muestrear colonias descendientes de la misma célula adaptada, solo se seleccionó para su posterior análisis una colonia resistente en cada ensayo. La inserción se valoró en primera estancia mediante PCR del *cassette* integrado con los cebadores T7 y M13 que hibridan en las secuencias flanqueantes del vector (ver Tabla S7). Normalmente se observan fragmentos de PCR de mayor tamaño junto con amplicones del tamaño original, estos últimos se deben a copias de la construcción inicial. En estos casos, para seleccionar células enriquecidas con el plásmido portador de la inserción, se realizaron tres ciclos adicionales de crecimiento (diluciones 1:5000, incubadas a 37° con agitación 24h) en medio LB con concentraciones crecientes de cloranfenicol (75  $\mu$ g/ml, 150  $\mu$ g/ml and 200  $\mu$ g/ml). Finalmente se extrajeron los plásmidos y se secuenciaron con el cebador M13R. El primer ciclo de crecimiento en el que se detectan clones adaptados proporciona una estimación de la eficiencia de adquisición.

Para calcular la probabilidad de que una integración no permita la traducción del gen *cat* debido a la presencia de codones de terminación en fase con el codón de inicio ATG en los plásmidos testigo se asumió que la frecuencia de los cuatro nucleótidos en el DNA donador es equivalente. Por ello la probabilidad de los tripletes de terminación (TAA, TAG o TGA) es  $3 \times 1/4^3 = 3/64$  y de que un codón no sea de terminación  $1 - 3/64 = 61/64$ . En cada espaciador de 33 nt insertado hay 10 trinucleótidos que pueden ser de parada de la traducción. Por ejemplo, en el caso de una única inserción, los tres nucleótidos que restan completarían tripletes en la CRISPR adyacente: el nucleótido en el extremo distal al líder del espaciador contribuye con el triplete TCN y el opuesto con 2 nucleótidos al codón NCG en los extremos de los duplicones CRISPR flanqueantes (nucleótidos subrayados). Por tanto la probabilidad de  $n$  inserciones sin que se interrumpa la traducción es  $(61/64)^{10n}$  (es decir, 0,62 para una sola inserción). Si se tiene en cuenta que solo el 95% de los espaciadores encontrados en la naturaleza tienen 33 nt de longitud, para una única inserción ( $n=1$ ), la probabilidad de detección de la integración es  $(61/64)^{10} \times 0.95 = 0.59$ .

Para comprobar la adquisición de espaciadores en el sistema CRISPR-Cas I-E de *E. coli*, desarrollamos un ensayo basado en la selección positiva de clones que se vuelven resistentes a cloranfenicol tras la expansión de una agrupación CRISPR en plásmidos recombinantes (pCSIR-A y pCSIR-T; ver Figura1). Ambos plásmidos llevaban un *cassette* consistente en una agrupación CRISPR2 de dos repeticiones seguidos por un fragmento del líder asociado (las 43 y 69 bases proximales en pCSIR-A y pCSIR-T respectivamente; ver Figura 1A). Por tanto la caja TATA -10 requerida para la transcripción de la agrupación está presente en pCSIR-T pero ausente en pCSIR-A (Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K. Transcription, processing and function of CRISPR *cassettes* in *Escherichia coli*. *Molecular Microbiology* 2010; 77:1367-79; Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R. Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Molecular Microbiology* 2010; 75:1495-512). Las repeticiones y líder en pCSIR-A pertenecen a la variante A (CRISPR-A y líder-A respectivamente), mientras que pCSIR-T contiene las variantes T de ambos elementos (CRISPR-T y líder T). El *cassette* CRISPR-espaciador-CRISPR-líder está traduccionalmente fusionado al comienzo de *lacZ- $\alpha$*  (*lacZ- $\alpha'$* ) y al gen (*cat*) de la proteína cloranfenicol

acetiltransferasa (o CAT) que se encuentra fuera de fase (+2 del codón de inicio de *lacZ-α* (Fig. 1A). La traducción iniciada en el codón AUG de *lacZ-α* se detiene en el líder. De modo similar, en el caso de una delección producida por recombinación entre las dos unidades CRISPR la traducción se pararía en otro triplete de la región líder (Ver Figura 1B). Para impedir la traducción de CAT debido a una duplicación (en vez de una inserción nueva) de una unidad espaciador-CRISPR, el espaciador existente contiene un triplete de terminación fuera de fase que estaría en fase en el caso de su duplicación (Fig. 1C).

Los experimentos de inserción se realizaron con varias cepas de *E. coli* con distintas combinaciones de pCSIR y plásmidos conteniendo genes *cas*. Para evitar la posibilidad de una adquisición aberrante debido a una gran sobreexpresión de los genes *cas* (los promotores T7-*lac*, inducibles por IPTG se encuentran localizados aguas arriba de los operones *cas* y el gen *lacI* codificante del represor *lac* está incluido en las construcciones (ver Figura 2), todos los ensayos fueron llevados a cabo en la ausencia de IPTG. Por ello, como mucho se esperan pequeños niveles de expresión de los genes *cas* (como en el caso de 'goteo' del promotor).

Los ensayos iniciales de inserción se realizaron con la cepa BL21-AI llevando pCSIR-T y el plásmido pCas1-2[K] (contiene los genes *cas1* y *cas2* de K12). Las cepas derivadas de BL21 carecen de genes *cas* propios y aunque BL21-AI aloja al gen de la RNA polimerasa de T7 bajo el control de un promotor inducible por arabinosa, no se añadió arabinosa al medio de cultivo. En estas condiciones, se encontró un único clon adaptado (la secuencia del espaciador es la SEQ ID NO: 104 idéntica a una secuencia cromosómica adyacente a CTT) tras varios intentos y crecimiento prolongado (seis ciclos de crecimiento). Sin embargo, cuando se utilizó el plásmido pWUR399 (el mismo vector que pCas1-2[K] pero con los genes de la Cascada fusionados transcripcionalmente a *cas1-cas2*) en vez de pCas1-2, se aislaron nueve clones adaptados tras 12 horas de incubación (un ciclo de crecimiento) (Tabla 3).

La combinación de estos resultados indicaron que los genes *cas* se transcribieron en ambas construcciones a pesar de la ausencia de inducción.

Tabla 3: Espaciadores integrados en pCSIR-T durante la propagación en BL21-AI que alberga a pWUR399.

Espaciador (5'-3')	Veces observada	Portador del protoespaciador	PAM
SEQ ID NO:1	2	pWUR399	CTT
SEQ ID NO:2	1	pWUR399	CTT
SEQ ID NO:3	1	pWUR399	AGT
SEQ ID NO:4	1	pWUR399	CTT
SEQ ID NO:5	1	pWUR399	ACA <sup>a</sup>
SEQ ID NO:6	1	pWUR399	CTT
SEQ ID NO:7	1	pWUR399	GCC
SEQ ID NO:8	1	pWUR399	CTT

5 <sup>a</sup>Encontramos Timina en la posición 4 detrás del protoespaciador, dando lugar al desplazamiento en una posición de la PAM putativa CAT.

En un análisis previo comparando el líder con las PAMs de 15 protoespaciadores putativos del CRISPR2 de *E. coli*, se propuso una preferencia por CAT o CTT para las agrupaciones CRISPR asociados con líderes A o T respectivamente. Ahora hemos  
 10 confirmado este sesgo mediante un estudio similar *in silico* realizado con 76 protoespaciadores putativos con identidad superior al 90% a espaciadores CRISPR2 de cepas ECOR (Ochman H, Selander RK. Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* 1984; 157:690-3.) y genomas de *E. coli* secuenciados: 11 protoespaciadores de un total de 21 asociados con el líder-T  
 15 mantienen el motivo PAM CTT y solo uno CAT, respetando la secuencia CTT en la PAM y 34 de los 55 protoespaciadores predichos asociados con un líder A se encuentran junto a la secuencia CAT, mientras que 9 lo están junto al triplete CTT, confirmando CWT como consenso PAM (Fig. 3B).

Aunque parece haber una conexión entre PAM y líder, sugiriendo que este último  
 20 puede tomar parte en el reconocimiento del motivo específico durante la adquisición de espaciadores, también podrían estar implicados otros elementos del sistema CRISPR-Cas. Como se muestra en la Figura 6, en agrupaciones CRISPR2.1 de *E. coli*, repeticiones del tipo A coexisten con un líder-A, y CRISPR-T con un líder-T. Además, en un estudio independiente (datos sin publicar) encontramos que cada tipo  
 25 de líder se corresponde con una determinada variante de genes *cas-E*, diferenciando

dos grupos en la especie definidos por la variante aquí referida como cas-EK (por la cepa K12), con CRISPR/líder tipo T, y la variante cas-EO (por O157:H7) asociada con el tipo A respectivamente.

Para identificar experimentalmente el verdadero motivo de adquisición (denominado SAM, *spacer acquisition motif*) reconocido por cada variante CRISPR/líder/Cas, realizamos ensayos de integración de espaciadores con combinaciones naturales de los elementos respectivos en cepas alojando un sistema CRISPR-Cas equivalente (de este modo se evitan interacciones entre elementos CRISPR-Cas de variantes distintas). Primero se valoró inserción en K12 conteniendo un plásmido testigo de integración (pCSIR-T) con líder-T y CRISPR-T, y un segundo plásmido recombinante con *cas1* (pCas1K), *cas2* (pCas2K) o *cas1* y *cas2* (pCas1-2[K]) de la misma cepa. Se observó integración eficiente de espaciadores (59 inserciones) con pCas1-2[K] (Tabla 4) con 12h de incubación, pero no con pCas1K o pCas2K tras 6 ciclos de crecimiento, probando que se requieren ambos genes *cas* en un plásmido multicopia para detectar adquisición de K12 en nuestro sistema.

Tabla 4: espaciadores integrados en pCSIR-T

Espaciador (5'-3')	Veces observada	Portador del protoespaciador	PAM
SEQ ID NO:1	2	pCas	CTT
SEQ ID NO:2	2	pCas	CTT
SEQ ID NO:4	2	pCas	CTT
SEQ ID NO:9	1	pCas	CTT
SEQ ID NO:10	1	pCas/Cromosoma	AGT
SEQ ID NO:11	1	pCas/Cromosoma	CTT
SEQ ID NO:12	1	Cromosoma	CTT
SEQ ID NO:13	1	pCSIR	CTA
SEQ ID NO:14	1	pCas/Cromosoma	CTT
SEQ ID NO:15	1	pCas	CCT <sup>a</sup>
SEQ ID NO:16	1	pCas/Cromosoma	TCT <sup>a</sup>
SEQ ID NO:17	1	pCas/Cromosoma	CTT
SEQ ID NO:18	1	pCas	CGG
SEQ ID NO:19	2	pCas/Cromosoma	CTT
SEQ ID NO:20	1	pCas	CTT
SEQ ID NO:21	1	pCas	TGG
SEQ ID NO:22	1	pCSIR	CTT
SEQ ID NO:23	1	pCas/Cromosoma	TTC
SEQ ID NO:24	3	pCas/Cromosoma	CTT
SEQ ID NO:25	1	pCas/Cromosoma	CTT
SEQ ID NO:26	2	pCas/Cromosoma	CTT
SEQ ID NO:27	1	pCas	CTT
SEQ ID NO:28	1	pCas/pCSIR/Cromosoma	CCG
SEQ ID NO:29	1	pCas	CTT

SEQ ID NO:30	1	pCas	CTT
SEQ ID NO:31	1	pCas/Cromosoma	ACC
SEQ ID NO:32	2	pCas/Cromosoma	CTT
SEQ ID NO:33	1	pCas/Cromosoma	CTT
SEQ ID NO:34	1	pCas	TTT
SEQ ID NO:35	2	pCas	CTT
SEQ ID NO:36	1	pCas	CTT
SEQ ID NO:37	1	pCas	CTT
SEQ ID NO:38	1	pCas	CTG
SEQ ID NO:39	2	pCas/Cromosoma	CTT
SEQ ID NO:40	1	pCas	CTT
SEQ ID NO:41	1	pCas/Cromosoma	CTT
SEQ ID NO:42	1	pCas	CTT
SEQ ID NO:43	1	pCas	CTT
SEQ ID NO:44	2	pCas	CTT
SEQ ID NO:45	1	pCas	CTT
SEQ ID NO:46	1	pCas	CTT
SEQ ID NO:47	1	Cromosoma	CTA
SEQ ID NO:48	1	pCas/Cromosoma	TGC
SEQ ID NO:49	1	pCas/Cromosoma	CTC
SEQ ID NO:50	1	pCSIR	GCT <sup>a</sup>
SEQ ID NO:51	1	pCas/Cromosoma	CTG
SEQ ID NO:52	1	pCas/Cromosoma	CTT
SEQ ID NO:53	1	pCas/Cromosoma	CTG

<sup>a</sup>Encontramos Timina en la posición 4 detrás del protoespaciador, dando lugar al desplazamiento en una posición de la PAM putativa CTT.

Además, el análisis de los precursores putativos de los nuevos espaciadores mostró que los plásmidos pCSIR-T tienden a adquirir espaciadores de secuencias junto al triplete CTT (42 inserciones de 59; nótese que no se detectó CAT; Tabla 4). Esta preferencia se ve apoyada por los experimentos descritos en la sección previa realizados en BL21-AI con pCSIR-T y pWUR399: 6 inserciones de 9 tenían CTT como PAM y sin observarse CAT.

Cabe destacar que cuando la adquisición se realizó en la cepa O157:H7 con pCSIR-A y pCas1-2[O] (contiene *cas1* y *cas2* de esa misma cepa), a diferencia de K12, se reveló la presencia de ambos CTT y CAT como motivo junto al protoespaciador (detectados 26 y 11 respectivamente en 45 inserciones; ver Tabla 5 y Figura 3D).

Tabla 5: Espaciadores integrados en pCSIR-A durante la propagación en O157:H7 que alberga a pCas1-2[O].

Secuencia Espaciador (5'-3')	Veces observado	Portador (carrier) del protoespaciador	PAM
SEQ ID NO:4	6	pCas	CTT
SEQ ID NO:8	1	pCas	CTT
SEQ ID NO:9	1	pCas	CTT

SEQ ID NO:14	2	pCas/Cromosoma	CTT
SEQ ID NO:22	1	pCSIR	CTT
SEQ ID NO:26	1	pCas	CTT
SEQ ID NO:27	1	pCas	CTT
SEQ ID NO:29	1	pCas	CTT
SEQ ID NO:35	1	pCas	CTT
SEQ ID NO:37	2	pCas	CTT
SEQ ID NO:54	1	pCas	CCT <sup>a</sup>
SEQ ID NO:55	1	pCas	CCA
SEQ ID NO:56	1	pCas/Cromosoma	CGT
SEQ ID NO:57	2	pCas	CAT
SEQ ID NO:58	1	Cromosoma	CAT
SEQ ID NO:59	2	pCas/Cromosoma	CTT
SEQ ID NO:60	1	pCas/Cromosoma	CAT
SEQ ID NO:61	1	pCas	CAT
SEQ ID NO:62	3	pCas	CTT
SEQ ID NO:63	1	pCSIR	GAA
SEQ ID NO:64	1	pCas/Cromosoma	CGC
SEQ ID NO:65	1	pCas	CCT <sup>a</sup>
SEQ ID NO:66	2	pCas	CAT
SEQ ID NO:67	1	pCas/Cromosoma	CAT
SEQ ID NO:68	1	pCas	CAT
SEQ ID NO:69	1	pCas	CTT
SEQ ID NO:70	1	pCas	CAT
SEQ ID NO:71	2	pCas	CTT
SEQ ID NO:72	1	Cromosoma	CCC
SEQ ID NO:73	1	pCas	CAT
SEQ ID NO:74	1	pCas	CGC
SEQ ID NO:75	1	pCSIR	CTT

<sup>a</sup>Encontramos Timina en la posición 4 detrás del protoespaciador, dando lugar al desplazamiento en una posición de la PAM putativa CTT.

5 Por lo tanto, nuestros datos demostraron que CTT es el motivo dominante (no se detectó CAT) reconocido para adquisición por el subtipo CRISPR-Cas I-E de K12 (comparar Figura 3A con 3C), pero CTT y CAT para las proteínas Cas-EO de O157:H7



(comparar Figura 3B con 3D), coincidiendo con motivos PAMs revelados por alineamientos de precursores putativos de espaciadores encontrados en las agrupaciones CRISPR2 de la especie. Estos resultados revelan que ambas variantes del sistema pueden diferir mecanísticamente, al menos en esta etapa de la expansión  
5 CRISPR.

A pesar de que el contenido de DNA por célula y la cantidad del trinucleótido CTT por Kb (Tabla 1) es mayor para pCSIR-T que para pWUR399, los nueve espaciadores adquiridos en células de BL21-AI alojando ambos plásmidos eran idénticos a secuencias presentes en pWUR399 pero ausentes en pCSIT-T (Tabla 3). De modo  
10 similar, aunque el DNA cromosómico es mucho más abundante que pWUR399, ninguno de los espaciadores se corresponde con un fragmento cromosómico (Tabla 3).

Al igual que con BL21-AI, los experimentos realizados con las cepas K12 o O157:H7 revelaron una preferencia por plásmidos conteniendo genes *cas* como fuente de  
15 espaciadores. Una razón de 39/104 espaciadores insertados eran idénticos a secuencias cromosómicas, y sólo 4 de ellos no se encuentran presentes en los plásmidos pCas1-2 (Tablas 4 y 5). Aunque el contenido de DNA por célula de plásmidos con genes *cas* era menor que el de los plásmidos testigo de integración (las 4 construcciones tienen un tamaño similar, y tanto pCas1-2 como pWUR399 tienen un  
20 origen de replicación CDF; Tabla 1), 7 inserciones eran idénticas a fragmentos de pCSIR frente a 94 que lo eran de pCas1-2 (Tablas 4 y 5). Además, la frecuencia del trinucleótido CWT en las construcciones no justifica la infrarrepresentación de protoespaciadores observada en los plásmidos testigo (CTT es más frecuente en pCSIR que en los plásmidos pCas1-2 y la incidencia de CAT es similar para todos  
25 ellos; Tabla 1). En este contexto, la similar relación CTT/CAT en los plásmidos pCas1-2 (0,93-0,99; Tabla 1) proporciona apoyo adicional a la preferencia por el motivo de adquisición CTT por *cas*-EK y CWT por los sistemas *cas*-EO. Por tanto, o algunos replicones son menos propensos a contribuir con espaciadores o existe alguna preferencia por ciertas moléculas como donadoras.

### 30 **Implicación de los elementos CRISPR-Cas en la orientación del espaciador**

Los espaciadores en una agrupación CRISPR tienden a estar igualmente orientados con respecto al motivo PAM. Además, esta conservación de la orientación se aplica al tipo de CRISPR-Cas: Los PAMEs (extremo de los espaciadores equivalente al borde

del protoespaciador adyacente al motivo PAM *-spacers ends equivalent to the protospacer edges adjacent to the PAM-*) de CRISPR-Cas del tipo I están orientados hacia el líder. Ello implica que los genes de adquisición (*cas1* y/o *cas2*) junto con elementos de la región de integración de espaciadores (bien en el líder o la secuencia CRISPR adyacente) están implicados en la determinación de la dirección de inserción.

A continuación se comprobó la inserción de espaciadores en K12 llevando una combinación CRISPR/líder/cas no convencional: pCSIR-A y pCas1-2[K]. Se obtuvieron adiciones de espaciadores (Tabla 6), mostrando que las proteínas Cas de K12 son capaces de funcionar con la variante CRISPR/líder2.1 alternativa.

10 Tabla 6: Espaciadores integrados en pCSIR-A durante la propagación en K12 que alberga a pCas1-2[K].

Secuencia Espaciador (5'-3')	Veces observado	Portador (carrier) del protoespaciador	PAM
SEQ ID NO:1	1	pCas	CTT
SEQ ID NO:76 #'	1	pCas	CTT
SEQ ID NO:77 #R	1	pCas	CCG <sup>c</sup>
SEQ ID NO:78 #R	1	pCas	TAC <sup>c</sup>
SEQ ID NO:24	1	pCas/Cromosoma	CTT
SEQ ID NO:79 #'	1	pCas/Cromosoma	CTT
SEQ ID NO:80 #R	1	pCas/Cromosoma	TAC <sup>c</sup>
SEQ ID NO:29	1	pCas	CTT
SEQ ID NO:81 #R	1	pCas	GTG <sup>c</sup>
SEQ ID NO:82	2	pCas/Cromosoma	CTT
SEQ ID NO:83	1	pCas/Cromosoma	CTT
SEQ ID NO:84	1	pCas	CGG
SEQ ID NO:85	1	pCas	CAA
SEQ ID NO:86	1	pCas	ACT <sup>d</sup>
SEQ ID NO:87	1	Cromosoma	CTT
SEQ ID NO:88	1	pCas/Cromosoma	CTT
SEQ ID NO:89	1	pCas	CTT

<sup>a</sup> Espaciadores identificados con #' se diferencian con el correspondiente espaciador # (integrado en este u otro experiment) por la presencia del dinucleótido AC en el extremo 5'. Espaciadores identificados con #R son idénticos a su respectivo espaciador # integrado en la orientación invertida (reversed) en este u otro experimento.

<sup>b</sup>Los desajustes en la secuencia más parecida (protoespaciador putative) se encuentran representados por letras minúsculas.

<sup>c</sup> El trinucleótido 5'-CTT-3' se encuentra contiguo al extremo opuesto del protoespaciador en la cadena complementaria.

<sup>d</sup> Encontramos Timina en la posición 4 detrás del protoespaciador, dando lugar al desplazamiento en una posición de la PAM putativa CTT.

A diferencia de lo que sucede en todas las inserciones convencionales con las combinaciones pCSIR y pCas1-2 nativas, 7 de 18 adiciones eran anómalas:

- I. tres insertos CRISPR-espaciador contenían espaciadores (identificados en la Tabla 6 como espaciadores identificados como SEQ ID NO: 76, 79 y 88) con dos nucleótidos adicionales, y la CRISPR duplicada con dos nucleótidos menos (ver Figura 4), y
- II. cuatro espaciadores se encontraban en orientación opuesta con respecto al líder (es decir, el PAME estaba localizado distal al líder, ver espaciadores etiquetados con #R en la Tabla 6).

Estos resultados indican un funcionamiento anómalo en la combinación líder-Cas artificial y la existencia de espaciadores invertidos apoya la implicación de tanto las proteínas Cas como del líder en la orientación del espaciador. Adicionalmente sugieren que, mientras se da una flexibilidad en el reconocimiento del líder por las proteínas Cas para producir la adquisición, la orientación del espaciador requiere interacciones específicas entre una variante determinada de proteínas Cas y los CRISPR/líder correspondientes. No obstante, el hecho de que los espaciadores de sistemas CRISPR-Cas I-E residentes en *E. coli* estén insertados adecuadamente en sus agrupaciones con líderes distintos (por ejemplo líder2.1 y 2.3) indica que se tolera cierta variabilidad en la secuencia líder. De acuerdo con estos resultados, se puede concluir que aunque Cas1-2K pueden producir la adición de espaciadores en la variante CRISPR2/líder2 alternativa, el reconocimiento está afectado. Sin embargo, no se puede descartar la posibilidad de que las inserciones anómalas sean debidas a un desequilibrio como consecuencia de la presencia de múltiples *cassettes* de inserción presentes en nuestro sistema experimental.

En concordancia con el reconocimiento de CTT por el sistema de K12, aunque el *cassette* testigo portara un líder tipo A, 15 de 18 espaciadores integrados corresponden a protoespaciadores con este motivo y, además, no se observó CAT

(Tabla 6). Ello evidencia que son los genes *cas* en vez del líder el principal factor implicado en la selección de PAM.

Al igual que en los ensayos realizados con la combinación silvestre de CRISPR, líder y Cas, la mayoría de los espaciadores (17/18) pertenecen a secuencias que se corresponden con fragmentos de pCas1-2 (7 de ellas presentes también en el cromosoma), y solo una exclusiva del cromosoma, mientras que ningún espaciador se originó de pCSIR-A (Tabla 6).

Los ensayos realizados demostraron que los plásmidos testigo de integración CRISPR-espaciador permitieron una detección sensible de insertos sin una elevada producción de proteínas Cas, o sin selección de células adaptadas dependiente del fenotipo de interferencia proporcionado por el espaciador nuevo. Las adiciones detectables estaban restringidas a secuencias que conducían a la síntesis de proteínas CAT funcionales. Por tanto, los espaciadores de 34 nt, observados en agrupaciones CRISPR en una proporción de entorno al 4%, quedaron excluidos, ya que sólo la integración de fragmentos de  $3n+1$  pb, como la unidad CRISPR-espaciador de la longitud canónica de 61 pb, pueden ser seleccionados (Fig. 1D). En torno al 62% de ellos fueron detectados debido a la ausencia de codones de terminación en fase con el gen testigo. Solo se obtuvieron adiciones simples lo cual se explica considerando que las integraciones deben ser secuenciales, un plásmido con múltiples espaciadores nuevos estaría en minoría con respecto a los que alojan un número más pequeño (debido a un mayor número de ciclos de replicación para este último). El aumento de la probabilidad de encontrar codones de terminación (0,38 para una inserción frente a 0,85 para cuatro) también puede contribuir a ello. Teniendo en cuenta estas consideraciones, se pueden sacar conclusiones sólidas sobre el proceso de adaptación.

Los datos demostraron que la PAM dominante en nuestros ensayos con el sistema CRISPR-Cas de K12 fue CTT, de acuerdo con las predicciones de análisis comparativos de las regiones de los protoespaciadores y con los estudios de adquisición dependientes de interferencia realizados bajo niveles altos de proteínas Cas. Por el contrario, en otro análisis con sobreexpresión de *cas1* y *cas2* de K12 no relacionado con interferencia, sólo 36 protoespaciadores de 94 adjuntaban CTT, posiblemente fue debido a los elevados niveles de Cas1 y Cas2 utilizados (Yosef I,

Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 2012 40(12):5569-76).

Por otro lado, en contraste con el sistema de K12 y en coincidencia con predicciones *in silico*, identificamos como PAM el triplete CWT para el sistema CRISPR-Cas I-E de *E. coli* O157:H7, revelándose diferencias en el mecanismo entre variante de un subtipo CRISPR-Cas dentro de una especie. A pesar de ello, el trinucleótido CAT es el motivo PAM observado en espaciadores presentes en agrupaciones CRISPR asociadas a un líder-A, pero CTT también es seleccionado por proteínas Cas-EO. La combinación de nuestros resultados bioinformáticos y experimentales demostraron que se requieren PAMs específicos para el reconocimiento del protoespaciador independientemente de las consecuencias en la interferencia, siendo los espaciadores más eficientes seleccionados subsecuentemente durante una carrera armamentística con el DNA invasor.

Los resultados obtenidos también demuestran el rechazo/preferencia por ciertos replicones como donadores de espaciadores y la existencia de un mecanismo de medición para cortar en el punto de inserción durante el proceso de integración del espaciador.

Los espaciadores se obtuvieron preferentemente de plásmidos conteniendo los genes *cas1* y *cas2*. Se dio una sobrerrepresentación de secuencias de plásmidos conteniendo genes *cas* en los protoespaciadores obtenidos (120/132). El hecho de que la mayoría de los protoespaciadores putativos del cromosoma también se encuentren en estos plásmidos (43/48), incluso aunque las secuencias compartidas fueran inferiores al 0,1% del cromosoma, proporcionan apoyo adicional al origen plasmídico de estos espaciadores.

La presencia de genes *cas* implicados en la adquisición, en los replicones preferidos como donadores de espaciadores, fue responsable de una adquisición más eficiente de las secuencias cercanas a sus regiones codificantes mediante el acoplamiento de la traducción y la transcripción. Esto fue corroborado por la detección de espaciadores en agrupaciones CRISPR, al igual que RNAs, en el complejo ribonucleoproteico Cascada, cuyas secuencias se corresponden con regiones en la proximidad de los *loci* CRISPR-Cas, incluyendo a los genes *cas*.

Con el uso de una combinación CRISPR/líder/cas no convencional (es decir CRISPR-A/líder-A/cas-EK) se observaron inserciones anómalas (cuatro casos de espaciadores invertidos y tres secuencias CRISPR-espaciador atípicas; ver Tabla 6), seguramente debidas a interacciones líder/Cas erróneas.

5 Sorprendentemente, las tres inserciones atípicas compartían las mismas características:

(i) se pierde el dinucleótido GG del extremo de la CRISPR duplicada adyacente al espaciador (Fig. 4) y

10 (ii) los espaciadores eran de 35 pb (en vez de 33) conteniendo AC en el extremo distal al líder (Fig. 4), independientemente de la secuencia en las posiciones correspondientes del protoespaciador putativo (ver Tabla 6).

Estas observaciones demostraron que la integración sucedió entre el segundo y tercer nucleótidos del líder (A1C2--T3) en vez de en la unión CRISPR-líder (C--A1; Figs. 1A y 4A). Como consecuencia de la inserción incorrecta, el extremo del líder proximal a la  
15 CRISPR incorpora el dinucleótido extra AC (Fig. 4A).

El hecho de que se recorte la nueva CRISPR en el extremo adyacente al espaciador sugiere que tras un corte inicial en el líder, el duplicón nuevo se genera tras un nick subsecuente a una distancia fija (a 28 nt del primer corte) en la CRISPR adyacente sin necesidad de reconocer la secuencia del extremo CRISPR (la incorporación de AC en  
20 el duplicón distal al líder se compensa con la pérdida de GG de la CRISPR proximal; ver Figura 4A). De todos modos, dado que una inserción de 63 pb no pudo ser detectada en los pCSIR (no restaura la fase de *cat*), no se puede descartar un segundo corte a esta distancia del primer nick. Sin embargo, las tres adiciones observadas demostraron que el extremo distal al líder de la CRISPR era innecesario  
25 para el corte, y la existencia de insertos de 61 pb, pero no de 58 o 64 pb apoya la hipótesis del corte a una distancia medida. Por tanto, teniendo en cuenta estas consideraciones, concluimos que el primer paso de la integración consistió en un corte secuencial en los puntos de inserción.

El primer corte, en el límite CRISPR/líder, depende de la secuencia (probablemente  
30 reconocido por Cas1 o Cas2). El segundo nick tuvo lugar en el extremo de la CRISPR distal al líder, muy probablemente en un punto determinado por la distancia al corte previo (Fig. 4B). A parte de Cas1 y Cas2, las adquisiciones deben requerir proteínas

no Cas con actividades como ligación de DNA y polimerización. Esto podría conseguirse mediante la implicación de mecanismos de reparación del DNA, lo que amplía las posibles implicaciones de las interacciones conocidas de Cas1 con componentes clave de sistemas de reparación, incluyéndose RecB, RecC y RuvB.

5 Aunque hemos considerado un modelo de inserción de un espaciador de doble cadena, no es descartable la integración de un protoespaciador monocatenario y su consiguiente replicación.

También se ha sugerido un mecanismo de corte a una distancia fija para la escisión del protoespaciador en otro sistema CRISPR-Cas de tipo I. Como consecuencia de  
10 esta estrategia común dirigiendo las dos etapas de la adquisición (escisión del protoespaciador e integración del espaciador) se conserva la periodicidad CRISPR.

## REIVINDICACIONES

1. Método para detectar inserciones de espaciadores mediante selección independiente de la interferencia a partir de estructuras artificiales basadas en repeticiones palindrómicas cortas agrupadas regularmente espaciadas (CRISPR) que  
5 comprende la inserción de al menos una unidad CRISPR-espaciador, en la estructura artificial dentro de la secuencia que codifica un gen testigo, que restaura la pauta de lectura de traducción, donde la expresión del gen testigo es indicativo de la inserción del espaciador.
2. Método según la reivindicación 1, caracterizado por que la unidad CRISPR-  
10 espaciador insertada comprende un número de nucleótidos no múltiplo de 3.
3. Método según cualquiera de las reivindicaciones anteriores, caracterizado por que el espaciador comprende un codón de terminación críptico.
4. Método según cualquiera de las reivindicaciones anteriores, caracterizado por que la estructura artificial comprende la secuencia identificada como SEQ ID NO: 103.
- 15 5. Método según cualquiera de las reivindicaciones anteriores, caracterizado por que el espaciador es seleccionado de entre las secuencias identificadas como SEQ ID NO: 1- SEQ ID NO: 89.
6. Estructura artificial identificada caracterizada por que comprende la secuencia identificada como SEQ ID NO: 103.
- 20 7. Uso de la estructura artificial según la reivindicación 6 para la detección de inserciones de espaciadores.



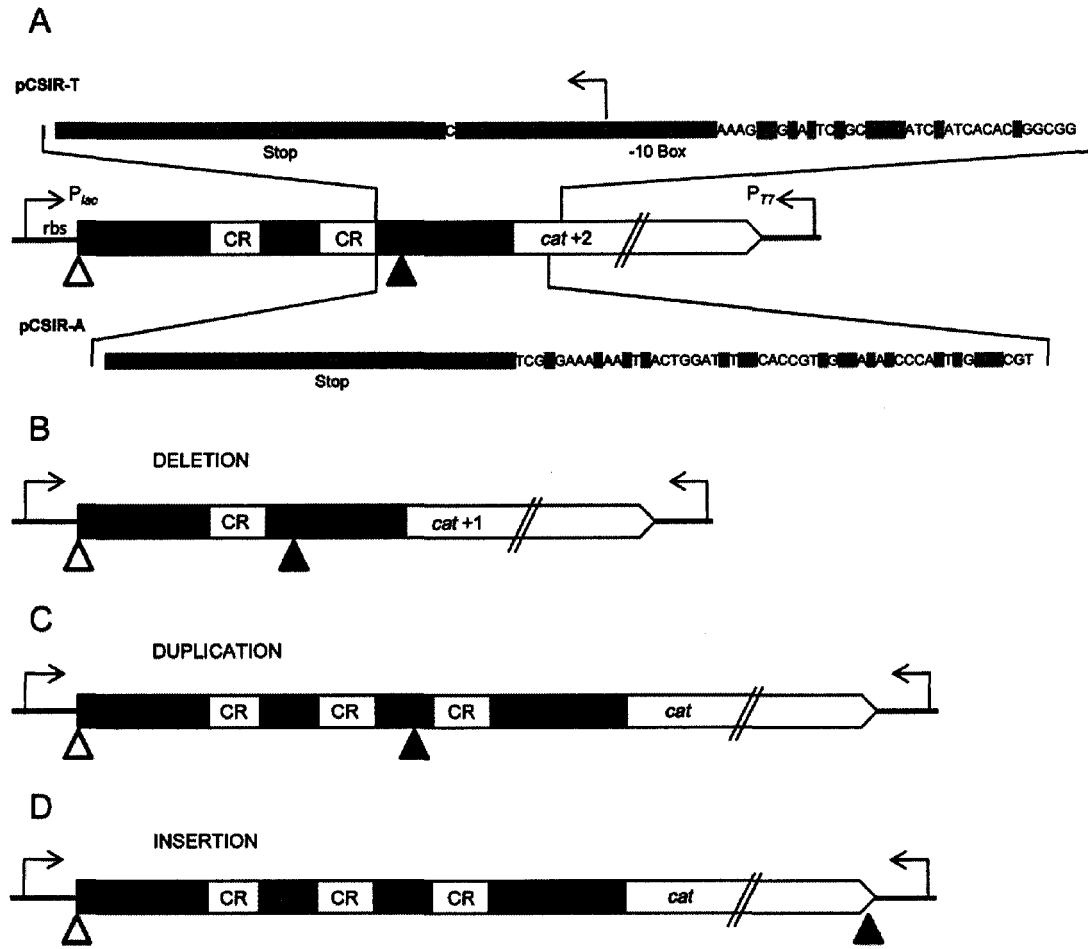


FIG.1

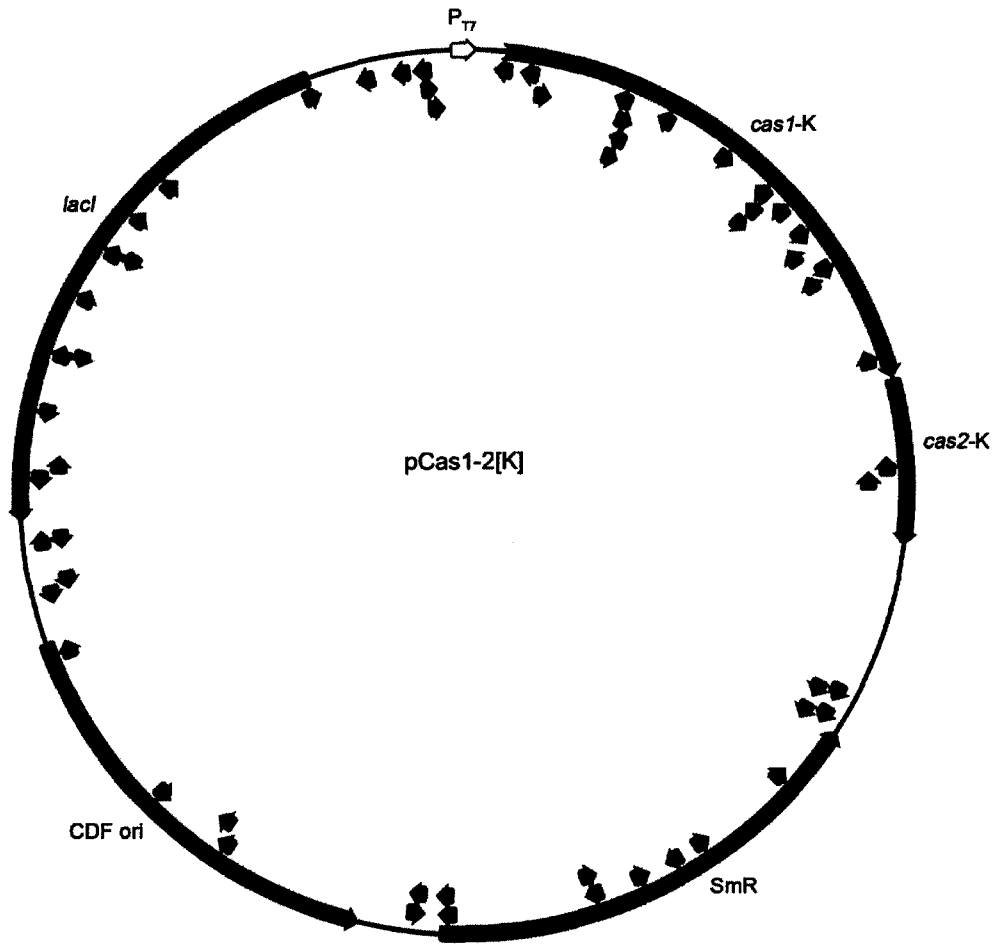


FIG.2

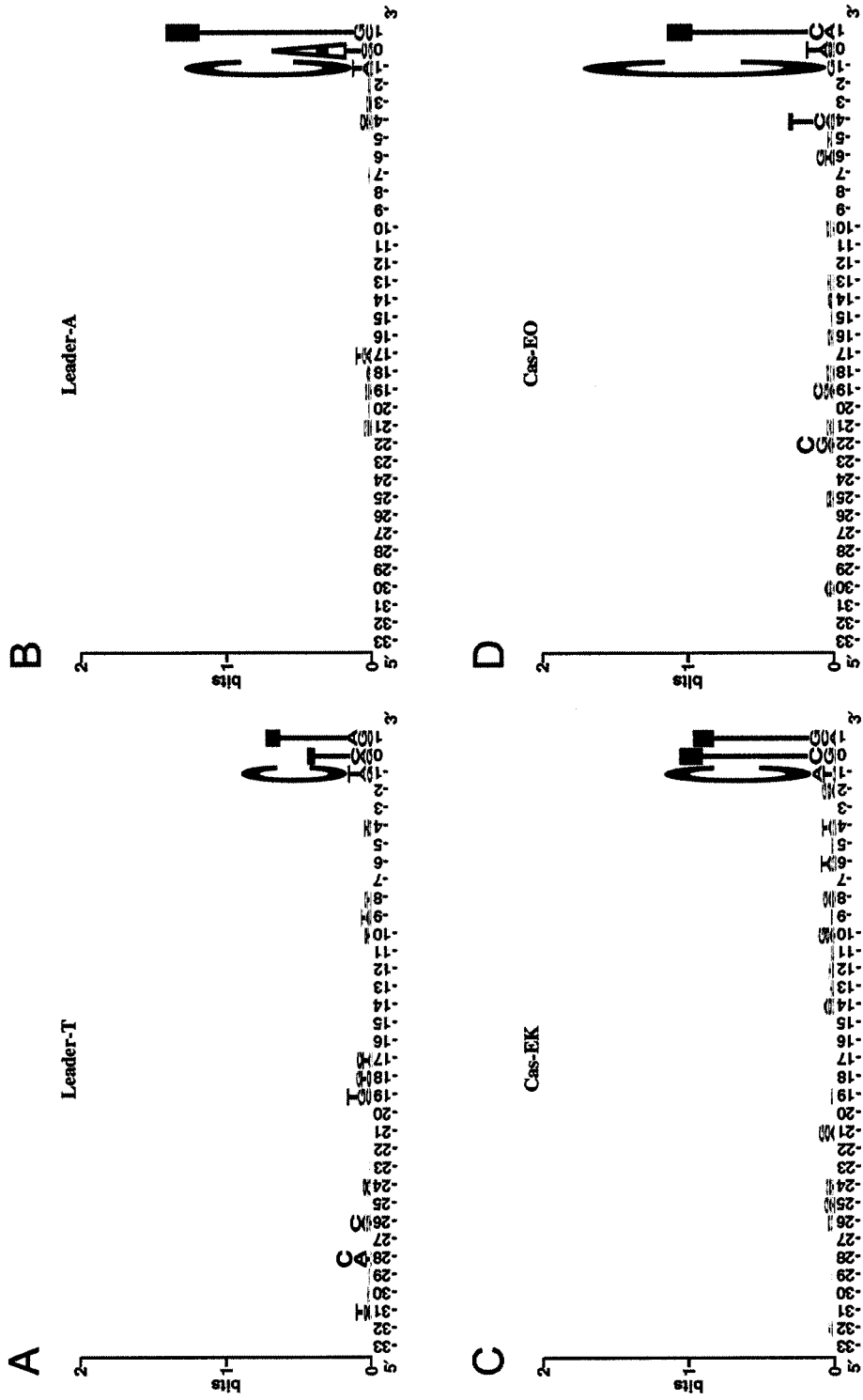


FIG.3





## SEQUENCE LISTING

<110> UNIVERSIDAD DE ALICANTE  
 <120> CRISPR  
 <130> PT0006/2013  
 <160> 104  
 <170> PatentIn version 3.5  
 <210> 1  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer1 pWUR399  
 <400> 1  
 gagatcccgg acaccatcga atggcgcaaa acc 33  
 <210> 2  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer2 pWUR399  
 <400> 2  
 attcgatggt gtccgggatc tcgacgctct ccc 33  
 <210> 3  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer3 pWUR399  
 <400> 3  
 atcgccgctt cagttacgcc gtataaacag gaa 33  
 <210> 4  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer pWUR399  
 <400> 4  
 gtcgagatcc cggtgcctaa tgagtgagct aac 33  
 <210> 5  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer5 pWUR399

<400> 5  
 tcttgcaggt atcttcgagc cagccacgat cga 33

<210> 6  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer6 pWUR399

<400> 6  
 taacacccca tgcaacgatg gttatctcag aac 33

<210> 7  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer7 pWUR399

<400> 7  
 ccctgccctg aaccgacgac cgggtcatcg tgg 33

<210> 8  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer8 pWUR399

<400> 8  
 ggcgcggggc atgactaaca tgagaattac aac 33

<210> 9  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer9 pCas

<400> 9  
 gcgcagcagt cgggctgaac ggggagttct tgc 33

<210> 10  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer10 pCas/Chromosome

<400> 10  
 tattggcgc cagggtgggt tttcttttca cca 33

<210> 11  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer11 pCas/Chromosome  
 <400> 11  
 cccatattgc agaaagatca tggagacgcg atc 33

<210> 12  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer12 Chromosome  
 <400> 12  
 tcgacaatgt tcaacagcca cggaatatct ttc 33

<210> 13  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer 13 pCSIR  
 <400> 13  
 aaaggaaggg aagaaagcga aaggagcggg cgc 33

<210> 14  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer14 pCas/Chromosome  
 <400> 14  
 gaccagatgc tccacgcca gtcgctacc gtc 33

<210> 15  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer15 pCas  
 <400> 15  
 tgaacaggat ctatttgagg cgctaaatga aac 33

<210> 16  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> pCas/Chromosome  
 <400> 16  
 tgcattggtt gatcgtatcg cccttttccc agt 33



<210> 17  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer 17 pCas/Chromosome  
  
 <400> 17  
 atgtctgcaa tatcgtaaac aaaggaaaga ggc 33

<210> 18  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer18 pCas  
  
 <400> 18  
 ttgcgctctc ctggtccgac cctgccgttt acc 33

<210> 19  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer19 pCas/Chromosome  
  
 <400> 19  
 cgtattcgtt gcccatgcca tcactacatt gcc 33

<210> 20  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer20 pCas  
  
 <400> 20  
 atccgctcat gagacaataa ccctgataaa tgc 33

<210> 21  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer21 pCas  
  
 <400> 21  
 gtccgcacca acgcgagcc cggactcggc aat 33

<210> 22  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer22 pCSIR

<400> 22  
 ccgagtacgt gctcgctcga tgcgatgttt cgc 33

<210> 23  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer23 pCas/Chromosome

<400> 23  
 acattgcctt cttccgccag tccagctatt tgt 33

<210> 24  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer24 pCas/Chromosome

<400> 24  
 atgggtgggg gaagcgggcg ttcgtgttta tgc 33

<210> 25  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer25 pCas/Chromosome

<400> 25  
 cgccattacc gagtccgggc tgcgcgttgg tgc 33

<210> 26  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer26 pCas/Chromosome

<400> 26  
 aagaaataac gccggaacat tagtgcaggc agc 33

<210> 27  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer27 pCas

<400> 27  
 gatacttcgg cgatcaccgc ttcctcata ctc 33

<210> 28  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer28 pCas/pCSIR/Chromosome  
 <400> 28  
 caatacgcaa accgcctctc cccgcgcggtt ggc 33

<210> 29  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer29 pCas  
 <400> 29  
 tgagaattac aacttatatc gtatggggct gac 33

<210> 30  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer30 pCas  
 <400> 30  
 ccgtttaccg gatacctggt cgcctttct ccc 33

<210> 31  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer31 pCas/Chromosome  
 <400> 31  
 tgtaccgaaa gcttttgaga tagcgcgtcg taa 33

<210> 32  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer32 pCas/Chromosome  
 <400> 32  
 tagaaggcag tcgcgtgcgg gcaacctacg cac 33

<210> 33  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer33 pCas/Chromosome  
 <400> 33  
 cgatacgatc aaccaatgca ttagcgtgc aac 33

<210> 34  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer34 pCas  
  
 <400> 34  
 gccctcggc ttgaacgaat tgtagacat tat 33

<210> 35  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer35 pCas  
  
 <400> 35  
 gaacgaattg ttagacatta tttgccgact acc 33

<210> 36  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer36 pCas  
  
 <400> 36  
 gtatcagccc gtcatacttg aagctagaca ggc 33

<210> 37  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer37 pCas  
  
 <400> 37  
 cggaatgatg tcgtcgtgca caacaatggt gac 33

<210> 38  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer38 pCas  
  
 <400> 38  
 ccgggatctc gacgctctcc cttatgacgac tcc 33

<210> 39  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer39 pCas/Chromosome

<400> 39  
 ctcaccaggg ttacgacgcg ctatctcaaa agc 33

<210> 40  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer40 pCas

<400> 40  
 gattaagggg aagccaggtc atggtatatc tcc 33

<210> 41  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer41 pCas/Chromosome

<400> 41  
 cgcttcccc acccatacca acaatgttcc aac 33

<210> 42  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer42 pCas

<400> 42  
 ccgctgacc taccaaggca acgctatggt ctc 33

<210> 43  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer43 pCas

<400> 43  
 ttacgcgag aaaaaaagga tctcaagaag atc 33

<210> 44  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer44 pCas

<400> 44  
 cttcggcgat caccgcttcc ctcatactct tcc 33

<210> 45  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer45 pCas  
 <400> 45  
 gttatacgat gtcgcagagt atgccggtgt ctc 33

<210> 46  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer46 pCas  
 <400> 46  
 attgggttg tgcatacagg aaagcctctt tcc 33

<210> 47  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer47 Chromosome  
 <400> 47  
 gtcaaagcc ccacttatac gctggtcgaa ccc 33

<210> 48  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer48 pCas/Chromosome  
 <400> 48  
 gtatgggtgg gggaagcggg cgttcgtggt tat 33

<210> 49  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer49 pCas/Chromosome  
 <400> 49  
 tttgcatggt ttgttgaaaa ccggacatgg cac 33

<210> 50  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer50 pCSIR  
 <400> 50  
 tcgaccaaga agcggttggt ggcgctctcg cgg 33

<210> 51  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer51 pCas/Chromosome  
  
 <400> 51  
 attaagttct gtctcggcgc gtctgcgtct ggc 33

<210> 52  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer52 pCas/Chromosome  
  
 <400> 52  
 agccagccag acgcagacgc gccgagacag aac 33

<210> 53  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer53 pCas/Chromosome  
  
 <400> 53  
 cccagtgaag caggaagcgg aatggcaaca ggc 33

<210> 54  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer54 pCas  
  
 <400> 54  
 cattcgatgg tgtccgggat ctcgacgctc tcc 33

<210> 55  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer55 pCas  
  
 <400> 55  
 acgaccatgc tcatccgccg cgccccggg gac 33

<210> 56  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer56 pCas/Chromosome

<400> 56  
 gcaatgcgcg ccattaccga gtccgggctg cgc 33

<210> 57  
 <211> 33  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> spacer57 pCas

<400> 57  
 cccacagatt ccgtggataa gcaggggact aac 33

<210> 58  
 <211> 33  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> spacer58 Chromosome

<400> 58  
 attacaagt ggatcggcat tggatgattac acc 33

<210> 59  
 <211> 33  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> spacer59 pCas/Chromosome

<400> 59  
 ttcaccagt agacgggcaa cagctgattg ccc 33

<210> 60  
 <211> 33  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> spacer60 pCas/Chromosome

<400> 60  
 ccacaaccg cgccgatat gttaccgct gcc 33

<210> 61  
 <211> 33  
 <212> DNA  
 <213> Artificial sequence

<220>  
 <223> spacer61 pCas

<400> 61  
 attaccgagt ccgggctgcg cgttggtgcg gac 33

<210> 62  
 <211> 33  
 <212> DNA  
 <213> Artificial sequence



<220>  
 <223> Spacer62 pCas  
 <400> 62  
 tcctcaatta acggtattaa tttgcccggt aac 33

<210> 63  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer63 pCSIR  
 <400> 63  
 ttcagtgaca acgtcgagca cagctgcgca agg 33

<210> 64  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer64 pCas/Chromosome  
 <400> 64  
 tgaaccacca caaccgcgc cggatagt acc 33

<210> 65  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer65 pCas  
 <400> 65  
 ggactcagtg gtacaaacgc catggtatat ctc 33

<210> 66  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer66 pCas  
 <400> 66  
 ttccataggc tccgccctcc tgccagagtt cac 33

<210> 67  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> Spacer67 pCas/Chromosome  
 <400> 67  
 aattgagga tcccgcgttc gccagaccta tgc 33

<210> 68  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer68 pCas  
  
 <400> 68  
 gcagtgggaa cgatgccctc attcagcatt tgc 33

<210> 69  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer69 pCas  
  
 <400> 69  
 gaatggcagc gcaatgacat tcttgacaggt atc 33

<210> 70  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer70 pCas  
  
 <400> 70  
 cagatgctcc acgcccagtc gcgtaccgtc ttc 33

<210> 71  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> spacer71 pCas  
  
 <400> 71  
 gtcccctgct tatccacgga atctgtgggt aac 33

<210> 72  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> spacer72 Chromosome  
  
 <400> 72  
 ctcatccgcc gcgccccggg tgaccactat cac 33

<210> 73  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> spacer73 pCas

<400> 73  
 agcatcgcag tgggaacgat gccctcattc agc 33

<210> 74  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer74 pCas

<400> 74  
 attgtgcacc gccgctttac aggcttcgac gcc 33

<210> 75  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer75 pCSIR

<400> 75  
 gatccttggc ggcaagaaag ccatccagtt tac 33

<210> 76  
 <211> 35  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer1' pCas

<400> 76  
 acgagatccc ggacaccatc gaatggcgca aaacc 35

<210> 77  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer4R pCas

<400> 77  
 gttagctcac tcattaggca ccgggatctc gac 33

<210> 78  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer20R pCas

<400> 78  
 gcattatca gggttattgt ctcagagcg gat 33

<210> 79  
 <211> 35  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer24' pCas/Chromosome  
 <400> 79  
 acatgggtgg gggaagcggg cgttcgtgtt tatgc 35

<210> 80  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer24R pCas/Chromosome  
 <400> 80  
 gcataaacac gaacgcccgc ttccccacc cat 33

<210> 81  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer37R pCas  
 <400> 81  
 gtcaccattg ttgtgcacga cgacatcatt ccg 33

<210> 82  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer76 pCas/Chromosome  
 <400> 82  
 tcatccagag caagttttgc ctgatagagc agc 33

<210> 83  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer77 pCas/Chromosome  
 <400> 83  
 ggcagatcga tgtaatagat ggcgcgtttg tac 33

<210> 84  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Spacer78 pCas  
 <400> 84  
 accgacgacc gggatcatcgt ggccgatct tgc 33

<210> 85  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer79 pCas  
  
 <400> 85  
 gcggtatggc atgatagcgc ccggaagaga gtc 33

<210> 86  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer80 pCas  
  
 <400> 86  
 gaatgtaatt cagctccgcc atcgccgctt cca 33

<210> 87  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer81 Chromosome  
  
 <400> 87  
 gttttccgta gcgtgcttcc tgccaggctg cgc 33

<210> 88  
 <211> 35  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer82 pCas/Chromosome  
  
 <400> 88  
 acgcagcgcct aatgcattgg ttgatcgtat cgccc 35

<210> 89  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Spacer83 pCas  
  
 <400> 89  
 tttcgctgc tggggcaaac cagcgtggac cgc 33

<210> 90  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Primer-CATF1

<400> 90  
 ctaaggaagc taaactcgag 20

<210> 91  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-CATR1

<400> 91  
 ttaaggcat ctagaactgc 20

<210> 92  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-CAS1EKF

<400> 92  
 gcgcgccatg gcctggcttc cccttaatcc 30

<210> 93  
 <211> 38  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-BG2375

<400> 93  
 ggccccggcg cgctcaaaca ggtaaaaaag acaccaac 38

<210> 94  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-CAS1EOF

<400> 94  
 cgcccatggc gttgtacca ctgagtcc 28

<210> 95  
 <211> 27  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-CAS2EOR

<400> 95  
 ccgcgccgc atcaacaatc taccaac 27

<210> 96  
 <211> 19  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-TS2F

<400> 96  
 cacctttggc ttcggctgc 19

<210> 97  
 <211> 34  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-TLR

<400> 97  
 ttgttacatt aaggttggtg ggttgtttgt atgg 34

<210> 98  
 <211> 24  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-AS2F

<400> 98  
 ggaaagctta cgcagcaaa ttac 24

<210> 99  
 <211> 28  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-ALR

<400> 99  
 ggtagactcg agatgtggaa taaatttg 28

<210> 100  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-T7t

<400> 100  
 ctagttattg ctcagcgg 18

<210> 101  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-T7

<400> 101  
 taatacgact cactataggg 20

<210> 102  
 <211> 17  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Primer-M13R

<400> 102  
 caggaaacag ctatgac

17

<210> 103  
 <211> 953  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> CRISPR-espaciador-CRISPR-leader-gencat

<400> 103  
 atgaccatga ttacgccaag cttggtaccg agctcggatc cactagtaac ggccgccagt 60  
 gtgctggaat tcggcttcac ctttggttc ggctgcgggt tatccccgct ggcgcgggga 120  
 actctgcgtg agcgtatcgc cgcgcgtctg cgaaagcggg ttatccccgc tggcgcgggg 180  
 aactctctaa aagtatacat ttgttcttaa agcatttttt cccatacaaa caaccaccca 240  
 accttaatgt aacaaaagcc gaattctgca gatatccatc acactggcgg ccgctcgaga 300  
 aaaaaatcac tggatatacc accgttgata tatcccaatc gcatcgtaaa gaacattttg 360  
 aggcatttca gtcagttgct caatgtacct ataaccagac cgttcagctg gatattacgg 420  
 cttttttaa gaccgtaaag aaaaataagc acaagtttta tccggccttt attcacattc 480  
 ttgcccgcct gatgaatgct catccggaat tccgtatggc aatgaaagac ggtgagctgg 540  
 tgatatggga tagtgttcac ctttgttaca ccgttttcca tgagcaaact gaaacgtttt 600  
 catcgctctg gagtgaatac cagcagcatt tccggcagtt tctacacata tattcgcaag 660  
 atgtggcgtg ttacggtgaa aacctggcct atttccctaa agggtttatt gagaatatgt 720  
 ttttctctc agccaatccc tgggtgagtt tcaccagttt tgattttaaac gtggccaata 780  
 tggacaactt cttcgcccc gttttcacca tgggcaaata ttatacgcaa ggcgacaagg 840  
 tgctgatgcc gctggcgatt caggttcac atgccgtctg tgatggcttc catgtcggca 900  
 gaatgcttaa tgaattaca cagtactgag atgagtggca gggcggggcg taa 953

<210> 104  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> spacer-clonadaptado

<400> 104  
 aaatacacag acacggagaa tcactatggt tac

33





②① N.º solicitud: 201300203

②② Fecha de presentación de la solicitud: 22.02.2013

②③ Fecha de prioridad:

## INFORME SOBRE EL ESTADO DE LA TECNICA

⑤① Int. Cl.: **C12Q1/68** (2006.01)  
C12N15/74 (2006.01)

### DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
A	ALMENDROS, C. et al., 'Target motifs affecting natural immunity by a constitutive CRISPR-Cas system in Escherichia coli', PLOS ONE, 2012, No. 7, No. 11, Página e50797, ISSN: 1932-6203 (electronic), doi: 10.1371/journal.pone.0050797, Epub: 26.11.2012, todo el documento.	1-7
A	SWARTS, D.C. et al., 'CRISPR interference directs strand specific spacer acquisition', PLOS ONE, 2012, Vol. 7, No. 4, Página e35888, ISSN: 1932-6203 (print), ISSN: 1932-6203 (electronic), doi: 10.1371/journal.pone.0035888, Epub: 27.04.2012, todo el documento.	1-7
A	SEMENOVA, E. et al., 'Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence', PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, 2011, Vol. 108, No. 25, Páginas 10098-10103, ISSN: 0027-8424 (print), ISSN: 1091-6490 (electronic), doi: 10.1073/pnas.1104144108, todo el documento.	1-7
A	DÍEZ-VILLASEÑOR, C. et al., 'Diversity of CRISPR loci in Escherichia coli', MICROBIOLOGY, 2010, Vol. 156, No. Pt 5, Páginas 1351-1361, ISSN: 1350-0872, doi: 10.1099/mic.0.036046-0, todo el documento.	1-7
A	POUGACH, K. et al., 'Transcription, processing and function of CRISPR cassettes in Escherichia coli', MOLECULAR MICROBIOLOGY, 2010, Vol. 77, No. 6, Páginas 1367-1379, ISSN: 0950-382X (print), ISSN: 1365-2958 (electronic), doi: 10.1111/j.1365-2958.2010.07265.x, todo el documento.	1-7
A	MOJICA, F.J. et al., 'Short motif sequences determine the targets of the prokaryotic CRISPR defence system', MICROBIOLOGY, 2009, Vol. 155, No. Pt 3, Páginas 733-740, ISSN: 1350-0872 (Print), doi: 10.1099/mic.0.023960-0, todo el documento.	1-7
A	EP 2341149 A1 (DANISCO) 06.07.2011, todo el documento.	1-7

#### 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  
06.11.2014

Examinador  
J. L. Vizán Arroyo

Página  
1/4

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

C12Q, C12N

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

INVENES, EPODOC, WPI, BIOSIS, MEDLINE, EMBASE, EMBL-EBI

Fecha de Realización de la Opinión Escrita: 06.11.2014

**Declaración**

<b>Novedad (Art. 6.1 LP 11/1986)</b>	Reivindicaciones 1-7	<b>SI</b>
	Reivindicaciones	<b>NO</b>
<b>Actividad inventiva (Art. 8.1 LP11/1986)</b>	Reivindicaciones 1-7	<b>SI</b>
	Reivindicaciones	<b>NO</b>

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

**Base de la Opinión.-**

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

**1. Documentos considerados.-**

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	ALMENDROS, C. et al., <i>PLoS One</i> , (2012), 7(11):e50797.	2012
D02	SWARTS, D.C. et al., <i>PLoS One</i> , (2012), 7(4):e35888.	2012
D03	SEMENOVA, E. et al., <i>Proc. Natl. Acad. Sci. U S A</i> , (2011), 108(25):10098-103.	2011
D04	DÍEZ-VILLASENOR, C. et al., <i>Microbiology</i> , (2010), 156(Pt 5): 1351-61.	2010
D05	POUGACH, K. et al., <i>Mol. Microbiol.</i> , (2010), 77(6): 1367-79.	2010
D06	MOJICA, F.J. et al., <i>Microbiology</i> , (2009), 155(Pt 3): 733-40.	2009
D07	EP 2341149 A1 (DANISCO)	06.07.2011

En D01-D07 se divulgan diferentes características funcionales y estructurales de los sistemas CRISPR-Cas de *E.coli*.

**2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración**

1. NOVEDAD (Art. 4.1. y Art. 6.1. de la Ley de Patentes) y ACTIVIDAD INVENTIVA (Art. 4.1. y Art. 8.1. de la Ley de Patentes).

1.1. Reivindicaciones independientes 1.

1.1.1. La solicitud reivindica un método para detectar inserciones de espaciadores mediante selección independiente de la interferencia a partir de estructuras artificiales localizadas dentro de la secuencia de un gen testigo y basadas en secuencias CRISPR ('*Clustered Regularly Interspaced Short Palindromic Repeats*') de repeticiones palindrómicas cortas que comprenden la inserción de al menos una unidad CRISPR-espaciador, en la estructura artificial dentro de la secuencia de un gen testigo, que restaura la pauta de lectura de traducción, donde la expresión del gen testigo es indicativo de la inserción del espaciador.

En el estado de la técnica, representado por los documentos D01-D07, se han descrito los elementos estructurales constitutivos de los sistemas CRISPR-Cas que comprenden secuencias casetes CRISPR consistentes en repeticiones palindrómicas cortas de ADN que están regularmente espaciadas por secuencias no reiteradas, denominadas espaciadores, y un conjunto diverso de genes asociados- CRISPR (*cas*) que flanquean a las secuencias casetes CRISPR. Los sistemas CRISPR-Cas funcionan como un sistema inmune adaptativo en bacterias que degrada el ADN exógeno portador de secuencias complementarias a la secuencia de los espaciadores (cf. D01-D07). Sin embargo, en el estado de la técnica referido no se ha divulgado ningún método para detectar inserciones de espaciadores en estructuras CRISPR con las características del método reivindicado en la solicitud. Además, dicho método no se deduce de una manera obvia del estado de la técnica señalado. Por todo ello se considera que el objeto de protección de la reivindicación independiente 1, y el de las reivindicaciones dependientes 2-7 es nuevo y tiene actividad inventiva sobre la base de los documentos D01-D07.

1.2. La presente solicitud satisface el criterio establecido en el Art. 4.1. de la Ley de Patentes, pues el objeto de las reivindicaciones 1-7 es nuevo y tiene actividad inventiva de acuerdo con los Arts. 6.1. y 8.1. de la Ley de Patentes.