

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 974 625**

51 Int. Cl.:

A01H 9/00 (2006.01)

C12N 15/82 (2006.01)

C12N 15/87 (2006.01)

C12N 5/00 (2006.01)

C12N 9/16 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.03.2015 PCT/US2015/020622**

87 Fecha y número de publicación internacional: **17.09.2015 WO15139008**

96 Fecha de presentación y número de la solicitud europea: **14.03.2015 E 15761154 (2)**

97 Fecha y número de publicación de la concesión europea: **06.12.2023 EP 3116305**

54 Título: **Procedimientos y composiciones para aumentar la eficacia de la modificación génica dirigida mediante reparación génica mediada por oligonucleótidos**

30 Prioridad:

14.03.2014 US 201461953333 P

17.09.2014 US 201462051579 P

05.11.2014 US 201462075816 P

05.11.2014 US 201462075811 P

13.03.2015 US 201562133129 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
28.06.2024

73 Titular/es:

CIBUS US LLC (50.0%)

6455 Nancy Ridge Drive

San Diego, California 92121, US y

CIBUS EUROPE B.V. (50.0%)

72 Inventor/es:

BEETHAM, PETER R.;

GOCAL, GREGORY F.W.;

SCHOPKE, CHRISTIAN;

SAUER, NOEL;

PEARCE, JAMES;

SEGAMI, ROSA E. y

MOZORUK, JERRY

74 Agente/Representante:

GONZÁLEZ PECES, Gustavo Adolfo

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 974 625 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Procedimientos y composiciones para aumentar la eficacia de la modificación génica dirigida mediante reparación génica mediada por oligonucleótidos

5 CAMPO DE LA INVENCION

La presente divulgación se refiere, al menos en parte, a mutaciones y modificaciones genéticas dirigidas, incluyendo procedimientos y composiciones para realizar dichas mutaciones y modificaciones.

ANTECEDENTES

10 La siguiente discusión se proporciona meramente para ayudar al lector en la comprensión y no se admite que describa o constituya el estado de la técnica anterior a la presente divulgación.

La Patente de los Estados Unidos No. 6,271,360 divulga procedimientos y composiciones para la introducción de cambios genéticos predeterminados en genes objetivo de una célula viva mediante la introducción de un oligodesoxinucleótido que codifica el cambio predeterminado. Los oligodesoxinucleótidos son eficaces en células de mamíferos, aves, plantas y bacterias.

15 La Patente de los Estados Unidos No. 8,771,945 divulga vectores y sistemas de vectores, algunos de los cuales codifican uno o más componentes de un complejo CRISPR, así como procedimientos para el diseño y uso de tales vectores.

20 La Patente de los Estados Unidos No. 8,470,973 "se refiere a procedimientos para reconocer selectivamente un par de bases en una secuencia de ADN por un polipéptido, a polipéptidos modificados que reconocen específicamente uno o más pares de bases en una secuencia de ADN y, a ADN que se modifica para que pueda ser reconocido específicamente por un polipéptido y a usos del polipéptido y el ADN en la orientación específica del ADN, así como a procedimientos de modulación de la expresión de genes objetivo en una célula."

25 Los documentos WO2009/046334A1 y WO2013/028188A1 divulgan procedimientos para introducir un cambio genético en una célula vegetal utilizando la tecnología de oligonucleobase reparadora de genes (GRON). Los GRONs pueden incluir opcionalmente una modificación como un enlace fosforotioato o un grupo Cy3. El documento WO2014/144951A1 también divulga procedimientos para introducir un cambio genético en una célula vegetal utilizando GRONs modificados. Ejemplo 5 del documento WO2014/144951A1 muestra el uso de CRISPR como concepto alternativo para inducir la recombinación.

30 SUMARIO

El objeto de la presente invención es un procedimiento para provocar un cambio genético en una célula vegetal. Las reivindicaciones dependientes se refieren a realizaciones particulares de las mismas.

35 El procedimiento de acuerdo con la presente invención comprende exponer una célula vegetal a un cortador de ADN y a una oligonucleobasa reparadora de genes (GRON). El cortador de ADN es CRISPR-Cas. El GRON incluye una o más alteraciones de los nucleótidos convencionales de ARN y ADN. Dichas alteraciones se seleccionan de uno o más nucleótidos 2'O-metilo en los extremos 5' y/o 3', una base inversa (idC) en el extremo 3', una modificación O-(2-metoxietil) (MOE) en los extremos 5' y/o 3', uno o más grupos Cy3 en los extremos 5' y/o 3', en donde Cy3 denota un fosforoamidito bloqueado que al incorporarse a un oligonucleótido produce un colorante de indomonocarbocianina sustituido con 3,3,3',3'-tetrametil N,N'-isopropil, y uno o más grupos 3PS en los extremos 5' y/o 3', donde 3PS denota 3 enlaces fosforotioato. Dicho GRON tiene entre 15 y 210 nucleótidos de longitud.

45 En el presente documento se proporcionan procedimientos y composiciones para efectuar un cambio genético dirigido en el ADN de una célula. Ciertas realizaciones se refieren a la mejora de la eficacia de la orientación de las modificaciones a lugares específicos en secuencias genómicas u otras secuencias de nucleótidos. Como se describe en el presente documento, los ácidos nucleicos que dirigen cambios específicos en el genoma pueden combinarse con diversos enfoques para aumentar la disponibilidad de componentes de los sistemas naturales de reparación presentes en las células que son objeto de modificación.

50 Un "oligonucleótido de reparación génica" o "GRON", tal como se utiliza en el presente documento, significa una oligonucleobase (por ejemplo, oligonucleótidos dúplex mixtos, moléculas que no contienen nucleótidos, oligodesoxinucleótidos monocatenarios, oligodesoxinucleótidos bicatenarios y otras moléculas de reparación génica) que puede, en determinadas condiciones, dirigir delecciones, inserciones o sustituciones de nucleótidos simples o, en algunas realizaciones, múltiples, en una secuencia de ADN. Esta edición de reparación génica del genoma mediada por oligonucleótidos puede comprender tanto sistemas de reparación no basados en homología (por ejemplo, unión de extremos no homólogos) como sistemas de reparación basados en homología (por ejemplo, reparación dirigida por homología). El GRON se diseña normalmente para alinearse en registro con un objetivo genómico, excepto por la(s) discordancia(s) diseñada(s). Estos desajustes pueden reconocerse y corregirse aprovechando uno o más de los sistemas endógenos de reparación del ADN de la célula. En algunas realizaciones, un GRON u oligonucleótido puede diseñarse para contener múltiples diferencias en comparación con la secuencia objetivo del organismo. Es posible que

no todas estas diferencias afecten a la secuencia proteica traducida a partir de dicha secuencia objetivo y que, en uno o más casos, se conozcan como cambios silenciosos. En el presente documento se describen numerosas variaciones de la estructura, la química y la función del GRON. Un GRON, tal y como se utiliza en el presente documento, tiene una o más modificaciones. Un GRON, tal como se utiliza en el presente documento, tiene una o más modificaciones que atraen la maquinaria de reparación del ADN al sitio objetivo (discordancia) y/o que impiden la recombinación de parte o de todo el GRON (distinto de la delección o delecciones, inserción o inserciones, sustitución o sustituciones deseadas o similares) en el ADN genómico de la secuencia de ADN objetivo y/o que aumentan la estabilidad del GRON.

- 10 Un GRON puede tener tanto nucleótidos de ARN como de ADN y/u otros tipos de nucleobases. Uno o más de los nucleótidos de ADN o ARN comprenden una modificación.

Los GRONs pueden dirigirse tanto a regiones no codificantes (NC) como codificantes (C) de un gen objetivo. A modo de ejemplo, las Figs. 27 y 28 representan respectivamente C-GRONs y NC-GRONs adecuados para introducir mutaciones en el genoma del arroz con el fin de introducir una o más de las siguientes sustituciones de aminoácidos en el gen ACCasa. La convención consiste en utilizar como referencia el sistema de numeración de aminoácidos de la ACCasa plastidial del pasto negro (*Alopecurus myosuroides*; Am). La numeración de la ACCasa utilizada en el presente documento se basa en la numeración de la secuencia de referencia de la proteína ACCasa del pasto negro (SEQ ID NO: 1) o en un residuo de aminoácido análogo en una ACCasa paralog (V=CY3; H=3'DMT dC CPG). La siguiente tabla enumera las mutaciones ACCasa que producen uno o más de aloxidim, butroxiidim, cletodim, cloproxiidim, cicloxiidim, setoxidim, tepraloxidim, tralkoxidim, clorazifop, clodinafop, clofop, diclofop, fenoxaprop, fenoxaprop-P, fentiaprop, fluazifop, fluazifop-P, haloxifop, haloxifop-P, isoxapirifop, propaquizafop, quizalofop, quizalofop-P, trifop, pinoxaden, sales y ésteres agronómicamente aceptables de cualquiera de estos herbicidas, y combinaciones de los mismos fenotipo resistente.

Cambio de aminoácidos	Cambio de codón		Cambio de aminoácidos	Cambio de codón
I1781A	ATA > GCT ATA > CCG ATA > GCA ATA > GCG		C2088F	TGC > TTT TGC > TTC
I1781L	ATA > CTT ATA > CTC ATA > CTA ATA > CTG ATA > TTA ATA > TTG		C2088G	TGC > GGT TGC > GGC TGC > GGA TGC > GGG
I1781M	ATA > ATG		C2088H	TGC > CAT TGC > CAC
I1781N	ATA > AAT ATA > CAA		C2088K	TGC > AAA TGC > AAG
I1781S	ATA > TCT ATA > TCC ATA > TCA ATA > TCG		C2088L	TGC > CTT TGC > CTC TGC > CTA TGC > CTG TGC > TTA TGC > TTG
I1781T	ATA > ACT ATA > ACC ATA > ACA ATA > ACG		C2088N	TGC > AAT TGC > CAA
			C2088P	TGC > CCT TGC > CCC

				TGC > CCA
Cambio de aminoácidos	Cambio de codón		Cambio de aminoácidos	Cambio de codón
I1781V	ATA > GTT			TGC > CCG
	ATA > GTC			
	ATA > GTA		C2088Q	TGC > CAA
	ATA > GTG			TGC > CAG
G1783C	GGA > TGT		C2088R	TGC > CGT
	GGA > TGC			TGC > CGC
				TGC > CGA
A1786P	GCT > CCT			TGC > CGG
	GCT > CCC			TGC > AGA
	GCT > CCA			TGC > AGG
	GCT > CCG			
			C2088S	TGC > TCT
D2078G	GAT > GGT			TGC > TCC
	GAT > GGC			TGC > TCA
	GAT > GGA			TGC > TCG
	GAT > GGG			
			C2088T	TGC > ACT
D2078K	GAT > AAA			TGC > ACC
	GAT > AAG			TGC > ACA
				TGC > ACG
D2078T	GAT > ACT			
	TAG > ACC		C2088V	TGC > GTT
	TAG > ACA			TGC > GTC
	GAT > ACG			TGC > GTA
				TGC > GTG
S2079F	AGC > TTT			
	AGC > TTC		C2088W	TGC > TGG
K2080E	AAG > GAA			
	AAG > GAG			

Del mismo modo, las Figs. 29 y 30 representan, respectivamente, C-GRONs (codificantes) y NC-GRONs (no codificantes) adecuados para introducir mutaciones en el genoma del lino con el fin de introducir una o más de las siguientes sustituciones de aminoácidos en el gen EPSPS (con toda la numeración relativa a la secuencia de aminoácidos de la proteína AroA de *E. coli* (equivalente de EPSPS procariota) (tal como las descritas en Patente de los Estados Unidos No. 8,268,622). (V=CY3; H=3'DMT dC CPG). La siguiente tabla enumera las mutaciones EPSPS que producen sales y ésteres agrónomicamente aceptables de glifosato de cualquiera de estos herbicidas, y combinaciones de fenotipo resistente a los mismos.

Cambio de aminoácidos	Cambio de codón
G96A	GGA > GCT
	GGA > GCC
	GGA > GCA

Cambio de aminoácidos	Cambio de codón
T97I	GGA > GCG ACA > ATT ACA > ATC ACA > ATA
P101A	CCG > GCT CCG > CCG CCG > GCA CCG > GCG
P101S	CCG > TCT CCG > TCC CCG > TCA CCG > TCG
P101T	CCG > ACT CCG > ACC CCG > ACA CCG > ACG

El término "CRISPR", tal como se utiliza en el presente documento, se refiere a elementos; es decir, un gen cas (asociado a CRISPR), un transcrito (por ejemplo, ARNm) o proteína y al menos una secuencia espaciadora CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats, también conocidas como SPIDRs--SPacer Interspersed Direct Repeats); que cuando están efectivamente presentes o expresadas en una célula podrían efectuar la escisión de una secuencia de ADN objetivo a través de la maquinaria celular CRISPR/CAS tal como se describe en por ejemplo, Cong, L. et al., Science, vol. 339 no 6121 pp. 819-823 (2013); Jinek et al, Science, vol. 337:816-821 (2013); Wang et al., RNA, vol. 14, pp. 903-913 (2008); Zhang et al., Plant Physiology, vol. 161, pp. 20-27 (2013), Zhang et al, solicitud PCT n.º PCT/US2013/074743y Charpentier y otros, solicitud PCT No. PCT/US2013/032589. En algunas realizaciones, tal como por ejemplo una CRISPR para uso en una célula eucariota, una CRISPR tal como se contempla en el presente documento también puede incluir un elemento adicional que incluye una secuencia para una o más señales funcionales de localización nuclear. Las CRISPR contempladas en el presente documento pueden expresarse en, administrarse a y/o estar presentes en una célula (tal como una célula vegetal) de muchas formas o manifestaciones. Por ejemplo, una CRISPR tal como se contempla en el presente documento puede incluir o implicar una o más de una CRISPR en un plásmido, una CRISPR nickasa en un plásmido, una CRISPRa en un plásmido o una CRISPRi en un plásmido, tal como se indica a continuación:

CRISPR en un plásmido: Un vector de expresión recombinante que comprende:

(i) una secuencia de nucleótidos que codifica un ARN dirigido al ADN (por ejemplo, ARN guía), en donde el ARN dirigido al ADN comprende:

- a. un primer segmento que comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN objetivo (por ejemplo, protoespaciador, espaciador o ARNcr); y
- b. un segundo segmento que interactúa con un polipéptido modificador dirigido al sitio (por ejemplo, ARNcr transactivador o ARNcrtra); y

(ii) una secuencia de nucleótidos que codifica el polipéptido modificador dirigido al sitio (por ejemplo, gen cas), en donde el polipéptido dirigido al sitio comprende:

- a. una porción de unión al ARN que interactúa con el ARN objetivo del ADN (por ejemplo, el lóbulo REC); y
- b. Una porción de actividad que provoca roturas de doble cadena en el ADN objetivo (por ejemplo, el lóbulo NUC), en donde el sitio de las roturas de doble cadena en el ADN objetivo está determinado por el ARN objetivo del ADN.

Midasa CRISPR en un plásmido. Un vector de expresión recombinante que comprende:

(i) una secuencia de nucleótidos que codifica un ARN dirigido al ADN (por ejemplo, ARN guía), en donde el ARN dirigido al ADN comprende:

- 5 a. un primer segmento que comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN objetivo (por ejemplo, protoespaciador, espaciador o ARNcr); y
- b. un segundo segmento que interactúa con un polipéptido modificador dirigido al sitio (por ejemplo, ARNcr transactivador o ARNcrtra); y

10 (ii) una secuencia de nucleótidos que codifica el polipéptido modificador dirigido al sitio (por ejemplo, gen cas), en donde el polipéptido dirigido al sitio comprende:

- a. una porción de unión al ARN que interactúa con el ARN objetivo del ADN (por ejemplo, el lóbulo REC); y
- 15 b. una porción de actividad que provoca roturas monocatenarias dentro del ADN objetivo (por ejemplo, el lóbulo NUC), en donde el sitio de las roturas monocatenarias dentro del ADN objetivo está determinado por el ARN ADN-objetivo.

CRISPRa en un plásmido. Un vector de expresión recombinante que comprende:

20 (i) una secuencia de nucleótidos que codifica un ARN dirigido al ADN (por ejemplo, ARN guía), en donde el ARN dirigido al ADN comprende:

- a. un primer segmento que comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN objetivo (por ejemplo, protoespaciador, espaciador o ARNcr); y
- 25 b. un segundo segmento que interactúa con un polipéptido modificador dirigido al sitio (por ejemplo, ARNcr transactivador o ARNcrtra); y

(ii) una secuencia de nucleótidos que codifica el polipéptido modificador dirigido al sitio (por ejemplo, gen cas), en donde el polipéptido dirigido al sitio comprende:

- 30 a. una porción de unión al ARN que interactúa con el ARN objetivo del ADN (por ejemplo, el lóbulo REC); y
- b. Una porción de actividad que modula la transcripción (por ejemplo, el lóbulo NUC; en ciertas realizaciones aumenta la transcripción) dentro del ADN objetivo, en donde el sitio de la modulación transcripcional dentro del ADN objetivo está determinado por el ARN ADN-objetivo.
- 35

CRISPRi en un plásmido. Un vector de expresión recombinante que comprende:

40 (i) una secuencia de nucleótidos que codifica un ARN dirigido al ADN (por ejemplo, ARN guía), en donde el ARN dirigido al ADN comprende:

- a. un primer segmento que comprende una secuencia de nucleótidos que es complementaria a una secuencia en un ADN objetivo (por ejemplo, protoespaciador, espaciador o ARNcr); y
- 45 b. un segundo segmento que interactúa con un polipéptido modificador dirigido al sitio (por ejemplo, ARNcr transactivador o ARNcrtra); y

(ii) una secuencia de nucleótidos que codifica el polipéptido modificador dirigido al sitio (por ejemplo, gen cas), en donde el polipéptido dirigido al sitio comprende:

- a. una porción de unión al ARN que interactúa con el ARN objetivo del ADN (por ejemplo, el lóbulo REC); y
- 50 b. Una porción de actividad que modula la transcripción/traducción (por ejemplo, el lóbulo NUC; en algunas realizaciones disminuye la transcripción/traducción) dentro del ADN objetivo, en donde el sitio de modulación transcripcional/traduccional dentro del ADN objetivo está determinado por el ARN objetivo del ADN.

55 Cada una de las CRISPR en un plásmido, CRISPR nickasa en un plásmido, CRISPRa en un plásmido, y CRISPRi en un plásmido puede en algunas realizaciones alternativamente tener uno o más elementos apropiados para ser administrados, expresados o presentes en una célula como un ARN (por ejemplo, ARNm) o una proteína en lugar de en un plásmido. El suministro de ARNm protegido puede realizarse como se describe en Kariko, et al, Patente de los Estados Unidos No. 8,278,036.

60 En algunas realizaciones, cada una de las CRISPRi y CRISPRa puede incluir una cas9 desactivada (dCas9). Una cas9 desactivada sigue uniéndose al ADN objetivo, pero no tiene actividad de corte. La nucleasa Cas9 deficiente puede ser el resultado de mutaciones puntuales D10A y H840A que inactivan sus dos dominios catalíticos.

En algunas realizaciones, un CRISPRi inhibe la iniciación o elongación de la transcripción a través del impedimento estérico de la ARN Polimerasa II. CRISPRi puede mejorarse opcionalmente (CRISPRei) mediante la fusión de un dominio represor fuerte con el extremo C-terminal de una proteína dCas9. En algunas realizaciones, un dominio represor recluta y emplea modificadores de la cromatina. En algunas realizaciones, el dominio represor puede incluir, entre otros, dominios como los descritos en Kagale, S. et al., *Epigenetics*, vol. 6 no 2 pp141-146 (2011):

1. LDLNRPPPVEN - Dominio represor OsERF3 (motivo LxLxPP)
 2. LRLFGVNM - Dominio represor AtBRD (motivo R/KLFGV)
 3. LKLFGVWL - Dominio represor de AtHsfB1 (motivo R/KLFGV)
 4. LDLELRLGFA - Dominio represor AtSUP (motivo EAR)
 5.
 ERSNSIELRNSFYGRARTSPWSYGDYDNCQQDHDYLLGFSWPPRSYTCSE
 CKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSSTPSPYPNPNSYST
 MANSPPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTKKSS
 LLVEAGEATRFTSKDACKILRNDEIISLELEIGLINESEQDLDLELRLGFA*-
 gen AtSUP completo que contiene un dominio represor (motivo EAR)

En algunas realizaciones, una activación CRISPRa de la transcripción lograda mediante el uso de la proteína dCas9 que contiene un activador transcripcional de extremo C-terminal fusionado. En algunas realizaciones, una activación puede incluir, entre otros, VP64 (4X VP16), dominio de activación AtERF98 o concámeros AtERF98x4 tal como los descritos en Cheng, AW et al., *Cell Research*, pp1-9 (2013); Perez-Pinera, P. et al., *Nature Methods*, vol. 10 pp 913-976 (2013); Maeder, ML. et al., *Nature Methods*, vol. 10 pp 977-979 (2013) y Mali, P., et al., *Nature Biotech.*, vol. 31 pp 833-838 (2013).

En algunas realizaciones, la CRISPR incluye una nickasa. En ciertas realizaciones, se utilizan dos o más nickasas CRISPR. En algunas realizaciones, las dos o más nickasas cortan cadenas opuestas del ácido nucleico objetivo. En otras realizaciones, las dos o más nickasas cortan la misma cadena de ácido nucleico objetivo.

Tal y como se utiliza en el presente documento, "proteína represora" o "represor" se refiere a una proteína que se une a un operador de ADN o a ARN para impedir la transcripción o la traducción, respectivamente.

Como se usa en el presente documento, "represión" se refiere a la inhibición de la transcripción o traducción por la unión de la proteína represora a un sitio específico en el ADN o ARNm. En algunas realizaciones, la represión incluye un cambio significativo en el nivel de transcripción o traducción de al menos 1,5 veces, en otras realizaciones al menos dos veces, y en otras realizaciones al menos cinco veces.

Tal como se utiliza en el presente documento, una "proteína activadora" o "activador" con respecto a la transcripción y/o traducción de genes, se refiere a una proteína que se une al operador de ADN o al ARN para mejorar o aumentar la transcripción o traducción, respectivamente.

Tal como se utiliza en el presente documento con respecto a la transcripción y/o traducción de genes, "activación" con respecto a la transcripción y/o traducción de genes, se refiere a la mejora o aumento de la transcripción o traducción mediante la unión de la proteína activadora a un sitio específico en el ADN o ARNm. En algunas realizaciones, la activación incluye un cambio significativo en el nivel de transcripción o traducción de al menos 1,5 veces, en algunas realizaciones de al menos dos veces, y en algunas realizaciones de al menos cinco veces.

Las condiciones que aumentan uno o más procedimientos celulares de reparación del ADN pueden incluir uno o más de los siguientes: introducción de uno o más sitios en el GRON o en el ADN de la célula vegetal que son objetivos para la reparación por escisión de bases, introducción de uno o más sitios en el GRON o en el ADN de la célula vegetal que son objetivos para la unión de extremos no homólogos, introducción de uno o más sitios en el GRON o en el ADN de la célula vegetal que son objetivos para la unión de extremos mediada por microhomología, la introducción de uno o más sitios en el GRON o en el ADN de la célula vegetal que sean objetivos de la recombinación homóloga, y la introducción de uno o más sitios en el GRON o en el ADN de la célula vegetal que sean objetivos para efectuar la reparación (por ejemplo, reparación por escisión de bases (BER); reparación por recombinación homóloga (HR); reparación por emparejamiento erróneo (MMR); reparación por unión de extremos no homólogos (NHEJ) que incluye NHEJ clásica y alternativa; y reparación por escisión de nucleótidos (NER)).

Como se describe en el presente documento, los GRON para su uso en el presente documento pueden incluir una o más de las siguientes alteraciones de los nucleótidos de ARN y ADN convencionales:

uno o más nucleótidos abásicos;
 uno o más nucleótidos 8'oxo dA y/o 8'oxo dG;

una base inversa en su extremo 3';
 uno o más 2'O-metil nucleótidos;
 uno o más nucleótidos de ARN;
 uno o más nucleótidos de ARN en el extremo 5' del mismo, y en algunas realizaciones 2, 3, 4, 5, 6, 7, 8, 9,
 10, o más; en las que uno o más de los nucleótidos de ARN pueden modificarse aún más; uno o más
 nucleótidos de ARN en el extremo 3' del mismo, y en algunas realizaciones 2, 3, 4, 5, 6, 7, 8, 9, 10, o más;
 en las que uno o más de los nucleótidos de ARN pueden modificarse aún más;
 uno o más nucleótidos 2'O-metil ARN en el extremo 5' del mismo, y en algunas realizaciones 2, 3, 4, 5, 6, 7,
 8, 9, 10, o más;
 un colorante intercalante;
 una tapa terminal 5';
 una modificación de la columna vertebral seleccionada del grupo que consiste en una modificación de
 fosfotioato, una modificación de metilfosfonato, una modificación de ácido nucleico bloqueado (LNA), una
 modificación de O -(2-metoxietil) (MOE), una modificación de di PS y una modificación de ácido nucleico
 peptídico (PNA);
 uno o más enlaces cruzados intracadena;
 uno o más colorantes fluorescentes conjugados con el mismo, y en algunas realizaciones en el extremo 5' o
 3' del GRON; y
 una o más bases que aumentan la energía de hibridación.

Esta lista no pretende ser limitativa.

El término "base tambaleante", tal como se utiliza en el presente documento, se refiere a un cambio en una o más
 bases nucleotídicas de una secuencia nucleotídica de referencia en la que el cambio no modifica la secuencia del
 aminoácido codificado por el nucleótido en relación con la secuencia de referencia.

El término "no nucleótido" o "nucleótido abásico", tal como se utiliza en el presente documento, se refiere a cualquier
 grupo o compuesto que puede incorporarse a una cadena de ácido nucleico en lugar de una o más unidades
 nucleotídicas, incluyendo sustituciones de azúcar y/o fosfato, y permite que las bases restantes exhiban su actividad
 enzimática. El grupo o compuesto es abásico en el sentido de que no contiene una base nucleotídica comúnmente
 reconocida, tal como adenosina, guanina, citosina, uracilo o timina. Puede tener sustituciones para un 2' o 3' H u OH
 como se describe en la técnica y en el presente documento.

Como se describe en el presente documento, la calidad y la eficiencia de conversión del GRON pueden mejorarse
 sintetizando la totalidad o una parte del GRON utilizando multímeros de nucleótidos, tales como dímeros, trímeros,
 tetrámeros, etc., mejorando su pureza.

La secuencia objetivo de ácido desoxirribonucleico (ADN) se encuentra dentro de una célula vegetal, por ejemplo, la
 secuencia objetivo de ADN se encuentra en el genoma de la célula vegetal. La célula vegetal puede ser no transgénica
 o transgénica, y la secuencia de ADN objetivo puede ser un transgén o un gen endógeno de la célula vegetal.

Las condiciones que aumentan uno o más procedimientos celulares de reparación del ADN comprenden la
 introducción de uno o más compuestos que inducen roturas de una o dos cadenas de ADN en la célula vegetal antes,
 coincidiendo o después de introducir el GRON en la célula vegetal. En el presente documento se describen
 compuestos ejemplares.

Los procedimientos y composiciones descritos en el presente documento son aplicables a las plantas en general. Sólo
 a título de ejemplo, una especie vegetal puede seleccionarse del grupo que consiste la colza, el girasol, el maíz, el
 tabaco, la remolacha azucarera, el algodón, el maíz, el trigo (incluyendo, entre otros, *Triticum spp*, *Triticum aestivum*,
Triticum durum, *Triticum timopheevii*, *Triticum monococcum*, *Triticum spelta*, *Triticum zhukovskyi* y *Triticum urartu* y sus
 híbridos), cebada (incluyendo, entre otros, *Hordeum vulgare* L., *Hordeum comosum*, *Hordeum depressum*, *Hordeum*
intercedens, *Hordeum jubatum*, *Hordeum marinum*, *Hordeum marinum*, *Hordeum parodii*, *Hordeum pusillum*, *Hordeum*
secalinum y *Hordeum spontaneum*), arroz (incluyendo, entre otros, *Oryza sativa* subsp. indica, *Oryza sa tiva* subsp.
 japonica, *Oryza sativa* subsp. javanica, *Oryza sativa* subsp. glutinosa (arroz glutinoso), grupo *Oryza sativa* Aromatica
 (por ejemplo, basmati), y *Oryza sa tiva* (grupo del arroz flotante)), alfalfa, cebada, sorgo, tomate, mango, melocotón,
 manzana, pera, fresa, plátano, melón, yuca, patata, zanahoria, lechuga, cebolla, soja, soja spp, caña de azúcar,
 guisante, garbanzo, guisante forrajero, haba, lentejas, nabo, colinabo, coles de Bruselas, altramuza, coliflor, col rizada,
 habas, álamo, pino, eucalipto, uva, cítricos, triticale, alfalfa, centeno (incluyendo, entre otros, *Secale sylvestre*, *Secale*
strictum, *Secale cereale*, *Secale vavilovii*, *Secale africanum*, *Secale ciliatoglume*, *Secale ancestrale* y *Secale*
montanum), avena, césped (incluyendo, entre otros, céspedes como *Zoysia japonica*, *Agrostis palustris*, *Poa*
pratensis, *Poa annua*, *Digitaria sanguinalis*, *Cyperus rotundus*, *Kyllinga brevifolia*, *Cyperus amuricus*, *Erigeron*
canadensis, *Hydrocotyle sibthorpioides*, *Kummerowia striata*, *Euphorbia humifusa* y *Viola arvensis*) y las gramíneas
 forrajeras, el lino, la colza, el algodón, la mostaza, el pepino, la gloria de la mañana, el bálsamo, el pimiento, la
 berenjena, la caléndula, el loto, la col, la margarita, el clavel, el tulipán, el iris, el lirio, las plantas productoras de frutos
 secos, siempre que no se mencionen ya específicamente. También pueden aplicarse, en su totalidad o en parte, a
 todos los demás sistemas biológicos, incluyendo, entre otros, bacterias, levaduras, hongos, algas y células de

mamíferos, e incluso a sus orgánulos (por ejemplo, mitocondrias y cloroplastos). En algunas realizaciones, el organismo o célula es de una especie seleccionada del grupo que consiste en *Escherichia coli*, *Mycobacterium smegmatis*, *Bacillus subtilis*, *Chlorella*, *Bacillus thuringiensis*, *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, *Chlamydomonas reinhardtii*, *Pichia pastoris*, *Corynebacterium*, *Aspergillus niger* y *Neurospora crassa*. En algunas realizaciones, la levadura es *Yarrowia lipolytica*. En otras realizaciones, la levadura no es *Saccharomyces cerevisiae*. En algunas realizaciones, la planta o célula vegetal es de una especie seleccionada del grupo que consiste en *Arabidopsis thaliana*, *Solanum tuberosum*, *Solanum phureja*, *Oryza sativa*, *Glycine max*, *Amaranthus tuberculatus*, *Linum usitatissimum* y *Zea mays*. La especie vegetal puede seleccionarse del grupo que consiste en plantas monocotiledóneas de la familia de las gramíneas Poaceae. La familia Poaceae puede dividirse en dos clados principales, el clado que contiene las subfamilias Bambusoideae, Ehrhartoideae y Pooideae (el clado BEP) y el clado que contiene las subfamilias Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Micrairoideae, Aristidoideae y Danthonioideae (el clado PACCMAD). La subfamilia Bambusoideae incluye la tribu Oryzeae. Las especies vegetales pueden referirse a plantas del clado BEP, en particular plantas de las subfamilias Bambusoideae y Ehrhartoideae. El clado BET incluye las subfamilias Bambusoideae, Ehrhartoideae y el grupo Triticodae y ningún otro grupo de la subfamilia Pooideae. Las plantas de cultivo BET son plantas cultivadas para alimentación o forraje que son miembros del subclado BET, por ejemplo, cebada, maíz, etc.

Como se describe en el presente documento, los procedimientos comprenden además la regeneración de una planta que tiene una mutación introducida por el GRON a partir de la célula vegetal, y pueden comprender la recolección de semillas de la planta.

Se divulgan además en el presente documento células vegetales que comprenden una modificación genómica introducida por un GRON de acuerdo con los procedimientos descritos en el presente documento, una planta que comprende una modificación genómica introducida por un GRON de acuerdo con los procedimientos descritos en el presente documento, o una semilla que comprende una modificación genómica introducida por un GRON de acuerdo con los procedimientos descritos en el presente documento; o la progenie de una semilla que comprende una modificación genómica introducida por un GRON de acuerdo con los procedimientos descritos en el presente documento.

Otras realizaciones de la divulgación se desprenderán de la siguiente descripción detallada y de las reivindicaciones.

BREVE DESCRIPCIÓN DE LAS FIGURAS

La Fig. 1 muestra la conversión de BFP a GFP mediada por GRONs marcados con fosfotioato (PS) (con 3 fracciones PS en cada extremo del GRON) y GRONs marcados con 5'Cy3/ 3'idC.

La Fig. 2 muestra GRONs que comprenden ARN/ADN, denominados en el presente documento "GRONs 2'-O-metilados".

La Fig. 3 es un esquema de la localización en el gen *bfp* donde se dirigen las CRISPRs BFP5.

La Fig. 4 muestra los resultados del efecto de las CRISPRs introducidas con los GRONs Cy3 o 3PS en diversas longitudes, sobre el porcentaje de conversión de BFP a GFP en un sistema modelo transgénico de *Arabidopsis thaliana* con BFP.

La Fig. 5 muestra los resultados del efecto de las CRISPRs introducidas con los GRONs 3PS en diversas longitudes, sobre el porcentaje de conversión de BFP a GFP en un sistema modelo transgénico de *Arabidopsis thaliana* con BFP.

La Fig. 6 es un esquema de los diseños GRON utilizados en el Ejemplo 9.

La Fig. 7 muestra la medición del porcentaje medio de protoplastos GFP positivos de un sistema modelo transgénico BFP de *Arabidopsis thaliana* determinado por citometría de flujo a partir de GRONs de 71-mer.

La Fig. 8 muestra la medición de protoplastos GFP positivos del sistema modelo transgénico BFP de *Arabidopsis thaliana*, determinada por citometría de flujo a partir de GRONs de 201-mer.

La Fig. 9 muestra el efecto de las CRISPRs introducidas con GRONs codificantes y no codificantes sobre el porcentaje promedio de células positivas a la GFP en un sistema modelo de *Arabidopsis thaliana* transgénica con BFP.

La Fig. 10 es un esquema de la unión de un GRON monocatenario o ADN bicatenario al complejo CRISPR/Cas.

La Fig. 11 muestra los resultados del efecto de CRISPRs y GRONs en la mediación de la conversión de BFP a GFP en un sistema modelo transgénico de *Arabidopsis thaliana* de espaciadores de diferentes longitudes.

La Fig. 12 muestra los resultados del efecto de CRISPRs y GRONs en la mediación de la conversión de BFP a GFP en un sistema modelo de *Arabidopsis thaliana* transgénica BFP de espaciadores se codificaron en un plásmido (plásmido ARNg) o utilizados como un amplicón (amplicón ARNg).

La Fig. 13 muestra los resultados del efecto de CRISPRs y GRONs en la mediación de la conversión de BFP a GFP en un sistema modelo de *Arabidopsis thaliana* transgénica BFP de GRONs de 41-mer no modificados frente a GRONs modificados con 3PS.

La Fig. 14 muestra los resultados de la secuenciación de nueva generación de microcallos de *Linum usitatissimum* (lino) de 3 y 6 semanas de edad derivados de protoplastos de punta de brote tratados con PEG con plásmido CRISPR a T=0.

La Fig. 15 muestra los resultados de la secuenciación de próxima generación de microcallos de *Linum usitatissimum* de 3 y 6 semanas de edad derivados de protoplastos de puntas de brotes tratados con PEG con plásmido CRISPR a T=0.

La Fig. 16 muestra la actividad de la nucleasa CRISPR-Cas y la edición de genes en protoplastos de *A. thaliana*. 16A: Distribución de indeles con base en el tamaño, determinada por secuenciación profunda, en protoplastos tratados con el plásmido CRISPR-Cas (BC-1) a las 72 h de su administración. Los indeles representaron el 0,79% del total de lecturas. 16B: Edición de *BFP* a *GFP* medida por el porcentaje de protoplastos con fluorescencia *GFP* identificados por citometría de flujo 72 h después de la entrega del plásmido (BC-1) y GRON (CG-6). Los datos representados no están normalizados en función de la eficacia de transfección. Las barras de error son m.e.s. ($n = 9$).

La Fig. 17 muestra la edición de genes en protoplastos de *A. thaliana* utilizando CRISPR-Cas combinado con GRONs de diferentes longitudes y modificaciones químicas. 17a: Comparación de 3PS y GRONs no modificados en la edición del gen *BFP* a *GFP* medida por citometría de flujo a las 72 h después de la entrega del plásmido (BC-1) y GRONs (CG-1) o (CG-2). 17b: Comparación de las longitudes de los GRON en la edición de genes *BFP* a *GFP*, medidas por citometría de flujo a las 72 horas de la entrega del plásmido (BC-2) y los GRON (CG-5) o (CG-8). 17c: Comparación de 3PS con GRONs 2'-O-Me para la edición génica de *BFP* a *GFP* medida por citometría de flujo a las 72 h de la entrega del plásmido (BC-1) y GRONs (CG-6), (CG-9) o (CG-10). 17d: Comparación de 3PS- a Cy3GRONs en la edición de genes *BFP* a *GFP* medida por citometría de flujo a las 72 h post entrega de plásmido (BC-3) y GRONs (CG-3) o (CG-4). Las barras de error son m.e.s. ($n = 3$). (CG-1): BFP antisentido 41 nb sin modificar; (CG-2): BFP antisentido 41 nb 3PS modificado; (CG-3): BFP sentido 41 nb 3PS modificado; (CG-4): BFP sentido 41 nb Cy3 modificado; (CG-5): BFP sentido 60 nb 3PS modificado; (CG-6): BFP antisentido 201 nb 3PS modificado; (CG-8): BFP sentido 201 nb 3PS modificado; (CG-9): BFP antisentido 201 nb modificación 2'-O-Me en la primera base 5' del ARN; (CG-10): BFP antisentido 201 nb modificaciones 2'-O-Me en las nueve primeras bases 5' del ARN.

La Fig. 18 muestra la actividad de la nucleasa TALEN y la edición de genes en protoplastos de *A. thaliana* y *L. usitatissimum*. 18a: Distribución de indeles basada en el tamaño determinado por secuenciación profunda en protoplastos de *Arabidopsis* tratados con plásmido TALEN (BT-1) a las 72 h de su administración. Los indeles representaron el 0,51% del total de lecturas. 18b: Edición génica de *BFP* a *GFP* medida por citometría de flujo a las 48 h de la entrega del plásmido (BT-1) y GRON (CG-7). 18c: Distribución representativa de indeles basada en la longitud de pb en protoplastos de *L. usitatissimum* tratados con un TALEN (LuET-1) dirigido a los genes EPSPS 7 d después de la entrega. La frecuencia total de indeles es del 0,50%. d, Edición del gen EPSPS de *L. usitatissimum* medida por secuenciación profunda a los 7 d de la introducción del plásmido (LuET-1) y GRON (CG-11) en protoplastos. El porcentaje del total de lecturas representa el número de lecturas que contienen ediciones T97I y P101A como porcentaje del total de lecturas. Las barras de error son m.e.s. ($n = 3$). (CG-7): BFP sentido 201 nb 3PS modificado; (CG-11): EPSPS sentido 144 nb Cy3 modificado.

La Fig. 19 muestra los efectos de los antibióticos inductores de rotura de doble cadena zeocina y fleomicina sobre la edición de *BFP* a *GFP* en protoplastos transgénicos de *A. thaliana*. Los protoplastos se trataron con zeocina o fleomicina durante 90 min antes de la introducción de GRON en PEG (CG2). La edición exitosa dio lugar a fluorescencia *GFP*. Los protoplastos fluorescentes verdes se cuantificaron utilizando un citómetro de enfoque acústico Attune.

La Fig. 20 muestra células de *A. thaliana* transgénicas de BFP convertidas cinco días después de la administración de GRON en protoplastos transgénicos de BFP, dirigiendo la conversión de BFP a GFP. La fluorescencia verde indica la edición de BFP-GFP. A, imagen de campo claro; B, el mismo campo de visión en luz azul. Las barras de error son m.e.s. ($n=4$); (CG2): BFP antisentido 41 nb 3PS modificada; (CG12) BFP antisentido 41 nb 3PS modificada no dirigida. Las imágenes se adquirieron con un sistema ImageXpress Micro (Molecular Devices, Sunnyvale, CA, EE. UU.) Barra de escala = 20 μ m

La Fig. 21 muestra los diseños CRISPR-Cas y TALEN. 21A: Esquema del plásmido CRISPR-Cas. El promotor de la manopina sintasa (Mas) impulsa la transcripción del gen *Cas9*, cuyo codón está optimizado para plantas superiores. El gen *Cas9* contiene dos señales de localización nuclear (NLS) SV40 en cada extremo del gen y una etiqueta epitópica 2X FLAG. El promotor U6 de *A. thaliana* impulsa la transcripción del andamiaje de ARNg y la transcripción se termina mediante una señal poli(T). 21B: Esquema del plásmido TALEN. El promotor Mas impulsa la transcripción de los brazos derecho e izquierdo del talo unidos por una secuencia 2A de omisión de ribosomas. Una endonucleasa Fok1 está unida al extremo 3' de cada brazo Tale. El extremo 5' del talo izquierdo contiene una señal de localización nuclear (NLS) y una etiqueta epitópica V5. rbcT es el terminador del gen RBCSE9 de *Pisum sativum*.

La Fig. 22 muestra las regiones objetivo de las nucleasas BFP y EPSPS. a, región objetivo del gen *BFP* para los protoespaciadores CRISPR-Cas, BC-1, BC-2 y BC-3 y el TALEN BT-1, brazos izquierdo y derecho. La secuencia PAM se muestra en rojo. Los sitios de unión de TALEN están en negrita y subrayados. El sitio de edición de *BFP* a *GFP* CAC→TAC (H66Y) está en negrita verde. b, región objetivo del gen *EPSPS* para los brazos TALEN, LuET-1, izquierdo y derecho. El lugar de las conversiones *EPSPS* ACA>ATA y CCG>GCG (T97I y P101A) aparecen en verde.

La Fig. 23 muestra la secuencia de aminoácidos del producto génico de la ACCasa de *Alopecurus myosuroides* (pasto negro) (SEQ ID NO: 1).

La Fig. 24 muestra la secuencia de aminoácidos del producto génico EPSPS de *Escherichia coli* (SEQ ID NO:2).

La Fig. 25 muestra posiciones análogas ejemplares de EPSPS.

La Fig. 26 muestra secuencias ejemplares de ACCasa.

DESCRIPCIÓN DETALLADA

La modificación genética dirigida mediada por oligonucleótidos es una técnica valiosa para su uso en la alteración específica de tramos cortos de ADN para crear delecciones, inserciones cortas y mutaciones puntuales. Estos procedimientos implican el emparejamiento/fusión del ADN, seguido de un evento de reparación/recombinación del ADN. En primer lugar, el ácido nucleico se adhiere a su cadena complementaria en el ADN de doble cadena en un procedimiento mediado por factores proteínicos celulares. Esta fusión crea un par de bases centralmente desajustadas (en el caso de una mutación puntual), lo que provoca una perturbación estructural que muy probablemente estimula la maquinaria proteica endógena para iniciar el segundo paso del procedimiento de reparación: la modificación específica de la secuencia cromosómica y/o la de los orgánulos (por ejemplo, mitocondrias y cloroplastos). Este desajuste recién introducido induce a la maquinaria de reparación del ADN a realizar un segundo evento de reparación, que conduce a la revisión final del sitio objetivo. Los presentes procedimientos y composiciones divulgados en el presente documento pueden mejorar los procedimientos proporcionando enfoques novedosos que aumentan la disponibilidad de los componentes de reparación del ADN, aumentando así la eficiencia y reproducibilidad de las modificaciones mediadas por la reparación génica de los ácidos nucleicos objetivo.

Los procedimientos eficientes para las modificaciones genómicas dirigidas al sitio son deseables para la investigación, la terapia génica clínica, la microbiología industrial y la agricultura. Un enfoque utiliza oligonucleótidos formadores de triplex (TFO) que se unen como terceras cadenas al ADN dúplex de forma específica para cada secuencia, para mediar en la mutagénesis dirigida. Estos TFO pueden actuar suministrando un mutágeno ligado, como el psoraleno o el clorambucilo (Havre et al., *Proc Nat'l Acad Sci, U.S.A.* 90:7879-7883, 1993; Havre et al., *J Virol* 67:7323-7331, 1993; Wang et al., *Mol Cell Biol* 15: 1759-1768, 1995; Takasugi et al., *Proc Nat'l Acad Sci, EE. UU.* 88:5602-5606, 1991; Belousov et al., *Nucleic Acids Res* 25:3440-3444, 1997), o uniéndose con suficiente afinidad como para provocar una reparación propensa a errores (Wang et al., *Science* 271:802-805, 1996).

Otra estrategia para la modificación genómica implica la inducción de recombinación homóloga entre un fragmento de ADN exógeno y el gen objetivo. Este procedimiento se ha utilizado con éxito para atacar y alterar genes seleccionados en células de mamíferos y ha permitido la producción de ratones transgénicos portadores de genes inactivador específicos (Capechi et al., *Science* 244:1288-1292, 1989; Wagner, *Pat. de EE. UU. No. 4,873,191*). Este enfoque implica la transferencia de marcadores seleccionables para permitir el aislamiento de los recombinantes deseados. Sin selección, la proporción de integración homóloga y no homóloga del ADN transfectado en los experimentos típicos de transferencia de genes es baja, normalmente del orden de 1:1000 o menos (Sedivy et al., *Gene Targeting*, W. H. Freeman and Co., Nueva York, 1992). Esta baja eficacia de la integración homóloga limita la utilidad de la transferencia de genes para uso experimental o terapia génica. La frecuencia de la recombinación homóloga puede verse aumentada por el daño producido en el sitio objetivo por la irradiación UV y determinados carcinógenos (Wang et al., *Mol Cell Biol* 8:196-202, 1988), así como por endonucleasas específicas del sitio (Sedivy et al., *Gene Targeting*, W. H. Freeman and Co., Nueva York, 1992; Rouet et al., *Proc Nat'l Acad Sci, EE. UU.* 91:6064-6068, 1994; Segal et al., *Proc Nat'l Acad Sci, EE. UU.* 92:806-810, 1995). Además, el daño en el ADN inducido por fotoaductos de psoraleno dirigidos por triplex puede estimular la recombinación dentro y entre vectores extracromosómicos (Segal et al., *Proc Nat'l Acad Sci, EE. UU.* 92:806-810, 1995; Faruqi et al., *Mol Cell Biol* 16:6820-6828, 1996; Glazer, *Pat. de EE. UU. No. 5,962,426*).

Los fragmentos donantes lineales son más recombinogénicos que sus homólogos circulares (Folger et al., *Mol Cell Biol* 2:1372-1387, 1982). En la recombinación también puede influir la longitud de la homología ininterrumpida entre los sitios donante y objetivo, ya que los fragmentos cortos a menudo parecen ser sustratos ineficaces para la recombinación (Rubnitz et al., *Mol Cell Biol* 4:2253-2258, 1984). No obstante, el uso de fragmentos cortos de ADN o de híbridos de ADN/ARN para la corrección génica es objeto de diversas estrategias. (Kunzelmann et al., *Gene Ther* 3:859-867, 1996).

"Secuencia de ácido nucleico", "secuencia de nucleótidos" y "secuencia de polinucleótidos", tal como se utilizan en el presente documento, se refieren a un oligonucleótido o polinucleótido, y a fragmentos o porciones de los mismos, y a ADN o ARN de origen genómico o sintético que puede ser de cadena simple o doble, y representan la cadena sentido o antisentido.

Tal como se utilizan en el presente documento, los términos "oligonucleótido" y "oligómero" se refieren a un polímero de nucleobases de al menos aproximadamente 10 nucleobases y hasta aproximadamente 1000 nucleobases.

Los términos "molécula modificadora del ADN" y "reactivo modificador del ADN" utilizados en el presente documento se refieren a una molécula que es capaz de reconocer y unirse específicamente a una secuencia de ácido nucleico en el genoma de una célula, y que es capaz de modificar una secuencia de nucleótidos objetivo dentro del genoma, en la que el reconocimiento y la unión específica de la molécula modificadora del ADN a la secuencia de ácido nucleico es independiente de proteínas. El término "independiente de proteínas" utilizado en el presente documento en relación con una molécula modificadora del ADN significa que la molécula modificadora del ADN no requiere la presencia y/o actividad de una proteína y/o enzima para el reconocimiento y/o la unión específica a una secuencia de ácido nucleico.

Se ejemplifican las moléculas modificadoras del ADN, pero sin limitarse a ellas, oligonucleótidos formadores de tríplex, ácidos nucleicos peptídicos, poliamidas y oligonucleótidos destinados a promover la conversión génica. Las moléculas modificadoras del ADN de la presente divulgación se distinguen de las secuencias de ácido nucleico de la técnica anterior que se utilizan para la recombinación homóloga (Wong & Capecchi, *Molec. Cell. Biol.* 7:2294-2295, 1987) en el sentido de que las secuencias de ácido nucleico del estado de la técnica que se utilizan para la recombinación homóloga dependen de proteínas. El término "dependiente de proteínas", tal como se utiliza en el presente documento en relación con una molécula, significa que la molécula requiere la presencia y/o actividad de una proteína y/o enzima para el reconocimiento y/o la unión específica de la molécula a una secuencia de ácido nucleico. Los procedimientos para determinar si una molécula modificadora del ADN requiere la presencia y/o actividad de una proteína y/o enzima para el reconocimiento de, y/o la unión específica a, una secuencia de ácido nucleico está dentro de la habilidad en la técnica (véase, por ejemplo, Dennis et al. *Nucl. Acids Res.* 27:4734-4742, 1999). Por ejemplo, la molécula modificadora del ADN puede incubarse *in vitro* con la secuencia de ácido nucleico en ausencia de proteínas y/o enzimas. La detección de la unión específica entre la molécula modificadora del ADN y la secuencia de ácido nucleico demuestra que la molécula modificadora del ADN es independiente de las proteínas. Por otra parte, la ausencia de unión específica entre la molécula modificadora del ADN y la secuencia de ácido nucleico demuestra que la molécula modificadora del ADN depende de proteínas y/o requiere factores adicionales.

El "oligonucleótido formador de tríplex" (TFO) se define como una secuencia de ADN o ARN capaz de unirse en el surco mayor de una hélice dúplex de ADN o ARN para formar una triple hélice. Aunque el TFO no está limitado a ninguna longitud en particular, una longitud preferida del TFO es de 250 nucleótidos o menos, 200 nucleótidos o menos, o 100 nucleótidos o menos, o de 5 a 50 nucleótidos, o de 10 a 25 nucleótidos, o de 15 a 25 nucleótidos. Aunque es necesario cierto grado de especificidad de secuencia entre el TFO y el ADN dúplex para la formación de la triple hélice, no se requiere ningún grado particular de especificidad, siempre que la triple hélice sea capaz de formarse. Del mismo modo, no se requiere un grado específico de avidez o afinidad entre el TFO y la hélice dúplex siempre que la triple hélice sea capaz de formarse. Sin pretender limitar la longitud de la secuencia de nucleótidos a la que se une específicamente el TFO, en un ejemplo, la secuencia de nucleótidos a la que se une específicamente el TFO es de 1 a 100, en algunos ejemplos de 5 a 50, en otros ejemplos de 10 a 25, y en otros ejemplos de 15 a 25, nucleótidos. Además, "triple hélice" se define como un ácido nucleico de doble hélice con un oligonucleótido unido a una secuencia objetivo dentro del ácido nucleico de doble hélice. El ácido nucleico "doble helicoidal" puede ser cualquier ácido nucleico bicatenario, incluido el ADN bicatenario, el ARN bicatenario y los dúplex mixtos de ADN y ARN. El ácido nucleico bicatenario no está limitado a ninguna longitud en particular. Sin embargo, en ejemplos preferidos tiene una longitud superior a 500 pb, en algunos ejemplos superior a 1 kb y en algunos ejemplos superior a aproximadamente 5 kb. En muchas aplicaciones, el ácido nucleico doble helicoidal es ácido nucleico celular, genómico. El oligonucleótido formador del tríplex puede unirse a la secuencia objetivo de forma paralela o antiparalela.

Los "ácidos nucleicos peptídicos", "poliamidas" o "PNA" son ácidos nucleicos en los que el esqueleto de fosfato se sustituye por una estructura de poliamida basada en N-aminoetilglicina. Los PNA tienen una mayor afinidad por los ácidos nucleicos complementarios que sus homólogos naturales siguiendo las reglas de emparejamiento de bases Watson-Crick. Los PNAs pueden formar estructuras de triple hélice altamente estables con ADN de la siguiente estequiometría: (PNA)₂.ADN. Aunque los ácidos nucleicos peptídicos y las poliamidas no están limitados a ninguna longitud particular, una longitud preferida de los ácidos nucleicos peptídicos y las poliamidas es de 200 nucleótidos o menos, en algunos ejemplos de 100 nucleótidos o menos, y en algunos ejemplos de 5 a 50 nucleótidos de longitud. Sin pretender limitar la longitud de la secuencia nucleotídica a la que se unen específicamente el ácido nucleico peptídico y la poliamida, en un ejemplo, la secuencia nucleotídica a la que se unen específicamente el ácido nucleico peptídico y la poliamida es de 1 a 100, en algunos ejemplos de 5 a 50, en otros ejemplos de 5 a 25, y en otros ejemplos de 5 a 20, nucleótidos.

El término "célula" se refiere a una sola célula. El término "células" se refiere a una población de células. La población puede ser una población pura que comprenda un tipo celular. Asimismo, la población puede comprender más de un tipo celular. En la presente divulgación, no hay límite en el número de tipos celulares que puede comprender una población celular. Una célula, tal como se utiliza en el presente documento, incluye sin limitación células de callo vegetal, células con y sin pared celular, células procariotas y células eucariotas.

El término "sincronizar" o "sincronizado", cuando se refiere a una muestra de células, o "células sincronizadas" o "población celular sincronizada" se refiere a una pluralidad de células que se han tratado para hacer que la población de células esté en la misma fase del ciclo celular. No es necesario que todas las células de la muestra estén sincronizadas. Un pequeño porcentaje de células puede no estar sincronizado con la mayoría de las células de la muestra. Un intervalo preferido de células sincronizadas está entre 10-100%. Un intervalo más preferido está entre 30-100%. Además, no es necesario que las células sean una población pura de un único tipo celular. La muestra puede contener más de un tipo de célula. En este sentido, sólo uno de los tipos celulares puede estar sincronizado o puede estar en una fase diferente del ciclo celular en comparación con otro tipo celular de la muestra.

El término "célula sincronizada" cuando se hace referencia a una sola célula significa que la célula se ha manipulado de tal manera que se encuentra en una fase del ciclo celular que es diferente de la fase del ciclo celular de la célula antes de la manipulación. Alternativamente, una "célula sincronizada" se refiere a una célula que ha sido manipulada para alterar (es decir, aumentar o disminuir) la duración de la fase del ciclo celular en la que se encontraba la célula

antes de la manipulación en comparación con una célula de control (por ejemplo, una célula en ausencia de la manipulación).

5 El término "ciclo celular" se refiere a la progresión fisiológica y morfológica de los cambios que sufren las células cuando se dividen (es decir, proliferan). En general, se reconoce que el ciclo celular se compone de fases denominadas "interfase", "profase", "metafase", "anafase" y "telofase". Además, las partes del ciclo celular pueden denominarse "M (mitosis)", "S (síntesis)", "G0", "G1 (brecha 1)" y "G2 (brecha2)". Además, el ciclo celular incluye periodos de progresión intermedios a las fases mencionadas.

10 El término "inhibición del ciclo celular" se refiere al cese de la progresión del ciclo celular en una célula o población de células. La inhibición del ciclo celular suele ser inducida por la exposición de las células a un agente (químico, proteínico o de otro tipo) que interfiere con aspectos de la fisiología celular para impedir la continuación del ciclo celular.

15 "Proliferación" o "crecimiento celular" se refiere a la capacidad de una célula progenitora de dividirse en dos células hijas de forma repetible, dando lugar a un aumento total de células en la población. La población celular puede estar en un organismo o en un aparato de cultivo.

20 El término "capaz de modificar el ADN" o "medios modificadores del ADN" se refiere a procedimientos, así como a agentes o reactivos endógenos o exógenos que tienen la capacidad de inducir, o pueden ayudar a inducir, cambios en la secuencia de nucleótidos de un segmento objetivo del ADN. Dichos cambios pueden realizarse mediante la supresión, adición o sustitución de una o más bases en el segmento de ADN objetivo. No es necesario que los cambios en la secuencia de ADN confieran cambios funcionales a cualquier gen codificado por la secuencia objetivo. Además, no es necesario que los cambios en el ADN se realicen en una porción o porcentaje concreto de las células.

25 El término "secuencia nucleotídica de interés" se refiere a cualquier secuencia nucleotídica, cuya manipulación puede ser considerada deseable por cualquier razón, por un experto en la técnica. Dichas secuencias de nucleótidos incluyen, entre otras, secuencias codificantes de genes estructurales (por ejemplo, genes reporteros, genes marcadores de selección, oncogenes, genes de resistencia a fármacos, factores de crecimiento, etc.) y secuencias reguladoras no codificantes que no codifican un ARNm o un producto proteico (por ejemplo, secuencia promotora, secuencia potenciadora, secuencia de poliadenilación, secuencia de terminación, ARN reguladores como ARNm, etc.).

30 "Secuencia de aminoácidos", "secuencia polipeptídica", "secuencia peptídica" y "péptido" se utilizan en el presente documento indistintamente para referirse a una secuencia de aminoácidos.

35 "Secuencia objetivo", como se utiliza en el presente documento, se refiere a un ácido nucleico doble helicoidal que comprende una secuencia de más de 8 nucleótidos de longitud, pero menos de 201 nucleótidos de longitud. En algunas realizaciones, la secuencia objetivo tiene entre 8 y 30 bases. La secuencia objetivo, en general, está definida por la secuencia de nucleótidos en una de las cadenas del ácido nucleico doble helicoidal.

40 Tal como se utiliza en el presente documento, una "secuencia rica en purinas" o "secuencia polipurina" cuando se hace en referencia a una secuencia de nucleótidos en una de las cadenas de una secuencia de ácido nucleico doble helicoidal se define como una secuencia contigua de nucleótidos en la que más del 50% de los nucleótidos de la secuencia objetivo contienen una base purina. Sin embargo, se prefiere que la secuencia objetivo rica en purina contenga más del 60% de nucleótidos de purina, en algunos ejemplos más del 75% de nucleótidos de purina, en otros ejemplos más del 90% de nucleótidos de purina y en otros ejemplos el 100% de nucleótidos de purina.

45 Tal como se utiliza en el presente documento, una "secuencia rica en pirimidina" o "secuencia de polipirimidina" cuando se hace referencia a una secuencia de nucleótidos en una de las cadenas de una secuencia de ácido nucleico doble helicoidal se define como una secuencia contigua de nucleótidos en la que más del 50% de los nucleótidos de la secuencia objetivo contienen una base de pirimidina. Sin embargo, se prefiere que la secuencia objetivo rica en pirimidina contenga más del 60% de nucleótidos de pirimidina y, en algunos ejemplos, más del 75% de nucleótidos de pirimidina. En algunos ejemplos, la secuencia contiene más del 90% de nucleótidos de pirimidina y, en otros ejemplos, es 100% de nucleótidos de pirimidina.

50 Una "variante" de una primera secuencia de nucleótidos se define como una secuencia de nucleótidos que difiere de la primera secuencia de nucleótidos (por ejemplo, por tener una o más delecciones, inserciones o sustituciones que pueden detectarse mediante ensayos de hibridación o mediante secuenciación del ADN). Se incluye en esta definición la detección de alteraciones o modificaciones de la secuencia genómica de la primera secuencia de nucleótidos. Por ejemplo, los ensayos de hibridación pueden utilizarse para detectar (1) alteraciones en el patrón de los fragmentos de enzimas de restricción capaces de hibridarse con la primera secuencia de nucleótidos cuando está comprendida en un genoma (es decir, análisis RFLP), (2) la incapacidad de una porción seleccionada de la primera secuencia de nucleótidos para hibridarse con una muestra de ADN genómico que contiene la primera secuencia de nucleótidos (por ejemplo, utilizando sondas de oligonucleótidos específicas de alelo), (3) hibridación inadecuada o inesperada, tal como la hibridación a un locus distinto del locus cromosómico normal para la primera secuencia de nucleótidos (por ejemplo, utilizando hibridación fluorescente in situ (FISH) a cromosomas metafásicos extendidos, etc.). Un ejemplo de variante es una secuencia mutada de tipo silvestre.

Los términos "ácido nucleico" y "ácido nucleico no modificado" utilizados en el presente documento se refieren a uno cualquiera de las cuatro bases de ácido desoxirribonucleico conocidas (es decir, guanina, adenina, citosina y timina). El término "ácido nucleico modificado" se refiere a un ácido nucleico cuya estructura está alterada en relación con la estructura del ácido nucleico no modificado. Ilustrativas de tales modificaciones serían las modificaciones covalentes de sustitución de las bases, tal como la alquilación de amino y nitrógenos de anillo, así como la saturación de dobles enlaces.

Tal como se utilizan en el presente documento, los términos "mutación" y "modificación" y sus equivalentes gramaticales cuando se utilizan en referencia a una secuencia de ácido nucleico se utilizan indistintamente para referirse a una delección, inserción, sustitución, rotura de cadena y/o introducción de un aducto. Una "delección" se define como un cambio en una secuencia de ácido nucleico en la que uno o más nucleótidos están ausentes. Una "inserción" o "adición" es aquel cambio en una secuencia de ácido nucleico que ha dado lugar a la adición de uno o más nucleótidos. Una "sustitución" resulta de la sustitución de uno o más nucleótidos por una molécula que es una molécula diferente de los nucleótidos sustituidos. Por ejemplo, un ácido nucleico puede ser sustituido por un ácido nucleico diferente, como se ejemplifica mediante la sustitución de una timina por una citosina, adenina, guanina o uridina. La pirimidina a pirimidina (por ejemplo, sustituciones de nucleótidos de C a T o de T a C) o de purina a purina (por ejemplo, sustituciones de nucleótidos de G a A o de A a G) se denominan transiciones, mientras que la pirimidina a purina o la purina a pirimidina (por ejemplo, de G a T o de G a C o de A a T o de A a C) se denominan transversiones. Alternativamente, un ácido nucleico puede ser sustituido por un ácido nucleico modificado, tal como por ejemplo la sustitución de una timina por timina glicol. Las mutaciones pueden dar lugar a un desajuste. El término "desajuste" se refiere a una interacción no covalente entre dos ácidos nucleicos, cada uno de los cuales reside en una secuencia de ácido polinucleico diferente, que no sigue las reglas de emparejamiento de bases. Por ejemplo, para las secuencias parcialmente complementarias 5'-AGT-3' y 5'-AAT-3', existe un desajuste G-A (una transición). Los términos "introducción de un aducto" o "formación de un aducto" se refieren a la unión covalente o no covalente de una molécula a uno o más nucleótidos en una secuencia de ADN de tal manera que la unión produce una reducción (en algunos ejemplos del 10% al 100%, en otros ejemplos del 50% al 100%, y en algunos ejemplos del 75% al 100%) en el nivel de replicación y/o transcripción del ADN.

El término "cortador de ADN" se refiere a una molécula que produce una rotura de la cadena. Ejemplos no limitados divulgados en el presente documento incluyen meganucleasas, TALEs/TALENs, antibióticos, dedos de zinc y CRISPRs o sistemas CRISPR/cas. De acuerdo con la presente invención, el cortador de ADN es CRISPR-Cas.

El término "rotura de cadena" cuando se hace referencia a una secuencia de ácido nucleico de doble cadena incluye una rotura de cadena simple y/o una rotura de doble cadena. Una rotura monocatenaria (una mella) se refiere a una interrupción en una de las dos cadenas de la secuencia de ácido nucleico bicatenario. Esto contrasta con una rotura de doble cadena, que se refiere a una interrupción en ambas cadenas de la secuencia de ácido nucleico de doble cadena, que puede dar lugar a extremos romos o escalonados. Las roturas de cadena pueden introducirse en una secuencia de ácido nucleico de doble cadena directamente (por ejemplo, mediante radiación ionizante o tratamiento con determinados productos químicos) o indirectamente (por ejemplo, mediante incisión enzimática en una base del ácido nucleico).

Los términos "célula mutante" y "célula modificada" se refieren a una célula que contiene al menos una modificación en la secuencia genómica de la célula.

El término "porción" cuando se utiliza en referencia a una secuencia de nucleótidos se refiere a fragmentos de esa secuencia de nucleótidos. El tamaño de los fragmentos puede oscilar entre 5 residuos de nucleótidos y toda la secuencia de nucleótidos menos un residuo de ácido nucleico.

Se dice que las moléculas de ADN tienen "extremos 5'" y "extremos 3'" porque los mononucleótidos reaccionan para formar oligonucleótidos de manera que el fosfato 5' de un anillo pentoso mononucleótido se une al oxígeno 3' de su vecino en una dirección mediante un enlace fosfodiéster. Por lo tanto, un extremo de un oligonucleótido se denomina "extremo 5'" si su fosfato 5' no está unido al oxígeno 3' de un anillo pentoso mononucleótido. Un extremo de un oligonucleótido se denomina "extremo 3'" si su oxígeno 3' no está unido a un fosfato 5' de otro anillo pentoso mononucleótido. Tal y como se utiliza en el presente documento, también puede decirse que una secuencia de ácido nucleico, aunque sea interna a un oligonucleótido mayor, tiene extremos 5' y 3'. En una molécula de ADN lineal o circular, los elementos discretos se denominan "corriente arriba" o 5' de los elementos "corriente abajo" o 3'. Esta terminología refleja que la transcripción avanza en dirección 5' a 3' a lo largo de la cadena de ADN. Los elementos promotores y potenciadores que dirigen la transcripción de un gen ligado suelen estar situados 5' o corriente arriba de la región codificante. Sin embargo, los elementos potenciadores pueden ejercer su efecto incluso cuando están situados a 3' del elemento promotor y de la región codificante. Las señales de terminación de la transcripción y de poliadenilación se sitúan a 3' o corriente abajo de la región codificante.

El término "molécula de ADN recombinante", tal como se utiliza en el presente documento, se refiere a una molécula de ADN compuesta por segmentos de ADN unidos mediante técnicas de biología molecular.

El término "proteína recombinante" o "polipéptido recombinante", tal como se utiliza en el presente documento, se refiere a una molécula de proteína que se expresa utilizando una molécula de ADN recombinante.

- 5 Tal como se utilizan en el presente documento, los términos "vector" y "vehículo" se usan indistintamente en referencia a moléculas de ácido nucleico que transfieren segmento(s) de ADN de una célula a otra.

10 Los términos "en combinación operable", "en orden operable" y "enlazados operablemente", tal como se utilizan en el presente documento, se refieren al enlace de secuencias de ácido nucleico de tal manera que se produzca una molécula de ácido nucleico capaz de dirigir la transcripción de un gen determinado y/o la síntesis de una molécula de proteína deseada. Los términos también se refieren al enlace de secuencias de aminoácidos de tal manera que se produzca una proteína funcional.

15 El término "transfección", tal como se utiliza en el presente documento, se refiere a la introducción de ADN extraño en las células. La transfección puede llevarse a cabo por diversos medios conocidos en la técnica, como la coprecipitación de fosfato cálcico-ADN, la transfección mediada por DEAE-dextrano, la transfección mediada por polibreno, la electroporación, la microinyección, la fusión de liposomas, la lipofectina, la fusión de protoplastos, la infección retroviral, la biolística (es decir, el bombardeo de partículas) y similares.

20 Tal como se utilizan en el presente documento, los términos "complementario" o "complementariedad" se utilizan en referencia a "polinucleótidos" y "oligonucleótidos" (que son términos intercambiables que se refieren a una secuencia de nucleótidos) relacionados por las reglas de apareamiento de bases. Por ejemplo, la secuencia "5'-CAGT-3'", es complementaria de la secuencia "5'-ACTG-3'". La complementariedad puede ser "parcial" o "total". La complementariedad "parcial" se da cuando una o más bases de ácido nucleico no coinciden de acuerdo con las reglas de emparejamiento de bases. La complementariedad "total" o "completa" entre ácidos nucleicos es aquella en la que todas y cada una de las bases del ácido nucleico se emparejan con otra base de acuerdo con las reglas de emparejamiento de bases. El grado de complementariedad entre las cadenas de ácido nucleico puede tener efectos significativos sobre la eficacia y la fuerza de la hibridación entre las cadenas de ácido nucleico. Esto puede ser especialmente importante en las reacciones de amplificación, así como en los procedimientos de detección que dependen de la unión entre ácidos nucleicos. Por comodidad, los términos "polinucleótidos" y "oligonucleótidos" incluyen moléculas que incluyen nucleósidos.

35 Los términos "homología" y "homólogo" utilizados en el presente documento en referencia a secuencias de nucleótidos se refieren a un grado de complementariedad con otras secuencias de nucleótidos. Puede haber homología parcial u homología completa (es decir, identidad). Cuando se utiliza en referencia a una secuencia de ácido nucleico bicatenario, tal como un ADNc o un clon genómico, el término "sustancialmente homólogo" se refiere a cualquier secuencia de ácido nucleico (por ejemplo, sonda) que pueda hibridar con una o ambas cadenas de la secuencia de ácido nucleico bicatenario en condiciones de baja rigurosidad tal como las descritas anteriormente. Una secuencia de nucleótidos parcialmente complementaria, es decir, "sustancialmente homóloga", a una secuencia de ácido nucleico es aquella que inhibe, al menos parcialmente, la hibridación de una secuencia completamente complementaria a una secuencia de ácido nucleico objetivo. La inhibición de la hibridación de la secuencia completamente complementaria a la secuencia objetivo puede examinarse utilizando un ensayo de hibridación (transferencia Southern o Northern, hibridación en solución y similares) en condiciones de baja rigurosidad. Una secuencia o sonda sustancialmente homóloga competirá e inhibirá la unión (es decir, la hibridación) de una secuencia completamente homóloga a una secuencia objetivo en condiciones de baja rigurosidad. Esto no quiere decir que las condiciones de baja rigurosidad permitan uniones no específicas; las condiciones de baja rigurosidad exigen que la unión de dos secuencias entre sí sea una interacción específica (es decir, selectiva). La ausencia de unión inespecífica puede comprobarse mediante el uso de una segunda secuencia objetivo que carezca incluso de un grado parcial de complementariedad (por ejemplo, menos de aproximadamente 30% de identidad); en ausencia de unión inespecífica, la sonda no hibridará con el segundo objetivo no complementario.

50 Las condiciones de baja rigurosidad comprenden condiciones equivalentes a la unión o hibridación a 68°C. En una solución consistente en 5×SSPE (43,8 g/l NaCl, 6,9 g/l NaH₂PO₄·H₂O y 1,85 g/l EDTA, pH ajustado a 7,4 con NaOH), 0,1% SDS, 5× reactivo de Denhardt (50× Denhardt contiene por 500 ml: 5 g de Ficoll (Tipo 400, Pharmacia), 5 g de BSA (Fracción V; Sigma)) y 100 µg/ml de ADN de espermatozoide de salmón desnaturalizado, seguido de un lavado en una solución que comprende 2,0×SSPE, 0,1% SDS a temperatura ambiente cuando se emplea una sonda de aproximadamente 100 a aproximadamente 1000 nucleótidos de longitud.

60 Además, las condiciones que promueven la hibridación en condiciones de alta rigurosidad (por ejemplo, el aumento de la temperatura de la hibridación y/o los pasos de lavado, el uso de formamida en la solución de hibridación, etc.) son bien conocidas en la técnica. Las condiciones de alta rigurosidad, cuando se utilizan en referencia a la hibridación de ácidos nucleicos, comprenden condiciones equivalentes a la unión o hibridación a 68°C. en una solución que consta de 5×SSPE, 1% SDS, 5×reactivo de Denhardt y 100 µg/ml de ADN de espermatozoide de salmón desnaturalizado, seguido de lavado en una solución que consta de 0,1×SSPE y 0,1% SDS a 68°C cuando se emplea una sonda de aproximadamente 100 a aproximadamente 1000 nucleótidos de longitud.

Es bien conocido en la técnica que pueden emplearse numerosas condiciones equivalentes para comprender condiciones de baja rigurosidad; factores tales como la longitud y la naturaleza (ADN, ARN, composición de bases) de la sonda y la naturaleza de la objetivo (ADN, ARN, composición de bases, presente en solución o inmovilizada, etc.) y la concentración de las sales y otros componentes (por ejemplo, la presencia o ausencia de formamida, sulfato de dextrano, polietilenglicol), así como los componentes de la solución de hibridación, pueden variarse para generar condiciones de hibridación de baja rigurosidad diferentes de, pero equivalentes a, las condiciones enumeradas anteriormente.

El término "equivalente" cuando se hace referencia a una condición de hibridación en relación con una condición de hibridación de interés significa que la condición de hibridación y la condición de hibridación de interés resultan en la hibridación de secuencias de ácido nucleico que tienen el mismo intervalo de homología porcentual (%). Por ejemplo, si una condición de hibridación de interés da lugar a la hibridación de una primera secuencia de ácido nucleico con otras secuencias de ácido nucleico que tienen de 50% a 70% de homología con la primera secuencia de ácido nucleico, entonces se dice que otra condición de hibridación es equivalente a la condición de hibridación de interés si esta otra condición de hibridación también da lugar a la hibridación de la primera secuencia de ácido nucleico con las otras secuencias de ácido nucleico que tienen de 50% a 70% de homología con la primera secuencia de ácido nucleico.

Tal como se utiliza en el presente documento, el término "hibridación" se usa en referencia al emparejamiento de ácidos nucleicos complementarios usando cualquier procedimiento por el cual una cadena de ácido nucleico se une con una cadena complementaria a través del emparejamiento de bases para formar un complejo de hibridación. La hibridación y la fuerza de la hibridación (es decir, la fuerza de la asociación entre los ácidos nucleicos) se ven afectadas por factores tales como el grado de complementariedad entre los ácidos nucleicos, la rigurosidad de las condiciones implicadas, la T_m del híbrido formado y la relación G:C dentro de los ácidos nucleicos.

Tal como se utiliza en el presente documento, el término "complejo de hibridación" se refiere a un complejo formado entre dos secuencias de ácido nucleico en virtud de la formación de enlaces de hidrógeno entre bases complementarias G y C y entre bases complementarias A y T; estos enlaces de hidrógeno pueden estabilizarse aún más mediante interacciones de apilamiento de bases. Las dos secuencias complementarias de ácido nucleico se unen por hidrógeno en una configuración antiparalela. Puede formarse un complejo de hibridación en solución (por ejemplo, análisis de Cot o Rot) o entre una secuencia de ácido nucleico presente en solución y otra secuencia de ácido nucleico inmovilizada en un soporte sólido (por ejemplo, una membrana de nailon o un filtro de nitrocelulosa como los empleados en transferencia Southern y Northern, transferencia dot o un portaobjetos de vidrio como los empleados en hibridación *in situ*, incluyendo FISH (hibridación *in situ* fluorescente)).

Tal como se utiliza en el presente documento, el término " T_m " se usa en referencia a la "temperatura de fusión" La temperatura de fusión es la temperatura a la que una población de moléculas de ácido nucleico de doble cadena se disocia a la mitad en cadenas simples. La ecuación para calcular la T_m de los ácidos nucleicos es bien conocida en la técnica. Como indican las referencias estándar, una estimación simple del valor de T_m puede calcularse mediante la ecuación: $T_m = 81,5 + 0,41(\% \text{ G+C})$, cuando un ácido nucleico está en solución acuosa a 1 M NaCl (véase, por ejemplo, Anderson y Young, Quantitative Filter Hybridization, en Nucleic Acid Hybridization, 1985). Otras referencias incluyen cálculos más sofisticados que tienen en cuenta las características estructurales y de secuencia para el cálculo de la T_m .

Tal como se utiliza en el presente documento, el término "rigurosidad" se emplea en referencia a las condiciones de temperatura, fuerza iónica y presencia de otros compuestos, tales como disolventes orgánicos, en las que se llevan a cabo las hibridaciones de ácidos nucleicos. "Estrictez" se produce típicamente en un intervalo de aproximadamente $T_m - 5^\circ\text{C}$ (5°C por debajo de la temperatura de fusión de la sonda) a aproximadamente 20°C a 25°C por debajo de T_m . Como comprenderán los expertos en la materia, una hibridación estricta puede utilizarse para identificar o detectar secuencias polinucleotídicas idénticas o para identificar o detectar secuencias polinucleotídicas similares o relacionadas.

Los términos "unión específica", "especificidad de unión" y sus equivalentes gramaticales, cuando se hacen en referencia a la unión de una primera secuencia de nucleótidos a una segunda secuencia de nucleótidos, se refieren a la interacción preferente entre la primera secuencia de nucleótidos con la segunda secuencia de nucleótidos en comparación con la interacción entre la segunda secuencia de nucleótidos con una tercera secuencia de nucleótidos. La unión específica es un término relativo que no requiere una especificidad absoluta de la unión; en otras palabras, el término "unión específica" no requiere que la segunda secuencia de nucleótidos interactúe con la primera secuencia de nucleótidos en ausencia de una interacción entre la segunda secuencia de nucleótidos y la tercera secuencia de nucleótidos. Más bien, basta con que el nivel de interacción entre la primera secuencia de nucleótidos y la segunda secuencia de nucleótidos sea mayor que el nivel de interacción entre la segunda secuencia de nucleótidos con la tercera secuencia de nucleótidos. La "unión específica" de una primera secuencia de nucleótidos con una segunda secuencia de nucleótidos también significa que la interacción entre la primera secuencia de nucleótidos y la segunda secuencia de nucleótidos depende de la presencia de una estructura particular sobre o dentro de la primera secuencia de nucleótidos; en otras palabras, la segunda secuencia de nucleótidos está reconociendo y uniéndose a una estructura específica sobre o dentro de la primera secuencia de nucleótidos en lugar de a ácidos nucleicos o a

secuencias de nucleótidos en general. Por ejemplo, si una segunda secuencia de nucleótidos es específica para la estructura "A" que se encuentra sobre o dentro de una primera secuencia de nucleótidos, la presencia de una tercera secuencia de ácido nucleico que contenga la estructura A reducirá la cantidad de la segunda secuencia de nucleótidos que se une a la primera secuencia de nucleótidos.

Tal como se utiliza en el presente documento, el término "ácido nucleico amplificable" se utiliza en referencia a los ácidos nucleicos que pueden amplificarse mediante cualquier procedimiento de amplificación. Se contempla que el "ácido nucleico amplificable" comprenderá normalmente la "plantilla de muestra"

Los términos "secuencia de ácido nucleico heteróloga" o "ADN heterólogo" se utilizan indistintamente para referirse a una secuencia de nucleótidos que está ligada a una secuencia de ácido nucleico a la que no está ligada en la naturaleza, o a la que está ligada en un lugar diferente en la naturaleza. El ADN heterólogo no es endógeno a la célula en la que se introduce, sino que se ha obtenido de otra célula. Generalmente, aunque no necesariamente, ese ADN heterólogo codifica ARN y proteínas que normalmente no produce la célula en la que se expresa. Entre los ejemplos de ADN heterólogo se incluyen genes informadores, secuencias reguladoras de la transcripción y la traducción, proteínas marcadoras seleccionables (por ejemplo, proteínas que confieren resistencia a los fármacos), etc.

La "amplificación" se define como la producción de copias adicionales de una secuencia de ácido nucleico y generalmente se lleva a cabo utilizando tecnologías de reacción en cadena de la polimerasa bien conocidas en la técnica (Dieffenbach C W y G S Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y.). Tal como se utiliza en el presente documento, el término "reacción en cadena de la polimerasa" ("PCR") se refiere al procedimiento de K. B. Mullis Pat. de EE. UU. Nos. 4,683,195 y 4,683,202 que describen un procedimiento para aumentar la concentración de un segmento de una secuencia objetivo en una mezcla de ADN genómico sin clonación ni purificación. La longitud del segmento amplificado de la secuencia objetivo deseada viene determinada por las posiciones relativas de dos cebadores oligonucleotídicos entre sí y, por lo tanto, esta longitud es un parámetro controlable. En virtud del aspecto repetitivo del procedimiento, el procedimiento se denomina "reacción en cadena de la polimerasa" ("PCR"). Dado que los segmentos amplificados deseados de la secuencia objetivo se convierten en las secuencias predominantes (en términos de concentración) en la mezcla, se dice que están "amplificados por PCR."

Con la PCR, es posible amplificar una sola copia de una secuencia objetivo específica en el ADN genómico a un nivel detectable mediante varias metodologías diferentes (por ejemplo, hibridación con una sonda marcada; incorporación de cebadores biotinilados seguida de detección con conjugado enzimático de avidina; incorporación de desoxinucleótidos trifosfatos marcados con ³²P, tal como dCTP o dATP, en el segmento amplificado). Además del ADN genómico, cualquier secuencia de oligonucleótidos puede amplificarse con el conjunto apropiado de moléculas cebadoras. En particular, los segmentos amplificados creados por el propio procedimiento de PCR son, en sí mismos, plantillas eficientes para posteriores amplificaciones por PCR.

Uno de estos procedimientos preferidos, en particular para aplicaciones comerciales, se basa en la tecnología de PCR en tiempo real TaqMan® ampliamente utilizada, y combina la PCR alelo-específica con un reactivo de bloqueo (ASB-PCR) para suprimir la amplificación del alelo de tipo silvestre. La ASB-PCR puede utilizarse para la detección de mutaciones somáticas o de la línea germinal en ADN o ARN extraído de cualquier tipo de tejido, incluyendo muestras tumorales fijadas en formol e incluidas en parafina. Se desarrolla un conjunto de reglas de diseño de reactivos que permiten la detección sensible y selectiva de sustituciones, inserciones o deleciones puntuales en un fondo de alelos de tipo silvestre en exceso de mil veces o más. (Morlan J, Baker J, Sinicropi D Mutation Detection by Real-Time PCR: A Simple, Robust and Highly Selective Method. PLoS ONE 4(2): e4584, 2009)

Los términos "reacción en cadena de la polimerasa de transcripción inversa" y "RT-PCR" se refieren a un procedimiento de transcripción inversa de una secuencia de ARN para generar una mezcla de secuencias de ADNc, seguido del aumento de la concentración de un segmento deseado de las secuencias de ADNc transcritas en la mezcla sin clonación ni purificación. Típicamente, el ARN se transcribe inversamente utilizando un único cebador (por ejemplo, un cebador oligo-dT) antes de la amplificación por PCR del segmento deseado del ADN transcrito utilizando dos cebadores.

Tal como se utiliza en el presente documento, el término "cebador" se refiere a un oligonucleótido, ya sea que se encuentre naturalmente como en un digerido de restricción purificado o producido sintéticamente, que es capaz de actuar como un punto de iniciación de la síntesis cuando se coloca en condiciones en las que se induce la síntesis de un producto de extensión de cebador que es complementario a una cadena de ácido nucleico, (es decir, en presencia de nucleótidos y de un agente inductor tal como la ADN polimerasa y a una temperatura y pH adecuados). En algunos ejemplos, el cebador es monocatenario para una máxima eficiencia en la amplificación, pero alternativamente puede ser bicatenario. Si es bicatenario, el cebador se trata primero para separar sus cadenas antes de utilizarlo para preparar los productos de extensión. En algunos ejemplos, el cebador es un oligodesoxirribonucleótido. El cebador debe ser suficientemente largo para cebar la síntesis de los productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, tales como la temperatura, la fuente del cebador y el uso del procedimiento.

Tal como se utiliza en el presente documento, el término "sonda" se refiere a un oligonucleótido (es decir, una secuencia de nucleótidos), ya sea que ocurra naturalmente como en un digerido de restricción purificado o producido sintéticamente, recombinantemente o por amplificación PCR, que es capaz de hibridarse a otro oligonucleótido de interés. Una sonda puede ser monocatenaria o bicatenaria. Las sondas son útiles en la detección, identificación y aislamiento de secuencias genéticas concretas. Se contempla que cualquier sonda utilizada en la presente divulgación estará marcada con cualquier "molécula informadora", de modo que sea detectable en cualquier sistema de detección, incluyendo, pero sin limitarse a, sistemas enzimáticos (por ejemplo, ELISA, así como ensayos histoquímicos basados en enzimas), fluorescentes, radiactivos y luminiscentes. No se pretende que la presente divulgación se limite a ningún sistema de detección o etiqueta en particular.

Tal como se utilizan en el presente documento, los términos "endonucleasas de restricción" y "enzimas de restricción" se refieren a enzimas bacterianas, cada una de las cuales corta o mella ADN de cadena doble o simple en o cerca de una secuencia de nucleótidos específica, por ejemplo, puede utilizarse un dominio de endonucleasa de una endonucleasa de restricción de tipo IIS (por ejemplo, FokI, tal como enseñan Kim et al., 1996, Proc. Nat'l. Acad. Sci. EE. UU., 6:1 156-60).

Tal como se utiliza en el presente documento, el término "un oligonucleótido que tiene una secuencia de nucleótidos que codifica un gen" significa una secuencia de ácido nucleico que comprende la región codificante de un gen, es decir, la secuencia de ácido nucleico que codifica un producto génico. La región codificante puede estar presente en forma de ADNc, ADN genómico o ARN. Cuando se presenta en forma de ADN, el oligonucleótido puede ser monocatenario (es decir, la cadena sensitiva) o bicatenario. Además, "un oligonucleótido que tiene una secuencia de nucleótidos que codifica un gen" puede incluir elementos de control adecuados, tales como potenciadores, promotores, uniones de empalme, señales de poliadenilación, etc., si son necesarios para permitir el inicio adecuado de la transcripción y/o el procesamiento correcto del transcrito primario de ARN. Además, la región codificante de la presente divulgación puede contener potenciadores endógenos, uniones de empalme, secuencias intermedias, señales de poliadenilación, etc.

Las señales de control transcripcional en eucariotas comprenden elementos "potenciadores". Los potenciadores consisten en conjuntos cortos de secuencias de ADN que interactúan específicamente con proteínas celulares implicadas en la transcripción (Maniatis, T. et al., Science 236:1237, 1987). Se han aislado elementos potenciadores de diversas fuentes eucariotas, incluyendo genes de plantas, levaduras, insectos, células de mamíferos y virus. La selección de un potenciador concreto depende del tipo de célula que se vaya a utilizar para expresar la proteína de interés.

La presencia de "señales de empalme" en un vector de expresión a menudo resulta en niveles más altos de expresión del transcrito recombinante. Las señales de empalme median en la eliminación de intrones del transcrito primario de ARN y consisten en un donante de empalme y un sitio aceptor (Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, 2ª ed., Cold Spring Harbor Laboratory Press, Nueva York, pp. 16.7-16.8, 1989). Un sitio donante y aceptor de empalme comúnmente utilizado es la unión de empalme del ARN 16S del SV40.

La expresión eficiente de secuencias de ADN recombinante en células eucariotas requiere la expresión de señales que dirijan la terminación y poliadenilación eficientes de la transcripción resultante. Las señales de terminación de la transcripción se encuentran generalmente corriente abajo de la señal de poliadenilación y tienen unos cientos de nucleótidos de longitud. El término "sitio poli A" o "secuencia poli A", tal como se utiliza en el presente documento, denota una secuencia de ADN que dirige tanto la terminación como la poliadenilación del transcrito de ARN naciente. Es deseable una poliadenilación eficiente del transcrito recombinante, ya que los transcritos que carecen de cola poli A son inestables y se degradan rápidamente. La señal poli A utilizada en un vector de expresión puede ser "heteróloga" o "endógena". Una señal de poli A endógena es la que se encuentra de forma natural en el extremo 3' de la región codificante de un gen determinado en el genoma. Una señal poli A heteróloga es la que se aísla de un gen y se coloca a 3' de otro gen.

El término "promotor", "elemento promotor" o "secuencia promotora", tal como se utiliza en el presente documento, se refiere a una secuencia de ADN que cuando se coloca en el extremo 5' de (es decir, precede a) una secuencia de oligonucleótidos es capaz de controlar la transcripción de la secuencia de oligonucleótidos en ARNm. Un promotor suele estar situado a 5' (es decir, corriente arriba) de una secuencia de oligonucleótidos cuya transcripción en ARNm controla, y proporciona un sitio para la unión específica de la ARN polimerasa y para el inicio de la transcripción.

El término "actividad promotora" cuando se hace referencia a una secuencia de ácido nucleico se refiere a la capacidad de la secuencia de ácido nucleico para iniciar la transcripción de una secuencia de oligonucleótidos en ARNm.

El término "específico de tejido" aplicado a un promotor se refiere a un promotor que es capaz de dirigir la expresión selectiva de una secuencia de oligonucleótidos a un tipo específico de tejido en ausencia relativa de expresión del mismo oligonucleótido en un tipo diferente de tejido. La especificidad tisular de un promotor puede evaluarse, por ejemplo, vinculando operablemente un gen informador a la secuencia promotora para generar un constructo informador, introduciendo el constructo informador en el genoma de una planta o un animal de modo que el constructo

informador se integre en cada tejido del animal transgénico resultante, y detectando la expresión del gen informador (por ejemplo, detectando ARNm, proteína o la actividad de una proteína codificada por el gen informador) en diferentes tejidos de la planta o el animal transgénicos. La selectividad no tiene por qué ser absoluta. La detección de un mayor nivel de expresión del gen informador en uno o más tejidos en relación con el nivel de expresión del gen informador en otros tejidos demuestra que el promotor es específico de los tejidos en los que se detectan mayores niveles de expresión.

El término "específico de tipo celular" aplicado a un promotor se refiere a un promotor que es capaz de dirigir la expresión selectiva de una secuencia de oligonucleótidos en un tipo específico de célula en ausencia relativa de expresión de la misma secuencia de oligonucleótidos en un tipo diferente de célula dentro del mismo tejido. El término "específico del tipo celular" cuando se aplica a un promotor también significa un promotor capaz de promover la expresión selectiva de un oligonucleótido en una región dentro de un único tejido. Una vez más, la selectividad no tiene por qué ser absoluta. La especificidad de tipo celular de un promotor puede evaluarse utilizando procedimientos bien conocidos en la técnica, por ejemplo, tinción inmunohistoquímica como se describe en el presente documento. Brevemente, las secciones de tejido se incrustan en parafina, y las secciones de parafina se hacen reaccionar con un anticuerpo primario que es específico para el producto polipeptídico codificado por la secuencia de oligonucleótidos cuya expresión está controlada por el promotor. Como alternativa a la sección en parafina, las muestras pueden crioseccionarse. Por ejemplo, las secciones pueden congelarse antes y durante el seccionamiento, evitando así las posibles interferencias de la parafina residual. Se deja que un anticuerpo secundario marcado (por ejemplo, conjugado con peroxidasa) que es específico para el anticuerpo primario se una al tejido seccionado y se detecta la unión específica (por ejemplo, con avidina/biotina) mediante microscopía.

Los términos "expresión selectiva", "expresar selectivamente" y sus equivalentes gramaticales se refieren a una comparación de niveles relativos de expresión en dos o más regiones de interés. Por ejemplo, "expresión selectiva" cuando se utiliza en relación con los tejidos se refiere a un nivel sustancialmente mayor de expresión de un gen de interés en un tejido particular, o a un número sustancialmente mayor de células que expresan el gen dentro de ese tejido, en comparación, respectivamente, con el nivel de expresión de, y el número de células que expresan, el mismo gen en otro tejido (es decir, la selectividad no tiene por qué ser absoluta). La expresión selectiva no requiere, aunque puede incluir, la expresión de un gen de interés en un tejido concreto y la ausencia total de expresión del mismo gen en otro tejido. Del mismo modo, la "expresión selectiva", tal y como se utiliza en el presente documento en referencia a los tipos celulares, se refiere a un nivel sustancialmente mayor de expresión de, o un número sustancialmente mayor de células que expresan, un gen de interés en un tipo celular particular, cuando se compara, respectivamente, con los niveles de expresión del gen y con el número de células que expresan el gen en otro tipo celular.

El término "contiguo" cuando se utiliza en referencia a dos o más secuencias de nucleótidos significa que las secuencias de nucleótidos se ligan en tándem, ya sea en ausencia de secuencias intermedias, o en presencia de secuencias intermedias que no comprenden uno o más elementos de control.

Tal como se utilizan en el presente documento, los términos "codificación de moléculas de ácido nucleico", "codificación de nucleótidos", "codificación de secuencias de ADN" y "codificación de ADN" se refieren al orden o secuencia de desoxirribonucleótidos a lo largo de una cadena de ácido desoxirribonucleico. El orden de estos desoxirribonucleótidos determina el orden de los aminoácidos a lo largo de la cadena polipeptídica (proteína). Así pues, la secuencia de ADN codifica la secuencia de aminoácidos.

El término "aislado" cuando se utiliza en relación con un ácido nucleico, como en "un oligonucleótido aislado" se refiere a una secuencia de ácido nucleico que está separada de al menos un ácido nucleico contaminante con el que está normalmente asociado en su fuente natural. El ácido nucleico aislado es el ácido nucleico presente en una forma o entorno diferente de aquel en donde se encuentra en la naturaleza. Por el contrario, los ácidos nucleicos no aislados son ácidos nucleicos tales como el ADN y el ARN que se encuentran en el estado en que existen en la naturaleza. Por ejemplo, una determinada secuencia de ADN (por ejemplo, un gen) se encuentra en el cromosoma de la célula huésped en proximidad a genes vecinos; las secuencias de ARN, tal como una secuencia específica de ARNm que codifica una proteína específica, se encuentran en la célula mezcladas con otros numerosos ARNm que codifican multitud de proteínas. Sin embargo, el ácido nucleico aislado que codifica un polipéptido de interés incluye, a modo de ejemplo, dicho ácido nucleico en células que expresan ordinariamente el polipéptido de interés cuando el ácido nucleico se encuentra en una localización cromosómica o extracromosómica diferente de la de las células naturales, o está flanqueado de otro modo por una secuencia de ácido nucleico diferente de la que se encuentra en la naturaleza. El ácido nucleico u oligonucleótido aislado puede estar presente en forma monocatenaria o bicatenaria. El ácido nucleico aislado puede identificarse fácilmente (si se desea) mediante diversas técnicas (por ejemplo, hibridación, transferencia de, etc.). Cuando se va a utilizar un ácido nucleico aislado u oligonucleótido para expresar una proteína, el oligonucleótido contendrá como mínimo la cadena sentido o codificante (es decir, el oligonucleótido puede ser monocatenario). Alternativamente, puede contener las cadenas sentido y antisentido (es decir, el oligonucleótido puede ser bicatenario).

Tal como se utiliza en el presente documento, el término "purificado" o "purificar" se refiere a la eliminación de uno o más componentes (no deseados) de una muestra. Por ejemplo, cuando los polipéptidos recombinantes se expresan

en células huésped bacterianas, los polipéptidos se purifican mediante la eliminación de las proteínas de la célula huésped, aumentando así el porcentaje de polipéptidos recombinantes en la muestra.

Tal como se utiliza en el presente documento, el término "sustancialmente purificado" se refiere a moléculas, ya sean secuencias nucleicas o de aminoácidos, que se retiran de su entorno natural, se aíslan o separan, y están libres al menos en 60%, por ejemplo, en 75%, tal como en 90%, de otros componentes con los que están asociadas de forma natural. Un "polinucleótido aislado" es, por tanto, un polinucleótido sustancialmente purificado.

Tal como se utiliza en el presente documento, el término "región codificante" cuando se utiliza en referencia a un gen estructural se refiere a las secuencias de nucleótidos que codifican los aminoácidos que se encuentran en el polipéptido naciente como resultado de la traducción de una molécula de ARNm. En los eucariotas, la región codificante está delimitada en el lado 5' generalmente por el triplete de nucleótidos "ATG", que codifica la metionina iniciadora, y en el lado 3' por uno de los tres tripletes que especifican los codones de parada (es decir, TAA, TAG, TGA).

Por "secuencia codificante" se entiende una secuencia de un ácido nucleico o su complemento, o una parte del mismo, que puede transcribirse y/o traducirse para producir el ARNm para y/o el polipéptido o un fragmento del mismo. Las secuencias codificantes incluyen exones en un ADN genómico o transcritos primarios inmaduros de ARN, que se unen mediante la maquinaria bioquímica de la célula para dar lugar a un ARNm maduro. La cadena antisentido es el complemento de dicho ácido nucleico, y la secuencia codificante puede deducirse de ella.

Por "secuencia no codificante" se entiende una secuencia de un ácido nucleico o su complemento, o una parte del mismo que no se transcribe en aminoácido *in vivo*, o donde el ARNt no interactúa para colocar o intentar colocar un aminoácido. Las secuencias no codificantes incluyen tanto secuencias intrónicas en el ADN genómico o en transcritos primarios inmaduros de ARN, como secuencias asociadas a genes, tales como promotores, potenciadores, silenciadores, etc.

Tal como se utiliza en el presente documento, el término "gen estructural" o "secuencia de nucleótidos estructural" se refiere a una secuencia de ADN que codifica un ARN o una proteína que no controla la expresión de otros genes. Por el contrario, un "gen regulador" o "secuencia reguladora" es un gen estructural que codifica productos (por ejemplo, factores de transcripción) que controlan la expresión de otros genes.

Tal como se utiliza en el presente documento, el término "elemento regulador" se refiere a un elemento genético que controla algún aspecto de la expresión de secuencias de ácido nucleico. Por ejemplo, un promotor es un elemento regulador que facilita el inicio de la transcripción de una región codificante operablemente enlazada. Otros elementos reguladores incluyen señales de empalme, señales de poliadenilación, señales de terminación, etc.

Tal como se utiliza en el presente documento, el término "sitio de unión a factores de transcripción peptídicos" o "sitio de unión a factores de transcripción" se refiere a una secuencia de nucleótidos que se une a factores de transcripción proteicos y, por lo tanto, controla algún aspecto de la expresión de secuencias de ácido nucleico. Por ejemplo, los sitios de unión de Sp-1 y AP1 (proteína activadora 1) son ejemplos de sitios de unión de factores de transcripción peptídicos.

Tal como se utiliza en el presente documento, el término "gen" significa las secuencias de desoxirribonucleótidos que comprenden la región codificante de un gen estructural. Un "gen" también puede incluir secuencias no traducidas situadas junto a la región codificante en los extremos 5' y 3', de forma que el gen corresponda a la longitud del ARNm completo. Las secuencias situadas a 5' de la región codificante y presentes en el ARNm se denominan secuencias no traducidas 5'. Las secuencias situadas a 3' o corriente abajo de la región codificante y presentes en el ARNm se denominan secuencias no traducidas 3'. El término "gen" abarca tanto el ADNc como las formas genómicas de un gen. Una forma genómica o clon de un gen contiene la región codificante interrumpida con secuencias no codificantes denominadas "intrones" o "regiones intermedias" o "secuencias intermedias". Los intrones son segmentos de un gen que se transcriben en ARN nuclear heterogéneo (ARNnh); los intrones pueden contener elementos reguladores tales como potenciadores. Los intrones se eliminan o "empalman" del transcrito nuclear o primario; por lo tanto, los intrones están ausentes en el transcrito del ARN mensajero (ARNm). Durante la traducción, el ARNm especifica la secuencia o el orden de los aminoácidos en un polipéptido naciente. Un gen es generalmente un único locus. En un organismo diploide normal, un gen tiene dos alelos. En la patata tetraploide, sin embargo, cada gen tiene 4 alelos. En la caña de azúcar, que es dodecaploide, puede haber 12 alelos por gen. Algunos ejemplos concretos son el lino, que tiene dos loci EPSPS con dos alelos cada uno, y el arroz, que tiene una única ACCasa plastidial homomérica con dos alelos.

Además de contener intrones, las formas genómicas de un gen también pueden incluir secuencias situadas en los extremos 5' y 3' de las secuencias que están presentes en el transcrito de ARN. Estas secuencias se denominan secuencias o regiones "flanqueantes" (estas secuencias flanqueantes se sitúan a 5' o 3' de las secuencias no traducidas presentes en el transcrito del ARNm). La región flanqueante 5' puede contener secuencias reguladoras tales como promotores y potenciadores que controlan o influyen en la transcripción del gen. La región flanqueante 3' puede contener secuencias que dirigen la terminación de la transcripción, la escisión postranscripcional y la poliadenilación.

Por "animal no humano" se entiende cualquier animal que no sea un ser humano e incluye vertebrados tales como roedores, primates no humanos, ovinos, bovinos, rumiantes, lagomorfos, porcinos, caprinos, equinos, caninos, felinos, aves, etc. Los animales no humanos preferidos se seleccionan del orden Rodentia. "Animal no humano" se refiere además a anfibios (por ejemplo, *Xenopus*), reptiles, insectos (por ejemplo, *Drosophila*) y otras especies animales no mamíferas.

Tal como se utiliza en el presente documento, el término "transgénico" se refiere a un organismo o célula que tiene ADN derivado de otro organismo insertado en donde se integra en el genoma, ya sea de células somáticas y/o de la línea germinal de la planta o animal. Por "transgén" se entiende una secuencia de ADN que es parcial o totalmente heteróloga (es decir, no presente en la naturaleza) para la planta o el animal en donde se encuentra, o que es homóloga a una secuencia endógena (es decir, una secuencia que se encuentra en el animal en la naturaleza) y se inserta en el genoma de la planta o el animal en un lugar que difiere del de la secuencia presente en la naturaleza. Las plantas transgénicas que incluyen uno o más transgenes están dentro del alcance de esta divulgación. Además, el término "transgénico" utilizado en el presente documento hace referencia a un organismo al que se le han modificado y/o "eliminado" uno o más genes (haciéndolos no funcionales o haciéndolos funcionar a un nivel reducido, es decir, una mutación "inactivada") mediante los procedimientos de la divulgación, por recombinación homóloga, mutación TFO o por procedimientos similares. Por ejemplo, un organismo o célula transgénica incluye ADN insertado que incluye un promotor y/o una región codificante extrañas.

Una "célula transformada" es una célula o línea celular que ha adquirido la capacidad de crecer en cultivo celular durante múltiples generaciones, la capacidad de crecer en agar blando y/o la capacidad de que el crecimiento celular no se vea inhibido por el contacto célula-célula. En este sentido, la transformación se refiere a la introducción de material genético extraño en una célula u organismo. La transformación puede llevarse a cabo mediante cualquier procedimiento conocido que permita la introducción con éxito de ácidos nucleicos en células y que dé lugar a la expresión del ácido nucleico introducido. "Transformación" incluye, entre otros, procedimientos como la transfección, la microinyección, la electroporación, la nucleofección y la lipofección (transferencia de genes mediada por liposomas). La transformación puede realizarse mediante el uso de cualquier vector de expresión. Por ejemplo, se contempla el uso de baculovirus para introducir ácido nucleico extraño en células de insecto. El término "transformación" también incluye procedimientos como la transformación de la línea germinal de insectos enteros mediada por elementos P. Además, la transformación se refiere a las células que se han transformado de forma natural, normalmente por mutación genética.

Tal como se utiliza en el presente documento, "exógeno" significa que el gen que codifica la proteína no se expresa normalmente en la célula. Además, "exógeno" se refiere a un gen transfectado en una célula para aumentar el nivel normal (es decir, natural) de expresión de ese gen.

Una secuencia peptídica y una secuencia nucleotídica pueden ser "endógenas" o "heterólogas" (es decir, "extrañas"). El término "endógena" se refiere a una secuencia que se encuentra de forma natural en la célula en la que se introduce, siempre que no contenga alguna modificación relativa a la secuencia natural. El término "heterólogo" se refiere a una secuencia que no es endógena a la célula en la que se introduce. Por ejemplo, el ADN heterólogo incluye una secuencia de nucleótidos que se liga a, o se manipula para ligarse a, una secuencia de ácido nucleico a la que no está ligada en la naturaleza, o a la que está ligada en un lugar diferente en la naturaleza. El ADN heterólogo también incluye una secuencia de nucleótidos que se encuentra de forma natural en la célula en la que se introduce y que contiene alguna modificación relativa a la secuencia natural. Generalmente, aunque no necesariamente, el ADN heterólogo codifica ARN heterólogo y proteínas heterólogas que normalmente no son producidas por la célula en la que se introduce. Entre los ejemplos de ADN heterólogo se incluyen genes informadores, secuencias reguladoras transcripcionales y traslacionales, secuencias de ADN que codifican proteínas marcadoras seleccionables (por ejemplo, proteínas que confieren resistencia a fármacos), etc.

Constructo

Las moléculas de ácido nucleico divulgadas en el presente documento (por ejemplo, nucleasas específicas de sitio, o ARN guía para CRISPRs) pueden utilizarse en la producción de constructos de ácido nucleico recombinante. En una realización, las moléculas de ácido nucleico de la presente divulgación pueden utilizarse en la preparación de constructos de ácido nucleico, por ejemplo, casetes de expresión para la expresión en la planta, microorganismo o animal de interés. Esta expresión puede ser transitoria, por ejemplo, cuando el constructo no está integrado en el genoma del huésped, o mantenerse bajo el control ofrecido por el promotor y la posición del constructo dentro del genoma del huésped si llega a integrarse.

Los casetes de expresión pueden incluir secuencias reguladoras enlazadas operablemente a la nucleasa específica del sitio o a las secuencias de ARN guía divulgadas en el presente documento. El casete puede contener además al menos un gen adicional para ser cotransformado en el organismo. Alternativamente, el gen o genes adicionales pueden proporcionarse en múltiples casetes de expresión.

Los constructos de ácido nucleico pueden estar provistas de una pluralidad de sitios de restricción para la inserción

de la secuencia codificante de la nucleasa específica del sitio que estará bajo la regulación transcripcional de las regiones reguladoras. Los constructos de ácido nucleico pueden contener además moléculas de ácido nucleico que codifican genes marcadores seleccionables.

Puede utilizarse cualquier promotor en la producción de los constructos de ácido nucleico. El promotor puede ser nativo o análogo, o extraño o heterólogo, a las secuencias de ácido nucleico del huésped vegetal, microbiano o animal divulgadas en el presente documento. Además, el promotor puede ser la secuencia natural o alternativamente una secuencia sintética. Cuando el promotor es "extraño" o "heterólogo" para la planta, microbio o animal huésped, se entiende que el promotor no se encuentra en la planta, microbio o animal nativo en donde se introduce el promotor. Tal como se utiliza en el presente documento, un gen quimérico comprende una secuencia codificante unida de forma operable a una región de iniciación de la transcripción que es heteróloga a la secuencia codificante.

Las secuencias de nucleasas de sitio dirigido descritas en el presente documento pueden expresarse utilizando promotores heterólogos.

Puede utilizarse cualquier promotor en la preparación de constructos para controlar la expresión de las secuencias nucleasa dirigidas al sitio, tales como promotores que proporcionan una expresión constitutiva, preferida por el tejido, inducible u otros promotores para la expresión en plantas, microbios o animales. Los promotores constitutivos incluyen, por ejemplo, el promotor del núcleo del promotor Rsyn7 y otros promotores constitutivos divulgados en el documento WO 99/43 838 y Patente de EE. UU. No. 6,072,050 el promotor central del CaMV 35S (Odell et al. Nature 313:810-812; 1985); actina del arroz (McElroy et al., Plant Cell 2:163-171, 1990); ubiquitina (Christensen et al., Plant Mol. Biol. 12:619-632, 1989 y Christensen et al., Plant Mol. Biol. 18:675-689, 1992); pEMU (Last et al., Theor. Appl. Genet. 81:581-588, 1991); MAS (Velten et al., EMBO J. 3:2723-2730, 1984); promotor ALS (Patente de EE. UU. No. 5,659,026), y similares. Otros promotores constitutivos incluyen, por ejemplo, la Patente de EE. UU. Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; y 6,177,611.

Los promotores preferidos por el tejido pueden utilizarse para dirigir la expresión de la nucleasa dirigida al sitio dentro de un tejido vegetal particular. Dichos promotores de tejidos preferentes incluyen, entre otros, promotores de hojas preferentes, promotores de raíces preferentes, promotores de semillas preferentes y promotores de tallos preferentes. Los promotores preferidos por el tejido incluyen Yamamoto et al., Plant J. 12(2):255-265, 1997; Kawamata et al., Plant Cell Physiol. 38(7):792-803, 1997; Hansen et al., Mol. Gen. Genet. 254(3):337-343, 1997; Russell et al., Transgenic Res. 6(2):157-168, 1997; Rinehart et al., Plant Physiol. 112(3):1331-1341, 1996; Van Camp et al., Plant Physiol. 112(2):525-535, 1996; Canevascini et al., Plant Physiol. 112(2): 513-524, 1996; Yamamoto et al., Plant Cell Physiol. 35(5):773-778, 1994; Lam, Results Probl. Cell Differ. 20:181-196, 1994; Orozco et al. Plant Mol Biol. 23(6):1129-1138, 1993; Matsuoka et al., Proc Nat'l. Acad. Sci. USA 90(20):9586- 9590, 1993 y Guevara-Garcia et al., Plant J. 4(3):495-505, 1993.

Los constructos de ácido nucleico también pueden incluir regiones de terminación de la transcripción. Cuando se utilizan regiones de terminación de la transcripción, puede utilizarse cualquier región de terminación en la preparación de los constructos de ácido nucleico. Por ejemplo, la región de terminación puede proceder de otra fuente (es decir, ajena o heteróloga al promotor). Ejemplos de regiones de terminación que están disponibles para su uso en los constructos de la presente divulgación incluyen los del plásmido Ti de *A. tumefaciens*, tal como las regiones de terminación de la octopina sintasa y la nopalina sintasa. Véase también Guérineau et al., Mol. Gen. Genet. 262:141-144, 1991; Proudfoot, Cell 64:671-674, 1991; Sanfacon et al., Genes Dev. 5:141-149, 1991; Mogen et al., Plant Cell 2:1261-1272, 1990; Munroe et al., Gene 91:151-158, 1990; Ballas et al., Nucleic Acids Res. 17:7891-7903, 1989 y Joshi et al., Nucleic Acid Res. 15:9627-9639, 1987.

En conjunción con cualquiera de los procedimientos y/o composiciones divulgados en el presente documento, los ácidos nucleicos pueden optimizarse para aumentar la expresión en la planta transformada. Es decir, los ácidos nucleicos que codifican las proteínas nucleasas dirigidas al sitio pueden sintetizarse utilizando codones preferidos por las plantas para mejorar la expresión. Véase, por ejemplo, Campbell y Gowri, (Plant Physiol. 92:1-11, 1990) para una discusión sobre el uso de codones preferido por el huésped. Existen procedimientos para sintetizar genes preferidos por las plantas. Véase, por ejemplo, Patentes de EE. UU. Nos. 5,380,831 y 5,436,391 y Murray et al., Nucleic Acids Res. 17:477-498, 1989. Véase también, por ejemplo, Lanza et al., BMC Systems Biology 8:33-43, 2014; Burgess-Brown et al., Protein Expr. Purif. 59:94-102, 2008; Gustafsson et al., Trends Biotechnol 22:346-353, 2004.

Además, se pueden realizar otras modificaciones de secuencia a las secuencias de ácido nucleico divulgadas en el presente documento. Por ejemplo, se sabe que las modificaciones de secuencia adicionales mejoran la expresión génica en un huésped celular. Esto incluye la eliminación de secuencias que codifican señales de poliadenilación espurias, señales de sitios de empalme exón/intrón, repeticiones tipo transposón y otras secuencias bien caracterizadas que pueden ser perjudiciales para la expresión génica. El contenido G-C de la secuencia también puede ajustarse a los niveles medios de un huésped celular objetivo, calculados por referencia a genes conocidos expresados en la célula huésped. Además, la secuencia puede modificarse para evitar las estructuras secundarias de ARNm en horquilla previstas.

También pueden utilizarse otras secuencias de ácido nucleico en la preparación de los constructos de la presente

divulgación, por ejemplo, para mejorar la expresión de la secuencia codificante de nucleasa dirigida al sitio. Tales secuencias de ácido nucleico incluyen los intrones del gen *Adhl*, intrón del maíz (Callis et al., Genes y Desarrollo 1:1183-1200, 1987), y secuencias líder (secuencia W) del virus del mosaico del tabaco (TMV), del virus del moteado clorótico del maíz y del virus del mosaico de la alfalfa (Gallie et al., Nucleic Acid Res. 15:8693-8711, 1987 y Skuzeski et al., Plant Mol. Biol. 15:65-79, 1990). Se ha demostrado que el primer intrón del locus reducido-1 del maíz aumenta la expresión de genes en constructos de genes quiméricos. Pat. de EE. UU. Nos. 5.424.412 y 5.593.874 divulgan el uso de intrones específicos en constructos de expresión génica, y Gallie et al. (Plant Physiol. 106:929-939, 1994) también han demostrado que los intrones son útiles para regular la expresión génica en función del tejido. Para mejorar aún más u optimizar la expresión del gen de la nucleasa dirigida al sitio, los vectores de expresión vegetal divulgados en el presente documento también pueden contener secuencias de ADN que contengan regiones de unión a la matriz (MAR). Las células vegetales transformadas con tales sistemas de expresión modificados, entonces, pueden exhibir sobreexpresión o expresión constitutiva de una secuencia nucleotídica de la divulgación.

Los constructos de expresión divulgados en el presente documento también pueden incluir secuencias de ácido nucleico capaces de dirigir la expresión de la secuencia de nucleasa dirigida al cloroplasto u otros orgánulos y estructuras tanto en procariotas como en eucariotas. Tales secuencias de ácido nucleico incluyen secuencias dirigidas al cloroplasto que codifican un péptido de tránsito del cloroplasto para dirigir el producto génico de interés a los cloroplastos de las células vegetales. Tales péptidos de tránsito son conocidos en la técnica. Con respecto a las secuencias dirigidas al cloroplasto, "operablemente enlazadas" significa que la secuencia de ácido nucleico que codifica un péptido de tránsito (es decir, la secuencia dirigida al cloroplasto) está enlazada a las moléculas de ácido nucleico de nucleasas dirigidas al sitio divulgadas en el presente documento de tal manera que las dos secuencias son contiguas y están en el mismo marco de lectura. Véase, por ejemplo, Von Heijne et al., Plant Mol. Biol. Rep. 9:104-126, 1991; Clark et al., J. Biol. Chem. 264:17544-17550, 1989; Della-Cioppa et al., Plant Physiol. 84:965-968, 1987; Romer et al., Biochem. Biophys. Res. Commun. 196:1414-1421, 1993 y Shah et al., Science 233:478-481, 1986.

Las secuencias dirigidas al cloroplasto son conocidas en la técnica e incluyen la subunidad pequeña del cloroplasto de la ribulosa-1,5-bisfosfato carboxilasa (Rubisco) (de Castro Silva Filho et al., Plant Mol. Biol. 30:769-780, 1996; Schnell et al., J. Biol. Chem. 266(5):3335-3342, 1991); 5-(enolpiruvil)shikimato-3-fosfato sintasa (EPSPS) (Archer et al., J. Bioenerg. Biomemb. 22(6):789-810, 1990); triptófano sintasa (Zhao et al., J. Biol. Chem. 270(11):6081-6087, 1995); plastocianina (Lawrence et al., J. Biol. Chem. 272(33):20357-20363, 1997); corismato sintasa (Schmidt et al., J. Biol. Chem. 268(36):27447-27457, 1993); y la proteína de unión a la clorofila a/b captadora de luz (LHBP) (Lamppa et al., J. Biol. Chem. 263:14996-14999, 1988). Véase también Von Heijne et al., Plant Mol. Biol. Rep. 9:104-126, 1991; Clark et al., J. Biol. Chem. 264:17544-17550, 1989; Della-Cioppa et al., Plant Physiol. 84:965-968, 1987; Romer et al., Biochem. Biophys. Res. Commun. 196:1414-1421, 1993; y Shah et al., Science 233: 478-481, 1986.

En conjunción con cualquiera de los procedimientos y/o composiciones divulgados en el presente documento, los constructos de ácido nucleico pueden prepararse para dirigir la expresión de la secuencia codificante de la nucleasa dirigida al sitio mutante del cloroplasto de la célula vegetal. Los procedimientos de transformación de cloroplastos son conocidos en la técnica. Véase, por ejemplo, Svab et al., Proc. Nat'l. Acad. Sci. USA 87:8526-8530, 1990; Svab y Maliga, Proc. Nat'l. Acad. Sci. USA 90:913-917, 1993; Svab y Maliga, EMBO J. 12:601-606, 1993. El procedimiento se basa en el suministro por pistola de partículas de ADN que contiene un marcador seleccionable y en la localización del ADN en el genoma plastidial mediante recombinación homóloga. Además, la transformación plastidial puede llevarse a cabo mediante la transactivación de un transgén plastidial silencioso a través de la expresión tisular preferente de una ARN polimerasa codificada en el núcleo y dirigida por el plásmido. Se ha informado de un sistema de este tipo en McBride et al. Proc. Nat'l. Acad. Sci. USA 91:7301-7305, 1994.

Los ácidos nucleicos de interés dirigidos al cloroplasto pueden optimizarse para su expresión en el cloroplasto para tener en cuenta las diferencias en el uso de codones entre el núcleo de la planta y este orgánulo. De este modo, los ácidos nucleicos de interés pueden sintetizarse utilizando codones preferidos por los cloroplastos. Véase, por ejemplo, Patente de EE. UU. No. 5,380,831.

Los constructos de ácido nucleico pueden utilizarse para transformar células vegetales y regenerar plantas transgénicas que comprendan las secuencias codificantes de nucleasas dirigidas al sitio. Existen numerosos vectores y procedimientos de transformación de plantas. Véase, por ejemplo, Patente de EE. UU. No. 6,753,458, An, G. et al., Plant Physiol., 81:301-305, 1986; Fry, J. et al., Plant Cell Rep. 6:321-325, 1987; Block, M., Theor. Appl Genet. 76:767-774, 1988; Hinchey et al., Stadler. Genet. Symp. 203:212-212, 1990; Cousins et al., Aust. J. Plant Physiol. 18:481-494, 1991; Chee, P. P. y Slightom, J. L., Gene. 118:255-260, 1992; Christou et al., Trends. Biotechnol. 10:239-246, 1992; D'Halluin et al., Bio/Technol. 10:309-314, 1992; Dhir et al., Plant Physiol. 99:81-88, 1992; Casas et al., Proc. Nat'l. Acad. Sci. USA 90:11212-11216, 1993; Christou, P., In Vitro Cell. Dev. Biol.-Plant 29P:119-124, 1993; Davies, et al., Plant Cell Rep. 12:180-183, 1993; Dong, J. A. y Mc Hughen, A., Plant Sci. 91:139-148, 1993; Franklin, C. I. y Trieu, T. N., Plant. Physiol. 102:167, 1993; Golovkin et al., Plant Sci. 90:41-52, 1993; Guo Chin Sci. Bull. 38:2072-2078, 1993; Asano, et al., Plant Cell Rep. 13, 1994; Ayeres N. M. y Park, W. D., Crit. Rev. Plant. Sci. 13:219-239, 1994; Barcelo et al., Plant. J. 5:583-592, 1994; Becker, et al., Plant. J. 5:299-307, 1994; Borkowska et al., Acta. Physiol Plant. 16:225-230, 1994; Christou, P., Agro. Food. Ind. Hi Tech. 5:17-27, 1994; Eapen et al., Plant Cell Rep. 13:582-586, 1994; Hartman et al., Bio-Technology 12:919-923, 1994; Ritala et al., Plant. Mol. Biol. 24:317-325, 1994 y Wan, Y.

C. y Lemaux, P. G., Plant Physiol. 104:3748, 1994. Los constructos también pueden transformarse en células vegetales mediante recombinación homóloga.

El término "tipo silvestre" cuando se hace referencia a una secuencia peptídica y secuencia nucleotídica se refiere a una secuencia peptídica y secuencia nucleotídica (locus/género/alelo), respectivamente, que tiene las características de esa secuencia peptídica y secuencia nucleotídica cuando se aísla de una fuente natural. Una secuencia peptídica y una secuencia nucleotídica de tipo silvestre es la que se observa con mayor frecuencia en una población y, por tanto, se designa arbitrariamente como la forma "normal" o "de tipo silvestre" de la secuencia peptídica y la secuencia nucleotídica, respectivamente. "Tipo silvestre" también puede referirse a la secuencia en una posición o posiciones específicas de nucleótidos, o a la secuencia en una posición o posiciones particulares de codones, o a la secuencia en una posición o posiciones particulares de aminoácidos.

"Secuencia de consenso" se define como una secuencia de aminoácidos o nucleótidos que contienen aminoácidos o nucleótidos idénticos o aminoácidos o nucleótidos funcionalmente equivalentes en al menos el 25% de la secuencia. No es necesario que los aminoácidos o nucleótidos idénticos o funcionalmente equivalentes sean contiguos.

El término "*Brassica*", tal como se utiliza aquí, se refiere a las plantas del género *Brassica*. Algunas especies ejemplares de *Brassica* son, entre otras, *B. carinata*, *B. elongate*, *B. fruticulosa*, *B. juncea*, *B. napus*, *B. narinosa*, *B. nigra*, *B. oleracea*, *B. perviridis*, *B. rapa* (syn *B. campestris*), *B. rupestris*, *B. septiceps* y *B. tournefortii*.

Una nucleobase es una base, que en ciertas realizaciones preferidas es una purina, una pirimidina o un derivado o análogo de las mismas. Los nucleósidos son nucleobases que contienen una fracción de pentosefuranosilo, por ejemplo, un ribosido opcionalmente sustituido o un 2'-desoxirribosido. Los nucleósidos pueden estar unidos por uno de varios enlaces, que pueden o no contener fósforo. Los nucleósidos unidos por enlaces fosfodiéster no sustituidos se denominan nucleótidos. El término "nucleobase", tal como se utiliza en el presente documento, incluye las nucleobases peptídicas, las subunidades de los ácidos nucleicos peptídicos y las nucleobases de morfina, así como los nucleósidos y los nucleótidos.

Una oligonucleobasa es un polímero que comprende nucleobases; en algunas realizaciones al menos una porción de la cual puede hibridarse por apareamiento de bases Watson-Crick a un ADN que tenga la secuencia complementaria. Una cadena oligonucleobásica puede tener un único extremo 5' y 3', que son las nucleobases finales del polímero. Una cadena oligonucleobásica concreta puede contener nucleobases de todo tipo. Un compuesto de oligonucleobase es un compuesto que comprende una o más cadenas de oligonucleobase que pueden ser complementarias e hibridarse por apareamiento de bases Watson-Crick. Las nucleobases de tipo ribo incluyen nucleobases que contienen pentosefuranosilo en las que el carbono 2' es un metileno sustituido con un hidroxilo, un alquilo o un halógeno. Las nucleobases de tipo desoxirribo son nucleobases distintas de las de tipo ribo e incluyen todas las nucleobases que no contienen una fracción de pentosefuranosilo.

En ciertas realizaciones, una cadena de oligonucleobase puede incluir tanto cadenas de oligonucleobase como segmentos o regiones de cadenas de oligonucleobase. Una cadena oligonucleobásica puede tener un extremo 3' y un extremo 5', y cuando una cadena oligonucleobásica es coextensiva con una cadena, los extremos 3' y 5' de la cadena son también los extremos 3' y 5' de la cadena.

Tal como se utiliza en el presente documento, el término "codón" se refiere a una secuencia de tres nucleótidos adyacentes (ya sea ARN o ADN) que constituye el código genético que determina la inserción de un aminoácido específico en una cadena polipeptídica durante la síntesis de proteínas o la señal para detener la síntesis de proteínas. El término "codón" también se utiliza para referirse a las secuencias correspondientes (y complementarias) de tres nucleótidos en el ARN mensajero en donde se transcribe el ADN original.

Tal como se utiliza en el presente documento, el término "homología" se refiere a la similitud de secuencia entre proteínas y ADN. Los términos "homología" u "homólogo" se refieren a un grado de identidad. Puede haber homología parcial u homología completa. Una secuencia parcialmente homóloga es aquella que tiene menos del 100% de identidad de secuencia cuando se compara con otra secuencia.

"Heterocigoto" se refiere a tener diferentes alelos en uno o más loci genéticos en segmentos cromosómicos homólogos. Tal como se utiliza en el presente documento, "heterocigoto" también puede referirse a una muestra, una célula, una población celular o un organismo en donde pueden detectarse diferentes alelos en uno o más loci genéticos. Las muestras heterocigotas también pueden determinarse mediante procedimientos conocidos en la técnica tales como, por ejemplo, la secuenciación de ácidos nucleicos. Por ejemplo, si un electroferograma de secuenciación muestra dos picos en un único locus y ambos picos son aproximadamente del mismo tamaño, la muestra puede caracterizarse como heterocigótica. O bien, si un pico es más pequeño que otro, pero tiene al menos 25% del tamaño del pico más grande, la muestra puede caracterizarse como heterocigótica. En algunos ejemplos, el pico más pequeño es al menos aproximadamente el 15% del pico más grande. En otros ejemplos, el pico más pequeño es al menos aproximadamente 10% del pico más grande. En otros ejemplos, el pico más pequeño es al menos aproximadamente 5% del pico más grande. En otros ejemplos, se detecta una cantidad mínima del pico más pequeño.

Tal como se utiliza en el presente documento, "homocigoto" se refiere a tener alelos idénticos en uno o más loci genéticos en segmentos cromosómicos homólogos. "Homocigoto" también puede referirse a una muestra, una célula, una población celular o un organismo en donde pueden detectarse los mismos alelos en uno o más loci genéticos. Las muestras homocigóticas pueden determinarse mediante procedimientos conocidos en la técnica, tales como, por ejemplo, la secuenciación de ácidos nucleicos. Por ejemplo, si un electroferograma de secuenciación muestra un único pico en un locus concreto, la muestra puede denominarse "homocigótica" con respecto a ese locus.

El término "hemizigoto" se refiere a un gen o segmento génico que está presente una sola vez en el genotipo de una célula o un organismo porque el segundo alelo está suprimido o no está presente en el segmento cromosómico homólogo. Tal y como se utiliza en el presente documento, "hemizigótico" también puede referirse a una muestra, una célula, una población celular o un organismo en donde un alelo en uno o más loci genéticos puede detectarse sólo una vez en el genotipo.

El término "estado de cigosidad", tal como se utiliza en el presente documento, se refiere a una muestra, una población celular o un organismo como heterocigoto, homocigoto o hemizigoto, tal como se determina mediante procedimientos de prueba conocidos en la técnica y descritos en el presente documento. Por "estado de cigosidad de un ácido nucleico" se entiende determinar si la fuente de ácido nucleico parece heterocigótica, homocigótica o hemizigótica. El "estado de cigosidad" puede referirse a diferencias en una única posición nucleotídica de una secuencia. En algunos procedimientos, el estado de cigosidad de una muestra con respecto a una única mutación puede clasificarse como homocigótico de tipo silvestre, heterocigótico (es decir, un alelo de tipo silvestre y un alelo mutante), homocigótico mutante o hemizigótico (es decir, una única copia del alelo de tipo silvestre o mutante).

Tal como se utiliza en el presente documento, el término "**RTDS**" se refiere a The Rapid Trait Development System™ (RTDS) desarrollado por Cibus. **El RTDS** es un sistema de modificación de genes en sitios específicos que es eficaz para realizar cambios precisos en una secuencia genética sin la incorporación de genes extraños o secuencias de control.

El término "aproximadamente", tal como se utiliza en el presente documento, significa en términos cuantitativos más o menos 10%. Por ejemplo, "aproximadamente 3%" englobaría 2,7-3,3% y "aproximadamente 10%" englobaría 9-11%. Además, cuando en el presente documento se utiliza "aproximadamente" junto con un término cuantitativo, se entiende que, además del valor más o menos 10%, también se contempla y describe el valor exacto del término cuantitativo. Por ejemplo, el término "aproximadamente 3%" contempla, describe e incluye expresamente exactamente el 3%.

RTDS y oligonucleótidos reparadores (GRON)

La presente divulgación se refiere en general a procedimientos novedosos para mejorar la eficacia de la orientación de modificaciones a lugares específicos en secuencias genómicas u otras secuencias de nucleótidos. Además, esta divulgación se refiere al ADN objetivo que ha sido modificado, mutado o marcado por los enfoques divulgados en el presente documento. La divulgación también se refiere a células, tejidos y organismos que han sido modificados por los procedimientos de la divulgación. La presente divulgación se basa en el desarrollo de composiciones y procedimientos relacionados en parte con el exitoso sistema de conversión, el Rapid Trait Development System (**RTDS**™, Cibus US LLC).

El RTDS se basa en la alteración de un gen objetivo utilizando el propio sistema de reparación génica de la célula para modificar específicamente la secuencia génica *in situ* y no insertar ADN extraño y secuencias de control de la expresión génica. Este procedimiento produce un cambio preciso en la secuencia genética, mientras que el resto del genoma permanece inalterado. A diferencia de los OMG transgénicos convencionales, no hay integración de material genético extraño ni queda material genético extraño en la planta. Los cambios en la secuencia genética introducidos por **el RTDS** no se insertan al azar. Dado que los genes afectados permanecen en su ubicación nativa, no se produce ningún patrón de expresión aleatorio, incontrolado o adverso.

El **RTDS** que efectúa este cambio es un oligonucleótido sintetizado químicamente (GRON), como se describe en el presente documento, que puede estar compuesto tanto de ADN como de bases de ARN modificadas, así como de otras moléculas químicas, y está diseñado para hibridarse en la ubicación del gen objetivo para crear un par o pares de bases no coincidentes. Este par de bases no coincidentes actúa como una señal para atraer al sistema natural de reparación de genes de la propia célula hacia ese lugar y corregir (sustituir, insertar o eliminar) el nucleótido o nucleótidos designados dentro del gen. Una vez completado el procedimiento de corrección, la molécula **RTDS** se degrada y el gen ahora modificado o reparado se expresa bajo los mecanismos de control endógenos normales de ese gen.

Los procedimientos y composiciones divulgados en el presente documento pueden practicarse o fabricarse con "oligonucleobases reparadoras de genes" (GRON) que tengan las conformaciones y químicas descritas en detalle en el presente documento y más adelante. Las "oligonucleobases reparadoras de genes" contempladas en el presente documento también se han descrito en la literatura científica y de patentes publicada con otros nombres, como

"oligonucleobases recombinogénicas"; "oligonucleótidos quiméricos ARN/ADN"; "oligonucleótidos quiméricos"; "oligonucleótidos dúplex mixtos" (MDONs); "oligonucleótidos ARN ADN (RDO)"; "oligonucleótidos dirigidos a genes"; "genoplastos"; "oligonucleótidos monocatenarios modificados"; "vectores mutacionales oligodeoxinucleótidos monocatenarios" (SSOMV); "vectores mutacionales dúplex"; y "vectores mutacionales heterodúplex". La oligonucleobasa de reparación génica puede introducirse en una célula vegetal mediante cualquier procedimiento comúnmente utilizado en la técnica, incluidos, entre otros, los microportadores (administración biolística), las microfibras, la captación mediada por polietilenglicol (PEG), la electroporación y la microinyección.

En una realización, la oligonucleobasa de reparación génica es un oligonucleótido dúplex mixto (MDON) en donde los nucleótidos de tipo ARN del oligonucleótido dúplex mixto se hacen resistentes a la RNasa sustituyendo el 2'-hidroxilo por una funcionalidad fluoro, cloro o bromo o colocando un sustituyente en el 2'-O. Los sustituyentes adecuados incluyen los sustituyentes enseñados por el Kmiec II. Los sustituyentes alternativos incluyen los sustituyentes enseñados por Pat. de EE. UU. No. 5,334,711 (Sproat) y los sustituyentes enseñados por las publicaciones de patentes EP 629 387 y EP 679 657 (colectivamente, las Solicitudes Martin). Tal como se utiliza en el presente documento, un derivado 2'-fluoro, cloro o bromo de un ribonucleótido o un ribonucleótido que tenga una T- OH sustituida con un sustituyente descrito en las Solicitudes Martin o Sproat se denomina "ribonucleótido sustituido con T". Tal como se utiliza en el presente documento, el término "nucleótido tipo ARN" significa un nucleótido T-hidroxilo o 2'-Sustituido que está unido a otros nucleótidos de un oligonucleótido dúplex mixto por un enlace fosfodiéster no sustituido o cualquiera de los enlaces no naturales enseñados por Kmiec I o Kmiec II. Tal como se utiliza en el presente documento, el término "nucleótido de tipo desoxirribosa" significa un nucleótido que tiene una T-H, que puede unirse a otros nucleótidos de una oligonucleobasa de reparación génica mediante un enlace fosfodiéster no sustituido o cualquiera de los enlaces no naturales enseñados por Kmiec I o Kmiec II.

En una realización particular de la presente divulgación, la oligonucleobasa de reparación génica es un oligonucleótido dúplex mixto (MDON) que está unido únicamente por enlaces fosfodiéster no sustituidos. En realizaciones alternativas, el enlace es por fosfodiésteres sustituidos, derivados de fosfodiéster y enlaces no basados en fósforo como enseña Kmiec II. En otra realización, cada nucleótido de tipo ARN en el oligonucleótido dúplex mixto es un Nucleótido 2'-Sustituido. Las formas de realización preferidas de los ribonucleótidos 2'-sustituidos son los ribonucleótidos 2'-fluoro, T-metoxi, 2'-propiloxi, 2'-aliloxi, 2'-hidroxietiloxi, 2'-metoxietiloxi, T-fluoropropiloxi y 2'-trifluoropropiloxi sustituidos. Las formas de realización más preferidas de los ribonucleótidos 2'-sustituidos son los nucleótidos 2'-fluoro, 2'-metoxi, 2'-metoxietiloxi y 2'-aliloxi sustituidos. En otra realización, el oligonucleótido dúplex mixto está unido por enlaces fosfodiéster no sustituidos,

Aunque los oligonucleótidos dúplex mixtos (MDONs) que tienen un solo tipo de nucleótido de tipo ARN 2'- sustituido se sintetizan más convenientemente, los procedimientos de la divulgación pueden practicarse con oligonucleótidos dúplex mixtos que tienen dos o más tipos de nucleótidos de tipo ARN. La función de un segmento de ARN puede no verse afectada por una interrupción causada por la introducción de un desoxinucleótido entre dos trinucleótidos de tipo ARN, en consecuencia, el término segmento de ARN engloba términos tales como "segmento de ARN interrumpido". Un segmento de ARN ininterrumpido se denomina segmento de ARN contiguo. En una realización alternativa, un segmento de ARN puede contener alternativamente nucleótidos 2'-OH resistentes a la RNasa y no sustituidos. Los oligonucleótidos dúplex mixtos en algunas realizaciones tienen menos de 100 nucleótidos y otras realizaciones menos de 85 nucleótidos, pero más de 50 nucleótidos. La primera y la segunda cadena están emparejadas por bases Watson-Crick. En una realización, las cadenas del oligonucleótido dúplex mixto están unidas covalentemente por un enlazador, tal como un hexa, penta o tetranucleótido monocatenario, de modo que la primera y la segunda cadenas son segmentos de una única cadena de oligonucleótidos con un único extremo 3' y un único extremo 5'. Los extremos 3' y 5' pueden protegerse mediante la adición de un "tapa de horquilla" por el que los nucleótidos terminales 3' y 5' se emparejan Watson-Crick con nucleótidos adyacentes. Además, puede colocarse una segunda horquilla en la unión entre la primera y la segunda cadena, alejada de los extremos 3' y 5', de modo que se establezca el emparejamiento Watson-Crick entre la primera y la segunda cadena.

La primera y la segunda cadena contienen dos regiones homólogas con dos fragmentos del gen/alelo objetivo, es decir, tienen la misma secuencia que el gen/alelo objetivo. Una región homóloga contiene los nucleótidos de un segmento de ARN y puede contener uno o más nucleótidos de tipo ADN de segmento de ADN de conexión y también puede contener nucleótidos de tipo ADN que no están dentro del segmento de ADN de intervención. Las dos regiones de homología están separadas por, y cada una es adyacente a, una región que tiene una secuencia que difiere de la secuencia del gen objetivo, denominada "región heteróloga". La región heteróloga puede contener uno, dos o tres nucleótidos no coincidentes. Los nucleótidos no coincidentes pueden ser contiguos o, alternativamente, pueden estar separados por uno o dos nucleótidos homólogos con el gen/alelo objetivo. Alternativamente, la región heteróloga también puede contener una inserción o uno, dos, tres o de cinco o menos nucleótidos. Alternativamente, la secuencia del oligonucleótido dúplex mixto puede diferir de la secuencia del gen/alelo objetivo sólo por la delección de uno, dos, tres o cinco o menos nucleótidos del oligonucleótido dúplex mixto. En este caso, se considera que la longitud y la posición de la región heteróloga es la longitud de la delección, aunque ningún nucleótido del oligonucleótido dúplex mixto se encuentre dentro de la región heteróloga. La distancia entre los fragmentos del gen objetivo que son complementarios a las dos regiones homólogas es idéntica a la longitud de la región heteróloga en la que se pretende realizar una sustitución o sustituciones. Cuando la región heteróloga contiene una inserción, las regiones homólogas quedan así separadas en el oligonucleótido dúplex mixto más de lo que lo están sus fragmentos homólogos complementarios en el gen/alelo, y lo contrario es aplicable cuando la región heteróloga codifica una delección.

Los segmentos de ARN de los oligonucleótidos dúplex mixtos son cada uno parte de una región homóloga, es decir, una región que es idéntica en secuencia a un fragmento del gen objetivo, cuyos segmentos juntos en algunas realizaciones contienen al menos 13 nucleótidos de tipo ARN y en algunas realizaciones de 16 a 25 nucleótidos de tipo ARN o en otras realizaciones de 18 a 22 nucleótidos de tipo ARN o en algunas realizaciones 20 nucleótidos. En una realización, los segmentos de ARN de las regiones homológicas están separados por y adyacentes a, es decir, "conectados por" un segmento de ADN intermedio. En una realización, cada nucleótido de la región heteróloga es un nucleótido del segmento de ADN intermedio. Un segmento de ADN intermedio que contiene la región heteróloga de un oligonucleótido dúplex mixto se denomina "segmento mutador"

En otra realización de la presente divulgación, la oligonucleobase de reparación génica (GRON) es un vector mutacional oligodeoxinucleótido monocatenario (SSOMV), como el divulgado en la Solicitud Internacional de Patente PCT/USOO/23457, Pat. de EE. UU. Nos. 6.271.360, 6,479,292y 7,060,500. La secuencia del SSOMV se basa en los mismos principios que los vectores mutacionales descritos en las Pat. de EE. UU. Nos. 5.756.325; 5,871,984; 5,760,012; 5,888,983; 5,795,972; 5,780,296; 5,945,339; 6,004,804y 6,010,907 y en la Publicación Internacional Nos. WO 98/49350; WO 99/07865; WO 99/58723; WO 99/58702y WO 99/40789. La secuencia del SSOMV contiene dos regiones homólogas con la secuencia objetivo separadas por una región que contiene la alteración genética deseada denominada región mutante. La región mutante puede tener una secuencia de la misma longitud que la secuencia que separa las regiones homólogas en la secuencia objetivo, pero con una secuencia diferente. Una región mutante de este tipo puede provocar una sustitución. Alternativamente, las regiones homólogas en el SSOMV pueden ser contiguas entre sí, mientras que las regiones en el gen objetivo que tienen la misma secuencia están separadas por uno, dos o más nucleótidos. Tal SSOMV causa una delección del gen objetivo de los nucleótidos que están ausentes del SSOMV. Por último, la secuencia del gen objetivo que es idéntica a las regiones homólogas puede ser adyacente en el gen objetivo, pero estar separada por uno, dos o más nucleótidos en la secuencia del SSOMV. Un SSOMV de este tipo provoca una inserción en la secuencia del gen objetivo. En ciertas realizaciones, una SSOMV no se recuece a sí misma.

Los nucleótidos del SSOMV son desoxirribonucleótidos que están unidos por enlaces fosfodiéster no modificados, excepto que el enlace internucleotídico terminal 3' y/o terminal 5' o alternativamente los dos enlaces internucleotídicos terminales 3' y/o terminales 5' pueden ser un fosforotioato o un fosfoamidato. Tal como se utiliza en el presente documento, un enlace internucleotídico es el enlace entre nucleótidos del SSOMV y no incluye el enlace entre el nucleótido del extremo 3' o el nucleótido del extremo 5' y un sustituyente bloqueante. En una realización específica, la longitud del SSOMV está comprendida entre 21 y 55 desoxinucleótidos y las longitudes de las regiones homológicas son, en consecuencia, una longitud total de al menos 20 desoxinucleótidos y al menos dos regiones homológicas deben tener cada una longitudes de al menos 8 desoxinucleótidos.

El SSOMV puede diseñarse para ser complementario a la cadena codificante o no codificante del gen objetivo. Cuando la mutación deseada es una sustitución de una sola base, se prefiere que tanto el nucleótido mutador como el nucleótido objetivo sean una pirimidina. En la medida en que sea compatible con la consecución del resultado funcional deseado, se prefiere que tanto el nucleótido mutador como el nucleótido objetivo de la cadena complementaria sean pirimidinas. Se prefieren especialmente los SSOMV que codifican mutaciones de transversión, es decir, un nucleótido mutador C o T no coincide, respectivamente, con un nucleótido C o T en la cadena complementaria.

Fragmento Okazaki / 2'-OME GRON Diseño. En diversas realizaciones, un GRON puede tener tanto nucleótidos de ARN como de ADN y/u otros tipos de nucleobases. En algunas realizaciones, uno o más de los nucleótidos de ADN o ARN comprenden una modificación. En ciertas realizaciones, el primer nucleótido 5' es un nucleótido de ARN y el resto de los nucleótidos son ADN. En otras realizaciones, el primer nucleótido 5' de ARN se modifica con un 2-O-Me. En otras realizaciones, los primeros dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más nucleótidos 5' son un nucleótido de ARN y el resto de los nucleótidos son ADN. En otras realizaciones, uno o más de los primeros dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más nucleótidos 5' de ARN están modificados con un 2-O-Me. En las células vegetales, los picos de doble cadena en el ADN suelen repararse mediante la vía de reparación del ADN NHEJ. Esta vía no requiere una plantilla para reparar el ADN y, por tanto, es propensa a errores. La ventaja de utilizar esta vía para reparar el ADN de una célula vegetal es que es rápida, ubicua y, lo que es más importante, puede producirse en momentos en los que la célula no está replicando el ADN. Otra vía de reparación del ADN que funciona en la reparación de roturas de doble cadena fuera de la horquilla de replicación en células vegetales se denomina recombinación homóloga (HR); sin embargo, a diferencia de la vía NHEJ, este tipo de reparación es precisa y requiere el uso de una plantilla de ADN (GRON). Dado que estos GRON imitan los fragmentos de Okazaki en la horquilla de replicación del ADN de los genes objetivo, no es obvio utilizarlos con un cortador de ADN de doble cadena para los expertos en la materia.

Mejorar la eficacia

La presente divulgación proporciona una serie de enfoques para aumentar la eficacia de la conversión de un gen objetivo utilizando oligonucleótidos de reparación, y que pueden utilizarse solos o en combinación unos con otros. Los enfoques divulgados en el presente documento incluyen:

1. Introducir modificaciones en los oligonucleótidos de reparación que atraigan a la maquinaria de reparación del ADN al sitio objetivo (desajuste).

A. La introducción de uno o más sitios abásicos en el oligonucleótido (por ejemplo, dentro de 10 bases, por ejemplo, dentro de 5 bases del sitio de desajuste deseado) genera una lesión que es un intermediario en la reparación por escisión de bases (BER), y que atrae la maquinaria BER a la vecindad del sitio objetivo para la conversión por el oligonucleótido de reparación. Los oligonucleótidos modificados con dEspaciador (furano abásico) pueden prepararse como se describe en, por ejemplo, Takeshita et al., J. Biol. Chem., 262:10171-79, 1987.

B. La inclusión de compuestos que inducen roturas de cadena simple o doble, ya sea en el oligonucleótido o junto con el oligonucleótido, genera una lesión que se repara mediante NHEJ, unión de extremos mediada por microhomología (MMEJ) y recombinación homóloga. A modo de ejemplo, los antibióticos de la familia de la bleomicina, los dedos de zinc, FokI (o cualquier clase de enzima de restricción de tipo IIS) y otras nucleasas pueden acoplarse covalentemente al extremo 3' o 5' de los oligonucleótidos de reparación, con el fin de introducir roturas de doble cadena en las proximidades del sitio objetivo de conversión por el oligonucleótido de reparación. Los antibióticos de la familia de la bleomicina son glicopéptidos desdobladores del ADN que incluyen la bleomicina, la zeocina, la fleomicina, la tallysomicina, la pepleomicina y otros.

C. La introducción de uno o más 8'oxo dA o dG incorporados en el oligonucleótido (por ejemplo, dentro de 10 bases, por ejemplo, dentro de 5 bases del sitio de desajuste deseado) genera una lesión que es similar a las lesiones creadas por especies reactivas de oxígeno. Estas lesiones inducen el llamado sistema de "reparación por empuje". Véase, por ejemplo, Kim et al., J. Biochem. Mol. Biol. 37:657-62, 2004.

2. Aumentar la estabilidad de los oligonucleótidos de reparación:

Introducción de una base inversa (idC) en el extremo 3' del oligonucleótido para crear un extremo 3' bloqueado en el oligonucleótido de reparación.

Introducción de uno o más 2'-O-metil nucleótidos o bases que aumentan la energía de hibridación (véase, por ejemplo, los documentos WO2007/073149) en el 5' y/o 3' del oligonucleótido de reparación.

Introducción de uno o una pluralidad de nucleótidos 2'-O-metil ARN en el extremo 5' del oligonucleótido de reparación, que conducen a bases de ADN que proporcionan el sitio de desajuste deseado, creando así una estructura de ácido nucleico similar al fragmento Okazaki.

Colorantes intercalantes conjugados (5' o 3') como la acridina, el psoraleno, el bromuro de etidio y las tinciones de Syber.

Introducción de una tapa terminal 5' tal como una abrazadera T/A, una fracción de colesterol, SIMA (HEX), riboC y amidita.

Modificaciones de la columna vertebral tales como fosfotioato, 2'-O metil, metil fosfonatos, ácido nucleico bloqueado (LNA), MOE (metoxietil), di PS y ácido nucleico peptídico (PNA).

Reticulación del oligonucleótido de reparación, por ejemplo, con agentes reactivos de reticulación intracadena como el cisplatino y la mitomicina C.

Conjugación con colorantes fluorescentes como Cy3, DY547, Cy3.5, Cy3B, Cy5 y DY647.

3. Aumentar la energía de hibridación del oligonucleótido de reparación mediante la incorporación de bases que aumentan la energía de hibridación (véase, por ejemplo, el documento WO2007/073149).

4. Aumente la calidad de la síntesis de oligonucleótidos de reparación utilizando multímeros de nucleótidos (dímeros, trímeros, tetrámeros, etc.) como bloques de constructo para la síntesis. El resultado es un menor número de pasos de acoplamiento y una separación más fácil de los productos de longitud completa de los bloques de constructo.

5. Uso de oligonucleótidos de reparación largos (es decir, de más de 55 nucleótidos de longitud, por ejemplo, como las longitudes descritas en el presente documento, por ejemplo, con una o más mutaciones o dos o más mutaciones dirigidas en el oligonucleótido de reparación).

En la Tabla 1 se proporcionan ejemplos de los enfoques anteriores.

Tabla 1. Química GRON ejemplar.

	Tipo de oligo	Modificaciones
5' mods	Abrazadera T/A	Abrazadera T/A
Modificaciones de la columna vertebral	Fosfotioato	PS

	Tipo de oligo	Modificaciones
Colorantes intercalantes	5' Acridina 3'	idC Acridina, idC
2'-O-metilo		ADN/ARN
Sustituciones Cy3		DY547
Facilitadores	oligos 2'-O-Me diseñados 5' y 3' del oligo conversor	2'-O-Me
Abásico	Sitio abásico colocado en varios lugares a 5' y 3' de la base de conversión. 44 mer	Abásico 2
Ayuda	Acercamiento de asistencia Solapamiento:	Cy3, idC en uno, ninguno en el otro:
	2 oligos: 1 con Cy3/idC, 1 oligo de reparación no modificado	
Ayuda	Acercamiento asistido No hay solapamiento:	sólo hacen el oligo no modificado
	2 oligos: 1 con Cy3/idC, 1 oligo de reparación no modificado	
Abásico	sitio THF colocado en diversos lugares 5 'y 3' a la base de conversión. 44 mer	Tetrahidrofurano (dsespaciador)
Modificaciones de la columna vertebral	9	2'-O-Me
Trímeros		Amiditas trímero, Cy3. idC
Reparación de empuje		8'oxo dA, 5' Cy3, idC
Reparación de empuje		8'oxo dA, 5' Cy3, idC
Rotura de doble cadena		Bleomicina
Reticulante		Cisplatino
Reticulante		Mitomicina C
Facilitadores	superbases 5' y 3' del oligo convertidor	2 amino dA y 2- tio T
Súper oligos		2'amino d, 5' Cy3, idC
Súper oligos		2-tio T, 5' Cy3, idC
Súper oligos		7-deaza A, 5' Cy3, idC
Súper oligos		7-deaza G, 5' Cy3, idC
Súper oligos		propanil dC, 5' Cy3, idC
Colorantes intercalantes	5' Psoraleno/3' idC	Psoraleno, idC
Colorantes intercalantes	5' Bromuro de etidio	Bromuro de etidio
Colorantes intercalantes	5' Manchas de siber	Manchas de siber
5' mods	5' Chol/3' idC	Colesterol
Doble mutación	Oligo largo (55+ bases) con 2 mutaciones	Cualquier modificación
5' mods	5' SIMA HEX/3'idC	SIMA HEX, idC
Modificaciones de la columna vertebral	9	Fosfonatos de metilo

(continuación)

	Tipo de oligo	Modificaciones
Modificaciones de la columna vertebral		LNA
Modificaciones de la columna vertebral	1	MOE (metoxietilo)
Sustituciones Cy3		Cy3.5
Sustituciones Cy3		Cy5
Modificaciones de la columna vertebral		di PS

5' mods		riboC para rama nm
Modificaciones de la columna vertebral		PNA
Sustituciones Cy3		DY647
5' mods	rama de 5	rama simétrica amidita/idC

Las modificaciones anteriores también pueden incluir modificaciones de nucleótidos conocidas tales como metilación, colorantes intercalantes 5', modificaciones de los extremos 5' y 3', modificaciones de la columna vertebral, reticulantes, ciclización y "tapas" y sustitución de uno o más de los nucleótidos naturales por un análogo como la inosina. Las modificaciones de nucleótidos incluyen la adición de acridina, amina, biotina, azul de cascada, colesterol, Cy3@, Cy5@, Cy5.5@ Daboyl, digoxigenina, dinitrofenil, Edans, 6-FAM, fluoresceína, 3'- gliceril, HEX, IRD-700, IRD-800, JOE, fosfato psoraleno, rodamina, ROX, tiol (SH), espaciadores, TAMRA, TET, AMCA-S", SE, BODIPY°, Marina Blue@, Pacific Blue@, Oregon Green@, Rhodamine Green@, Rhodamine Red@, Rhodol Green@ y Texas Red@. Las modificaciones de la columna vertebral de los polinucleótidos incluyen metilfosfonato, 2'-OMe-metilfosfonato ARN, fosforotiorato, ARN, 2'-OMeARN. Las modificaciones de las bases incluyen 2-amino-dA, 2-aminopurina, 3'- (ddA), 3'dA (cordicepina), 7-deaza-dA, 8-Br-dA, 8- oxo-dA, N6-Me-dA, sitio abásico (dEspaciador), biotina dT, 2'-OMe-5Me-C, 2'-OMe-propinil-C, 3'- (5-Me-dC), 3'- (ddC), 5-Br-dC, 5-1-duc, 5-Me-dC, 5-F-dC, carboxi-dT, dA convertible, dC convertible, dG convertible, dT convertible, dU convertible, 7-deaza-dG, 8-Br-dG, 8- oxo-dG, O6-Me-dG, S6-DNP-dG, 4-metilindol, 5-nitroindol, 2'-OMe-inosina, 2'-dl, o6- fenil-dl, 4-metilindol, 2'-desoxinebularina, 5-nitroindol, 2-aminopurina, dP (análogo de purina), dK (análogo de la pirimidina), 3-nitropirrol, 2-tio-dT, 4-tio-dT, biotina-dT, carboxi-dT, 04-Me-dT, 04-triazol dT, 2'-OMe-propinil-U, 5-Br-dU, 2'-dU, 5-F-dU, 5-I-dU, 04-triazol dU. Dichos términos también abarcan los ácidos nucleicos peptídicos (PNA), un análogo del ADN en donde la columna vertebral es un pseudopéptido formado por unidades de N-(2-aminoetil)-glicina en lugar de un azúcar. Los PNA imitan el comportamiento del ADN y se unen a cadenas complementarias de ácido nucleico. El esqueleto neutro del PNA da como resultado una unión más fuerte y una mayor especificidad de la que se consigue normalmente. Además, las propiedades químicas, físicas y biológicas únicas del PNA se han explotado para producir potentes herramientas biomoleculares, agentes antisentido y antigénicos, sondas moleculares y biosensores.

Las oligonucleobases pueden tener mellas, brechas, nucleótidos modificados tales como esqueletos de oligonucleótidos modificados, nucleótidos abásicos u otras moléculas químicas. En otro ejemplo divulgado en el presente documento, al menos una cadena de la oligonucleobasa incluye al menos un nucleótido modificado adicional, por ejemplo, un nucleótido modificado con 2'-O-metilo, tal como un MOE (metoxietilo), un nucleótido con un grupo 5'-fosforotioato, un nucleótido terminal unido a un derivado colesterílico, un nucleótido modificado con 2'-desoxi-2'-fluoro, un nucleótido modificado con 2'-desoxi, un nucleótido bloqueado, un nucleótido abásico (la nucleobase carece o tiene un grupo hidroxilo en su lugar (véase, por ejemplo, Glen Research, <http://www.glenresearch.com/GlenReports/GR21-14.html>)), un nucleótido modificado con 2'-amino, un nucleótido modificado con 2'-alquilo, un nucleótido morfolino, un fosforamidito y un nucleótido con una base no natural. También se incluyen diversas sales, sales mixtas y formas ácidas libres.

Las columnas vertebrales oligonucleotídicas modificadas preferidas incluyen, por ejemplo, fosforotioatos, fosforotioatos quirales, fosforoditioatos, fosfortriesteres, aminoalquilfosfortriesteres, metil y otros alquilfosfonatos incluyendo 3'-alquilfosfonatos, 5'-alquilfosfonatos y fosfonatos quirales, fosfinatos, fosforamidatos, incluyendo el 3'-aminofosforamidato y los aminoalquilfosforamidatos, tionofosforamidatos, tionoalquilfosfonatos, tionoalquilfosfortriesteres, selenofosfatos y boranofosfatos con enlaces normales 3'-5', análogos de éstos con enlaces 2'-5', y aquellos con polaridad invertida en los que uno o más enlaces internucleotídicos son enlaces 3' a 3', 5' a 5' o 2' a 2'. Los oligonucleótidos preferidos que tienen polaridad invertida comprenden un único enlace 3' a 3' en el enlace internucleotídico más 3', es decir, un único residuo nucleósido invertido que puede ser abásico (la nucleobase falta o tiene un grupo hidroxilo en su lugar). El uso más común de una inversión de enlace es añadir un enlace 3'-3' al final de un oligonucleótido antisentido con un esqueleto de fosforotioato. El enlace 3'-3' estabiliza aún más el oligonucleótido antisentido frente a la degradación por exonucleasas al crear un oligonucleótido con dos extremos 5'-OH y ningún extremo 3'-OH. Las inversiones de enlace pueden introducirse en lugares específicos durante la síntesis de oligonucleótidos mediante el uso de "fosforamiditos invertidos". Estos reactivos tienen los grupos fosforamidita en la posición 5'-OH y el grupo protector dimetoxitritilo (DMT) en la posición 3'-OH. Normalmente, el grupo protector DMT está en el 5'-OH y el fosforamidito en el 3'-OH.

Ejemplos de bases modificadas incluyen, pero no se limitan a, 2-aminopurina, 2'-amino-butilil pireno-uridina, 2'-aminouridina, 2'-desoxiuridina, 2'-fluoro-citidina, 2'-fluoro-uridina, 2,6-diaminopurina, 4-tio-uridina, 5-bromo-uridina, 5-fluoro-citidina, 5-fluorouridina, 5-indo-uridina, 5-metil-citidina, inosina, N3-metil-uridina, 7-deaza-guanina, 8-aminohexil-amino-adenina, 6-tio-guanina, 4-tio-timina, 2-tio-timina, 5-yodo-uridina, 5-yodo-citidina, 8-bromo-guanina, 8-bromo-adenina, 7-deaza-adenina, 7-diaza-guanina, 8-oxo-guanina, 5,6-dihidro-uridina y 5-hidroximetil-uridina. Estas unidades sintéticas están disponibles en el mercado (por ejemplo, compradas a Glen Research Company) y pueden incorporarse al ADN mediante síntesis química.

Ejemplos de modificación de la fracción de azúcar son la 3'-desoxilación, la 2'-fluoración y la arabanosidación, sin

embargo, esto no debe interpretarse como una limitación. Su incorporación al ADN también es posible mediante síntesis química.

Ejemplos de modificación del extremo 5' son 5'-aminación, 5'-biotinilación, 5'-fluoresceinilación, 5'-tetrafluoro-fluoreceinilación, 5'-tionación, y 5'-dabsilación, sin embargo, no debe interpretarse como limitado a ello.

Ejemplos de modificación del extremo 3' son 3'-aminación, 3'-biotinilación, 2,3-dideoxidación, 3'-tionación, 3'-dabsilación, 3'-carboxilación, y 3'-coesterilación, sin embargo, no debe interpretarse como limitado a ello.

En una realización preferida, la oligonucleobasa puede contener un sustituyente bloqueante 5' que se une a los carbonos terminales 5' a través de un enlazador. La composición química del enlazador no es crítica, salvo su longitud, que en algunas realizaciones debe ser de al menos 6 átomos y que el enlazador debe ser flexible. Pueden utilizarse diversos sustituyentes no tóxicos, tales como biotina, colesterol u otros esteroides, o un colorante fluorescente catiónico no intercalante. Reactivos particularmente preferidos para hacer oligonucleobases son los reactivos vendidos como Cy3[™] y Cy5[™] por Glen Research, Sterling Va. (ahora GE Healthcare), que son fosforoamiditas bloqueadas que al incorporarse a un oligonucleótido producen 3,3',3',3'-tetrametil N,N'-isopropil sustituido indomonocarbocianina y colorantes indodicarbocianina, respectivamente. Se prefiere especialmente Cy3. Cuando la indocarbocianina está sustituida por N-oxialquil, puede unirse convenientemente al terminal 5' del oligodesoxinucleótido como un fosfodiéster con un fosfato terminal 5'. Cuando la fosforamidita Cy3 disponible comercialmente se utiliza de acuerdo con las instrucciones, la modificación 5' resultante consiste en un sustituyente bloqueante y un enlazador juntos que son una N-hidroxiopropil, N'-fosfatidilpropil 3,3',3',3'-tetrametil indomonocarbocianina. Otros colorantes contemplados incluyen Rodamina6G, Tetrametilrhodamina, Sulforhodamina 101, Merocianina 540, Atto565, Atto550 26, Cy3.5, Dy547, Dy548, Dy549, Dy554, Dy555, Dy556, Dy560, mStrawberry y mCherry.

En una realización preferida, el colorante de indocarbocianina está tetra sustituido en las posiciones 3 y 3' de los anillos de indol. Sin limitaciones teóricas, estas sustituciones impiden que el colorante sea un colorante intercalante. La identidad de los sustituyentes en estas posiciones no es crítica.

Los diseños de oligos descritos en el presente documento también podrían utilizarse como plantillas donantes más eficientes en combinación con otras tecnologías de edición o recombinación de ADN, incluyendo, pero sin limitarse a, la orientación génica mediante recombinación homóloga específica de sitio por nucleasas de dedos de zinc, nucleasas efectoras similares a activadores de transcripción (TALEN) o repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR).

La presente divulgación se refiere en general a procedimientos para la modificación eficiente del ADN genómico celular y/o la recombinación de ADN en el ADN genómico de las células. Aunque no se limitan a ningún uso en particular, algunos procedimientos proporcionados en el presente documento pueden ser útiles para, por ejemplo, introducir una modificación en el genoma de una célula con el fin de determinar el efecto de la modificación en la célula. Por ejemplo, se puede introducir una modificación en la secuencia de nucleótidos que codifica una enzima para determinar si la modificación altera la actividad enzimática de la enzima, y/o determinar la localización de la región catalítica de la enzima. Alternativamente, la modificación puede introducirse en la secuencia codificante de una proteína de unión al ADN para determinar si se altera la actividad de unión al ADN de la proteína, y así delimitar la región concreta de unión al ADN dentro de la proteína. Otra alternativa consiste en introducir una modificación en una secuencia reguladora no codificante (por ejemplo, un promotor, un potenciador, una secuencia de ARN regulador (ARNmi), etc.) para determinar el efecto de la modificación en el nivel de expresión de una segunda secuencia que está vinculada de forma operativa a la secuencia reguladora no codificante. Esto puede ser deseable para, por ejemplo, definir la secuencia particular que posee actividad reguladora.

Cortadores de ADN

Una estrategia para producir disrupción génica dirigida es a través de la generación de roturas de ADN de cadena simple o doble usando un cortador de ADN tal como una endonucleasa de sitio específico. Las endonucleasas se utilizan con mayor frecuencia para la interrupción selectiva de genes en organismos que tradicionalmente han sido refractarios a los procedimientos de selección de genes más convencionales, tal como las algas, las plantas y los modelos animales de gran tamaño, incluyendo el ser humano. Por ejemplo, actualmente se están realizando ensayos clínicos en humanos con nucleasas de dedos de zinc para el tratamiento y la prevención de la infección por VIH. Además, la modificación de endonucleasas se está utilizando actualmente en intentos de interrumpir genes que producen fenotipos indeseables en los cultivos. De acuerdo con la presente invención, el cortador de ADN es un sistema CRISPR-Cas. Otros cortadores de ADN divulgados en el presente documento incluyen dedos de zinc, TALENs, meganucleasas y antibióticos.

Dedos de zinc

Una clase de endonucleasas artificiales son las endonucleasas de dedos de zinc. Las endonucleasas de dedo de zinc combinan un dominio de corte inespecífico, normalmente el de la endonucleasa FokI, con dominios de proteína de dedo de zinc diseñados para unirse a secuencias de ADN específicas. La estructura modular de las endonucleasas de dedos de zinc las convierte en una plataforma versátil para producir roturas de doble cadena en sitios específicos

del genoma. Dado que la endonucleasa FokI se escinde como un dímero, una estrategia para evitar que se produzcan eventos de escisión fuera del objetivo ha sido diseñar dominios de dedos de zinc que se unan en sitios adyacentes de 9 pares de bases. Véase también Patentes de EE. UU. Nos. 7,285,416; 7,521,241; 7,361,635; 7,273,923; 7,262,054; 7,220,719; 7,070,934; 7,013,219; 6,979,539; 6,933,113; 6,824,978.

TALENs

Los TALENs son nucleasas direccionables que se utilizan para inducir roturas de cadena simple y doble en sitios específicos de ADN, que luego son reparados por mecanismos que pueden ser explotados para crear alteraciones de secuencia en el sitio de escisión.

El bloque de constructo fundamental que se utiliza para diseñar la región de unión al ADN de TALENs es un dominio repetido altamente conservado derivado de TALEs naturales codificadas por proteobacterias *Xanthomonas spp.* La unión al ADN de un TALEN está mediada por matrices de repeticiones de 33-35 aminoácidos altamente conservadas que están flanqueadas por dominios adicionales derivados de TALE en los extremos aminoterminal y carboxiterninal de las repeticiones.

Estas repeticiones TALE se unen específicamente a una sola base de ADN, cuya identidad está determinada por dos residuos hipervariables que se encuentran típicamente en las posiciones 12 y 13 de la repetición, correspondiendo el número de repeticiones en una matriz a la longitud del ácido nucleico objetivo deseado, seleccionándose la identidad de la repetición para que coincida con la secuencia del ácido nucleico objetivo. El ácido nucleico objetivo puede tener entre 15 y 20 pares de bases para maximizar la selectividad del sitio objetivo. La escisión del ácido nucleico objetivo suele producirse en los 50 pares de bases siguientes a la unión de TALEN. En la técnica se han descrito programas informáticos para el diseño de sitios de reconocimiento TALEN. Véase, por ejemplo, Cermak et al., *Nucleic Acids Res.* 2011 julio; 39(12): e82.

Una vez diseñados para coincidir con la secuencia objetivo deseada, los TALENs pueden expresarse recombinantemente e introducirse en protoplastos como proteínas exógenas, o expresarse a partir de un plásmido dentro del protoplasto o administrarse como ARNm.

Meganucleasas

Las endonucleasas homólogas, también conocidas como meganucleasas, son endonucleasas específicas de secuencia que generan roturas de doble cadena en el ADN genómico con un alto grado de especificidad debido a sus grandes sitios de corte (por ejemplo, >14 pb). Aunque la especificidad de las endonucleasas homólogas para sus sitios objetivo permite dirigir con precisión las roturas de ADN inducidas, los sitios de escisión de las endonucleasas homólogas son raros y la probabilidad de encontrar un sitio de escisión natural en un gen objetivo es baja.

Otra clase de endonucleasas artificiales son las meganucleasas modificadas. Las endonucleasas homólogas modificadas se generan modificando la especificidad de las endonucleasas homólogas existentes. En uno de los enfoques, se introducen variaciones en la secuencia de aminoácidos de las endonucleasas homólogas naturales y, a continuación, se examinan las endonucleasas homólogas diseñadas resultantes para seleccionar proteínas funcionales que escindan un sitio de unión específico. En otro enfoque, las endonucleasas homólogas químicas se diseñan combinando los sitios de reconocimiento de dos endonucleasas homólogas diferentes para crear un nuevo sitio de reconocimiento compuesto por medio sitio de cada endonucleasa homóloga. Véase, por ejemplo, Patente de EE. UU. Nos. 8,338,157

Sistemas CRISPR o CRISPR/cas

El sistema CRISPR-Cas contiene tres componentes básicos de diseño: 1) gen Cas, transcrito (por ejemplo, ARNm) o proteína; 2) ARN guía (ARNg); y 3) ARNcr (ARN CRISPR) son segmentos de ARN procesados a partir de transcritos de ARN que codifican los arreglos de repeticiones CRISPR, que albergan una región "protoespaciadora" complementaria a un sitio de ADN extraño (por ejemplo, una región objetivo de ADN endógeno) y una parte de la repetición CRISPR. Véase, por ejemplo, Solicitud de PCT Nos. WO/2014/093661 y WO/2013/176772.

Gen, transcrito (por ejemplo, ARNm) o proteína Cas (asociado a CRISPR)

Expresión transitoria de Cas a partir de un vector plasmídico, liberación directa de proteína Cas y/o liberación directa de ARNm Cas en células vegetales. Los genes Cas tienen un codón optimizado para su expresión en plantas superiores, algas o levaduras, y están controlados por un promotor constitutivo, inducible, específico de tejido o específico de especie, según proceda. Las señales de terminación y poliadenilación del transcrito Cas son NosT, RBCT, HSP18.2T u otros terminadores específicos de genes o especies. Los casetes del gen Cas o el ARNm pueden contener intrones, ya sean nativos o en combinación con promotores específicos del gen y/o promotores sintéticos. La proteína Cas puede contener una o más secuencias señal de localización nuclear (NLS), mutaciones, delecciones, alteraciones o truncamientos. En los sistemas de expresión transitoria, los casetes del gen Cas pueden combinarse

con otros componentes del sistema CRISPR-Cas, tal como los casetes de ARNg, en el mismo vector de expresión transitoria. Alternativamente, los casetes del gen Cas pueden localizarse y expresarse a partir de constructos independientes de los casetes de ARNg o de otros componentes del sistema CRISPR-Cas. Gen asociado a CRISPR (Cas): codifica para proteínas con diversas actividades previstas de manipulación de ácidos nucleicos, tales como nucleasas, helicasas y polimerasas. Los genes Cas incluyen el cas9. Cas9 es un gen que codifica una proteína de gran tamaño que contiene dominios predichos de tipo RuvC y endonucleasas HNH, y está asociado al sistema de inmunidad adaptativa CRISPR, presente en la mayoría de las arqueas y en muchas bacterias. La proteína Cas9 consta de dos lóbulos:

1) Lóbulo de reconocimiento (REC)- consta de tres dominios:

- a) BH (hélice puente)
- b) REC1- facilita la orientación del ADN guiada por ARN
- c) REC2- facilita la orientación del ADN guiada por ARN

2) Lóbulo de la nucleasa (NUC)- consta de tres dominios:

- a) RuvC- facilita la orientación del ADN guiada por ARN; actividad endonucleasa
- b) HNH - actividad endonucleasa
- c) Interacción PI-PAM

En otras realizaciones, el gen cas puede ser un homólogo de cas9 en donde los dominios RuvC, HNH, REC y BH están altamente conservados. En algunas realizaciones, los genes cas son los de las siguientes especies.

ARN guía (ARNg)

El ARNg o ARNsg (ARN guía único) se diseña como una fusión entre un ARNcr y parte de la secuencia del ARN transactivador CRISPR (ARNtracr), que guía el Cas9 a una secuencia de ADN objetivo específica que es complementaria a la región protoespaciador. El ARN guía puede incluir un casete de expresión que contenga un diseño de ARN quimérico con un híbrido de ARN trazador largo, un híbrido de ARN trazador corto o una conformación nativa de arreglo de CRISPR + ARNtracr. El ARNg quimérico combina la especificidad del ARNcr con las propiedades de andamiaje del ARNtracr en un único transcrito. La expresión transitoria del ARNg está controlada por promotores de ARN polimerasa III específicos de plantas superiores, tal como los de la familia de genes de ARNsn U6 o U3 (Wang et al 2008). La terminación del transcrito del ARNg está controlada por un tracto de 6-20 nucleótidos de poli dT, de acuerdo con Wang et al 2008. Los casetes de expresión del ARNg están ubicados en los mismos o diferentes vectores de expresión transitoria de otros componentes del sistema CRISPR-Cas. Los transcritos del ARNg pueden sintetizarse *in vitro* y administrarse directamente a las células vegetales, independientemente de los vectores de expresión transitoria del ARNg o en combinación con ellos.

Región objetivo

Los ARN guía contienen dos componentes que definen la especificidad a una región objetivo del ADN, un protoespaciador y un motivo adyacente al protoespaciador (PAM). La secuencia protoespaciadora, normalmente de 20 nucleótidos pero que puede variar en función del objetivo de ADN, proporciona especificidad de secuencia de ADN para el complejo CRISPR-Cas. Los objetivos de ADN también contienen una secuencia de trinucleótidos NNG o NAG (PAM), donde N denota cualquier nucleótido, inmediatamente 3' o corriente abajo del protoespaciador.

Enfoque de un componente

De forma similar a Le Cong *et al.* (2013) y otros, un "enfoque de un componente" simplificado para la edición de genes CRISPR-Cas, en donde un único constructo de expresión transitoria contiene todos los componentes del complejo CRISPR-Cas, es decir, tanto el ARNg como los casetes de expresión Cas. Esto permite un diseño modular sencillo para dirigirse a uno o varios loci en cualquier planta o cultivo. La selección de múltiples loci puede lograrse simplemente intercambiando los casetes de ARNg específicos del objetivo. Además, los promotores específicos de cada especie, los terminadores u otros elementos potenciadores de la expresión pueden introducirse y extraerse fácilmente de los vectores transitorios de "enfoque de un componente", lo que permite una expresión óptima tanto del ARNg como de la proteína Cas de forma específica para cada especie.

Enfoque de dos componentes

En el enfoque de dos componentes, los casetes de expresión de Cas y ARNg se ubican en diferentes vectores de expresión transitoria. Esto permite la entrega de componentes de edición CRISPR-Cas por separado, permitiendo diferentes proporciones de ARNg a Cas dentro de la misma célula. Al igual que el enfoque de un componente, el enfoque de dos componentes también permite alterar fácilmente los promotores, terminadores u otros elementos que

afectan a la expresión de los componentes CRISPR-Cas y permite dirigir el ADN de forma específica para cada especie.

Antibióticos

Otra clase de endonucleasas son los antibióticos que son glicopéptidos que escinden el ADN, tal como la familia de antibióticos de la bleomicina que son glicopéptidos que escinden el ADN que incluyen bleomicina, zeocina, fleomicina, tallysomicina, pepleomicina y otros que se describen en el presente documento.

En la recombinación génica dirigida pueden utilizarse otras moléculas modificadoras del ADN. Por ejemplo, los ácidos nucleicos peptídicos pueden utilizarse para inducir modificaciones en el genoma de la célula o células objetivo (véase, por ejemplo, Ecker, Pat. de EE. UU. No. 5,986,053). En sumario, se utilizan nucleótidos sintéticos que comprenden, al menos, un esqueleto peptídico parcial para dirigirse a una secuencia nucleotídica genómica homóloga. Al unirse al ADN doble helicoidal, o a través de un mutágeno ligado al ácido nucleico peptídico, se induce la modificación de la secuencia de ADN objetivo y/o la recombinación. La especificidad del objetivo viene determinada por el grado de homología entre la secuencia objetivo y la secuencia genómica.

Además, la presente divulgación no se limita a los procedimientos particulares que se utilizan en el presente documento para ejecutar la modificación de secuencias genómicas. De hecho, se contemplan varios procedimientos. Por ejemplo, los genes pueden dirigirse utilizando oligonucleótidos formadores de triple hélice (TFO). Los TFO pueden generarse sintéticamente, por ejemplo, por PCR o mediante el uso de un aparato sintetizador de genes. Además, los TFO pueden aislarse del ADN genómico si se encuentran secuencias naturales adecuadas. Los TFO pueden utilizarse de varias maneras, incluyendo, por ejemplo, mediante la unión a un mutágeno tal como, pero no limitado a, psoraleno o clorambucilo (véase, por ejemplo, Havre et al., Proc Natl Acad Sci, U.S.A. 90:7879-7883, 1993; Havre et al., J Virol 67:7323-7331, 1993; Wang et al., Mol Cell Biol 15:1759-1768, 1995; Takasugi et al., Proc Natl Acad Sci, U.S.A. 88:5602-5606, 1991; Belousov et al., Nucleic Acids Res 25:3440-3444, 1997). Además, por ejemplo, los TFO pueden estar unidos a ADN dúplex donante (véase, por ejemplo, Chan et al., J Biol Chem 272:11541-11548, 1999). Los TFO también pueden actuar uniéndose con suficiente afinidad para provocar una reparación propensa a errores (Wang et al., Science 271:802-805, 1996).

Los procedimientos divulgados en el presente documento no están necesariamente limitados a la naturaleza o tipo de reactivo modificador del ADN que se utilice. Por ejemplo, estos reactivos modificadores del ADN liberan radicales que provocan la rotura de la cadena de ADN. Alternativamente, los reactivos alquilan el ADN para formar aductos que bloquearían la replicación y la transcripción. En otra alternativa, los reactivos generan enlaces cruzados o moléculas que inhiben las enzimas celulares que conducen a roturas de cadena. Ejemplos de reactivos modificadores del ADN que se han unido a oligonucleótidos para formar TFO incluyen, entre otros, indolocarbazoles, diimida de naftaleno (NDI), transplatino, bleomicina, análogos del ciclopropapirrolindol y fenantridioxinas. En particular, los indolocarbazoles son inhibidores de la topoisomerasa I. La inhibición de estas enzimas da lugar a roturas de la cadena y a la formación de aductos proteicos del ADN (Arimondo et al., Bioorganic and Medicinal Chem. 8, 777, 2000). El NDI es un fotooxidante que puede oxidar las guaninas, lo que podría causar mutaciones en sitios de residuos de guanina (Nunez, et al., Biochemistry, 39, 6190, 2000). Se ha demostrado que el transplatino reacciona con el ADN en una objetivo triplex cuando el TFO está unido al reactivo. Esta reacción provoca la formación de aductos de ADN que serían mutagénicos (Columbier, et al., Nucleic Acids Research, 24: 4519, 1996). La bleomicina es un rompedor del ADN, ampliamente utilizado como mimético de la radiación. Se ha unido a oligonucleótidos y se ha demostrado su actividad como rompedor en ese formato (Sergeyev, Nucleic Acids Research 23, 4400, 1995; Kane, et al., Biochemistry, 34, 16715, 1995). Se han unido análogos del ciclopropapirrolindol a los TFO y se ha demostrado que alquilan el ADN en una secuencia objetivo triplex. El ADN alquilado contendría entonces aductos químicos que serían mutagénicos (Lukhtanov, et al., Nucleic Acids Research, 25, 5077, 1997). Las fenantodihidrodioxinas son quinonas enmascaradas que liberan especies radicales tras la fotoactivación. Se han vinculado a los TFO y se ha demostrado que introducen roturas en el ADN dúplex en la fotoactivación (Bendinskas et al., Bioconjugate Chem. 9, 555, 1998).

La presente divulgación contempla otros procedimientos para inducir modificaciones y/o recombinación. Por ejemplo, otro ejemplo divulgado en el presente documento implica la inducción de recombinación homóloga entre un fragmento de ADN exógeno y el gen objetivo (véase, por ejemplo, Capecchi et al., Science 244:1288-1292, 1989) o mediante el uso de ácidos nucleicos peptídicos (PNA) con afinidad por el sitio objetivo. Otros procedimientos incluyen el reconocimiento del ADN específico de la secuencia y la orientación mediante poliamidas (véase, por ejemplo, Dervan et al, Dervan et al., Curr Opin Chem Biol 3:688-693, 1999; Biochemistry 38:2143-2151, 1999) y el uso de nucleasas con actividad específica (por ejemplo, proteínas de dedos de zinc, TALENs, Meganucleasas y/o CRISPRs).

La presente divulgación no se limita a ninguna frecuencia particular de modificación y/o recombinación. En algunos ejemplos, los procedimientos divulgados en el presente documento dan como resultado una frecuencia de modificación en la secuencia nucleotídica objetivo del 0,2% al 3%. No obstante, cualquier frecuencia (es decir, entre 0% y 100%) de modificación y/o recombinación se contempla dentro del alcance de la presente divulgación. La frecuencia de modificación y/o recombinación depende del procedimiento utilizado para inducir la modificación y/o recombinación, del tipo de célula utilizada, del gen específico al que se dirige y del reactivo de mutación del ADN utilizado, en su caso. Además, el procedimiento utilizado para detectar la modificación y/o recombinación, debido a limitaciones en el

procedimiento de detección, puede no detectar todas las ocurrencias de modificación y/o recombinación. Además, algunos eventos de modificación y/o recombinación pueden ser silenciosos, no dando ninguna indicación detectable de que la modificación y/o recombinación ha tenido lugar. La incapacidad de detectar modificaciones silenciosas y/o eventos de recombinación da una estimación artificialmente baja de modificación y/o recombinación. Por estas y otras razones, la divulgación no se limita necesariamente a ninguna modificación y/o frecuencia de recombinación en particular. En un ejemplo, la frecuencia de modificación y/o recombinación está comprendida entre el 0,01% y el 100%. En otro ejemplo, la frecuencia de modificación y/o recombinación está comprendida entre el 0,01% y el 50%. En otro ejemplo, la frecuencia de modificación y/o recombinación está comprendida entre el 0,1% y el 10%. En otro ejemplo, la frecuencia de modificación y/o recombinación está comprendida entre el 0,1% y el 5%.

El término "frecuencia de mutación", tal como se utiliza en el presente documento en referencia a una población de células tratadas con una molécula modificadora del ADN capaz de introducir una mutación en un sitio objetivo del genoma de las células, se refiere al número de células de la población tratada que contienen la mutación en el sitio objetivo en comparación con el número total de células tratadas con la molécula modificadora del ADN. Por ejemplo, con respecto a una población de células tratadas con la molécula modificadora del ADN TFO unida al psoraleno, diseñada para introducir una mutación en un sitio objetivo del genoma de las células, una frecuencia de mutación del 5% significa que de un total de 100 células tratadas con TFO-psoraleno, 5 células contienen una mutación en el sitio objetivo.

Aunque la presente divulgación no está necesariamente limitada a cualquier grado de precisión en la modificación y/o recombinación del ADN en la célula, se contempla que algunas realizaciones de la presente divulgación requieran mayores grados de precisión, dependiendo del resultado deseado. Por ejemplo, los cambios de secuencia específicos necesarios para la reparación del gen (por ejemplo, cambios de base particulares) requieren un mayor grado de precisión en comparación con la producción de un gen inactivo en donde sólo es necesaria la interrupción del gen. Con los procedimientos de la presente divulgación, el logro de mayores niveles de precisión en las técnicas de modificación y/o recombinación homóloga es mayor que con los procedimientos de la técnica anterior.

Distribución de oligonucleobases reparadoras de genes en células vegetales

Cualquier procedimiento comúnmente conocido utilizado para transformar una célula vegetal puede utilizarse para entregar las oligonucleobases reparadoras de genes. A continuación, se enumeran procedimientos ilustrativos. Los procedimientos y composiciones descritos en el presente documento pueden implicar cualquiera de los muchos procedimientos para transfectar las células con el reactivo o reactivos modificadores del ADN. Los procedimientos para la introducción de reactivos modificadores del ADN en una célula o células son bien conocidos en la técnica e incluyen, entre otros, la microinyección, la electroporación, la adsorción pasiva, la coprecipitación de fosfato cálcico-ADN, la transfección mediada por DEAE-dextrano, la transfección mediada por polibreno, la fusión de liposomas, la lipofectina, la nucleofección, la fusión de protoplastos, la infección retroviral, la biolística (es decir, el bombardeo de partículas) y similares.

El uso de microportadores metálicos (microesferas) para introducir grandes fragmentos de ADN en células vegetales con paredes celulares de celulosa mediante la penetración de proyectiles es bien conocido por los expertos en la técnica (en adelante, entrega biolística). Pat. de EE. UU. Nos. 4.945.050; 5,100,792 y 5,204,253 describen técnicas generales de selección de microportadores y dispositivos para proyectarlos.

Las condiciones específicas para el uso de microportadores en los procedimientos divulgados en el presente documento pueden incluir las condiciones descritas en la Publicación Internacional WO 99/07865. En una técnica ilustrativa, se añaden microportadores helados (60 mg/mL), oligonucleótidos dúplex mezclados (60 mg/mL) 2,5 M CaCl_2 y 0,1 M espermidina en ese orden; la mezcla se agita suavemente, por ejemplo, mediante vórtex, durante 10 minutos y luego se deja a temperatura ambiente durante 10 minutos, tras lo cual los microportadores se diluyen en 5 volúmenes de etanol, se centrifugan y se resuspenden en etanol al 100%. Pueden obtenerse buenos resultados con una concentración en la solución adherente de 8-10 $\mu\text{g}/\mu\text{L}$ de microportadores, 14-17 $\mu\text{g}/\text{mL}$ de oligonucleótido dúplex mixto, 1,1-1,4 M CaCl_2 y 18-22 mM de espermidina. Los resultados óptimos se observaron en las condiciones de 8 $\mu\text{g}/\mu\text{L}$ de microportadores, 16,5 $\mu\text{g}/\text{mL}$ de oligonucleótido dúplex mixto, 1,3 M CaCl_2 y 21 mM de espermidina.

Las oligonucleobases reparadoras de genes también pueden introducirse en células vegetales utilizando microfibras para penetrar la pared celular y la membrana celular. Pat. de EE. UU. No. 5,302,523 a Coffee et al describe el uso de fibras de carburo de silicio para facilitar la transformación de cultivos de maíz en suspensión de Black Mexican Sweet. Cualquier técnica mecánica que pueda utilizarse para introducir ADN para la transformación de una célula vegetal mediante microfibras puede utilizarse para suministrar oligonucleobases reparadoras de genes para su transmutación.

Una técnica ilustrativa para la entrega por microfibras de una oligonucleobase de reparación génica es la siguiente: Las microfibras estériles (2 μg) se suspenden en 150 μL de medio de cultivo vegetal que contiene aproximadamente 10 μg de un oligonucleótido dúplex mezclado. Se deja sedimentar un cultivo en suspensión y se mezclan en vórtex volúmenes iguales de células empaquetadas y de la suspensión estéril de fibras/nucleótidos durante 10 minutos y se colocan en placas. Los medios selectivos se aplican inmediatamente o con un retraso de hasta aproximadamente 120 h, según convenga para el rasgo concreto.

En una realización alternativa, las oligonucleobases reparadoras de genes pueden entregarse a la célula vegetal por electroporación de un protoplasto derivado de una parte de la planta. Los protoplastos se forman mediante tratamiento enzimático de una parte de la planta, en particular una hoja, de acuerdo con técnicas bien conocidas por los expertos en la técnica. Véase, por ejemplo, Gallois et al., 1996, en *Methods in Molecular Biology* 55:89-107, Humana Press, Totowa, N.J.; Kipp et al., 1999, en *Methods in Molecular Biology* 133:213-221, Humana Press, Totowa, NJ. No es necesario cultivar los protoplastos en medios de crecimiento antes de la electroporación. Las condiciones ilustrativas para la electroporación son 300.000 protoplastos en un volumen total de 0,3 mL con una concentración de oligonucleobasa reparadora de genes de entre 0,6-4 µg/mL.

En una realización alternativa, los protoplastos vegetales absorben los ácidos nucleicos en presencia del agente modificador de membrana polietilenglicol, de acuerdo con técnicas bien conocidas por los expertos en la técnica. En otra realización alternativa, las oligonucleobases reparadoras de genes pueden administrarse inyectándolas con un microcapilar en células vegetales o en protoplastos.

En una realización alternativa, los ácidos nucleicos se incrustan en microesferas compuestas de alginato de calcio y son absorbidos por protoplastos de plantas en presencia del agente modificador de membrana polietilenglicol (véase, por ejemplo, Sone et al., 2002, Liu et al., 2004).

En una realización alternativa, los ácidos nucleicos congelados en agua e introducidos en las células vegetales mediante bombardeo en forma de micropartículas (véase, por ejemplo, Gilmore, 1991, Patente de EE. UU. 5,219,746; Brinegar et al.).

En una realización alternativa, los ácidos nucleicos unidos a nanopartículas se introducen en células vegetales intactas incubando las células en una suspensión que contiene la nanopartícula (véase, por ejemplo, Pasupathy et al., 2008) o introduciéndolos en células intactas mediante bombardeo de partículas o en protoplastos mediante coincubación (véase, por ejemplo, Torney et al., 2007).

En una realización alternativa, los ácidos nucleicos complejados con péptidos penetrantes y administrados a las células mediante coincubación (véase, por ejemplo, Chugh et al., 2008, WO 2008148223 A1 eudes y Chugh).

En una realización alternativa, los ácidos nucleicos se introducen en células intactas mediante electroporación (véase, por ejemplo, He et al., 1998, US 2003/0115641 A1 Dobres et al.).

En una realización alternativa, los ácidos nucleicos se introducen en las células de embriones secos sumergiéndolas en una solución con ácidos nucleicos (véase, por ejemplo, Töpfer et al., 1989, Senaratna et al., 1991) o en otras realizaciones se introducen mediante Cellsqueeze (SQZ Biotech).

Selección de plantas

En diversas realizaciones, las plantas divulgadas en el presente documento pueden ser de cualquier especie de planta dicotiledónea, monocotiledónea o gimnosperma, incluyendo cualquier especie de planta leñosa que crezca como árbol o arbusto, cualquier especie herbácea, o cualquier especie que produzca frutos comestibles, semillas o vegetales, o cualquier especie que produzca flores coloridas o aromáticas. Por ejemplo, la planta puede seleccionarse de una especie vegetal del grupo que consiste en canola, girasol, maíz, tabaco, remolacha azucarera, algodón, maíz, trigo, cebada, arroz, alfalfa, cebada, sorgo, tomate, mango, melocotón, manzana, pera, fresa, plátano, melón, mandioca, patata, zanahoria, lechuga, cebolla, soja, soja spp, caña de azúcar, guisante, garbanzo, guisante forrajero, haba, lenteja, nabo, colinabo, coles de Bruselas, altramuza, coliflor, col rizada, habas, álamo, pino, eucalipto, uva, cítricos, triticale, alfalfa, centeno, avena, césped y gramíneas forrajeras, lino, colza, mostaza, pepino, gloria de la mañana, bálsamo, pimienta, berenjena, caléndula, loto, col, margarita, clavel, tulipán, iris, lirio y plantas productoras de frutos secos, en la medida en que no estén ya específicamente mencionadas.

Las plantas y células vegetales pueden ser probadas para resistencia o tolerancia a un herbicida usando procedimientos comúnmente conocidos en la técnica, por ejemplo, cultivando la planta o célula vegetal en presencia de un herbicida y midiendo la tasa de crecimiento en comparación con la tasa de crecimiento en ausencia del herbicida.

Tal como se utiliza en el presente documento, el crecimiento sustancialmente normal de una planta, órgano vegetal, tejido vegetal o célula vegetal se define como una tasa de crecimiento o tasa de división celular de la planta, órgano vegetal, tejido vegetal o célula vegetal que es al menos 35%, al menos 50%, al menos 60% o al menos 75% de la tasa de crecimiento o tasa de división celular en una planta, órgano vegetal, tejido vegetal o célula vegetal correspondiente que expresa la proteína de interés de tipo silvestre.

Tal como se utiliza en el presente documento, el desarrollo sustancialmente normal de una planta, órgano vegetal, tejido vegetal o célula vegetal se define como la aparición de uno o más eventos de desarrollo en la planta, órgano vegetal, tejido vegetal o célula vegetal que son sustancialmente iguales a los que ocurren en una planta, órgano vegetal, tejido vegetal o célula vegetal correspondiente que expresa la proteína de tipo silvestre.

En ciertas realizaciones, los órganos vegetales proporcionados en el presente documento incluyen, entre otros, hojas, tallos, raíces, yemas vegetativas, yemas florales, meristemas, embriones, cotiledones, endospermo, sépalos, pétalos, pistilos, carpelos, estambres, anteras, microsporas, polen, tubos polínicos, óvulos, ovarios y frutos, o secciones, rodajas o discos tomados de los mismos. Los tejidos vegetales incluyen, entre otros, tejidos de callo, tejidos del suelo, tejidos vasculares, tejidos de almacenamiento, tejidos meristemáticos, tejidos foliares, tejidos de brotes, tejidos radiculares, tejidos de agallas, tejidos de tumores vegetales y tejidos reproductivos. Las células vegetales incluyen, entre otras, células aisladas con paredes celulares, agregados de distintos tamaños y protoplastos.

Las plantas son sustancialmente "tolerantes" a un herbicida relevante cuando son sometidas a él y proporcionan una curva dosis/respuesta que se desplaza hacia la derecha cuando se compara con la proporcionada por una planta similar no tolerante sometida de forma similar. Dichas curvas de dosis/respuesta tienen la "dosis" trazada en el eje X y el "porcentaje de muerte", el "efecto herbicida", etc., trazados en el eje Y. Las plantas tolerantes necesitarán más herbicida que las plantas similares no tolerantes para producir un efecto herbicida determinado. Las plantas que son sustancialmente "resistentes" al herbicida presentan pocas lesiones necróticas, líticas, cloróticas o de otro tipo, si es que presentan alguna, cuando se someten al herbicida en las concentraciones y dosis empleadas habitualmente por la comunidad agroquímica para eliminar las malas hierbas en el campo. Las plantas resistentes a un herbicida también lo toleran.

Generación de plantas

Se conoce el cultivo de tejidos de diversas especies vegetales y la regeneración de plantas a partir de ellos. Por ejemplo, la propagación de un cultivar de colza mediante cultivo de tejidos se describe en cualquiera de las siguientes, pero sin limitarse a ellas: Chuong et al., "A Simple Culture Method for Brassica hypocotyls Protoplasts," Plant Cell Reports 4:4-6, 1985; Barsby, T. L., et al., "A Rapid and Efficient Alternative Procedure for the Regeneration of Plants from Hypocotyl Protoplasts of Brassica napus," Plant Cell Reports (Primavera, 1996); Kartha, K., et al., "In vitro Plant Formation from Stem Explants of Rape," Physiol. Plant, 31:217-220, 1974; Narasimhulu, S., et al., "Species Specific Shoot Regeneration Response of Cotyledonary Explants of Brassicas," Plant Cell Reports (Spring 1988); Swanson, E., "Microspore Culture in Brassica," Methods in Molecular Biology, Vol. 6, Capítulo 17, p. 159, 1990.

La reproducción ulterior de la variedad puede producirse por cultivo de tejidos y regeneración. El cultivo de tejidos de diversos tejidos de soja y la regeneración de plantas a partir de ellos son bien conocidos y han sido objeto de numerosas publicaciones. Por ejemplo, se puede hacer referencia a Komatsuda, T. et al., "Genotype X Sucrose Interactions for Somatic Embryogenesis in Soybeans," Crop Sci. 31:333-337, 1991; Stephens, P. A., et al., "Agronomic Evaluation of Tissue-Culture-Derived Soybean Plants," Theor. Appl. Genet. 82:633-635, 1991; Komatsuda, T. et al., "Maturation and Germination of Somatic Embryos as Affected by Sucrose and Plant Growth Regulators in Soybeans Glycine gracilis Skvortz and Glycine max (L.) Merr" Plant Cell, Tissue and Organ Culture, 28:103-113, 1992; Dhir, S. et al., "Regeneration of Fertile Plants from Protoplasts of Soybean (Glycine max L. Merr.); Genotypic Differences in Culture Response," Plant Cell Reports 11:285-289, 1992; Pandey, P. et al., "Plant Regeneration from Leaf and Hypocotyl Explants of Glycine wightii (W. and A.) VERDC. var. longicauda," Japan J. Breed. 42:1-5, 1992; Shetty, K., et al., "Stimulation of In Vitro Shoot Organogenesis in Glycine max (Merrill.) by Allantoin and Amides," Plant Science 81:245-251, 1992. Las divulgaciones de Pat. de US. No. 5,024,944 expedido el 18 de junio de 1991 a Collins et al. y Pat. de US. No. 5,008,200 expedido el 16 de abril de 1991 a Ranch et al.

EJEMPLOS

Los siguientes son ejemplos, que ilustran procedimientos para practicar los procedimientos y composiciones descritos en el presente documento. Estos ejemplos no deben interpretarse como limitativos.

Ejemplo 1 (no de acuerdo con la invención): Longitud del GRON

Sommer et al., (Mol Biotechnol. 33:115-22, 2006) describe un sistema reportero para la detección de la conversión génica *in vivo* que se basa en un único cambio de nucleótido para convertir entre fluorescencia azul y verde en variantes de la proteína verde fluorescente (GFP). Este sistema reportero se adaptó para su uso en los siguientes experimentos utilizando *Arabidopsis thaliana* como especie modelo con el fin de evaluar la eficiencia de la conversión de GRON tras la modificación de la longitud de GRON.

En sumario, para este ejemplo y los siguientes se creó una línea de *Arabidopsis thaliana* con múltiples copias de un gen de proteína fluorescente azul mediante procedimientos conocidos por los expertos en la técnica (véase, por ejemplo, Clough y Brent, 1998). Con esta línea se establecieron cultivos de tejidos meristemáticos derivados de la raíz, que se utilizaron para el aislamiento y cultivo de protoplastos (véase, por ejemplo, Mathur et al., 1995). El suministro de GRON a los protoplastos se consiguió mediante la absorción de GRON en los protoplastos mediada por polietilenglicol (PEG). Se utilizó un procedimiento en formato de 96 pocillos, similar al descrito por Fujiwara y Kato (2007). A continuación, se describe brevemente el protocolo. Los volúmenes indicados son los aplicados a pocillos individuales de una placa de 96 pocillos.

1. Mezclar 6,25 µl de GRON (80 µM) con 25 µl de protoplastos derivados de tejido meristemático de raíz transgénica de *Arabidopsis thaliana* BFP a 5×10^6 células/ml en cada pocillo de una placa de 96 pocillos.

2. Se añadieron 31.25 µl de una solución de PEG al 40% y se mezclaron los protoplastos.

3. Las células tratadas se incubaron en hielo durante 30 minutos.

4. A cada pocillo se añadieron 200 µl de solución W5 y se mezclaron las células.

5. Las placas se dejaron incubar en hielo durante 30 minutos para que los protoplastos se depositaran en el fondo de cada pocillo.

6. Se retiraron 200 µl del medio por encima de los protoplastos asentados.

7. Se añadieron 85 µl de medio de cultivo (MSAP, véase Mathur et al., 1995).

8. Las placas se incubaron a temperatura ambiente en la oscuridad durante 48 horas. La concentración final de GRON tras añadir el medio de cultivo es de 8 µM.

Cuarenta y ocho horas después de la administración de GRON las muestras fueron analizadas por citometría de flujo para detectar protoplastos cuya fluorescencia verde y amarilla es diferente de la de los protoplastos de control (BFP0 indica GRONs no-objetivo sin cambio en comparación con el objetivo BFP; C es el diseño de la cadena codificante y NC es el diseño de la cadena no codificante). Una única diferencia de nucleótidos de C a T (cadena codificante) o una mutación dirigida de nucleótidos de G a A (cadena no codificante) en el centro de las moléculas BFP4. La fluorescencia verde se debe a la introducción de una mutación dirigida en el gen BFP, que da lugar a la síntesis de GFP. Los resultados se muestran en la Figura 1.

La Tabla 2 muestra las secuencias de GRONs BFP4/NC 5'-3PS/ 3'-3PS de 101-mer y 201-mer ejemplares diseñados para la conversión de un gen de proteína azul fluorescente (BFP) a fluorescencia verde. "3PS" denota 3 enlaces fosfotioato en cada uno de los extremos 5' y 3' del oligo.

Tabla 2: Secuencias de nucleótidos GRON ejemplares para la conversión de BFP a GFP

Nombre GRON	Secuencia de nucleótidos de GRON
BFP4/NC 101-mer	G* T*C*G TGC TGC TTC ATG TGG TCG GGG TAG CGG CTG AAG CAC TGC ACG CCG TAG GTG AAG GTG GTC ACG AGG GTG GGC CAG GGC ACG GGC AGC TTG CCG G*T*G* G
BFPO/NC 101-mer	G* T*C*G TGC TGC TTC ATG TGG TCG GGG TAG CGG CTG AAG CAC TGC ACG CCG TGG GTG AAG GTG GTC ACG AGG GTG GGC CAG GGC ACG GGC AGC TTG CCG G*T*G *G
BFP4/C 101-mer	C *C*A*C CGG CAA GCT GCC CGT GCC CTG GCC CAC CCT CGT GAC CAC CTT CAC CTA CGG CGT GCA GTG CTT CAG CCG CTA CCC CGA CCA CAT GAA GCA GCA C*G*A *C
BFP0/C 101-mer	C*C*A*CCGGCAAGCTGCCCGTGCCCTGGCCCAACCCTCGTGACCAC CTTACCCACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATG AAGCAGCAC*G*A* C

Nombre GRON	Secuencia de nucleótidos de GRON
BFP4/NC 201-mer	A*A*G*ATGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACT TGAAGAAGTCGTGCTGCTTCATGTGGTCTGGGTAGCGGCTGAAGC ACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGGCCAGGGCA CGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGC CG TAGGTGGCATCGCCCTCG *C*C*C
BFP0/NC 201-mer	A*A*G*TGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACTT GAAGAAGTCGTGCTGCTTCATGTGGTTCGGGGTAGCGGCTGAAGCA CTGCACGCCGTGGGTGAAGGTGGTCACGAGGGTGGGCCAGGGCAC GGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCC G TAGGTGGCATCGCCCTCG *C*C*C
BFP4/C 201-mer	G*G*G*CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC ATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGA CCACCTTCACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCA CATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTA CGTCCAGGA GCGCACCAT *C*T*T
BFP0/C 201-mer	G*G*G*CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC ATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGA CCACCTTCACCCACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCA CATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTA CGTCCAGGA GCGCACCAT *C*T*T
* = enlace PS (fosfotioato)	

Ejemplo 2 (no de acuerdo con la invención): Tasas de conversión utilizando GRONs marcados con 5'Cy3/ 3'idC

5 El propósito de esta serie de ejemplos es comparar las eficiencias de los GRONs marcados con fosfotioato (PS) (que tienen 3 moléculas PS en cada extremo del GRON) con los GRONs marcados con 5'Cy3/ 3'idC. Los GRON marcados con 5'Cy3/ 3'idC tienen un fluoróforo 5' Cy3 (amidita) y una base inversa 3' idC. La eficacia se evaluó mediante la conversión de la proteína azul fluorescente (BFP) en fluorescencia verde.

10 En los tres ejemplos, realizados mediante la administración de GRON con PEG en protoplastos en tubos Falcon individuales (etiquetados como "Tubos") o en placas de 96 pocillos (etiquetadas como "Placa de 96 pocillos"), no hubo diferencias significativas entre las diferentes químicas de GRON en la eficiencia de conversión de BFP a GFP determinada por citometría (Figura 1).

15 Ejemplo 3 (no de acuerdo con la invención): Comparación entre los GRONs de 41-mer BFP4/NC 5'-3PS/ 3'-3PS y los GRONs 2'-O-Me

20 El propósito de esta serie de ejemplos es comparar las eficiencias de conversión de los GRONs marcados con fosfotioato (PS) con moléculas 3PS en cada extremo del GRON a 2'-O-Me o "GRONs de fragmentos Okazaki" en presencia y ausencia de un miembro de la familia de la bleomicina, Zeocina™ (1 mg/ml) para inducir roturas de ADN. Los diseños de estos GRON se representan en la Figura 2. Los GRON se introdujeron en protoplastos BFP de *Arabidopsis thaliana* mediante tratamiento con PEG y la conversión de BFP a GFP se determinó a las 24 h del tratamiento mediante citometría. Las muestras tratadas con zeocina (1 mg/ml) se incubaron con zeocina durante 90 min en hielo antes del tratamiento con PEG.

25 En general, la presencia de zeocina (1 mg/ml) incrementó la conversión de BFP a GFP determinada por citometría (Tabla 3). Tanto en presencia como en ausencia de zeocina, el GRON NC Okazaki que contenía un grupo 2'-O Me en la primera base de ARN en el extremo 5' del GRON fue más eficaz en la conversión de BFP a GFP en comparación con el GRON NC Okazaki que contenía un grupo 2'-O Me en cada una de las primeras nueve bases 5' de ARN (Figura 2 y Tabla 3).

En todos los ejemplos, no hubo diferencia significativa entre el 41-mer BFP4/NC 5'3PS/ 3'3PS y el 71-mer Fragmento Okazaki BFP4/NC GRON que contiene un grupo 5' 2'-O me en la primera base 5' de ARN (denotado como BFP4 71-mer (1) NC) en la conversión de BFP a GFP tanto en presencia como en ausencia de 1 mg/ml de zeocina como se determinó por citometría (Figura 2 y Tabla 3). Es importante señalar que en presencia de zeocina (y es de esperar para bleomicina, fleomicina, tallysomicina, pepleomicina y otros miembros de esta familia de antibióticos) esa conversión se vuelve independiente de la cadena (es decir, tanto los GRON codificantes (C) como los no codificantes (NC) con los diseños probados en estos ejemplos muestran una actividad aproximadamente igual).

Tabla 3: Comparación de un diseño GRON estándar con diseños GRON de fragmentos Okazaki en presencia y ausencia de un antibiótico glucopéptido zeocina.

Zeocina (+)								
Exp. Nombre	BFP4 41-mer		BFP4 71-mer (0)		BFP4 71-mer (1)		BFP4 71-mer (9)	
	NC	C	NC	C	NC	C	NC	C
APT043	0,13	0,0875	0,2275	0,2075	0,355	0,2275	0,2325	0,195
APT066	1,9	0,713	0,762	0,683	1,318	0,7103	0,769	0,883
Media	1,015	0,40025	0,49475	0,44525	0,8365	0,4689	0,50075	0,539
Desviación estándar	1,251579	0,442295	0,377949	0,336229	0,680944	0,341391	0,379363	0,486489
SE	0,885134	0312797	0,26729	0,237786	0,481573	0,241436	0,268291	0,344052
Zeocina (-)								
Exp. Nombre	BFP4 41-mer		BFP4 71-mer (0)		BFP4 71-mer (1)		BFP4 71-mer (9)	
	NC	C	NC	C	NC	C	NC	C
APT043	nd	nd	0,1875	0,0175	0,21	0,025	0,1	0,0225
APT066	0,109	0,007	0,112	0,005	0,141	0,023	0,065	0,021
Media	0,109	0,007	0,14975	0,01125	0,1755	0,024	0,0825	0,02175
Desviación estándar	na	na	0,053387	0,008839	0,04879	0,001414	0,024749	0,001061
SE	na	na	0,037756	0,006251	0,034505	0,001	0,017503	0,00075
BFP4 71-mer (0)		5' primeros 10bp son ARN y GRON no tiene protección						
NC	C							
BFP4 71-mer (1)		los primeros 10 pb 5' son ARN y el primer pb del extremo 5' tiene un O-Me 2'						
NC	C							
BFP4 71-mer (9)		los primeros 10 pb 5' son ARN y los primeros nueve pb en el extremo 5' tienen un 2' O-Me						
NC	C							

Ejemplo 4 (no de acuerdo con la invención): Comparación entre los GRONs BFP4/NC 5'-3PS/ 3'-3PS de 41-mer, 101-mer y 201-mer

El propósito de esta serie de ejemplos era comparar las eficiencias de conversión (en presencia y ausencia de zeocina) de los GRONs marcados con fosfotioato (PS) con moléculas 3PS en cada extremo del GRON de diferentes longitudes: 41-mer, 101-mer y 201-mer mostrados en la Tabla 2. De nuevo, la presencia de zeocina (1 mg/ml) incrementó las tasas de conversión de BFP a GFP determinadas por citometría (Tabla 4). La tendencia general en los tres ejemplos fue lineal con el aumento de la longitud de NC GRON tanto en presencia como en ausencia de zeocina. Excepto en el caso de la BFP-4/NC/101 y la BFP-4/C/101 en presencia de zeocina, se obtuvieron tasas de conversión casi iguales pero inferiores a las del GRON de 41-mer NC. Esto contrasta con todos los ejemplos anteriores en los que se utilizaron los GRON codificantes y no codificantes BFP-4/41, en los que el no codificante siempre fue muy superior al GRON codificante. Esta asimetría en la frecuencia de conversión también se aplica a los GRON BFP-4/201 utilizados en esta serie de ejemplos.

Tabla 4:

Zeocina (+)						
Exp. Nombre	BFP4 41-mer		BFP4 101-mer		BFP4 201-mer	
	NC	C	NC	C	NC	C
APT038	0,2425	0,1275	0,3025	0,2575	0,97	0,245
APT043	0,13	0,0875	0,185	0,2275	0,66	0,1875
APT047	0,3975	0,145	0,19	0,125	0,235	0,085
APT052	0,3275	nd	0,17	0,21	0,585	0,225
APT052	nd	nd	0,3225	0,3175	0,5075	0,3125
APT058	1,4275	nd	1,2	nd	1,9	nd
APT066	1,9	0,713	0,992	1,05	1,7	0,916
Media	0,7375	0,26825	0,480286	0,364583	0,936786	0,3285
Desviación estándar	0,7382801	0,297475	0,428968	0,341634	0,630943	0,297412
SE	0,30146186	0,148738	0,162119	0,139499	0,238452	0,121442
APT038	0,05	0,01	0,1025	0,025	0,5725	0,025
APT066	0,109	0,007	0,214	0,047	0,566	0,035
Media	0,0795	0,0085	0,15825	0,036	0,56925	0,03
Desviación estándar	0,0417193	0,002121	0,078842	0,015556	0,004596	0,007071
SE	0,02950446	0,0015	0,055758	0,011002	0,00325	0,005001

Ejemplo 5 (no de acuerdo con la invención): Introducción de la proteína Cas9 en las plantas

- 5 Este ejemplo hace uso de la entrega directa de proteína Cas9 recombinante a células vegetales como alternativa a la entrega de plásmidos de expresión CRISPR-Cas. Este procedimiento emplea portadores tales como péptidos penetrantes de células (CPP), reactivos liposomales de transfección, polietilenglicol (PEG) ya sea solos o en combinación para permitir la entrega de la proteína Cas9 recombinante activa a las células.

10 Procedimientos

- Los protoplastos transgénicos de *Arabidopsis thaliana* BFP derivados de tejido radicular inducido se siembran en una placa de 96 pocillos de fondo plano a 250.000 células por pocillo a una densidad celular de 1×10^7 células/ml. La proteína Cas9 recombinante marcada con fluorescencia (1 μ g) se recubre previamente con CPP a 20:1, 10: o 5:1 y otras relaciones CPP/carga (TAT, Penetratin, Chariot™, PEP-1 u otras, por ejemplo) o se encapsula con reactivos liposomales y se mezcla con los protoplastos sembrados y se incuba a 23 °C durante 2-6 h para permitir que los complejos Cas9/transportador penetren en las células. En otra serie de tratamientos se introduce en los protoplastos la proteína Cas9 recombinante marcada con fluorescencia (1 μ g), ya sea pre-recubierta con CPPs como se ha descrito anteriormente o no recubierta, utilizando la metodología PEG. A continuación, se analizaron los protoplastos mediante citometría de flujo 24 h después del tratamiento para determinar el porcentaje de protoplastos Cas9 positivos dentro de un tratamiento determinado.

Ejemplo 6: CRISPR con GRONs de 201-mer $\pm$ base de cabeceo

- 25 El propósito de esta serie de ejemplos es demostrar la conversión de BFP a GFP en nuestro sistema modelo transgénico de BFP de *Arabidopsis thaliana* utilizando CRISPRs para crear roturas de doble cadena en el gen *bfp* y los GRONs de 201-mer para mediar la conversión. El CRISPR BFP se dirige a la cadena codificante del gen *bfp* y el sitio de conversión se encuentra 27 pb aguas arriba de la secuencia PAM (Figura 3). El GRON se utiliza como plantilla para reparar la rotura de doble cadena en el gen *bfp* creada por la CRISPR y, además de convertir el gen objetivo de BFP a GFP, introduce una base de cabeceo en el gen *bfp* que corresponde también a la secuencia PAM del CRISPR BFP. Se cree que una base de cabeceo en la secuencia PAM del CRISPR BFP minimiza la recortadura del gen *bfp* por la CRISPR una vez que se ha producido la conversión. Esta serie de ejemplos ayudará a determinar si la introducción de una base de cabeceo en la secuencia PAM del CRISPR BFP en genes *bfp* convertidos aumentará o no las eficiencias de conversión.

35 Procedimientos

- Los protoplastos transgénicos de *Arabidopsis thaliana* BFP derivados de tejido radicular inducido se sembraron en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1×10^7 células/ml. Los plásmidos codificados CRISPR contienen el promotor de la manopina sintasa (MAS) que impulsa la secuencia

codificadora Cas9 con un terminador rbcSE9 y el promotor U6 de *Arabidopsis* que impulsa el ARNsg con un terminador poly-T10. Los plásmidos CRISPR junto con GRON se introdujeron en los protoplastos mediante entrega mediada por PEG a una concentración final de 0,05 µg/µl y 0,16 µM respectivamente. Los protoplastos se incubaron en la oscuridad a 23 °C durante 72 horas y, a continuación, se analizaron mediante citómetro de flujo para determinar el porcentaje de protoplastos positivos para GFP dentro de un tratamiento determinado.

La CRISPR consta de dos componentes: la planta de codón optimizado *Streptococcus pyogenes* Cas9 (SpCas9) y ARNsg, ambos expresados a partir del mismo plásmido. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar la nucleasa Cas9 hasta el gen objetivo BFP. En este ejemplo, la BFP se dirige al gen *bfp* (Figura 3). Se utilizaron GRONs de 201-mer dirigidos a BFP con o sin bases de cabeceo para determinar su efecto sobre la tasa de conversión de BFP a GFP. La Tabla 5 da una lista de los GRON y sus secuencias correspondientes.

Tabla 5. Lista de los GRON utilizados en estos ejemplos

Nombre GRON	GRON Química	Secuencia GRON
BFP4/NC 201-mer	3PS	5'AAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTTCATGTGCTCTGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGAGATGAACCTCAGGGTCAGCTTGCCGTAGGTGGCATGCCCTCGCCC3'
BFP4/C 201-mer	3PS	5'GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACACCGGCAAGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACTTCACCTACGGCGTGCACTGCTTCAGCCGCTACCCGACCATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGACCATCTT3'
BFP4/NC 201-mer (1 cabeceo)	3PS	5' AAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGCTCTGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGAGATGAACCTCAGGGTCAGCTTGCCGTAGGTGGCATGCCCTCGCCC 3'
BFP4/C 201-mer (1 cabeceo)	3PS	5' GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACACCGGCAAGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACTTCACCTACGGCGTGCACTGCTTCAGCCGCTACCCGACCATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGACCATCTT 3'

Resultados

Usando el CRISPR BFP, el GRON BFP4/C con las bases de cabeceo es hasta 3.5 veces más eficaz en la conversión de BFP a GFP cuando se compara con el GRON BFP4/C sin las bases de cabeceo (Tabla 6). Hay un aumento de hasta 5,9 veces en la conversión de BFP a GFP cuando se utiliza el GRON BFP4/C con la base tambaleante en lugar del GRON BFP4/NC con la base de cabeceo (Tabla 6). Por lo tanto, el GRON BFP4/C con la base de cabeceo es más eficaz en la conversión de BFP a GFP cuando se utiliza con el CRISPR BFP.

Conclusiones

La inclusión de una base de cabeceo en el GRON que cambia la secuencia PAM del CRISPR BFP en el gen objetivo convertido aumenta la conversión de BFP a GFP. La conversión de BFP a GFP mediante el CRISPR BFP junto con el GRON basado en cabeceo se confirmó mediante secuenciación de próxima generación (datos no mostrados). Además, la capacidad del CRISPR BFP para escindir el ADN y producir indeles en el gen *bfp* fue confirmada por Next Generation Sequencing (datos no mostrados).

Tabla 6. Porcentaje de conversión de BFP a GFP determinado por citometría a las 72 h tras la administración por PEG de la BFP CRISPR y GRON en protoplastos derivados de la línea transgénica de BFP 21-15-09 de *Arabidopsis thaliana*.

Exp. Nombre	Porcentaje de células positivas para GFP (desviación estándar)			
	CRISPR: BFP5			
	BFP4/C GRON		BFP4/NC GRON	
	(-) Cabeceos	(+) 1 Cabeceo	(-) Cabeceos	(+) 1 Cabeceo
Exp 1	0,46(0,07)	1,59(0,06)	0,08(0,02)	0,27(0,04)
Exp 2	0,24(0,03)	0,61(0,05)	0,04(0,01)	0,16(0,04)

Ejemplo 7: CRISPR con GRONs modificados con Cy3

El propósito de esta serie de ejemplos es demostrar la conversión de BFP a GFP en nuestro sistema modelo transgénico de BFP de *Arabidopsis thaliana* utilizando CRISPRs para crear roturas de doble cadena en el gen *bf*p y GRONs para mediar la conversión. El CRISPR BFP6 (CR:BFP6) utilizado en estos ejemplos se dirige al gen *bf*p y provoca una rotura de doble cadena en el ADN cerca del lugar de conversión. Los GRONs utilizados con el CRISPR BFP6, contienen la secuencia codificante del gen *bf*p alrededor del sitio de conversión y están etiquetados en el extremo 5' con Cy3 y en el extremo 3' con una base inversa idC y se denominan en el presente documento GRONs Cy3. Estos GRONs se prueban en tres longitudes diferentes de 41-mers, 101-mers y 201-mers y se comparan directamente con los GRONs 3PS que sólo difieren de los GRONs Cy3 en que tienen 3 enlaces fosforotato en los extremos 5' y 3' del GRON. Estos GRON se denominan en lo sucesivo GRON 3PS. Véase en la Tabla 7 la lista de GRON utilizados en estos ejemplos.

Procedimientos

Los protoplastos transgénicos de *Arabidopsis thaliana* BFP derivados de tejido radicular inducido se sembraron en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1x10⁷ células/ml. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador *rbcSE9* y el promotor U6 de *Arabidopsis thaliana* que impulsa el ARNsg con un terminador poly-T10. Los plásmidos CRISPR junto con GRON se introdujeron en protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para la CRISPR y 8,0 µM para los GRONs de 41-mer, 0,32 µM para los GRONs de 101-mer y 0,16 µM para los GRONs de 201-mer. Los tratamientos con GRON solo recibieron una concentración final de GRON tras la administración de PEG de 8,0 µM para el 41-mer, 5,0 µM para el 101-mer y 2,5 µM para el 201-mer. Los protoplastos se incubaron en la oscuridad a 23 °C durante 72 horas y, a continuación, se analizaron mediante citómetro de flujo para determinar el porcentaje de protoplastos positivos para GFP dentro de un tratamiento determinado.

La CRISPR consta de dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo plásmido. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar la nucleasa Cas9 hasta el gen objetivo BFP. En este experimento se utilizó el CRISPR BFP6 (5'GGTGCCGCACGTCACGAAGTCGG 3') dirigido al gen *bf*p. Los GRON contienen la secuencia codificadora del gen *bf*p cerca del lugar de conversión. La Tabla 7 ofrece una lista de los GRON utilizados.

Tabla 7. Lista de los GRON utilizados en estos ejemplos

Nombre GRON	GRON Química	Secuencia GRON	CRISPR
BFP4/C 41-mer	Cy3 o 3PS	5' CCCTCGTGACCACCTTCACCTACGGCGTGCAAGTGCCTTCAGC 3'	BFP6
BFP4/C 101-mer	Cy3 o 3PS	5'CCACCGGCAAGCTGCCCCGTGCCCTGGCCACCTCGTGACCACCTTCACCTACGGCGTGCAAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCACGAC 3'	BFP6
BFP4/C 201-mer	Cy3 o 3PS	5'GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCCCTGGCCACCTCGTGACCACCTTCACCTACGGCGTGCAAGTGCCTTCAGCCGCTACCCCGACCATGAAGCAGCACGACTTCTCAAGTCCGCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTT3'	BFP6

Resultados

Usando el CRISPR BFP6, los GRONs Cy3 en todas las longitudes probadas son capaces de mediar la conversión de BFP a GFP tan eficientemente como los GRONs 3PS (Figura 4). En general, las muestras que contienen la CRISPR

BFP6 y GRON tienen niveles más altos de conversión de BFP a GFP en comparación con las muestras que sólo contienen GRON (Figura 4), lo que demuestra el impacto positivo que tienen las CRISPR en el aumento de las tasas de conversión.

5 Ejemplo 8: CRISPR con GRONs de tamaño variable

El propósito de esta serie de ejemplos es demostrar la conversión de BFP a GFP en nuestro sistema modelo transgénico de BFP de *Arabidopsis thaliana* utilizando CRISPRs para crear roturas de doble cadena en el gen *bfp* y GRONs de diferentes longitudes para mediar la conversión. El CRISPR BFP utilizado en estos ejemplos se dirige al gen *bfp* y provoca una rotura de doble cadena en el ADN cerca del lugar de conversión. Los GRONs utilizados con el CRISPR BFP, contienen la secuencia codificante del gen *bfp* aproximadamente sitio de conversión y están etiquetados tanto en el extremo 5' como en el extremo 3' con 3 enlaces fosfotioato y se denominan aquí GRONs 3PS. Estos GRON se prueban en tres longitudes diferentes de 60, 101 y 201 moléculas y se comparan directamente con los tratamientos de GRON solo. Véase en la Tabla 8 la lista de GRON utilizados en estos ejemplos.

Procedimientos

Los protoplastos transgénicos de *Arabidopsis thaliana* BFP derivados de tejido radicular inducido se sembraron en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1×10^7 células/ml. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador *rbcSE9* y el promotor U6 de *Arabidopsis thaliana* que impulsa el ARNsg con un terminador poly-T10. Los plásmidos CRISPR junto con GRON se introdujeron en protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para la CRISPR y 0,547 µM para los GRONs de 60-mer, 0,32 µM para los GRONs de 101-mer y 0,16 µM para los GRONs de 201-mer. Los tratamientos con GRON solo recibieron una concentración final de GRON tras la administración de PEG de 7,5 µM para el 60-mer, 5,0 µM para el 101-mer y 2,5 µM para el 201-mer. Los protoplastos se incubaron en la oscuridad a 23 °C durante 72 horas y, a continuación, se analizaron mediante citometría de flujo para determinar el porcentaje de protoplastos positivos para GFP dentro de un tratamiento determinado.

La CRISPR consta de dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo plásmido. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar la nucleasa Cas9 hasta el gen objetivo BFP. La secuencia espaciadora CRISPR BFP es 5' GTCGTGCTGCTTCATGTGGT3'. En este ejemplo se utilizó el CRISPR BFP, dirigido al gen *bfp*. Los GRON contienen la secuencia codificadora del gen *bfp* cerca del lugar de conversión. La Tabla 8 ofrece una lista de los GRON utilizados.

Tabla 8. Lista de los GRON utilizados en estos ejemplos.

Nombre GRON	GRON Química	Secuencia GRON
BFP4/C 60-mer (1 cabeceo)	3PS	5'GTGACCACCTTCACCTACGGCGTGCAGTGCTTCAGCCGCTACCCAGACCACATGAAGCAG 3'
BFP4/C 101-mer (1 cabeceo)	3PS	5'CCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCTTCACCTACGGCGTGCGAGTGCTTCAGCCGCTACCCAGACCACATGAAGCAGCAGCAGAC 3'
BFP4/C 201-mer (1 cabeceo)	3PS	5' GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCTTCACCTACGGCGTGCGAGTGCTTCAGCCGCTACCCAGACCACATGAAGCAGCAGCAGTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCAGGA GCGCACCATCTT 3'

Resultados

Usando el CRISPR de BFP, los GRONs de longitudes ≥ 101 nt son mejores en la mediación de la conversión de BFP a GFP cuando se comparan directamente con los GRONs de 60 nt de longitud (Figura 5). En general, las muestras que contienen la CRISPR de BFP y GRON tienen niveles más altos de conversión de BFP a GFP en comparación con las muestras de GRON solamente (Figura 5), lo que demuestra el impacto positivo que tienen las CRISPR en el aumento de las tasas de conversión. Estos datos demuestran además que la longitud del GRON que es más eficaz en la mediación de la conversión de BFP a GFP, cuando se utiliza junto con el CRISPR, tiene que ser ≥ 101 nt de longitud.

Ejemplo 9: CRISPR con 2'-O-Me GRONs

El propósito de estos ejemplos es demostrar la conversión de BFP a GFP en nuestro sistema modelo transgénico de BFP de *Arabidopsis thaliana* utilizando CRISPRs para crear roturas de doble cadena en el gen *bfp* y GRONs para mediar la conversión. El CRISPR BFP utilizado en este ejemplo se dirige al gen *bfp* y provoca una rotura de doble cadena en el ADN cerca del lugar de conversión. Los GRONs utilizados con el CRISPR BFP, contienen la secuencia codificante o no codificante del gen *bfp* aproximadamente sitio de conversión con las primeras diez bases 5' del GRON siendo bases de ARN en lugar de bases de ADN. Estas bases de ARN están marcadas con grupo(s) 2'-O-Me en la primera base 5' de ARN o en las primeras nueve bases 5' de ARN, como se muestra en la Figura 6. Estos GRONs se denominan en el presente documento GRONs 2'-O-Me y se comparan directamente con los GRONs 3PS de longitudes similares que contienen bases de ADN con 3 enlaces fosfotioato en los extremos 5' y 3' del GRON. Estos GRON se denominan en lo sucesivo GRON 3PS. Véase en la Tabla 9 la lista de GRON utilizados en estos ejemplos.

Procedimientos

Los protoplastos transgénicos de *Arabidopsis thaliana* BFP derivados de tejido radicular inducido se sembraron en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1x10⁷ células/ml. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador rbcSE9 y el promotor U6 de *Arabidopsis* que impulsa el ARNsg con un terminador poly-T10. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). Los plásmidos CRISPR junto con GRON se introdujeron en protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para el CRISPR, 0,5 µM para el GRON de 71-mer y 0,16 µM para el GRON de 201-mer. Los tratamientos con GRON solo recibieron una concentración final de GRON tras la administración de PEG de 5,0 µM para el 71-mer y de 2,5 µM para el 201-mer. Los protoplastos se incubaron en la oscuridad a 23 °C durante 72 horas y, a continuación, se analizaron mediante citómetro de flujo para determinar el porcentaje de protoplastos positivos para GFP dentro de un tratamiento determinado.

La CRISPR consta de dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo plásmido. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar la nucleasa Cas9 hasta el gen objetivo BFP. La secuencia espaciadora CRISPR BFP es 5'-CTCGTGACCACCTTCACCCA 3'. En este ejemplo se utilizó el CRISPR BFP, dirigido al gen *bfp*. Los GRON contienen la secuencia codificante o no codificante del gen *bfp* cerca del lugar de conversión. La Tabla 9 muestra una lista de los GRON utilizados.

Tabla 9. Lista de los GRON utilizados en estos ejemplos

Nombre GRON	GRON Química		Secuencia GRON
BFP4/C 71-mer	3PS	5'GCGUGCCCGUGCCCT GCTACCCCG3'	GGCCCACCCTCGTGACCACCTTCACCTACGGCGTGCAAGTGCCTTCAGCC
BFP4/NC 71-mer	3PS	5'TTCATGTGGTCGGG. GGGCCAGGG3'	3TAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGT
BFP4/C 201-mer	3PS	5'GGGCGAGGGCGAT TGCCCGTGCCCTGGCC CCCCGACCATGAAC CGACCATCTT3'	3CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACACCGGCAAGC CACCCCTGTCACACCTTCACCTACGGCGTGCAAGTGCCTTCAGCCGCTA CAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAG
BFP4/NC 201-mer	3P	5'AAGATGGTGCGCTC CATGTGGTCTGGGTAA CCAGGGCACGGGCAC ATCGCCCTCGCCC3'	CTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTT 3CGGCTGAAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGGTGGG CTTGCCGGTGGTGAGATGAAGTTCAGGGTCAGCTTCCGTAGGTGGC
BFP4/C 71-mer	2'-O-Me	5'gcuugcccgugCCCTGc TACCCCG 3'	CCCACCCTCGTGACCACCTTCACCTACGGCGTGCAAGTGCCTTCAGCCCG

Nombre GRON	GRON Química		Secuencia GRON
BFP4/NC 71-mer		5'uucaugugguCGGGG' GGGCCAGGG3'	TAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGT
Nombre GRON	GRON Química		Secuencia GRON
BFP4/C 201-mer	2'-O-Me	5'gggcgagggcgATGC: CCGTGCCCTGGCCC: CCGACCACATGAAG'	CACCTACGGCAAGCTGACCCGTAAGTTCATCTGCACCACGGCAAGCTG :CCCTCGTGACCACCTTCACCTACGGCGTGCACTGCTTCAGCCGCTACC AGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCG CACCATCTT3'
BFP4/NC 201-mer	2'-O-Me	5'aagauggugcGCTCC	GGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCA TGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGGC CAGGGCACGGGCAGCTTCCGGTGGTGACAGATGAAGTTCAGGGTCAGCTTCCGTAGGTGGCA TCGCCCTCGCCC3'

Resultados

- 5 Los GRONs de 71-mer y 201-mer 2'-O-Me tuvieron una conversión similar de BFP a GFP cuando se compararon con los diferentes tipos de protecciones GRON de (0), (1) o (9) usando las CRISPRs de BFP (Figuras 7 y 8). Los GRONs 2'-O-Me son más eficaces que sus homólogos 3PS GRON para mediar en la conversión de BFP a GFP utilizando las CRISPRs BFP (Figuras 7 y 8).

Ejemplo 10: Nickasas CRISPR con Introducción de GRONs

- 10 El propósito de estos ejemplos es demostrar la conversión de BFP a GFP en nuestro sistema modelo transgénico de BFP de *Arabidopsis thaliana* utilizando CRISPRs para crear muescas monocatenarias dirigidas en el gen *bfp* y GRONs para mediar la conversión. El CRISPR BFP1 (CR:BFP1) utilizado en este ejemplo tiene como objetivo el gen *bfp* y contiene mutaciones en los residuos catalíticos (D10A en RuvC y H840A en HNH) que provocan muescas monocatenarias en el ADN del gen *bfp* cerca del sitio de conversión en las cadenas de ADN complementarias o no complementarias al ARN guía, respectivamente. Estas CRISPR se denominan en el presente documento BFP1 CRISPR nickasa D10A y BFP1 CRISPR nickasa H840A y se utilizan solas o con BFP5 ARNsg en un plásmido separado. En este ejemplo, cuando se utilizan varias nickasas CRISPR juntas, pueden mellar la misma cadena de ADN o cadenas opuestas. Cuando las dos proteínas Cas9 que contienen las mismas mutaciones, D10A o H840A, se utilizan juntas, mellan la misma cadena de ADN. Por el contrario, cuando se utilizan dos proteínas Cas9 juntas y una de ellas contiene la mutación D10A y la otra contiene la mutación H840A, mellan cadenas opuestas del ADN. Los GRONs utilizados con las nickasas CRISPRs, contienen la secuencia codificante o no codificante del gen *bfp* alrededor del sitio de conversión con una base de cabeceo localizada en la secuencia PAM de BFP5 CRISPR. Estos GRONs tienen 3 enlaces fosfotioato tanto en el extremo 5' como en el 3' y se denominan aquí GRONs 3PS. Véase en la Tabla 10 la lista de GRON utilizados en estos ejemplos. Las CRISPR nickasas se comparan directamente con sus homólogas CRISPR capaces de provocar roturas selectivas de doble cadena en el ADN del gen *bfp*.

Procedimientos

- 30 Los protoplastos transgénicos BFP de *Arabidopsis thaliana* derivados de tejido radicular inducido se siembran en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1x10⁷ células/ml. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador *rbcSE9* y el promotor U6 de *Arabidopsis thaliana* que impulsa el ARNsg con un terminador poly-T10. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). El gen Cas9 contiene mutaciones en los residuos catalíticos, ya sea D10A en RuvC o H840A en HNH. Los plásmidos CRISPR junto con GRON se introducen en los protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para la CRISPR y 0,16 µM para el 201-mer. Los tratamientos con GRON solo recibieron una concentración final de GRON tras la administración de PEG de 2,5 µM para el 201-mer. Los protoplastos se incubaron en la oscuridad a 23 °C durante 72 horas y, a continuación, se analizaron mediante citómetro de flujo para determinar el porcentaje de protoplastos positivos para GFP dentro de un tratamiento determinado.

- 45 La CRISPR consta de dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo plásmido. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar la nucleasa Cas9 hasta el gen objetivo BFP. En este ejemplo se utilizaron los ARNsg BFP1 y BFP5 que se dirigen a

diferentes regiones del gen *bfp* cerca del sitio de conversión. El espaciador BFP1 (5'CTCGTGACCACCTTCACCCA3') se dirige a la cadena codificante, mientras que el espaciador BFP5 (5'GTCGTGCTGCTTCATGTGGT3') se dirige a la cadena no codificante del gen *bfp*. Los GRON contienen la secuencia codificante o no codificante del gen *bfp* cerca del lugar de conversión. La Tabla 10 muestra una lista de los GRON utilizados.

5

Tabla 10. Lista de GRONs usados en estos ejemplos

Nombre GRON	Química de GRON	Secuencia de GRON	CRISPR
BFP4/NC 201-mer (1 cabeceo; BFP5)	3PS	5' AAGATGGTGGCTCTGGAGTGTAGCTTCGGCATGGCGGACTTGAAGAAGTCG TGCTGCTTCATGTGGTCTGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTG GGCCAGGGCACGGGAGCTTGCCGGTGGTGCAGATGAACCTCAGGGTCAGCTTCCC TAGGTGGCATCGCCCTCGCCC 3'	BFP1 y BFP5
BFP4/C 201-mer (1 cabeceo; BFP5)	3PS	5' GGSCGAGGGCGATGCCACTACGGCAAGTGAACCTGAAGTTCATCTGCACCACCS GCAAGCTGCCCGTGGCTGGCCCAECCCTGTGACCACTTCACCTACGGCGTCAAGTGTTCAGCCGCTAC CCAGACCATGAAGCAGCAGATTTCAAGTCCCATGCCGAAGGCTACGTCCAGGA GGGCACCATCTT 3'	BFP1 y BFP5

Resultados

- 10 Ambas nickasas CRISPR (D10A y H840A) son más eficientes en mediar la conversión de BFP a GFP cuando la CRISPR BFP1 y la CRISPR BFP5 se usaron juntas en lugar de separadas (Figura 9). Además, cuando las nickasas CRISPR BFP1 y BFP5 D10A se utilizan junto con el GRON C/201 1W, la conversión de BFP a GFP es significativamente mayor en comparación con los tratamientos en los que estas nickasas CRISPR se utilizan con el GRON NC/201 1W (Figura 9). Cuando las nickasas CRISPR BFP1 y BFP5 H840A se utilizan juntas, se observa
- 15 aproximadamente el mismo nivel de conversión de BFP a GFP con los GRONs C/201 o NC/201 1W (Figura 9). Estos niveles de conversión de BFP a GFP son ligeramente más altos que cuando se utiliza la CRISPR BFP5 sola y ligeramente más bajos que cuando se utiliza la CRISPR BFP1 sola (Figura 9).

Ejemplo 11: Uso de CRISPR para dirigirse a múltiples genes

- 20 El propósito de este ejemplo es demostrar la conversión de múltiples genes simultáneamente en una población dada de protoplastos derivados del sistema modelo *Arabidopsis thaliana* utilizando CRISPRs para crear roturas de doble cadena en genes objetivo y GRONs para mediar la conversión. Las CRISPR utilizadas en este ejemplo se dirigen a los genes BFP y acetohidroxiácido sintasa (AHAS) en el genoma de *Arabidopsis thaliana* mediante la introducción en
- 25 protoplastos de plásmidos que codifican el gen Cas9 y múltiples ARNsg dirigidos a estos dos genes diferentes. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). Esto permitirá a Cas9 causar roturas de doble cadena en los genes BFP y AHAS en presencia de GRONs que mediarán su conversión.

Procedimientos

- 30 Los protoplastos de *Arabidopsis thaliana* derivados de tejido radicular inducido se siembran en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1x10⁷ células/ml. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador *rbcSE9* y el promotor U6 de *Arabidopsis thaliana* que impulsa múltiples ARNsgs diferentes con un terminador poly-T10. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). Los plásmidos CRISPR junto con GRON se introducen en los protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para la CRISPR y 0,16 µM para el 201-mer. Los tratamientos con GRON solo reciben una concentración final de GRON tras la administración de PEG de 2,5 µM para el 201-mer. Los protoplastos se incubarán en la oscuridad a 23°C durante 72 horas, y después se analizarán mediante citómetro de flujo y un ensayo PCR específico de alelo para
- 40 determinar el porcentaje de protoplastos convertidos de BFP a GFP y AHAS respectivamente dentro de un tratamiento dado.

En el ensayo PCR específico de un alelo se utilizaron 10-16 réplicas de 5.000 equivalentes genómicos de ADN genómico en las reacciones PCR primarias.

- 45 La CRISPR consiste en dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo o de múltiples plásmidos. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar a la nucleasa Cas9 hasta los genes objetivo. En este ejemplo se utilizan diferentes ARNsg y GRONs para dirigirse a múltiples genes cerca de sus sitios de conversión; espaciador BFP (5'CTCGTGACCACCTTCACCCA 3') y espaciador AHAS (5' TGGTTATGCAATTGGAAGATCGC 3'). La Tabla 11 describe los GRON utilizados.
- 50

Tabla 11. Lista de GRON utilizados en este ejemplo

Nombre GRON	Gen objetivo	GRON Química	Secuencia GRON
BFP/C 201-mer	BFP	3PS	5'GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTTCATCTGCACCA CCGGCAAGCTGCCCGTCCCTGGCCACCCCTCGTGACCACCTTCACCTACGGCGT GCAGTGCTTCAGCCGCTACCCGACCACATGAAGCAGCAGCACTTCTCAAGTCC GCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATTCTT3'
AHAS(W) 574/NC 201-mer	AHAS	3PS	5'AGCTGCTGCAAAACAGCAACATGTTCCGGGAATATCTCGTCTCTGAGCCGGATC CCCGAGAAATGTGTGAGCTCGGTTAGCTTTGTAGAAGCGATCTTCAATTGCATA ACCATGCCAAGATGCTGGTTGTTTAATAAAAGTACCTTCACTGGAAGATTCTCTAC ACGAATAGTGGCTAGCTCTTGACATTATTATAAA3'

Resultados

La conversión de BFP a GFP y AHAS se determinó a las 144 h post entrega por PEG de los plásmidos CRISPR de BFP y AHAS y los GRONs BFP/C 201-mer y AHAS(W)574/NC 201-mer en la línea transgénica BFP de *Arabidopsis thaliana*. Los datos de citometría de flujo revelaron que el Tratamiento 1 produjo una conversión de 0,20% de BFP a GFP (Tabla 12). El ensayo PCR específico de alelo reveló que el Tratamiento 1 dio lugar a protoplastos convertidos en AHAS en un 0,01% (Tabla 12). Los tratamientos sólo con GRON tuvieron una conversión mínima en ambos ensayos (Tabla 12). Este ejemplo demuestra el éxito de la conversión simultánea de dos genes objetivo independientes (BFP y AHAS) dentro de una población determinada de protoplastos derivados del sistema modelo BFP de *Arabidopsis thaliana*.

Tabla 12. Medición de la conversión de los genes BFP y AHAS a las 144 h posteriores a la entrega por PEG de plásmidos CRISPR y GRONs en una población determinada de protoplastos derivados del sistema modelo BFP de *Arabidopsis thaliana*, ya sea por: (1) citometría de flujo que determina el porcentaje de protoplastos GFP positivos o (2) PCR específica de alelo que determina el porcentaje de protoplastos AHAS convertidos.

Tratamiento	CRISPR	GRONs	Conversión de BFP a GFP	AHAS - Conversión W574L
			Citometría de flujo	PCR alelo específica
1	CR-BFP y CR-AHAS	BFP/C 201-mer y AHAS(W)574/NC 201-mer	0,20%	~0,01%
2	Ninguno	BFP/C 201-mer y AHAS(W)574/NC 201-mer	0,01%	~0,001%
3	Ninguno	Ninguno	0,01%	~0,001%

Ejemplo 12 (no de acuerdo con la invención): Liberación de ARNm Cas9 en células vegetales

Este ejemplo hace uso de la entrega directa de ARNm Cas9 recombinante en células vegetales como alternativa a la entrega de plásmidos de expresión CRISPR-Cas. Este procedimiento incluye (1) la síntesis in vitro de ARNm modificado y (2) el suministro de este ARNm modificado a las células vegetales.

Procedimientos

Un ARNm Cas9 se transcribirá in vitro utilizando una ARN polimerasa como T7, T3 o SP6 a partir de una plantilla de plásmido linealizada que incluya componentes de una 5'UTR, la secuencia codificante (CDS) para la proteína y una 3'UTR. Una ARN polimerasa puede incorporar un determinado nucleósido modificado mejor que otra. El 5' UTR puede contener elementos que mejoren su estabilidad, como el MiR-122 del virus de la hepatitis C (Shimakami et al., 2012). La síntesis in vitro incorporará nucleósidos que sean protectores y garanticen una buena traducción en las células vegetales objetivo. El ARNm recombinante de Cas9 estará capado y contendrá una cola de poliA.

Los protoplastos transgénicos BFP de *Arabidopsis thaliana* derivados de tejido radicular inducido se siembran en una placa de 96 pocillos de fondo plano a 250.000 células por pocillo a una densidad celular de 1x10⁷ células/ml. El ARNm recombinante Cas9 se introducirá en las células vegetales por uno de los siguientes medios (lista no exhaustiva): péptidos penetrantes de células (PPC), reactivos liposomales de transfección, poli(etilenglicol) (PEG), solos o en combinación, para permitir la introducción del ARNm recombinante activo Cas9 en las células.

Ejemplo 13 (no de acuerdo con la invención): CRISPR-Cas para la fijación de ADN, ARN o proteínas.

Este ejemplo hace uso de un casete modificado de ARN guía único (ARNsg) en donde se incluye una secuencia enlazadora (también puede denominarse secuencia de anclaje) en el extremo 3' del ARNtracr, pero a 5' de la señal de terminación de la ARN polimerasa III, como se muestra en el ejemplo siguiente (Figura 10). El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). Aunque es preferible, la colocación del enlazador no se limita al extremo 3' del ARN trazador, sino que se investigará en varias posiciones dentro del casete de ARNsg. La secuencia enlazadora puede variar en longitud nucleotídica o contener una estructura secundaria que mejore la unión o aumente el número de moléculas unidas mediante interacciones triplex.

El enlazador permitirá el apareamiento de bases Watson-Crick con un ADN, ARN o proteínas que contengan la secuencia complementaria (Figura 10). Además, la secuencia enlazadora de los casetes de ARNsg se diseñará para que contenga una estructura secundaria y terciaria avanzada que permita regiones de interacción polifacética más complejas que amarren múltiples moléculas de ADN, ARN o proteínas.

El concepto general es que un complejo CRISPR-Cas unirá moléculas biológicas al sitio de la actividad nucleasa, aumentando así la probabilidad de edición de genes. Estas moléculas biológicas incluyen el GRON que mediará la conversión del gen o genes objetivo. Los enlazadores de fijación pueden añadirse a los ARNsg simplemente utilizando, por ejemplo, Gene Strings u oligos recocidos.

Procedimientos

Los protoplastos transgénicos BFP de *Arabidopsis thaliana* derivados de tejido radicular inducido se siembran en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1×10^7 células/ml. Los plásmidos de anclaje CRISPR-Cas contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador rbcSE9 y

Promotor U6 de *Arabidopsis thaliana* que impulsa ARNsg con una secuencia enlazadora que es complementaria a un tracto polinucleotídico de 15-30 pb localizado en un GRON de edición de 201-mer dirigido a *bfp* (como se muestra en la Figura 10). El casete de anclaje del ARNsg está terminado por un terminador poli-T10. Los plásmidos CRISPR-Cas junto con GRON se introducen en los protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para la CRISPR y 0,16 µM del GRON de 201-mer. Los tratamientos con GRON solo recibieron una concentración final de GRON tras la administración de PEG de 2,5 µM para el 201-mer. Los protoplastos se incubaron en la oscuridad a 23 °C durante 72 horas y, a continuación, se analizaron mediante citómetro de flujo para determinar el porcentaje de protoplastos positivos para GFP dentro de un tratamiento determinado.

La CRISPR consiste en dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo o de diferentes plásmidos. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr) y enlazador. La región ARNcr contiene la secuencia espaciadora utilizada para guiar la nucleasa Cas9 hasta el gen objetivo BFP. En este ejemplo se utiliza CRISPR, cuyo objetivo es el gen *bfp*. Los GRON contienen la secuencia codificante o no codificante del gen *bfp* cerca del lugar de conversión.

Ejemplo 14: CRISPRs con ARNg truncado.

El propósito de este ejemplo es demostrar la conversión de BFP a GFP en protoplastos derivados del sistema modelo BFP de *Arabidopsis thaliana* utilizando CRISPRs para crear roturas de doble cadena en genes objetivo y GRONs para mediar la conversión. Las CRISPR utilizadas en este ejemplo se dirigen al gen *bfp* en el genoma de *Arabidopsis thaliana* mediante la introducción en protoplastos de plásmidos que codifican el gen Cas9 y un ARNsg de dos longitudes diferentes. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). El ARNcr que guía al Cas9 hasta los genes objetivo se denomina espaciador y suele tener una longitud de 20 nt (CR:BFP1 20-nt), sin embargo, en estos ejemplos hemos probado la eficacia de utilizar un espaciador de menor longitud de 17 nt (CR:BFP1 17-nt) para mediar en la conversión de BFP a GFP.

Procedimientos

Los protoplastos de *Arabidopsis thaliana* derivados de tejido radicular inducido se siembran en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1×10^7 células/ml. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador rbcSE9 y el promotor U6 de *Arabidopsis thaliana* que impulsa múltiples ARNsgs diferentes con un terminador poly-T10. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). Los plásmidos CRISPR junto con GRON se introducen en los protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para la CRISPR y 0,16 µM para el 201-mer. Los tratamientos con GRON solo reciben una concentración final de GRON tras la administración de PEG de 2,5 µM para el 201-mer. Los protoplastos se incubarán en la oscuridad a 23°C durante 72 horas, y después se analizarán mediante citómetro de flujo para determinar el porcentaje de BFP a GFP en un tratamiento determinado.

La CRISPR consiste en dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo o de múltiples plásmidos. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar a la nucleasa Cas9 hasta los genes objetivo. En estos ejemplos, se probaron dos espaciadores BFP1 de longitud diferente de 20-nt (5'CTCGTGACCACCTTCACCCA3') frente a 17-nt (5'GTGACCACCTTCACCCA 3'). La Tabla 13 describe el GRON utilizado

Tabla 13. Lista de GRON utilizados en este ejemplo

Nombre GRON	GRON Química	Secuencia GRON
BFP4/NC 201-mer 3W	3PS	5'AAGATGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCG TGCTGCTTCATGTGGTTCGGGTAGCGGCTGAAGCACTGCACGCCGTACGTAAACGTGGTCACG AGGTTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGACATGAACCTCAGGGTCAGCTTGCC 3' TAGGTGGCATCGCCCTCGCCC 3'

Resultados

La reducción de la longitud del protoespaciador de BFP 1 de 20bp a 17bp tuvo niveles similares de conversión de BFP a GFP de 0,163% vs. 0,177% respectivamente a las 72 h post entrega de PEG de plásmidos y GRONs en el sistema modelo de BFP de *Arabidopsis thaliana* (Figura 11).

Ejemplo 15: CRISPRs con ARNg amplicón.

El propósito de este Ejemplo es demostrar la conversión de BFP a GFP en protoplastos derivados del sistema modelo BFP de *Arabidopsis thaliana* utilizando CRISPRs para crear roturas de doble cadena en genes objetivo y GRONs para mediar la conversión. Las CRISPR utilizadas en este ejemplo se dirigen al gen *bfp* del genoma de *Arabidopsis thaliana* introduciendo en protoplastos plásmidos que codifican el gen *Cas9* y un ARNsg codificado en un plásmido o introducido en protoplastos como amplicón. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). El ARNcr guía al Cas9 hasta los genes objetivo, donde el Cas9 crea una rotura de doble cadena y el GRON se utiliza como plantilla para convertir la BFP en GFP de forma dirigida.

Procedimientos

Los protoplastos de *Arabidopsis thaliana* derivados de tejido radicular inducido se siembran en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1×10^7 células/ml. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador rbcSE9 y el promotor U6 de *Arabidopsis* que impulsa múltiples ARNsg diferentes con un terminador poly-T₁₀. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). Los plásmidos CRISPR junto con GRON se introducen en los protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para la CRISPR y 0,16 µM para el 201-mer. Los tratamientos con GRON solo reciben una concentración final de GRON tras la administración de PEG de 2,5 µM para el 201-mer. Los protoplastos se incubarán en la oscuridad a 23°C durante 72 horas, y después se analizarán mediante citómetro de flujo para determinar el porcentaje de BFP a GFP en un tratamiento determinado.

La CRISPR consiste en dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo o de múltiples plásmidos. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar a la nucleasa Cas9 hasta los genes objetivo. En estos ejemplos, el mismo ARNg de BFP6 (5'GGTGCCGCACGTCACGAAGTCGG 3') se introdujo en protoplastos como amplicón o codificado en un plásmido. La Tabla 14 describe los GRON utilizados.

Tabla 14. Lista de GRON utilizados en este ejemplo

Nombre GRON	GRON Química	Secuencia GRON
BFP4/NC 201-mer	3PS	5'AAGATGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTCGCTTCATGTGGTCTGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGACATGAACCTCAGGGTCAGCTTGCCGTAGGTGGCATCGCCCTCGCCC 3'

Nombre GRON	GRON Química	Secuencia GRON
BFP4/C 201-mer	3PS	5'GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACGGCAAGC TGCCCGTGCCCTGGCCACCCCTCGTGACCACTTCACCTACGGCGTGAGTGCTTCAGCCGCTA CCCCGACCACATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGA GCGCACCATCTT3'

Resultados

La entrega del ARNg BFP6 como un amplicón (CR:BFP6 (amplicón de ARNg)) junto con un plásmido que contenía sólo Cas9 tuvo tasas similares de conversión de BFP a GFP cuando se comparó con tratamientos tanto con el ARNg (plásmido de ARNg) como con Cas9 codificados en plásmidos separados a las 72 h posteriores a la entrega por PEG de plásmidos y GRONs en el sistema modelo de BFP de *Arabidopsis thaliana* (Figura 12).

Ejemplo 16: CRISPRs con GRONs no modificados.

El propósito de este ejemplo es demostrar la conversión de BFP a GFP en nuestro sistema modelo transgénico de BFP de *Arabidopsis thaliana* utilizando CRISPRs para crear roturas de doble cadena en el gen *bfp* y GRONs para mediar la conversión. El CRISPR BFP utilizado en este ejemplo se dirige al gen *bfp* y provoca una rotura de doble cadena en el ADN cerca del lugar de conversión. Los GRONs 3PS contienen bases de ADN con 3 enlaces fosfotioato tanto en el extremo 5' como en el 3' del GRON y se denominan en el presente documento GRONs 3PS. Los GRONs 3PS se compararon directamente con sus homólogos GRON no modificados en la conversión mediada de BFP a GFP utilizando las CRISPRs BFP en nuestro sistema modelo transgénico BFP de *Arabidopsis thaliana*. Véase en la Tabla 15 la lista de GRON utilizados en estos ejemplos

Tabla 15. Lista de GRONs usados en estos ejemplos

Nombre GRON	Química de GRON	Secuencia de GRON
BFP4/C 41-mer	3PS	5' CCCTCGTGACCACCTTCACCTACGGCGTGAGTGCTTCAGC 3'
BFP4/NC 41-mer	Ninguna	5'GCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGG3'

Procedimientos

Los protoplastos transgénicos de *Arabidopsis thaliana* BFP derivados de tejido radicular inducido se sembraron en una placa de 96 pocillos de fondo plano, a 250.000 células por pocillo a una densidad celular de 1x10⁷ células/ml. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador rbcSE9 y el promotor U6 de *Arabidopsis* que impulsa el ARNsg con un terminador poly-T10. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). Los plásmidos CRISPR junto con los GRON se introdujeron en protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl para el CRISPR, 0,16 µM para los GRON de 41-mer. Los tratamientos con GRON solo recibieron una concentración final de GRON tras la administración de PEG de 0,8 µM para el 41-mer. Los protoplastos se incubaron en la oscuridad a 23 °C durante 72 horas y, a continuación, se analizaron mediante citómetro de flujo para determinar el porcentaje de protoplastos positivos para GFP dentro de un tratamiento determinado.

La CRISPR consta de dos componentes: la planta de codón optimizado *Streptococcus pyogenes* Cas9 (SpCas9) y ARNsg, ambos expresados a partir del mismo plásmido. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar la nucleasa Cas9 hasta el gen objetivo BFP. La secuencia espaciadora CRISPR BFP es 5'CTCGTGACCACCTTCACCCA 3'. En este ejemplo se utilizó el CRISPR BFP, dirigido al gen *bfp*. Los GRON contienen la secuencia no codificante del gen *bfp* cerca del lugar de conversión. La Tabla 16 muestra una lista de los GRON utilizados.

Tabla 16. Lista de GRONs usados en estos ejemplos

Nombre GRON	Química de GRON	Secuencia de GRON
BFP4/C 41-mer	3PS	5' CCCTCGTGACCACCTTCACCTACGGCGTGAGTGCTTCAGC 3'
BFP4/NC 41-mer	Ninguna	5'GCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGG3'

Resultados

Los GRONs de 41-mer 3PS son más eficaces que sus contrapartes GRON no modificadas para mediar la conversión de BFP a GFP usando las CRISPRs BFP (Figura 13).

Ejemplo 17 (no de acuerdo con la invención): TALENs y GRONs en Lino.

El propósito de este ejemplo es demostrar la conversión EPSPS en lino tanto a las 24 horas en protoplastos como a las 3 semanas en microcallos después de la entrega de plásmidos TALEN y GRONs. Los TALEN utilizados en este ejemplo se dirigen al gen *epsps* en el genoma de *Linum usitatissimum* introduciendo en protoplastos derivados de puntas de brotes plásmido(s) que codifican los TALEN crean una rotura de doble cadena y el GRON se utiliza como plantilla para convertir el gen *epsps* de una manera dirigida al sitio.

Procedimientos

Se aislaron protoplastos de lino a partir de puntas de brotes obtenidas de plántulas germinadas *in vitro*. Los plásmidos TALEN junto con los GRON se introdujeron en los protoplastos mediante la administración mediada por PEG a una concentración final de 0,05 µg/µl y 0,5 µM respectivamente. Los protoplastos se incubaron en la oscuridad a 25 °C durante un máximo de 48 h en medio líquido, o se incrustaron en perlas de alginato (5×10^5 células/ml), y se cultivaron en medio líquido para inducir la división celular y la formación de microcallos. Las muestras de protoplastos o microcallos obtenidas 24 h o 3 semanas después de la administración de ADN se analizaron mediante NGS para determinar el porcentaje de células (lecturas de ADN) portadoras de la mutación objetivo en un tratamiento determinado. También se estimó el porcentaje de indeles generados por una reparación imperfecta del ADN mediada por NHEJ.

Los constructos TALEN incluyen dos brazos (izquierdo y derecho), cada uno consistente en un dominio de unión a ADN tipo efector TAL enlazado a un dominio catalítico de escisión de ADN de FokI. El dominio de unión al ADN similar al efector TAL guía los brazos TALEN a sitios específicos del ADN, lo que permite a las endonucleasas plegadas de cada brazo dimerizarse juntas y escindir el ADN de doble cadena. Los plásmidos codificados con TALEN contienen MasP:LuEPSPS_(brazo izquierdo)-T2A-LuEPSPS_(brazo derecho) con un terminador rbcSE9. La secuencia LuEPSPS_(brazo izquierdo) es 5'TGGAACAGCTATGCGTCCG 3' y la secuencia LuEPSPS_(brazo derecho) es 5'TGAGTTGCCCTCCAGCGGCT 3'. Se utilizaron GRONs (144-mers) dirigidos a LuEPSPS con o sin bases de cabeceo para determinar su efecto sobre la tasa de conversión.

Resultados

Los protoplastos de 24 horas y los microcallos de 3 semanas tienen una conversión EPSPS de 0,067% y 0,051% respectivamente, de acuerdo con lo determinado por la Secuenciación de Próxima Generación (Figura 14). Además, estos datos muestran que el TALEN es activo y capaz de escindir el gen objetivo *epsps* en *Linum usitatissimum* y formar indeles de 2,60% y 1,84% respectivamente a las 24 horas en protoplastos y hasta 3 semanas en microcallos. Además, la conversión EPSPS y los indeles se mantienen hasta 3 semanas después de la introducción del plásmido TALEN y GRON.

Ejemplo 18 (no de acuerdo con la invención): CRISPRs en el lino.

El propósito de este ejemplo es demostrar la actividad de Cas9 en microcallos de lino tres y seis semanas después de la entrega de un plásmido Cas9. Las CRISPR utilizadas en este ejemplo se dirigen al gen *epsps* del genoma del *Linum usitatissimum* mediante la introducción en protoplastos derivados de puntas de brotes de plásmidos que codifican el gen *Cas9* y un ARNsg. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). El ARNcr guía a Cas9 hasta los genes objetivo, donde Cas9 crea una rotura de doble cadena en el gen *epsps* de forma dirigida. Las roturas de doble cadena en el gen *epsps*, al ser reparadas por la vía ubicua NHEJ, provocarán la formación de indeles alrededor del sitio de escisión.

Procedimientos

Se aislaron protoplastos de lino a partir de puntas de brotes obtenidas de plántulas germinadas *in vitro*. Los plásmidos codificados CRISPR contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador rbcSE9 y el promotor U6 de *Arabidopsis thaliana* que impulsa el ARNsg con un terminador poly-T₁₀. Los plásmidos CRISPR se introdujeron en los protoplastos mediante entrega mediada por PEG a una concentración final de 0,05 µg/µl. Los protoplastos se incrustaron en perlas de alginato (5×10^5 células/ml), se cultivaron en medio líquido y se incubaron en un agitador rotatorio (30 rpm) en la oscuridad a 25°C. Los microcallos desarrollados a partir de células individuales se analizaron mediante NGS, 3 y 6 semanas después de la entrega del plásmido CRISPR, para determinar el porcentaje de células (lecturas de ADN) portadoras de indeles generados por la vía de reparación del ADN mediada por NHEJ, propensa a errores.

La CRISPR consta de dos componentes: la planta de codón optimizado *Streptococcus pyogenes* Cas9 (SpCas9) y ARNsg, ambos expresados a partir del mismo plásmido. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora utilizada para guiar a la nucleasa Cas9 hasta el gen objetivo. En este ejemplo, el CRISPR se dirige al gen *epsps*.

Resultados

Los microcallos de 3 y 6 semanas de edad tienen una formación de indeles del 46,5% y 54,7% respectivamente, de acuerdo con lo determinado por Secuenciación de Próxima Generación (Figura 15). Estos datos demuestran que Cas9 es activa y capaz de escindir el gen objetivo *EPSPS* en *Linum usitatissimum* y formar indeles. Además, estos indeles se mantienen hasta 6 semanas después de la introducción del plásmido CRISPR.

Ejemplo 19: Constructo de nucleasas modificadas

CRISPR-Cas

Para el constructo de plásmidos de expresión CRISPR-Cas9 transitorios, se sintetizó un gen SpCas9 optimizado para codones de plantas superiores que contenía un SV40 NLS en los extremos N y C y una etiqueta 2x FLAG en el extremo N como una serie de GeneArt® Strings™ (Life Technology, Carlsbad, CA), luego se clonó corriente abajo del promotor de la manopina sintasa (MAS) y corriente arriba del terminador de la ribulosa bisfosfato carboxilasa del guisante (rbcE9) por el procedimiento de Gibson. A continuación, se sintetizó un casete ARNg consistente en un ARNg quimérico cuya expresión está dirigida por el promotor U6 de *Arabidopsis*, como GeneArt® Strings™, y luego se introdujo en el constructo que contenía Cas9 utilizando el procedimiento de Gibson para formar pBCRISPR. Para especificar el ARNg quimérico para la secuencia objetivo respectiva, se recogieron pares de oligonucleótidos de ADN que codificaban las secuencias objetivo variables de ARNg de 20 nt (Tabla 2 de datos ampliados) para generar fragmentos cortos de doble cadena con salientes de 4 pb. Los fragmentos se ligaron en pBCrispr digerido con BbsI para obtener los constructos CRISPR-Cas BC-1, BC-2 y BC-3.

TALEN (no de acuerdo con la invención)

El diseño y construcción de los constructos de expresión TALEN BT-1 y LuET-1 se basó en las reglas descritas en Cermak et al., Nucleic Acids Res. 39, e82 (2011). La secuencia objetivo se seleccionó en función del sitio de edición del gen y del residuo i variable repetido (RVD) siguiendo las reglas de acuerdo con las cuales NG, HD, NI y NN reconocen T, C, A y G, respectivamente. El ensamblaje del dominio efector TAL unido a los dominios FokI heterodiméricos se completó mediante un servicio comercial (GeneArt; Life Technologies). Los monómeros TALEN se clonaron corriente abajo del promotor MAS y corriente arriba del terminador rbcE9 utilizando el procedimiento de Gibson y se expresaron como una unidad acoplada 2A.

Cultivo celular y aislamiento de protoplastos

Semillas de *Arabidopsis* esterilizadas en superficie se germinaron en medio sólido ½ MS (Minerales y vitaminas de acuerdo con XX; ½ concentrado; 87,7 mM de sacarosa) a 28°C bajo un ciclo de luz/oscuridad de 12 h. Se recogió material radicular de plántulas de 2 a 3 semanas de edad y se mantuvo en medio líquido ½ MS en condiciones de poca luz a 28°C. Los cultivos de raíces se transfirieron y mantuvieron en MSAR [0,22% 1/2 MS, 87,7 mM de sacarosa, 11,4 µM de IAA, 2,3 µM de 2,4-D, 1,5 µM de 2iP, pH 5,8] tres semanas antes del aislamiento de protoplastos. Las raíces se cortaron en segmentos de aproximadamente 6 mm y se incubaron en una solución MSAP [0,22% 1/2 MS, 87,7 mM sacarosa, 11,4 µM IAA, 2,3 µM 2,4-D, 1,5 µM 2iP y 400 mM manitol, pH 5,8] que contenía enzimas de digestión de la pared celular [1,25% Cellulase RS, 0,25% Macerozyme R-10, 0,6 M manitol, 5 mM MES, 0,1% BSA] durante 3-4 h en la oscuridad con agitación suave. Los protoplastos liberados se recogieron y se pasaron por un filtro estéril de 100 µm y otro de 35 µm. El filtrado de protoplastos se mezcló con 0,8 veces el volumen de Optiprep™ Density Gradient Medium (Sigma) y se mezcló suavemente. Una solución W5 al 60% [154 mM NaCl, 5 mM KCl, 125 mM CaCl₂·2H₂O, 5 mM glucosa, 10 mM MES, (pH 5,8)] / 40% Optiprep seguida de una solución W5 al 90%/10% Optiprep se colocaron lentamente en capas sobre la solución filtrada/Optiprep para formar un gradiente, que se centrifugó a 198 RCF durante 10 min. Se recogió la capa blanca de protoplastos y se mezcló con 2 veces el volumen de W5. Los protoplastos se centrifugaron a 44 RCF durante 10 minutos y se volvieron a suspender en solución TM [14,8 mM MgCl₂·6H₂O, 5 mM MES, 572 mM manitol, (pH 5,8)] a una densidad de 1×10^7 células/ml. Para los experimentos con Zeocina™ (Life Technologies, Carlsbad, CA) y fleomicina (InvivoGen, San Diego, CA), los protoplastos se mantuvieron en TM ajustada a pH 7,0 durante 90 min en hielo antes de la transfección. Para las concentraciones de antibióticos, véase la Figura 1 de datos ampliados.

Se aislaron protoplastos de lino a partir de puntas de brotes obtenidas de plántulas de 3 semanas germinadas *in vitro*. Las puntas de los brotes se picaron finamente con un bisturí, se preplasmolizaron durante 1 h a temperatura ambiente en medio B [sales y vitaminas B5 (Gamborg et al., 1968), 4 mM CaCl₂, 0,1 M glucosa, 0,3 M manitol, 0,1 M glicina, 250 mg/l hidrolizado de caseína, 10 mg/l L-cisteína-HCL, 0,5% polivinilpirrolidona (MW 10.000), 0,1% BSA, 1 mg/l BAP, 0,2 mg/l NAA, y 0,5 mg/l 2,4-D], y se incubaron en una solución de enzimas digestoras de la pared celular que contenía B-medium suplementado con 0,66% Cellulase YC y 0,16% de Macerozyme R-10 en un agitador rotatorio (50 rpm) a 25 °C durante 5 h. Los protoplastos liberados se tamizaron y purificaron mediante centrifugación en gradiente de densidad utilizando capas Optiprep (Sigma), se contaron con un hemocitómetro y se mantuvieron estacionarios durante la noche en la oscuridad a una densidad de $0,5 \times 10^6$ protoplastos/ml en medio B.

Transfección de protoplastos

En una placa de fondo plano de 96 pocillos, se transfectaron $2,5 \times 10^5$ células por pocillo con 10 pmol GRON, 10 pmol GRON más 3,25 µg de constructo de expresión CRISPR-Cas o TALEN o simulacro utilizando PEG [270 mM manitol, 67,5 mM $\text{Ca}(\text{NO}_3)_2$, 38,4% PEG 1500, (pH 5,8)]. Las células y el ADN se incubaron con PEG en hielo durante 30 minutos, seguidos de un lavado con 200 µl de solución W5. Por último, se añadieron 85 µl de MSAP++ [MSAP que contiene 50 nM de fitosulfokina-α y 20 µM de galato de n-propilo] y las células se cultivaron en condiciones de poca luz a 28°C.

Después de aproximadamente 18 h de cultivo, los protoplastos fueron transfectados con el plásmido TALEN junto con GRONs (20 µg de plásmido y 0.2 nmol GRON/10⁶ protoplastos) usando entrega mediada por PEG. Los protoplastos tratados se incubaron en la oscuridad a 25 °C durante un máximo de 48 h en medio B, o se incrustaron en perlas de alginato 24 h después de la transfección, y se cultivaron en medio líquido V-KM para inducir la división celular y la formación de microcallos. Para los experimentos con antibióticos, se transfectaron $1,25 \times 10^5$ células por pocillo con 8 µM de GRON CG13 utilizando la solución de PEG descrita anteriormente.

Citometría

Setenta y dos h después de la transfección, las células se analizaron por citometría utilizando el citómetro Attune® Acoustic Focusing (Applied Biosystems®) con excitación y detección de emisión de acuerdo con correspondencia para GFP. El nivel de fondo se basó en protoplastos tratados con PEG sin suministro de ADN. Para los experimentos con antibióticos, los protoplastos tratados con Zeocina o fleomicina antes de la transfección se analizaron mediante citometría 48 h después de la transfección.

Análisis de secuenciación

El ADN genómico se extrajo de protoplastos tratados con CRISPR-Cas o TALEN utilizando el kit NucleoSpin® Plant II de acuerdo con las recomendaciones del fabricante (Machery-Nagel, Bethlehem, PA). Los amplicones que flanquean la región objetivo TALEN o CRISPR se generaron utilizando la polimerasa Phusion® con 100 ng de ADN genómico y los cebadores BFPF-1 (5'-GGACGACGGCAACTACAAGACC-3')/BFPR-1 (5'-TAAACGGCCACAAGTTCAGC-3') para *Arabidopsis* CRISPR y TALEN; o LuEPF-1 (5'-GCATAGCAGTGAGCAGAAGC-3')/LuEPR-1 (5'-AGAAGCTGAAAGGCTGGAAG-3') para *L. usitatissimum* TALEN. Los amplicones se purificaron y concentraron utilizando columnas Qiaquick MinElute (Qiagen, Valencia, CA). La secuenciación profunda de los amplicones fue realizada por GeneWiz (South Plainfield, NJ) utilizando una serie MiSeq de 2 × 250 pb (Illumina, San Diego, CA). Para el análisis de datos, los archivos fastq de las lecturas 1 y 2 se importaron a CLC Genomics Workbench 7.0.4 (CLCBio, Boston, MA). Las lecturas emparejadas se fusionaron en una única secuencia si sus secuencias se solapaban. Se identificó una secuencia para un amplicón si ésta o su secuencia inversa y complementada contenían secuencias de cebadores tanto directa como inversa. La presencia de una secuencia única en una muestra se registró como su abundancia. El porcentaje de indel o edición génica se calculó dividiendo el número de lecturas con la edición o indel por el número total de lecturas secuenciadas y multiplicando por 100.

Secuencia de los fotoespaciadores CRISPR-Cas

Nombre	Secuencia (5' a 3')	Figura de referencia
BC-1	CTCGTGACCACCTTCACCCA	1a;1b;2a;2c
BC-2	GTCGTGCTGCTTCATGTGGT	2b
BC-3	GGCTGAAGCACTGCACGCCG	2d

Secuencias de dominio de unión TALEN

Nombre del papel	Secuencia (5' a 3')	Figura de referencia
BT-1	Brazo izquierdo: TGGTCGGGGTAGCGGCTGA	3a;3b
	Brazo derecho: TCGTGACCACCTTCACCCA	
LuET-1	Brazo izquierdo: TGAACAGCTATGCGTCCG	3c;3d
	Brazo derecho: TGAGTTGCCTCCAGCGGCT	

Secuencias GRON utilizadas

[illegible]

Análisis estadístico

La significación estadística se determinó mediante una prueba t de Student con distribución de dos colas. Los valores $p < 0,05$ se consideraron significativos. Los datos se muestran como media y SEM.

Resultados

Actividad nucleasa CRISPR-Cas y edición génica en protoplastos de *A. thaliana*.

Las nucleasas diseñadas como las nucleasas efectoras TAL (TALENs) y la endonucleasa asociada a repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (CRISPR) Cas9 (CRISPR-Cas9) pueden programarse para dirigirse y escindir ADN de doble cadena con alta especificidad. Los TALEN consisten en dos brazos, ambos con un dominio de unión al ADN similar al efector TAL (TALE) unido a un dominio catalítico de nucleasa de ADN de FokI. Los dominios TALE guían los brazos TALEN a sitios específicos del ADN permitiendo la dimerización de las endonucleasas FokI y la subsiguiente generación de una rotura de doble cadena (DSB). El sistema CRISPR-Cas9 consiste en dos componentes: una nucleasa Cas9 de *Streptococcus pyogenes* (SpCas9) y una fusión quimérica de dos ARN (ARNcr y ARNtracr) denominada guía única modificada (ARNsg). El ARNsg favorece la especificidad del ácido nucleico objetivo para Cas9 mediante el emparejamiento de bases de sus primeras veinte bases 5' con el ADN objetivo, dando lugar posteriormente a un DSB específico del sitio.

En plantas, los DSBs son típicamente reparados por la vía de reparación del ADN de unión de extremos no homólogos (NHEJ), propensa a errores, resultando en delecciones y/o inserciones aleatorias (indeles) en el sitio de reparación. Por el contrario, la edición genómica de precisión se basa en DSB inducidos por nucleasas cerca del cambio que se desea reparar mediante la reparación dirigida por homología (HDR), una vía de reparación más precisa que la NHEJ debido a la necesidad de un ADN molde homólogo, en la mayoría de los casos cromátidas hermanas. Aprovechando la vía de la HDR, es posible utilizar un oligonucleótido modificado como plantilla de reparación para editar el ADN de forma específica, cuando es escindido por nucleasas o de forma no específica cuando se utiliza en combinación con antibióticos inductores de roturas de doble cadena.

La Fig. 16 muestra la actividad nucleasa CRISPR-Cas9 en protoplastos de *Arabidopsis thaliana* derivados del sistema modelo transgénico BFP en donde un gen *BFP* integrado de forma estable puede convertirse en *GFP* editando el codón que codifica H66 (CAC→TAC H66Y). Cuando las células se trataron con CRISPR-Cas9 (BC-1), se produjeron indeles inducidos por NHEJ con una frecuencia del 0,79% cerca del locus H66 del gen *BFP* mediante secuenciación profunda (Fig. 16a). La mayoría de los indeles eran de un solo pb y ninguno superaba los 9 pb. Por el contrario, las células tratadas únicamente con GRON o con un tratamiento simulado no mostraron indeles (datos no mostrados). Estos resultados muestran que la nucleasa CRISPR-Cas9 puede dirigirse activamente al gen *BFP* en este sistema modelo transgénico.

Con respecto a la eficacia de CRISPR-Cas9 en combinación con GRON para mediar la edición génica de *BFP* a *GFP* en protoplastos derivados de nuestro sistema modelo transgénico, se observó una mejora de 7,4 veces en la edición de *BFP* a *GFP* cuando se introducen concurrentemente tanto CRISPR-Cas9 como GRONs modificados con fosforotioato (PS) (CG6) en comparación con tratamientos de GRON solo o CRISPR-Cas9 solo (Fig. 16b). Estos resultados demuestran que la introducción de CRISPR-Cas9 con GRONs modificados con PS en protoplastos de *Arabidopsis* aumenta significativamente la frecuencia de edición génica de *BFP* a *GFP*.

Los GRONs que contienen tres modificaciones PS adyacentes (en adelante 3PS) en los extremos 5' y 3' tienen un efecto positivo en la edición de *BFP* a *GFP* en comparación con un GRON no modificado. El GRON modificado con 3PS (CG2), cuando se combina con CRISPR-Cas9 (BC-1), es más eficaz en la edición de *BFP* a *GFP* en comparación con una plantilla GRON no modificada (CG1; Fig. 17a). Además, se observó una correlación positiva entre la edición y la longitud del GRON (Fig. 17b). En conjunto, estos resultados muestran que tanto la modificación como la longitud del GRON pueden mejorar en gran medida la frecuencia de edición de genes en plantas como *Arabidopsis* en presencia de CRISPR-Cas9.

Cuando el GRON modificado con 201 nucleobases (nb) 3PS (CG6), o los GRON modificados con 201 nb 2'-O-metil (CG9) o (CG10), que consisten en las primeras diez bases 5' como ARN con la primera base de ARN o las primeras 9 bases de ARN modificadas con 2'-O-metil se introducen junto con CRISPR-Cas9 (BC-1) en protoplastos de *Arabidopsis*, no se observó ninguna diferencia estadística en la edición de *BFP* a *GFP* entre ellos (Fig. 17c). De forma similar, cuando el GRON modificado 201 nb 3PS (CG3) o el GRON modificado 201 nb Cy3 (CG4), compuesto por una base 5' Cy3 y una 3' idC inversa, se introdujeron junto con CRISPR-Cas9 (BC-3) en protoplastos de *Arabidopsis*, no se observaron diferencias estadísticas en las frecuencias de edición (Fig 17d). En general, estos datos muestran que diversas modificaciones GRON pueden mejorar en gran medida la frecuencia de edición de genes en *Arabidopsis* en presencia de CRISPR-Cas9.

Basándose en estos resultados con CRISPR-Cas9 y GRONs modificados, se determinó si los GRONs modificados acoplados con pares TALEN dirigidos al gen *BFP* resultan también en una edición mejorada del gen *BFP* a *GFP*. Para mostrar por primera vez la actividad efectiva de las nucleasas en el locus *BFP*, se trataron protoplastos

de *Arabidopsis* con TALENs (BT-1) y se encontró un 0,51 % de indeles en o cerca del sitio de escisión esperado mediante secuenciación profunda, lo que indica que los TALENs son tan activos como CRISPR-Cas9 en este sistema modelo (Fig 18a). La mayoría de las deleciones eran >10 pb, pero inferiores a 50 pb, mientras que las inserciones, significativamente menos abundantes que las deleciones, eran de tres pb o menos. A continuación, examinamos la eficacia de TALENs acoplados con GRON modificado para editar *BFP* a *GFP*. Se observó una mejora de 9,2 veces en la frecuencia de edición de *BFP* a *GFP* cuando se introducen tanto el TALEN BFP (BT-1) como el 3PS GRON (CG7) en comparación con el 3PS GRON solo (Fig. 18b). De forma similar a los experimentos CRISPR-Cas descritos anteriormente, estos resultados demuestran que la introducción de TALENs con GRONs modificados con 3PS en protoplastos de *Arabidopsis* también aumenta significativamente la frecuencia de edición génica de *BFP* a *GFP*.

El loci *EPSPS* (5'-enolpiruvilsiquimato-3-fosfato sintasa) en *Linum usitatissimum* (Lino común) también se utilizó como objetivo en este sistema. Los genes *EPSPS* codifican una enzima de la vía del shikimato que participa en la biosíntesis de los aminoácidos aromáticos fenilalanina, tirosina y triptófano. En las plantas, *EPSPS* es un objetivo del herbicida glifosato, que actúa como inhibidor competitivo del sitio de unión del fosfoenolpiruvato.

Se seleccionaron TALENs dirigidos a un sitio cercano a dos loci (T97I y P101A) en *L. usitatissimum* que cuando se editan hacen a *EPSPS* tolerante al glifosato (Datos Extendidos Fig. 18b). Al introducir TALEN (LuET-1) junto con un GRON modificado con Cy3 5' de 144 nb (CG11) que contenía los cambios objetivo en T97I y P101A en protoplastos, se observaron frecuencias de edición génica del 0,19% en ambos loci y una frecuencia de indel del 0,5% siete días después de la introducción (Fig. 18c, 18d). La mayoría de los indeles eran de 10 pb o menos (Fig. 18c). Estos resultados demuestran que la introducción de TALENs con GRONs modificados con Cy3 en protoplastos de *L. usitatissimum* incrementa significativamente la frecuencia de edición del gen *EPSPS* y además que se pueden realizar múltiples ediciones de nucleótidos con un único GRON.

Ejemplo 20 (no de acuerdo con la invención): Efecto de dos miembros de la familia de antibióticos bleomicina en la conversión.

El propósito de esta serie de ejemplos era evaluar el efecto de los antibióticos en las eficiencias de conversión.

Procedimientos

Protoplastos de una línea de *Arabidopsis thaliana* con múltiples copias de un gen de proteína fluorescente azul se trataron con GRON como se describe en el Ejemplo 1, con la siguiente modificación: antes de la adición de GRON, los protoplastos se mantuvieron durante 90 minutos en hielo en una solución de TM (14.8 mM $MgCl_2 \times 6H_2O$, 5 mM ácido 2-(N-morfolino)etanosulfónico, 572 mM manitol), suplementada con 0, 250, o 1000 µg/ml de Zeocina™ o fleomicina. El pH de las soluciones se ajustó a 7,0. El porcentaje de células con fluorescencia verde resultante de la conversión de BFP a GFP se evaluó mediante citometría de flujo, tal como se describe en el Ejemplo 1.

Resultados

Zeocina y fleomicina en ambas concentraciones usadas (250 y 1000 µg/ml) resultaron en un incremento en la edición del gen BFP a GFP (ver Figura 19). Se observaron células con fluorescencia verde resultantes de la edición génica de BFP a GFP cinco días después de la administración de GRON (Figura 20).

Referencias

1. LeCong et al 2013 Science : vol. 339 no. 6121 pp. 819-823.
2. Jinek et al 2012 Science. 337:816-21
3. Wang et al 2008 RNA 14: 903-913
4. Zhang et al 2013. Plant Physiol. 161: 20-27

Ejemplo 21: CRISPRs y GRONs en el arroz.

El propósito de este experimento es demostrar la conversión ACCasa en *Oryza sativa* a las 120 horas después de la entrega de PEG de plásmidos CRISPR-Cas y GRONs en protoplastos. El CRISPR-Cas utilizado en este experimento se dirige al gen ACCasa del genoma del arroz mediante la introducción en protoplastos de plásmidos que codifican el gen Cas9 y un ARNsg. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). El ARNcr guía al Cas9 hasta los genes objetivo, donde el Cas9 crea una rotura de doble cadena en el gen de *accasa* y el GRON se utiliza como plantilla para convertir el gen de *accasa* de forma dirigida al sitio.

Procedimientos

Se aislaron protoplastos de arroz a partir de callos. Los plásmidos codificados con CRISPR-Cas contienen un promotor de ubiquitina de maíz que impulsa la secuencia codificante de Cas9 con un terminador *rbcSE9* y un promotor U6 de arroz que impulsa el ARNsg con un terminador *poly-T*₁₀. Los plásmidos CRISPR-Cas se introdujeron en los protoplastos mediante entrega mediada por PEG a una concentración final de 0,05 µg/µl. GRONs con la secuencia

siguiente, 5' V C TGA CCT GAA CTT GAT CTC AAT TAA CCC TTG CGG TTC CAG AAC ATT GCC TTT TGC AGT CCT CTC AGC ATA GCA CTC AAT GCG GTC TGG GTT TAT CTT GCT TCC AAC GAC AAC CCA AGC CCC TCC TCG TAG CTC TGC AGC CAT GGG AAT GTA GAC AAA GGC AGG CTG ATT GTA TGT CCT AAG GTT CTC AAC AAT AGT CGA GCC H 3', se utilizaron a una concentración final de 0,8 µM. Los protoplastos se incrustaron en agarosa (2,5 × 10⁶ células/ml), se cultivaron en medio líquido y se incubaron en un agitador rotatorio (60 rpm) en la oscuridad a 28°C. Se analizaron muestras individuales mediante NGS, 120 horas después de la entrega del plásmido CRISPR-Cas y/o GRON, para determinar el porcentaje de células (lecturas de ADN) portadoras de la conversión ACCasa y que tenían indeles en el gen ACCasa.

La CRISPR consta de dos componentes: el *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir del mismo plásmido. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene la secuencia espaciadora (5'-ACGAGGAGGGGCTTGGGTTGTGG-3') utilizada para guiar a la nucleasa Cas9 hasta el gen objetivo. En este experimento, el CRISPR se dirige al gen ACCasa.

Resultados

A las 120 h, los protoplastos de arroz tienen un 0,026% de conversión de ACCasa de acuerdo con lo determinado por Next Generation Sequencing. Los controles sólo GRON sin CRISPR-Cas mostraron una conversión mínima de ACCasa del 0,002% a las 120 horas y los controles no tratados no mostraron ninguna conversión. Además, estos datos muestran que el CRISPR-Cas es activo y capaz de escindir el gen objetivo ACCasa y formar indeles del 8,0%.

Referencias

5. LeCong et al 2013 Science : vol. 339 no. 6121 pp. 819-823.
6. Jinek et al 2012 Science. 337:816-21
7. Wang et al 2008 RNA 14: 903-913
8. Zhang et al 2013. Plant Physiol. 161: 20-27

Ejemplo 22: CRISPR y GRON en el arroz

Sumario: Se han identificado mutaciones específicas de *la* ACCasa en callos de diecinueve semanas mediante análisis de PCR y secuenciación de ADN.

Introducción

El propósito de este experimento es demostrar la conversión ACCasa en callos de *Oryza sativa* después de la entrega de PEG de plásmidos CRISPR-Cas y GRONs en protoplastos. El CRISPR-Cas utilizado en este experimento se dirige al gen ACCasa del genoma del arroz mediante la introducción en protoplastos de plásmidos que codifican el gen Cas9 y un ARNsg. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). El ARNcr guía al Cas9 hasta el gen objetivo, donde el Cas9 crea una rotura de doble cadena o una mella en el lugar deseado del gen de *la* ACCasa y los GRON se utilizan como plantilla para convertir el gen de *la* ACCasa de forma dirigida.

Resultados

Las mutaciones dirigidas de *OsACCasa* descritas en las tablas siguientes, se han identificado en callos de diecinueve semanas mediante PCR y análisis de secuenciación de ADN.

Procedimientos

Se aislaron protoplastos de arroz a partir de cultivos en suspensión iniciados a partir de callos embriogénicos derivados de semillas maduras. Los plásmidos CRISPR-Cas con GRONs a una concentración final de 0,05 µg/µl y 0,8 µM, respectivamente, se introdujeron en protoplastos mediante el procedimiento de entrega mediado por PEG. Un intervalo ejemplar para los plásmidos CRISPR-Cas en una concentración final incluye, pero no se limita a 0,01 a 0,2 µg/µl. Un intervalo ejemplar para el GRON en una concentración final incluye, pero no se limita a 0,01 a 4 µM. Los plásmidos codificados con CRISPR-Cas contienen un promotor de ubiquitina de maíz que impulsa la secuencia codificante de Cas9 con un terminador *rbcSE9* y un promotor U6 de arroz que impulsa el ARNsg con un terminador *poly-T₁₀*. La información sobre la secuencia de los GRON se describe en la Tabla 3. Tras el tratamiento con PEG, los protoplastos se incrustaron en agarosa (1,25 × 10⁶ células/ml), se cultivaron en medio líquido y se incubaron en un agitador rotatorio (60 rpm) en la oscuridad a 28°C. Un intervalo ejemplar para incrustar protoplastos en agarosa incluye, entre otros, de 0,625 × 10⁶ a 2,5 × 10⁶ células/ml. Las muestras de cada tratamiento se analizaron mediante secuenciación de próxima generación después de 4 semanas tras el tratamiento con plásmido CRISPR-Cas y/o GRON para determinar la proporción de células (lecturas de ADN) portadoras de la conversión ACCasa. Los microcallos de los tratamientos convertidos se liberaron en un medio de selección sólido que contenía cletodim (0,25-0,5 µM) o setoxidim (2,5 µM). Las líneas de callos individuales que crecieron en este medio de selección tras 19 semanas de cultivo se analizaron

mediante procedimientos de cribado internos, así como mediante secuenciación del ADN, con el fin de identificar los callos individuales que contenían las conversiones de ACCasa objetivo.

El CRISPR consta de dos componentes: el Cas9, tal como un *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales y el ARNsg, ambos expresados a partir de plásmido(s). Cas9 y ARNsg también se pueden liberar mediante ARNm/ARN o proteína/ARN, respectivamente. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene el espaciador con las secuencias descritas en la Tabla 2, que se utilizaron para guiar a la nucleasa Cas9 hasta el gen objetivo. En este experimento, el CRISPR se dirige al gen ACCasa del arroz.

Lista de conversiones en OsACCasa en dos lugares diferentes dentro del gen, Sitio 1 y Sitio 2. Para cada sitio, todas las combinaciones de eventos de conversión son posibles.

OsACCasa	
Sitio1	Sitio2
I1781A	D2078G
I1781L	D2078K
I1781M	D2078T
I1781N	S2079F
I1781S	K2080E
I1781T	C2088F
I1781V	C2088G
G1783C	C2088H
A1786P	C2088K
	C2088L
	C2088N
	C2088P
	C2088Q
	C2088R
	C2088S
	C2088T
	C2088V
	C2088W

Lista de secuencias espaciadoras de ARNg CRISPR-Cas utilizadas en este experimento. La longitud del espaciador puede variar hasta ± 20 pb. Los desajustes en la secuencia espaciadora pueden tolerarse hasta 10 pb.

Muestra ID	Secuencia de ARN espaciador (5' a 3')	OsACCasa	
		Sitio1	Sitio2
1	CAUAAGAUGCAGCUAGACAG	X	
2	AGCUAGACAGUGGUGAAAUU	X	
3	UAGACAGUGGUGAAAUUAGG	X	
4	AGACAGUGGUGAAAUUAGGU	X	
5	GUGGGUUAUUGAUUCUGUUG	X	
6	UGGGUUAUUGAUUCUGUUGU	X	
7	UAUUGAUUCUGUUGUGGGCA	X	
8	UCUGUUGUGGGCAAGGAAGA	X	
9	GUGGGCAAGGAAGAUGGACU	X	
10	CAAGGAAGAUGGACUUGGUG	X	
11	CUUGGUGUGGAGAAUAUACA	X	

Muestra ID	Secuencia de ARN espaciador (5' a 3')	OsACCasa	
		Sitio1	Sitio2
12	CUAUUGCCAGUGCUUAUUCU	X	
13	UGCUUAUUCUAGGGCAUAUA	X	
14	UUUACACUUACAUUUGUGAC	X	
15	UUUGUGACUGGAAGAACUGU	X	
16	ACUGGAAGAACUGUUGGAAU	X	
17	GGAGCUUAUCUUGCUCGACU	X	
18	AUAUGCCCUAGAAUAAGCAC	X	
19	GAUAAGAUGCAGCUAGACAG	X	
20	GGCUAGACAGUGGUGAAAUU	X	
21	GAGACAGUGGUGAAAUUAGG	X	
22	GGACAGUGGUGAAAUUAGGU	X	
24	GGGGUUAUUGAUUCUGUUGU	X	
25	GAUUGAUUCUGUUGUGGGCA	X	
26	GCUGUUGUGGGCAAGGAAGA	X	
28	GAAGGAAGAUGGACUUGGUG	X	
29	GUUGGUGUGGAGAAUAUACA	X	
30	GUAUUGCCAGUGCUUAUUCU	X	
31	GGCUUAUUCUAGGGCAUAUA	X	
32	GUUACACUUACAUUUGUGAC	X	
33	GUUGUGACUGGAAGAACUGU	X	
34	GCUGGAAGAACUGUUGGAAU	X	
36	GUAUGCCCUAGAAUAAGCAC	X	
37	CGACUAUUGUUGAGAACCUU		X
38	UGCCUUUGUCUACAUUCCCA		X
39	CCCAUGGCUGCAGAGCUACG		X
40	AUGGCUGCAGAGCUACGAGG		X
41	UGGCUGCAGAGCUACGAGGA		X
42	GGCUGCAGAGCUACGAGGAG		X
43	CAGAGCUACGAGGAGGGGCU		X
44	AGAGCUACGAGGAGGGGCUU		X
45	ACGAGGAGGGGCUUGGGUUG		X
46	GCAUUGAGUGCUAUGCUGAG		X
47	UAUGCUGAGAGGACUGCAAA		X
48	GACUGCAAAAGGCAAUGUUC		X
49	GGCAAUGUUCUGGAACCGCA		X
50	GCAAUGUUCUGGAACCGCAA		X
51	GGUUAUUGAGAUCAAGUUC		X
52	UGAGAUCAAGUUCAGGUCAG		X
53	GUUCAGGUCAGAGGAACUCC		X
54	AACUCCAGGAUUGCAUGAGU		X
55	CAAGCCGACUCAUGCAAUCC		X
56	UUGAUCUCAAUUAACCCUUG		X
57	UCUCAGCAUAGCACUCAUG		X

Muestra ID	Secuencia de ARN espaciador (5' a 3')	OsACCasa	
		Sitio1	Sitio2
58	GCAUAGCACUCA AUGCGGUC		X
59	CAUAGCACUCA AUGCGGUCU		X
60	UCCUCGUAGCUCUGCAGCCA		X
61	AGCCAUGGGAAUGUAGACAA		X
62	AUGGGAAUGUAGACAAAGGC		X
63	CAGGCUGAUUGUAUGUCCUA		X
64	GGACUAUUGUUGAGAACCUU		X
65	GGCCUUUGUCUACAUUCCCA		X
66	GCCAUGGCUGCAGAGCUACG		X
67	GUGGCUGCAGAGCUACGAGG		X
68	GGGCUGCAGAGCUACGAGGA		X
69	GAGAGCUACGAGGAGGGGCU		X
70	GGAGCUACGAGGAGGGGCUU		X
71	GCGAGGAGGGGCUUGGGUUG		X
72	GAUGCUGAGAGGACUGCAAA		X
73	GGAGAUCAAGUUCAGGUCAG		X
74	GACUCCAGGAUUGCAUGAGU		X
75	GAAGCCGACUCAUGCAAUCC		X
76	GUGAUCUCAAUUAACCCUUG		X
77	GCUCAGCAUAGCACUCA AUG		X
78	GAUAGCACUCA AUGCGGUCU		X
79	GCCUCGUAGCUCUGCAGCCA		X
80	GGCCAUGGGAAUGUAGACAA		X
81	GUGGGAAUGUAGACAAAGGC		X
82	GAGGCUGAUUGUAUGUCCUA		X

Una lista de secuencias GRON adecuadas para su uso en este experimento se proporcionan en la siguiente tabla (V=CY3; H=3' DMT dC CPG).

1	VGTCATAGCACATAAGATGCAGCTAGACAGTGGTGAAATTAGGTGGGTTAT TGATTCTGTTGTGGGCAAGGAAGATGGACTTGGTGTGGAGAATGCTCATGG AAGTGCTGCTATTGCCAGTGCTTATTCTAGGGCATATAAGGAGACATTTAC ACTTACATTTGTGACTGGAAGAACTGTTGGAATAGGAGCTTATCTTGCTCH
2	VGAGCAAGATAAGCTCCTATTCCAACAGTTCTTCCAGTCACAAATGTAAGT GTAAATGTCTCCTTATATGCCCTAGAATAAGCACTGGCAATAGCAGCACTT CCATGCAGATTCTCCACACCAAGTCCATCTTCCTTGCCACAACAGAATCA ATAACCCACCTAATTTCACTGTCTAGCTGCATCTTATGTGCTATGACH
3	VGTCATAGCACATAAGATGCAGCTAGACAGTGGTGAAATTAGGTGGGTTATTGATT CTGTTGTGGGCAAGGAAGATGGACTTGGTGTGGAGAATATACATTGCAGTGCTGCT ATTGCCAGTGCTTATTCTAGGGCATATAAGGAGACATTTACACTTACATTTGTGAC TGAAGAACTGTTGGAATAGGAGCTTATCTTGCTCH

(continuación)

4	VGAGCAAGATAAGCTCCTATTCCAACAGTTCTTCCAGTCACAAATGTAAGTGTAAT TGTCTCCTTATATGCCCTAGAATAAGCACTGGCAATAGCAGCACTGCAATGTATAT TCTCCACACCAAGTCCATCTTCCCTTGCCCACAACAGAATCAATAACCCACCTAATT TCACCACTGTCTAGCTGCATCTTATGTGCTATGACH
5	VGTCATAGCACATAAGATGCAGCTAGACAGTGGTGAAATTAGGTGGGTAT TGATTCTGTTGTGGGCAAGGAAGATGGACTTGGTGTGGAGAATATACATGG AAGTGCTCCAATTGCCAGTGCTTATTCTAGGGCATATAAGGAGACATTTAC ACTTACATTTGTGACTGGAAGAAGTGTGGAATAGGAGCTTATCTTGCTCH
6	VGAGCAAGATAAGCTCCTATTCCAACAGTTCTTCCAGTCACAAATGTAAGT GTAAATGTCTCCTTATATGCCCTAGAATAAGCACTGGCAATGGTAGCACTT CCATGTATATTCTCCACACCAAGTCCATCTTCCCTTGCCCACAACAGAATCAA TAACCCACCTAATTTACCACTGTCTAGCTGCATCTTATGTGCTATGACH
7	VCTGACCTGAACTTGATCTCAATTAACCCTTGCGGTTCCAGAACATTGCCTT TTGCAGTCCTCTCAGCATAGCACTCAATGCGGTCTGGGTTTATCTTGCTTCC AACGACAACCCAAGCCCCTCCTCGTAGCTCTGCAGCCATGGGAATGTAGAC AAAGGCAGGCTGATTGTATGTCCTAAGGTTCTCAACAATAGTCGAGCCH
8	VGGCTCGACTATTGTTGAGAACCTTAGGACATACAATCAGCCTGCCTTTGTC TACATTTCCATGGCTGCAGAGCTACGAGGAGGGGCTTGGGTTGTGGTTGGT AGCAAGATAAACCAGACCGCATTGAGTGCTATGCTGAGAGGACTGCAAA AGGCAATGTTCTGGAACCGCAAGGGTTAATTGAGATCAAGTTCAGGTCAGH
9	VCTGACCTGAACTTGATCTCAATTAACCCTTGCGGTTCCAGAACATTGCCTT TTGCAGTCCTCTCAGCATAGCACTCAATGCGGTCTGGGTTTATCTTAAAATC AACGACAACCCAAGCCCCTCCTCGTAGCTCTGCAGCCATGGGAATGTAGAC AAAGGCAGGCTGATTGTATGTCCTAAGGTTCTCAACAATAGTCGAGCCH
10	VGGCTCGACTATTGTTGAGAACCTTAGGACATACAATCAGCCTGCCTTTGTC TACATTTCCATGGCTGCAGAGCTACGAGGAGGGCGCTTGGGTTGTGGTTGAT AGCAAGATAAACCAGACCGCATTGAGAGGTATGCTGAGAGGACTGCAAA
	AGGCAATGTTCTGGAACCGCAAGGGTTAATTGAGATCAAGTTCAGGTCAGH
11	VGGCTCGACTATTGTTGAGAACCTTAGGACATACAATCAGCCTGCCTTTGTC TACATTTCCATGGCTGCAGAGCTACGAGGAGGGGCTTGGGTTGTGGTTGAT AGCGAAATAAACCAGACCGCATTGAGTGCTATGCTGAGAGGACTGCAAA AGGCAATGTTCTGGAACCGCAAGGGTTAATTGAGATCAAGTTCAGGTCAGH
12	VCTGACCTGAACTTGATCTCAATTAACCCTTGCGGTTCCAGAACATTGCCTT TTGCAGTCCTCTCAGCATAGCACTCAATGCGGTCTGGGTTTATTTGCTATC AACCACAACCCAAGCGCCTCCTCGTAGCTCTGCAGCCATGGGAATGTAGAC AAAGGCAGGCTGATTGTATGTCCTAAGGTTCTCAACAATAGTCGAGCCH

(continuación)

13	VGGCTCGACTATTGTTGAGAACCTTAGGACATACAATCAGCCTGCCTTTGTC TACATTCCCATGGCTGCAGAGCTACGAGGAGGGGCTTGGGTTGTGGTTGAT AGCAAGATAAACCCAGACCGCATTGAGCGTTATGCTGAGAGGACTGCAAA AGGCAATGTTCTGGAACCGCAAGGGTTAATTGAGATCAAGTTCAGGTCAGH
14	VCTGACCTGAACTTGATCTCAATTAACCCTTGCGGTTCCAGAACATTGCCTT TTGCAGTCCTCTCAGCATATTGCTCAATGCGGCTCTGGGTTTATCTTGCTATC AACGACAACCCAAGCCCCTCCTCGTAGCTCTGCAGCCATGGGAATGTAGAC AAAGGCAGGCTGATTGTATGTCCTAAGGTTCTCAACAATAGTCGAGCCH

Ejemplo 23: CRISPR y GRON en el lino

- 5 **Sumario:** Se han identificado mutaciones específicas de *LuEPSPS* en callos de cuatro semanas mediante análisis de PCR y secuenciación de ADN.

Introducción

- 10 El propósito de este experimento es demostrar la conversión de los genes *EPSPS* en el genoma de *Linum usitatissimum* en protoplastos derivados de puntas de brotes mediante la entrega mediada por PEG de plásmidos CRISPR-Cas y GRONs. Los CRISPR-Cas y GRON utilizados en este experimento se dirigen a los genes *EPSPS* del genoma del lino. El CRISPR consta de dos componentes: un Cas9, tal como un *Streptococcus pyogenes* Cas9 (SpCas9) optimizado para codones vegetales, y un ARNsg, ambos expresados a partir de plásmido(s). Cas9 y ARNsg también se pueden liberar mediante ARNm/ARN o proteína/ARN, respectivamente. El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). El ARNcr guía al Cas9 hasta los genes objetivo, donde el Cas9 crea una rotura o mella de doble cadena en los genes *EPSPS* y los GRON se utilizan como plantilla para convertir los genes *EPSPS* de forma dirigida.

Resultados

Las mutaciones *LuEPSPS* dirigidas (T97I y/o la P101A, P101T o P101S y/o la G96A) se han identificado en callos de cuatro semanas de edad mediante PCR y análisis de secuenciación de ADN. Se han regenerado brotes a partir de estos callos convertidos.

Procedimientos

- Se aislaron protoplastos de lino a partir de puntas de brotes obtenidas de plántulas germinadas *in vitro*. Los plásmidos codificados CRISPR-Cas contienen el promotor MAS que impulsa la secuencia codificadora Cas9 con un terminador rbcSE9 y el promotor U6 de *Arabidopsis thaliana* que impulsa el ARNsg con un terminador poly-T₁₀. Los plásmidos CRISPR-Cas se introdujeron en los protoplastos mediante entrega mediada por PEG a una concentración final de 0,05 µg/µl. Se utilizaron GRONs dirigidos a cada uno de los dos genes *LuEPSPS* del lino (Tabla 2) a una concentración final de 4,0 µM. Un intervalo ejemplar para los plásmidos CRISPR-Cas en una concentración final incluye, pero no se limita a 0,01 a 0,2 µg/µl. Un intervalo ejemplar para el GRON en una concentración final incluye, pero no se limita a 0,01 a 4 µM. Los protoplastos se incrustaron en microesferas de alginato (5 × 10⁵ células/ml), se cultivaron en medio líquido y se incubaron en un agitador rotatorio (30 rpm) en la oscuridad a 25°C. Un intervalo ejemplar para incrustar protoplastos en perlas de alginato incluye, pero no se limita a 3,0 × 10⁵ a 7,5 × 10⁵ células/ml. Los microcallos desarrollados a partir de células individuales se analizaron mediante secuenciación de próxima generación, 3 y 7 semanas después de la administración del plásmido CRISPR-Cas y GRON, para determinar la proporción de células (lecturas de ADN) portadoras de las mutaciones objetivo en los genes *LuEPSPS*. Se cultivaron callos más grandes a partir de microcallos convertidos de 8 semanas de edad colocados sobre un medio de regeneración sólido, y los brotes empezaron a diferenciarse a partir de los callos regenerados al cabo de unas 4-8 semanas. Los callos y brotes convertidos con las mutaciones del gen *EPSPS* objetivo se identificaron mediante análisis de PCR y secuenciación del ADN.

- El CRISPR consta de dos componentes: la planta de codón optimizado *Streptococcus pyogenes* Cas9 (SpCas9) y ARNsg, ambos expresados a partir de plásmido(s). El ARNsg es una fusión de ARN CRISPR (ARNcr) y ARNcr transactivador (ARNtracr). La región ARNcr contiene el espaciador con las secuencias descritas en la tabla siguiente, que se utilizaron para guiar la nucleasa Cas9 a los genes objetivo *EPSPS*.
- Lista de secuencias espaciadoras de ARNsg CRISPR-Cas utilizadas en este experimento. La longitud del espaciador puede variar hasta ±20 pb. Los desajustes en la secuencia espaciadora pueden tolerarse hasta 10 pb.

Muestra ID	Secuencia de ARN espaciador (5' a 3')	Genes LuEPSPS	
		1	2
1	CAGAAGCGCGCCAUUGUUGA	X	X
2	CGCGCCAUUGUUGAAGGUUG	X	
3	CGCGCCAUUGUUGAAGGUCG		X
4	GCCAUUGUUGAAGGUUGUGG	X	
5	GCCAUUGUUGAAGGUCGUGG		X
6	AGGUUGUGGUGGUGUGUUUC	X	
7	AGGUCGUGGUGGUGUGUUUC		X
8	UGUGGUGGUGUGUUUCCGGU	X	
9	CGUGGUGGUGUGUUUCCGGU		X
10	UGUGUUUCCGGUCGGUAAAC	X	X
11	UGUUUCCGGUCGGUAAACUG		X
12	AACGAUAUUGAACUUUCCU	X	
13	AACGAUAUCGAACUUUCCU		X
14	GAACUUUCCUUGGAAAUGC	X	X
15	ACAGCUGCUGUAACAGCCGC	X	X
16	GCUGCUGUAACAGCCGCUGG	X	X
17	AACUCAAGCUACAUACUCGA	X	
18	AACUCAAGCUACAUACUCGA		X
19	CGAAUGAGAGAGAGACCAAU	X	
20	CGAAUGAGAGAGAGACCGAU		X
21	AGAGAGACCAAUUGGAGAUU	X	
22	CCAAUUGGAGAUUUGGUUGU	X	
23	CCGAUUGGAGAUUAGUUGU		X
24	CCAACAACCAAUCUCCAAU	X	
25	CCAACAACUAAAUCUCCAAU		X
26	AUUGGUCUCUCUCUCAUUCG	X	
27	AUCGGUCUCUCUCUCAUUCG		X
28	GUAGCUUGAGUUGCCUCCAG	X	X
29	GCUGUUACAGCAGCUGUCAG	X	X
30	UAGCUGUCCAGCAUUUCCA	X	X
31	UUCUUCGCCAGUUUACCGAC	X	
32	UUCUUCCCCAGUUUACCGAC		X
33	ACCACCACAACCUUCAACAA	X	
34	ACCACCACGACCUUCAACAA		X
35	GAGAAGCGCGCCAUUGUUGA	X	X
36	GGCGCCAUUGUUGAAGGUUG	X	
37	GGCGCCAUUGUUGAAGGUCG		X
38	GGGUUGUGGUGGUGUGUUUC	X	
39	GGGUCGUGGUGGUGUGUUUC		X
40	GGUGGUGGUGUGUUUCCGGU	X	
41	GGUGGUGGUGUGUUUCCGGU		X
42	GGUGUUUCCGGUCGGUAAAC	X	X
43	GGUUUCCGGUCGGUAAACUG		X

(continuación)

Muestra ID	Secuencia de ARN espaciador (5' a 3')	Genes LuEPSPS	
		1	2
44	GACGAUUAUUGAACUUUUCU	X	
45	GACGAUAUCGAACUUUUCU		X
46	GCAGCUGCUGUAACAGCCGC	X	X
47	GACUCAAGCUACAUACUCGA	X	
48	GACUCAAGCUACAUACUCGA		X
49	GGAAUGAGAGAGAGACCAU	X	
50	GGAAUGAGAGAGAGACCGAU		X
51	GGAGAGACCAAUUGGAGAUU	X	
52	GCAAUUGGAGAUUUGGUUGU	X	
53	GCGAUUGGAGAUUUGUUGU		X
54	GCAACAACCAAUCUCCAAU	X	
55	GCAACAACUAAAUCUCCAAU		X
56	GUUGGUCUCUCUCUCAUUCG	X	
57	GUCGGUCUCUCUCUCAUUCG		X
58	GAGCUGUCCAGCAUUUCCA	X	X
59	GUCUUCGCCAGUUUACCGAC	X	
60	GUCUUCCCCAGUUUACCGAC		X
61	GCCACCACAACCUUCAACAA	X	
62	GCCACCACGACCUUCAACAA		X

Una lista de secuencias GRON adecuadas para su uso en este experimento se proporcionan en la siguiente tabla (V=CY3; H=3'DMT dC CPG).

5

1	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAATGCT GCTATAGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
2	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGGACGCATAGCTGTGGCAGCATT TCCAAGGAAAAGTTCAATATCGTTCTTCGCCAGTTTACCGACCGH
3	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAATGCTGGAAT AGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAACTCAAGGTCC CTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
4	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGCACGCATAGCTATTCCAGCATT TCCAAGGAAAAGTTCAATATCGTTCTTCGCCAGTTTACCGACCGH
5	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAATGCT GGAATTGCTATGCGTTCTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH

(continuación)

6	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGTGAACGCATAGCAATTCCAGCATT TCCAAGGAAAAGTTCAATATCGTTCCTTCGCCAGTTTACCGACCGH
7	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAATGCT GGAATCGCTATGCGTACTCTGACAGCTGCTGTAAACAGCCGCTGGAGGCAAC TCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
8	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGTGTACGCATAGCAATTCCAGCATT TCCAAGGAAAAGTTCAATATCGTTCCTTCGCCAGTTTACCGACCGH
9	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAATGCT GGAATTGCTATGCGTGCGCTGACAGCTGCTGTAAACAGCCGCTGGAGGCAAC TCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
10	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGCACGCATAGCTATTCCAGCATT TCCAAGGAAAAGTTGATATCGTTCCTTCCCCAGTTTACCGACCGH
11	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAATGCT GGAATAGCTATGCGTGCTCTGACAGCTGCTGTAAACAGCCGCTGGAGGCAAC TCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
12	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGCACGCATAGCTGTTCCAGCATT TCCAAGGAAAAGTTCAATATCGTTCCTTCGCCAGTTTACCGACCGH
13	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAATGCT GCTATAGCTATGCGTGCGCTGACAGCTGCTGTAAACAGCCGCTGGAGGCAAC TCAAGGTCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
14	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGGACGCATAGCTGTTGCAGCATT TCCAAGGAAAAGTTGATATCGTTCCTTCCCCAGTTTACCGACCGH
15	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAATGCT GGAATAGCTATGCGTGCGCTGACAGCTGCTGTAAACAGCCGCTGGAGGCAAC TCAAGGTCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
16	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGCACGCATAGCTATTCCAGCATT TCCAAGGAAAAGTTGATATCGTTCCTTCCCCAGTTTACCGACCGH

(continuación)

17	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATTGCTATGCGTTCTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
18	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGTGAACGCATAGCGATTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
19	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATCGCTATGCGTACTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
20	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGTGTACGCATAGCTATTCCAGCATTT CCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
21	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATCGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
22	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGGACGCATAGCTATTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
23	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATAGCTATGCGTGCTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
24	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGGGAACGCATAGCTGTTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH

REIVINDICACIONES

1. Un procedimiento para provocar un cambio genético en una célula vegetal, dicho procedimiento comprende exponer dicha célula a un cortador de ADN y a una oligonucleobasa reparadora de genes (GRON),
 5 en donde el cortador de ADN es CRISPR-Cas,
 en donde el GRON incluye una o más alteraciones de los nucleótidos convencionales de ARN y ADN, seleccionándose dichas alteraciones entre uno o más nucleótidos 2'O-metilo en los extremos 5' y/o 3', una base inversa (idC) en el extremo 3', una modificación O-(2-metoxietilo) (MOE) en los extremos 5' y/o 3', uno
 10 o más grupos Cy3 en los extremos 5' y/o 3', donde Cy3 denota un fosforoamidito bloqueado que al incorporarse a un oligonucleótido produce un colorante de indomonocarbocianina sustituido con 3,3,3',3'-tetrametil N,N'-isopropil, y uno o más grupos 3PS en los extremos 5' y/o 3', en donde 3PS denota 3 enlaces fosfotioato, y en donde dicho GRON tiene entre 15 y 210 nucleótidos de longitud.
- 15 2. El procedimiento de la reivindicación 1, en donde dicho GRON es monocatenario.
3. El procedimiento de cualquiera de las reivindicaciones 1 o 2, en donde el GRON tiene un par de bases de cabeceo en relación con la secuencia objetivo para el cambio genético.
- 20 4. El procedimiento de cualquiera de las reivindicaciones precedentes, en donde dicha planta se selecciona del grupo que consiste en canola, girasol, maíz, tabaco, remolacha azucarera, algodón, maíz, trigo, cebada, arroz, alfalfa, cebada, sorgo, tomate, mango, melocotón, manzana, pera, fresa, plátano, melón, patata, zanahoria, lechuga, cebolla, soja, soja spp, caña de azúcar, guisante, garbanzo, guisante forrajero, habas, lentejas, nabo, colinabo, coles de Bruselas, altramuza, coliflor, col rizada, haboncillos, álamo, pino, eucalipto, uva, cítricos, triticale, alfalfa, centeno, avena,
 25 césped y gramíneas forrajeras, lino, colza, mostaza, pepino, gloria de la mañana, bálsamo, pimienta, berenjena, caléndula, loto, col, margarita, clavel, tulipán, iris, yuca, lirio y plantas productoras de frutos secos.
5. El procedimiento de cualquiera de las reivindicaciones precedentes, en donde dicha planta es canola, o en donde dicha planta es maíz, o en donde dicha planta es trigo, o en donde dicha planta es arroz, o en donde dicha planta es sorgo, o en donde dicha planta es patata, o en donde dicha planta es soja, o en donde dicha planta es lino, o en donde
 30 dicha planta es colza, o en donde dicha planta es mandioca, o en donde dicha planta es girasol.
6. El procedimiento de cualquiera de las reivindicaciones anteriores, en donde dicha planta es una planta del género Brassica seleccionada de B. carinata, B. elongate, B. fruticulosa, B. juncea, B. napus, B. narinosa, B. nigra, B. oleracea, B. perviridis, B. rapa (syn B. campestris), B. rupestris, B. septiceps y B. tournefortii, en particular en donde la planta es
 35 B. carinata.
7. El procedimiento de cualquiera de las reivindicaciones anteriores en donde se realizan múltiples cambios genéticos.
- 40 8. El procedimiento de cualquiera de las reivindicaciones anteriores en donde se utilizan dos o más ARN guía.
9. El procedimiento de la reivindicación 8, en donde cada uno de los más de un ARN guía es complementario a un objetivo diferente para el cambio genético.
- 45 10. El procedimiento de cualquiera de las reivindicaciones anteriores, en donde el CRISPR-Cas incluye una nickasa.
11. El procedimiento de cualquiera de las reivindicaciones precedentes, en donde el CRISPR-Cas incluye dos o más nickasas, y/o en donde dos o más cortes de nickasas están en cadenas opuestas de la secuencia de ácido nucleico objetivo, o en donde dos o más cortes de nickasas están en la misma cadena de la secuencia de ácido nucleico
 50 objetivo.

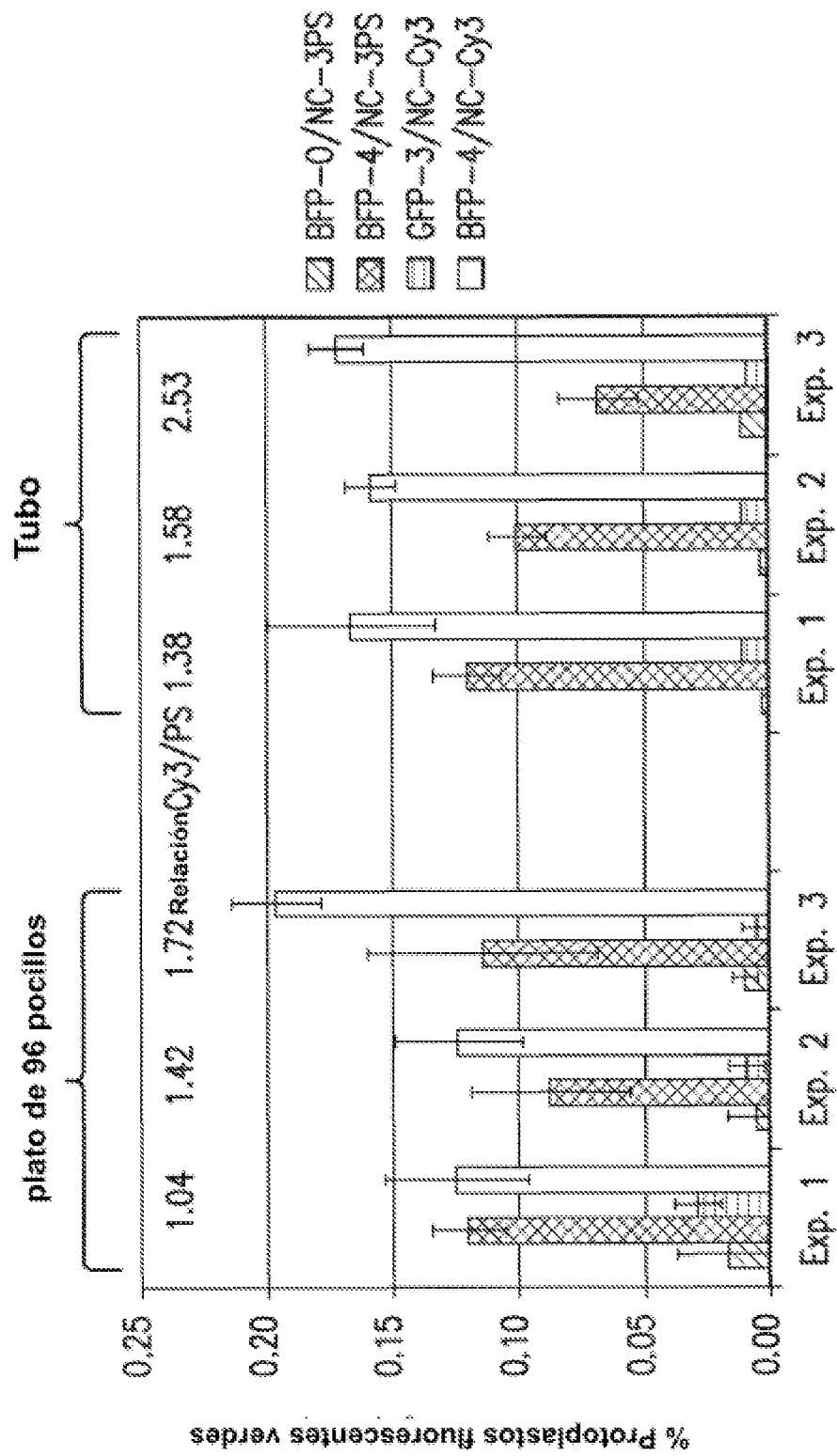


FIG.1

A. Fragmento Okazaki GRONs (BFP4 o 0/C o NC/71/5' 2' -O-Me (1))

BFP4/NC: U UCA UGU GG U CGG GGT AGC G GC TGA AGC ACT GCA CGC CGT AGG TGA AGG
TGG TCA CGA GGG TGG GCC AGG G (71 -mer)

BFP0/NC: U UCA UGU GG U CGG GGT AGC G GC TGA AGC ACT GCA CGC CGT GGG TGA AGG
TGG TCA CGA GGG TGG GCC AGG G (71 -mer)

BFP4/C: G CUG CCC GUG CCC TGG CCC A CC CTC GTG ACC ACC TTC ACC IAC GGC GTG
CAG TGC TTC AGC CGC TAC CCC G (71 -mer)

BFP0/C: G CUG CCC GUG CC C TGG CCC A CC CTC GTG ACC ACC TTC ACC CAC GGC GTG
CAG TGC TTC AGC CGC TAC CCC G (71 -mer)

Bases ARN

Grupo 2' - O -Me en primer base ARN 5'

(bases de ARN sin los grupos 2' - O -Me están en cajas)

B. Fragmento Okazaki GRONs (BFP4 or 0 /C or NC/71/5' 2' -O-Me (9))

BFP4/NC: U UCA UGU GG U CGG GGT AGC G GC TGA AGC ACT GCA CGC CGT AGG TGA AGG
TGG TCA CGA GGG TGG GCC AGG G (71 -mer)

BFP0/NC: U UCA UGU GG U CGG GGT AGC G GC TGA AGC ACT GCA CGC CGT GGG TGA AGG
TGG TCA CGA GGG TGG GCC AGG G (71 -mer)

BFP4/C: G CUG CCC GUG CC C TGG CCC A CC CTC GTG ACC ACC TTC ACC IAC GGC GTG
CAG TGC TTC AGC CGC TAC CCC G (71 -mer)

BFP0/C: G CUG CCC GUG CCC TGG CCC A CC CTC GTG ACC ACC TTC ACC CAC GGC GTG
CAG TGC TTC AGC CGC TAC CCC G (71 -mer)

Bases ARN

Grupo 2'-O-Me en todas las bases de ARN excepto en la base más cercana a la primera base de ADN del GRON

(bases de ARN sin los grupos 2' - O -Me están en cajas)

FIG.2

Gen BFP

5' CCTTCACCCACGGCGTGCAGTGCTTCAGCCGCTACCccgaccacatgAAGCAGCACGAC 3'
 3' GGAAGTGGGTGCCGCACGTCACGAAGTCGGCGATGGGCTGGTGTACTTCGTCGTGCTG 5'

Secuencia espaciadora BFP CRISPR

Secuencia BFP CRISPR PAM

Sitio de conversión: BFP C>T

Bases de cabeceo (secuencia wt)

FIG.3

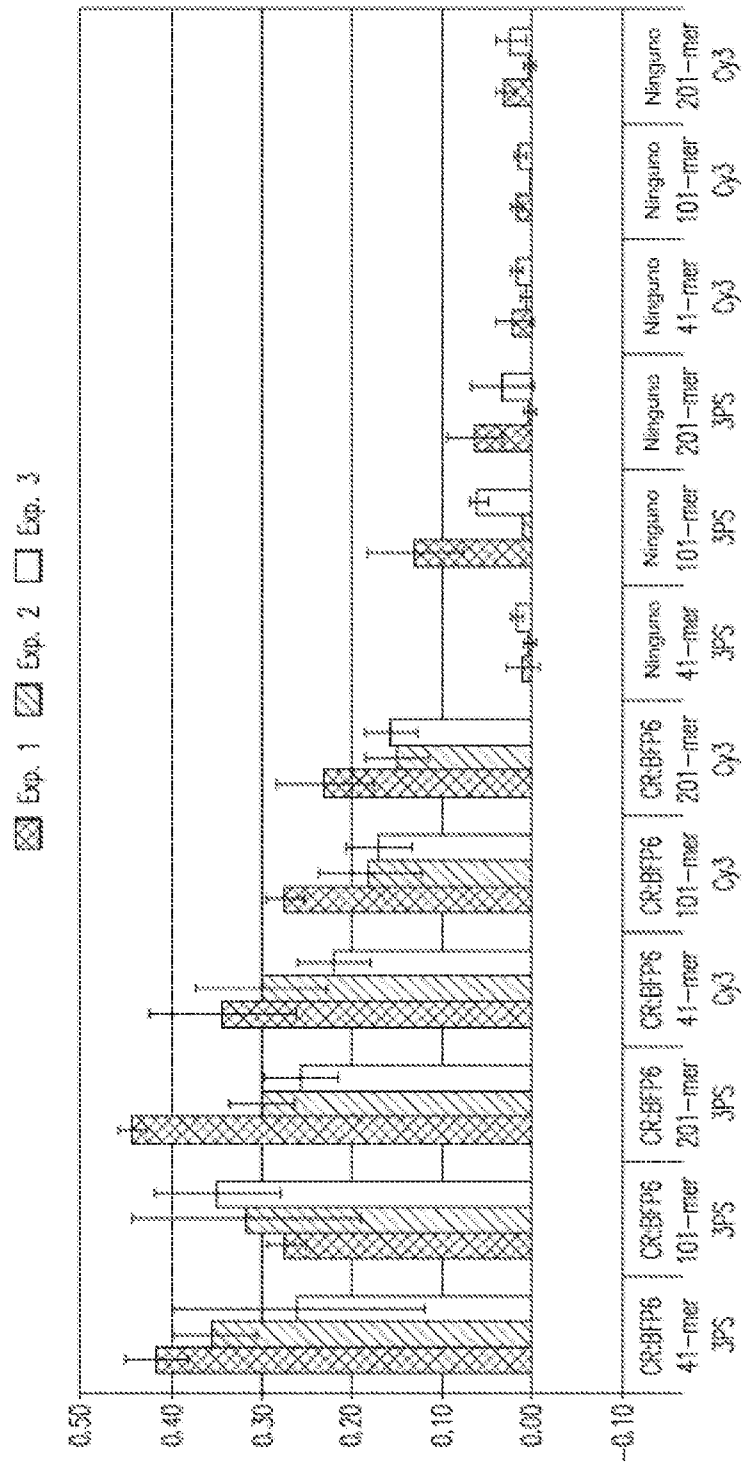


FIG.4

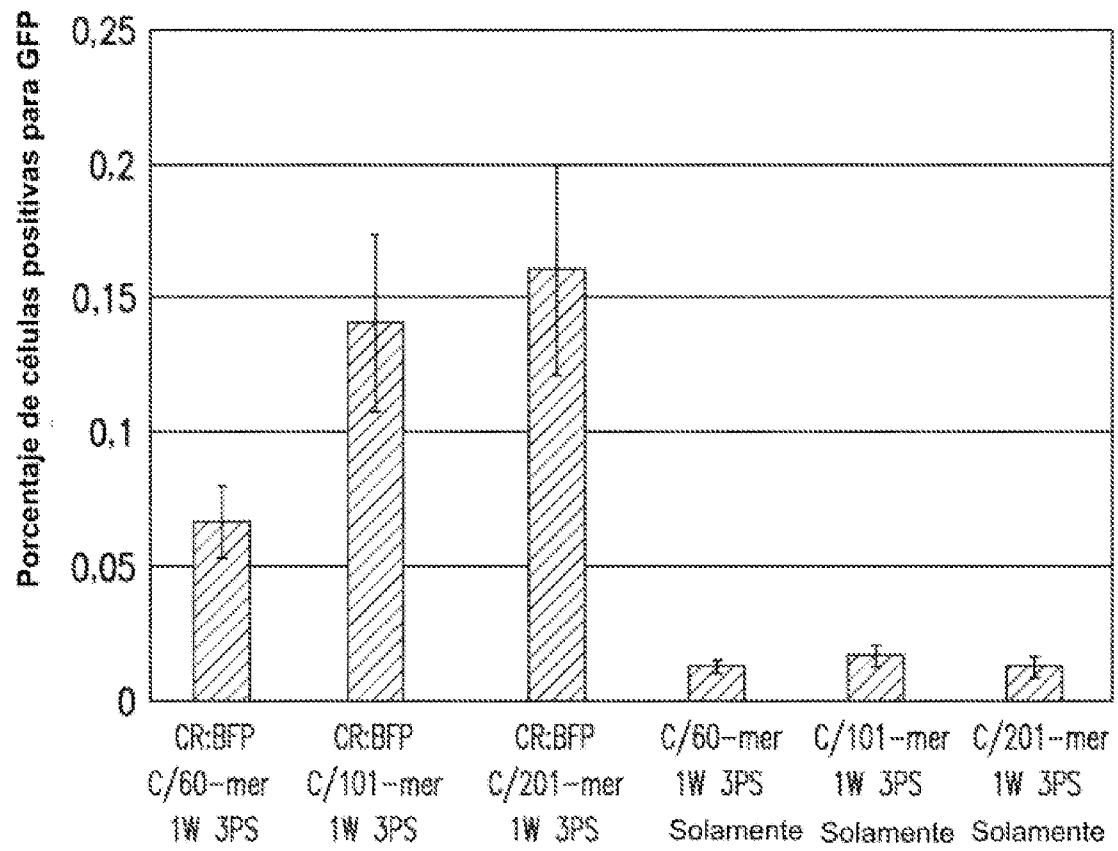


FIG.5

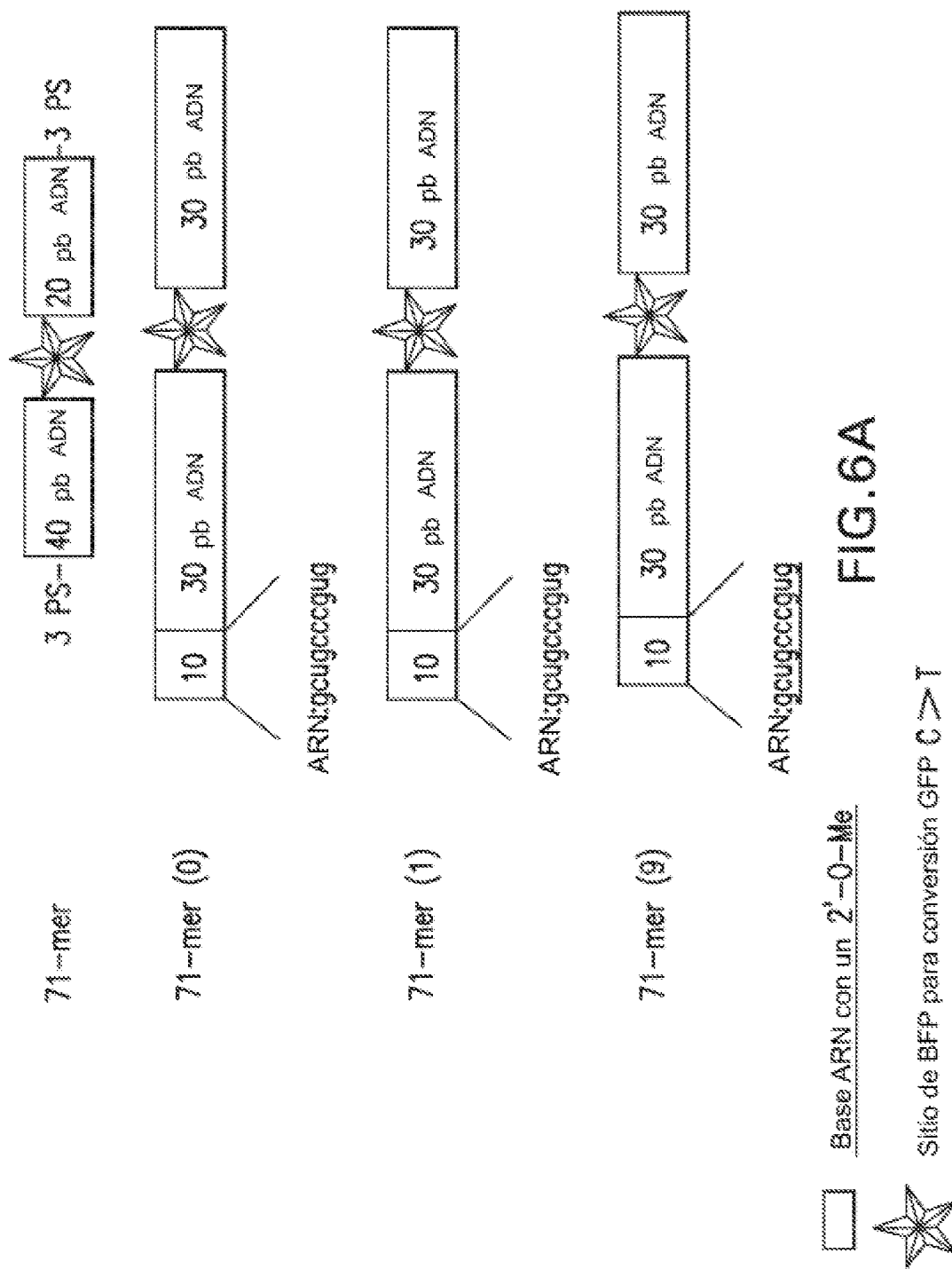


FIG.6A

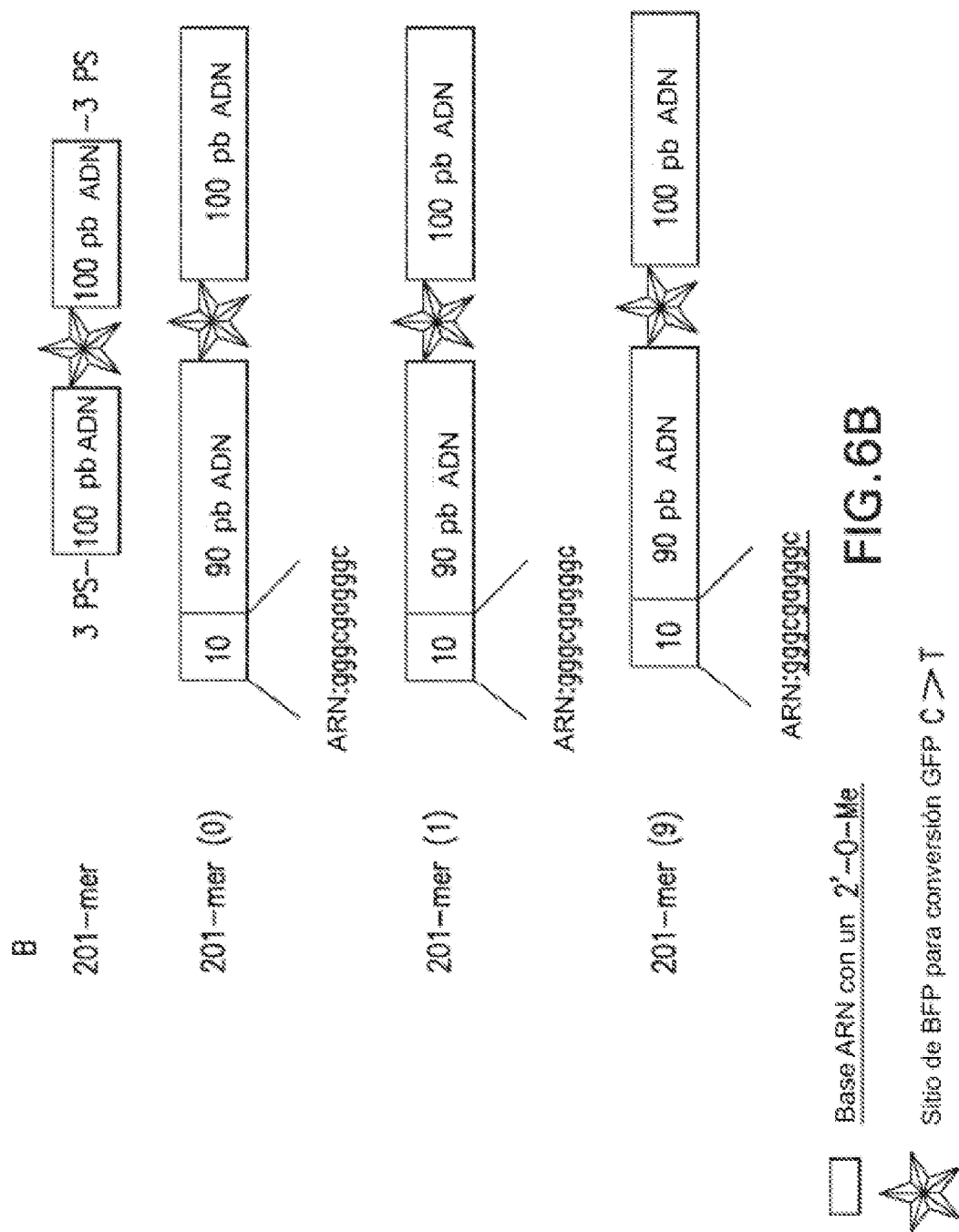


FIG.6B

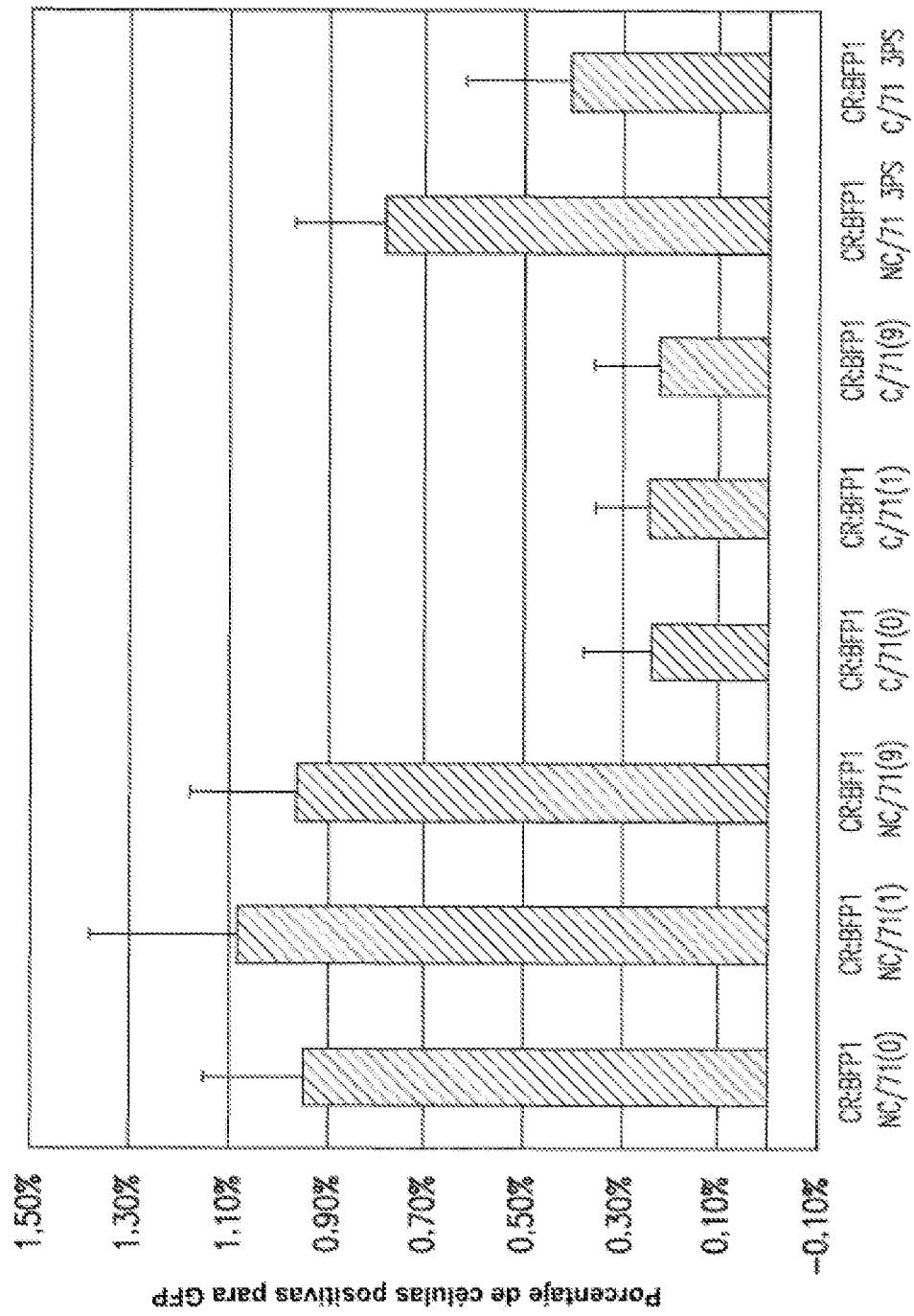


FIG. 7

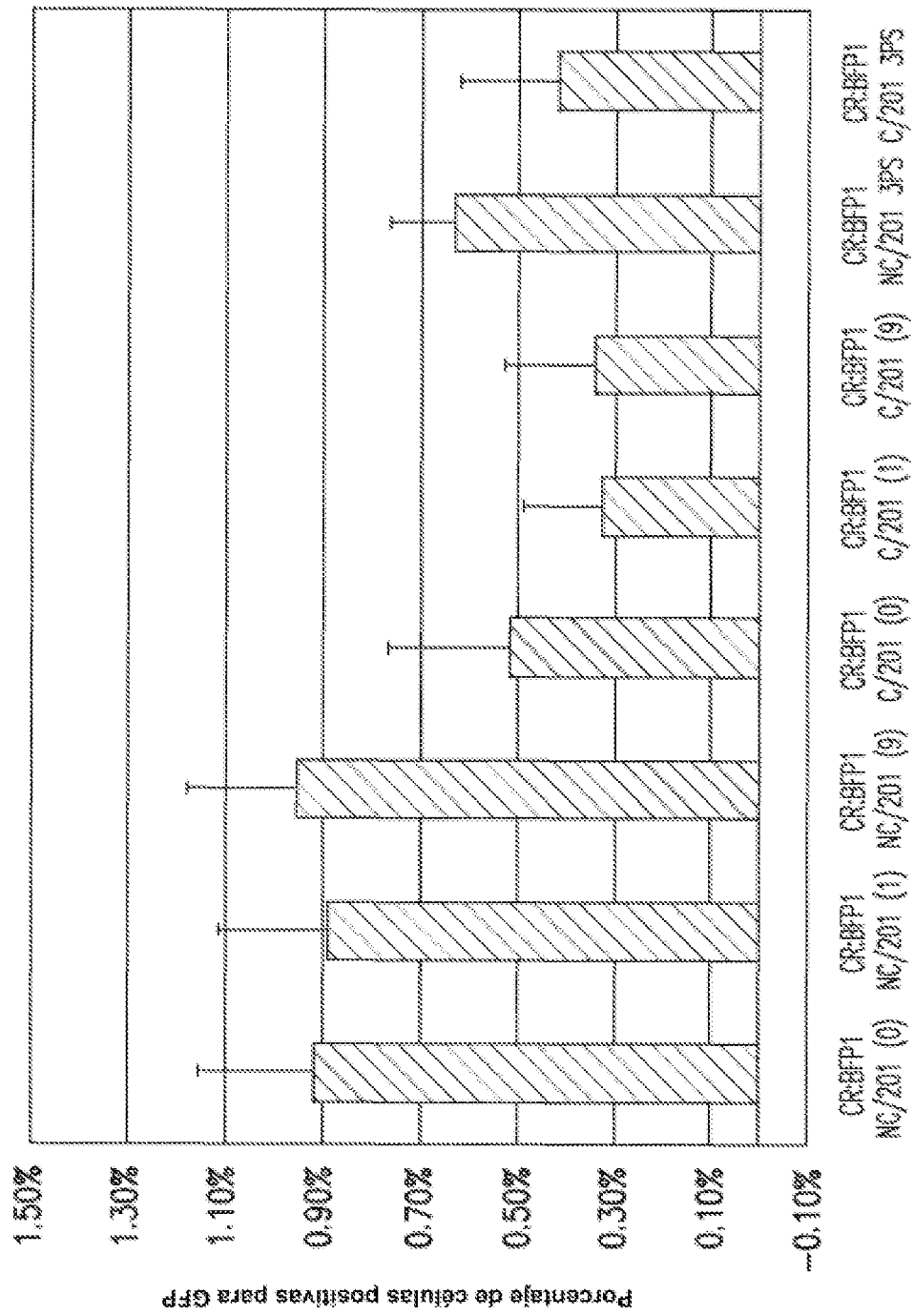


FIG.8



Ejemplo de casete de ARNsg

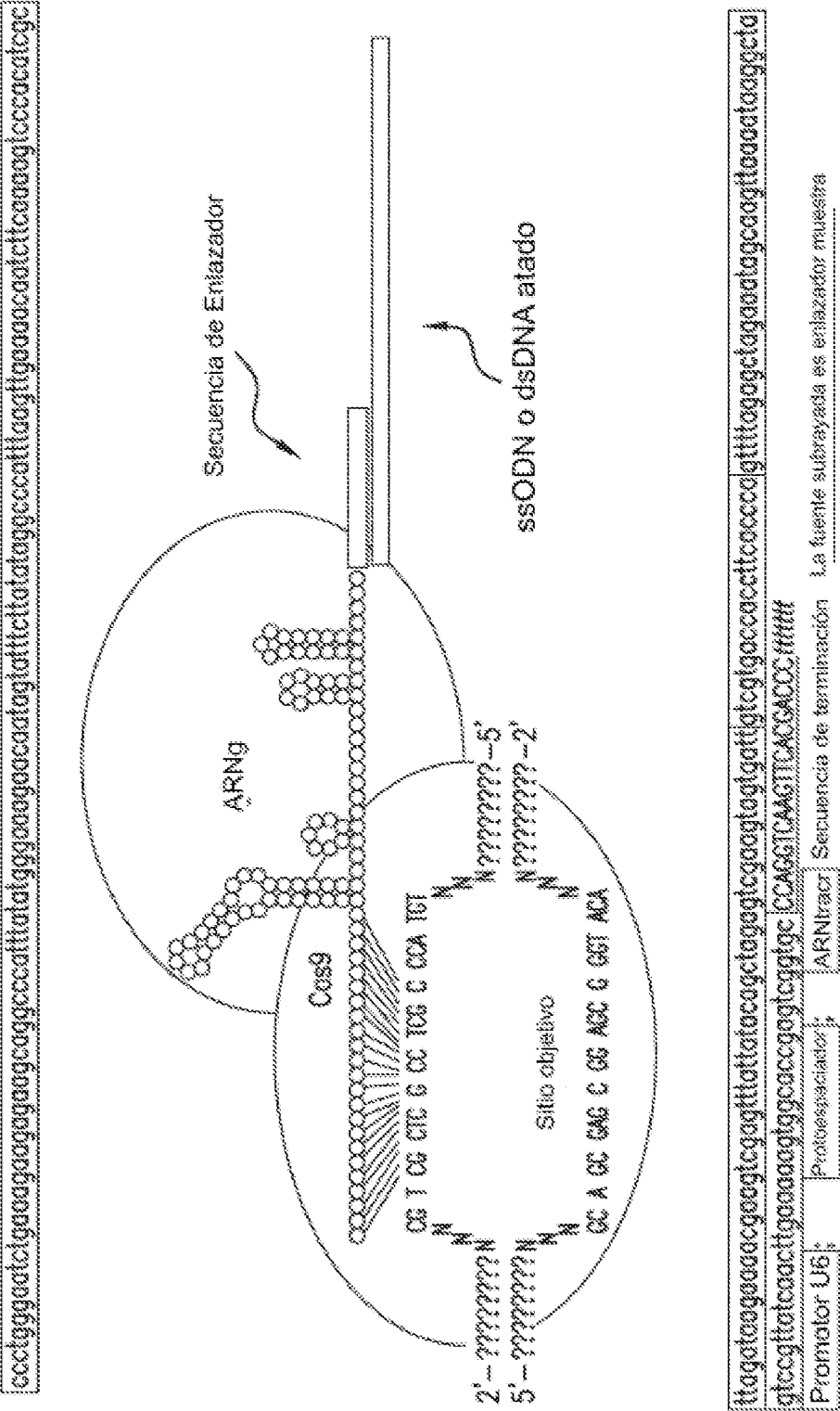


FIG.10

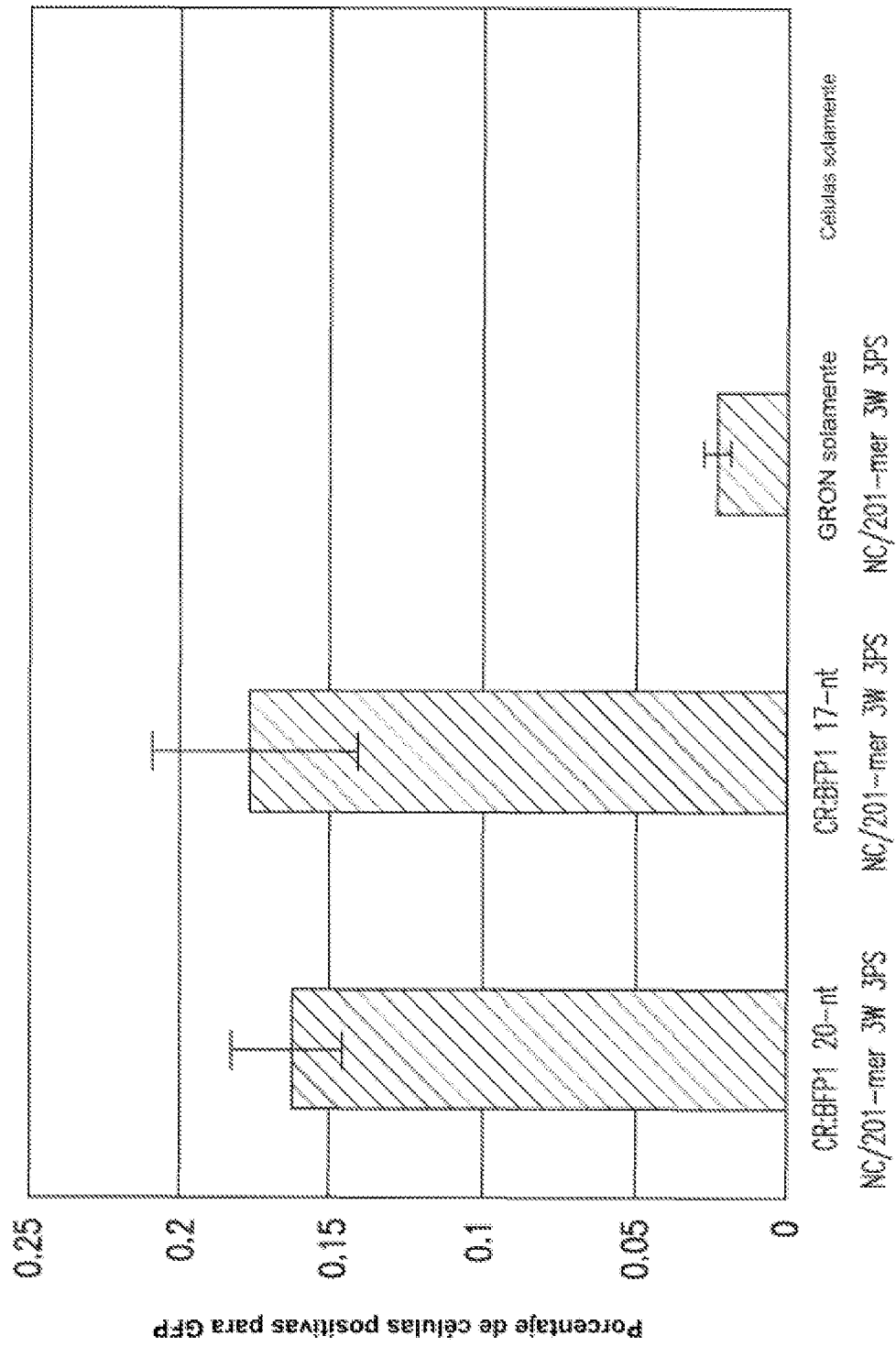


FIG.11

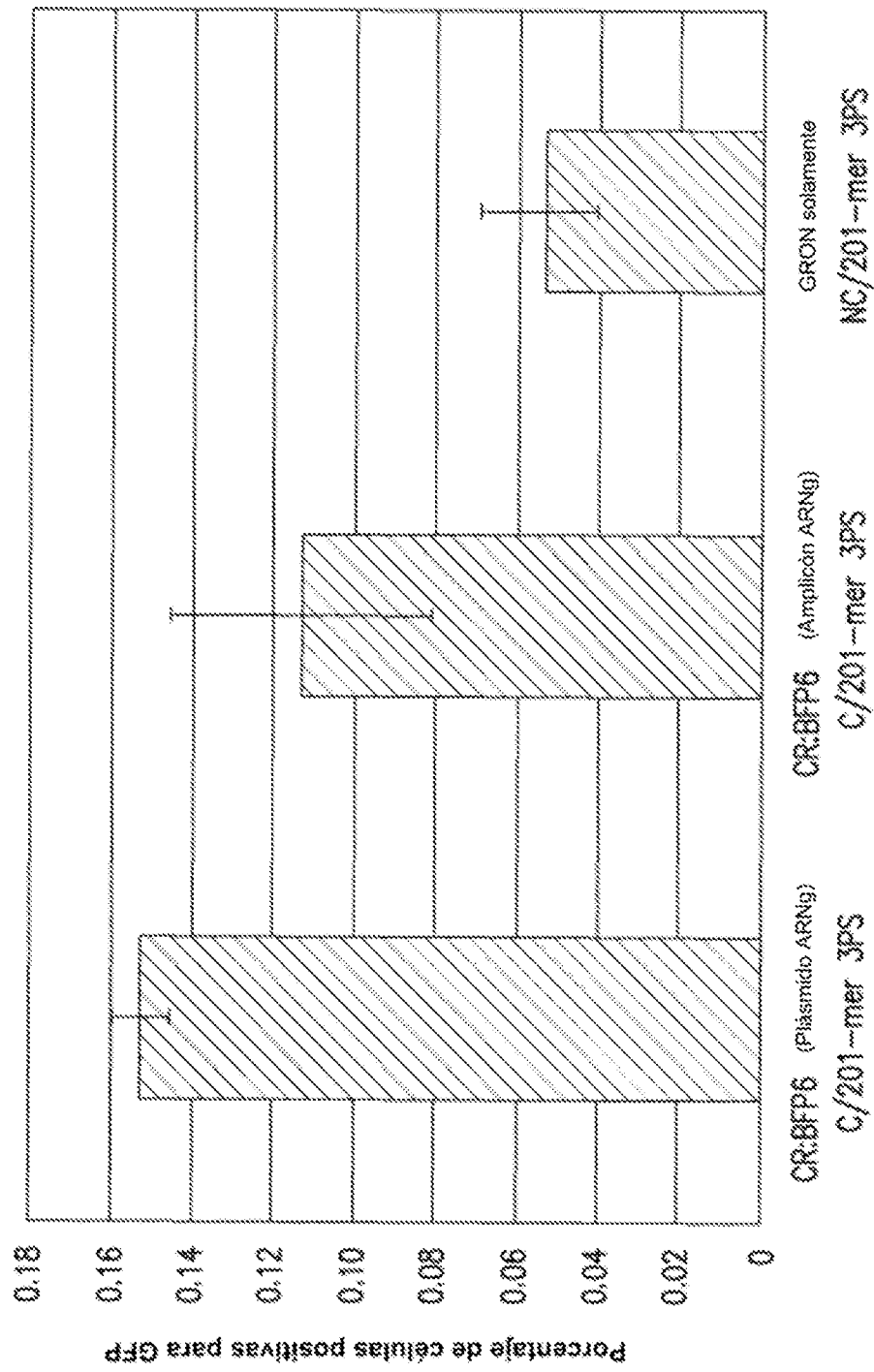


FIG.12

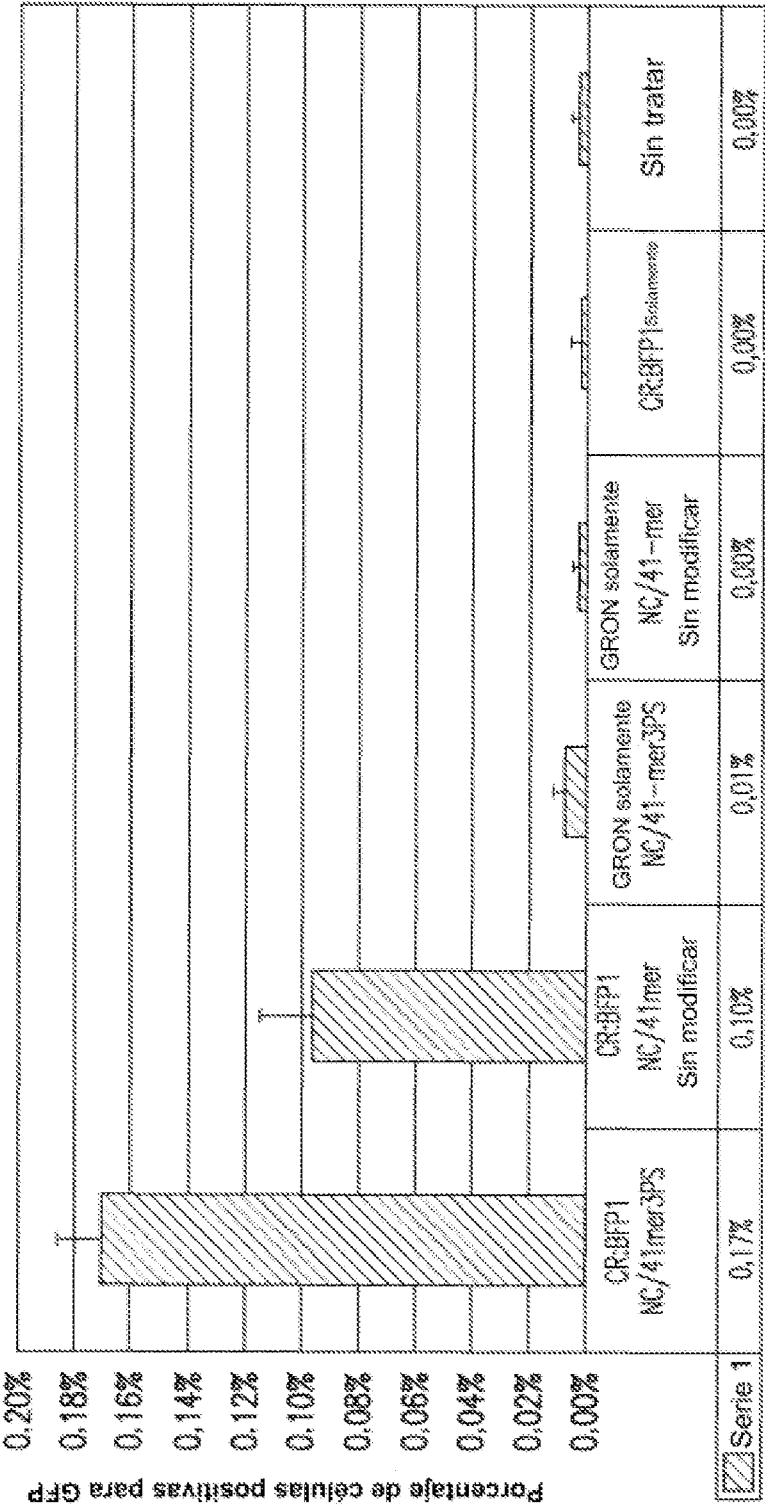


FIG.13

Tratamiento	Tiempo*	GRON	TALEN	% Conversion	% indeles
Proplastos**	24 h	Sí	Sí	0,067	2,60
Microcallos**	3 W	Sí	Sí	0,051	1,84

*indica el tiempo después de la administración por PEG de los plásmidos TALEN y GRON.

**denota que los protoplastos y los microcallos no son del mismo experimento

FIG.14

Tratamiento **	Tiempo*	CRISPR	% indeles
Microcallos	3W	Sí	46,5
Microcallos	6W	Sí	54,7

*Tiempo después de la administración por PEG de plásmidos CRISPR

**Los tratamientos no son del mismo experimento.

FIG.15

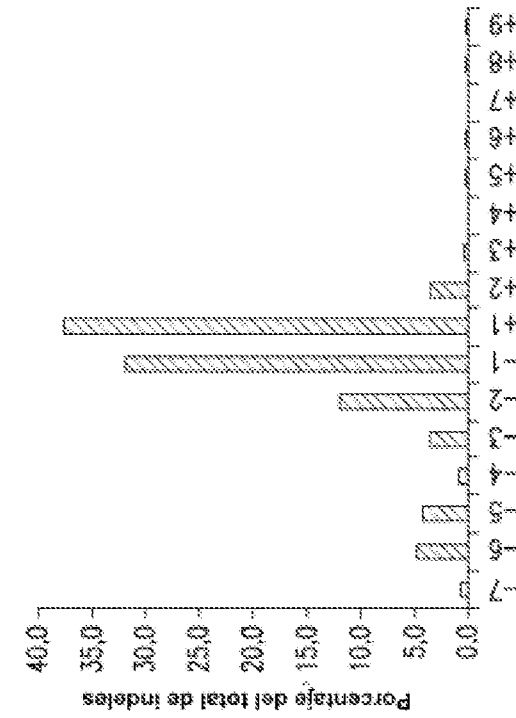


FIG. 16A

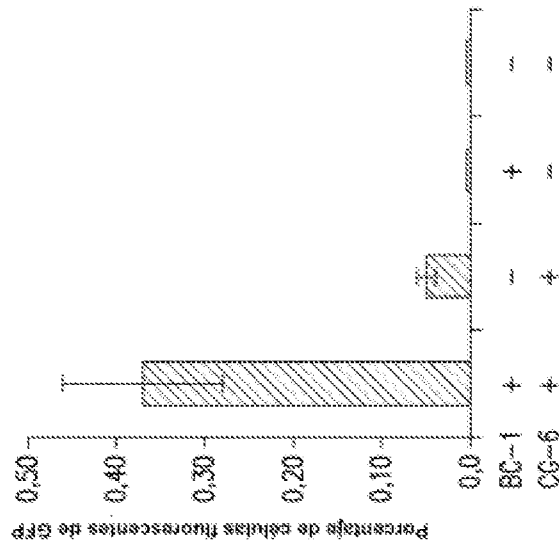


FIG. 16B

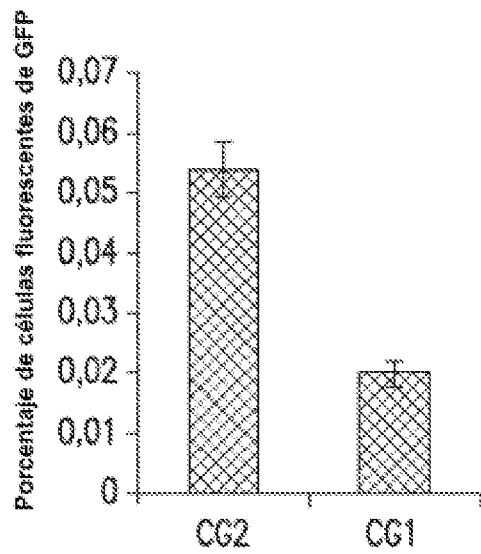


FIG.17A

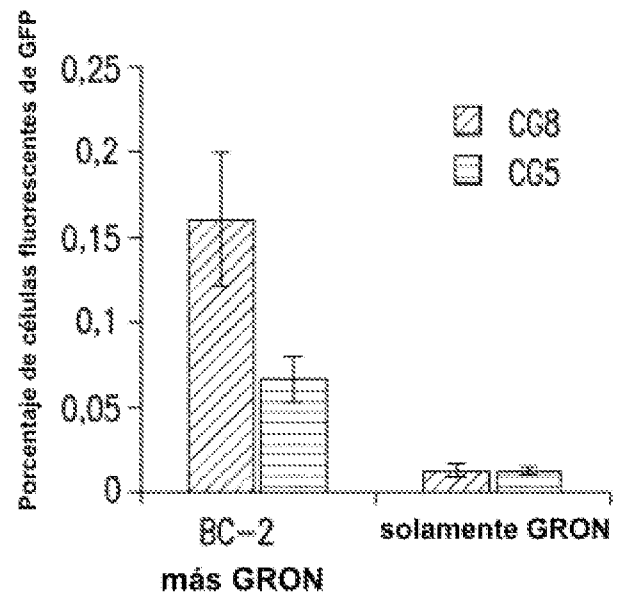


FIG.17B

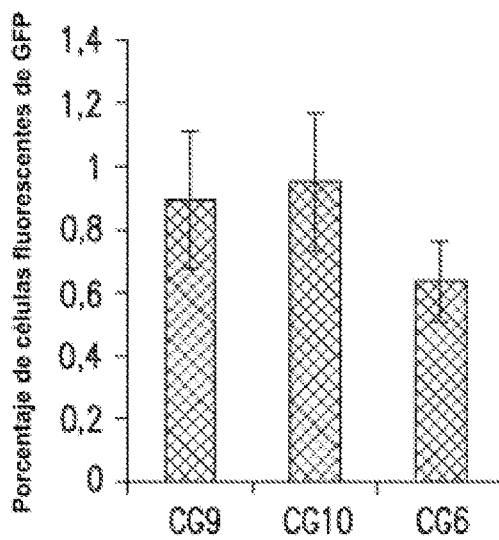


FIG.17C

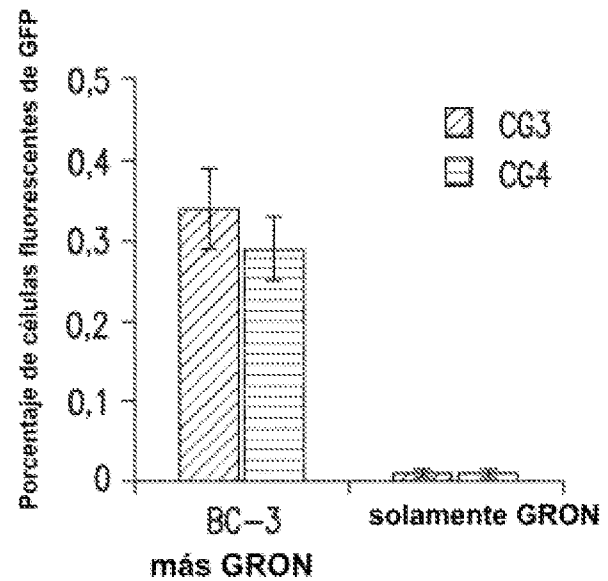


FIG.17D

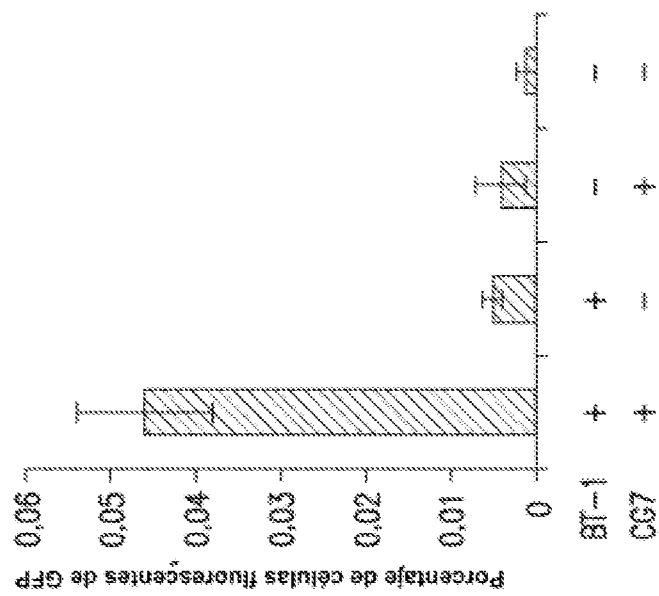


FIG.18B

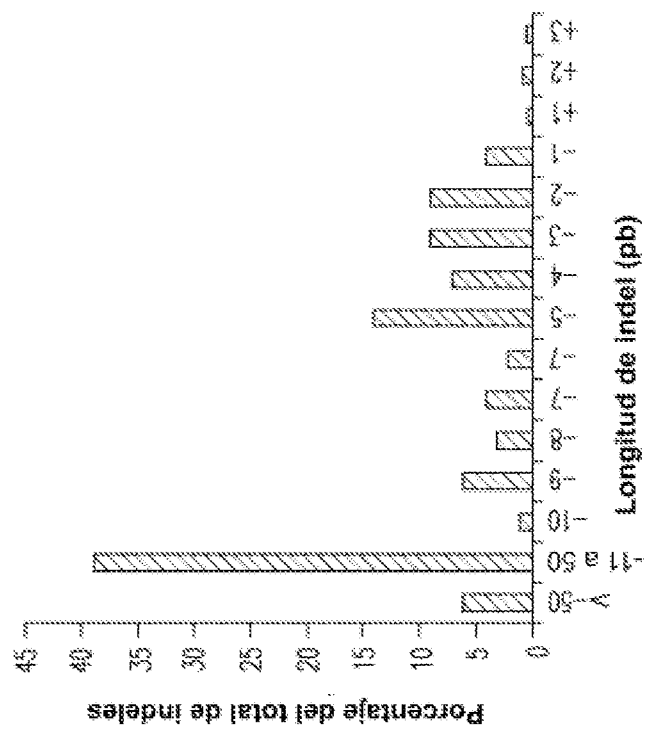
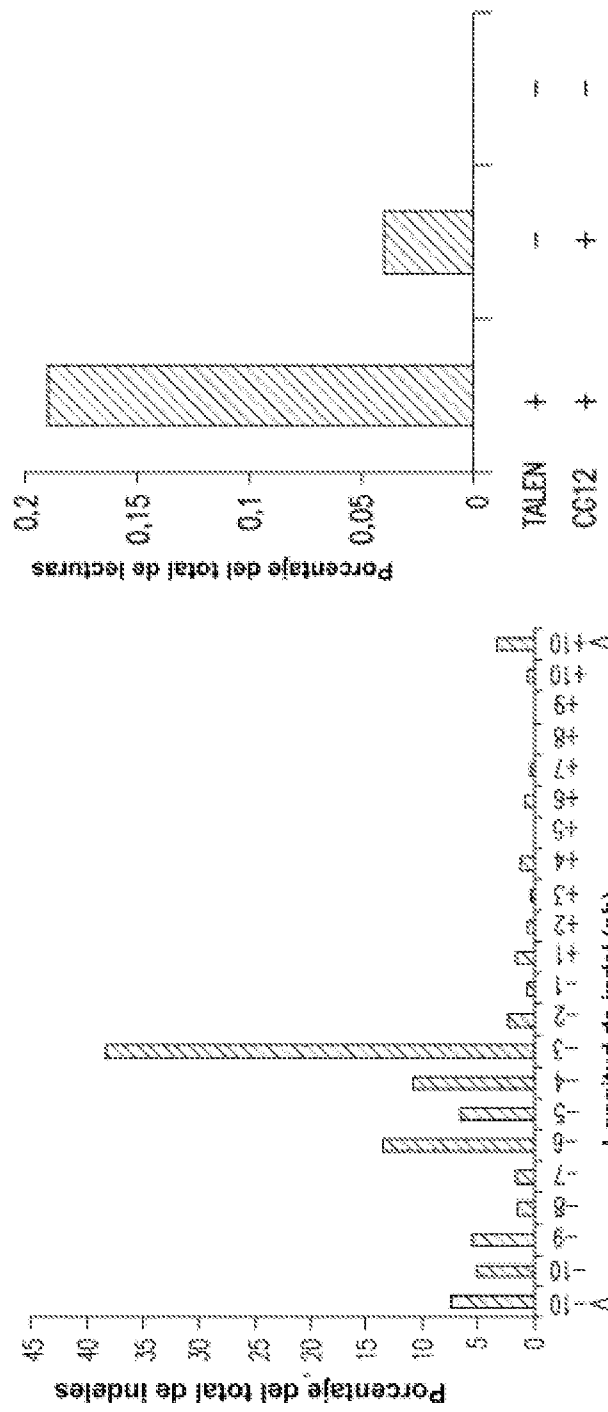


FIG.18A



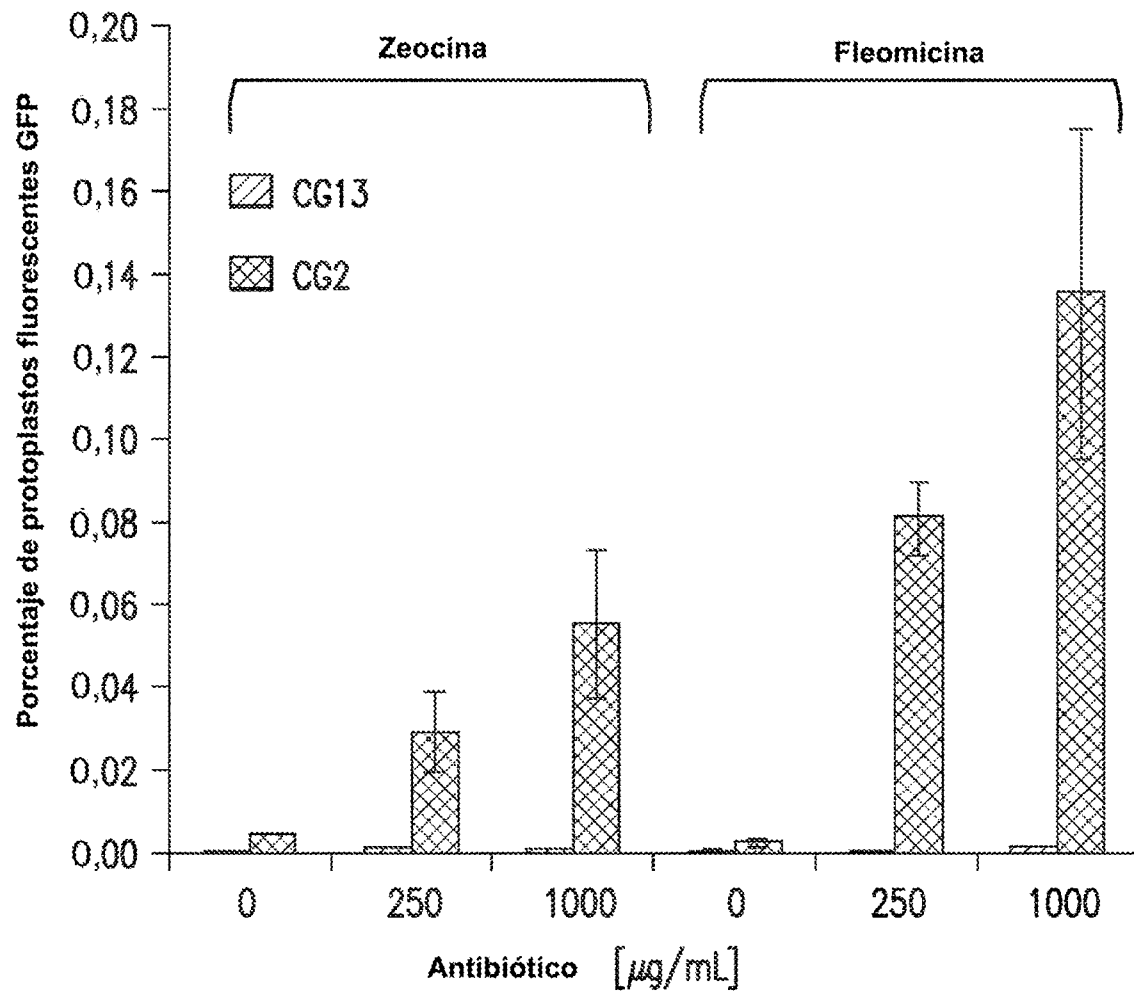


FIG.19

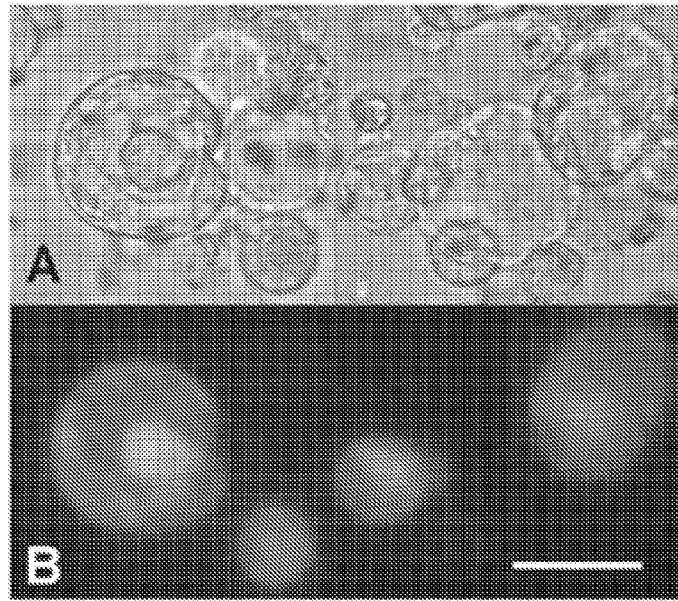


FIG. 20

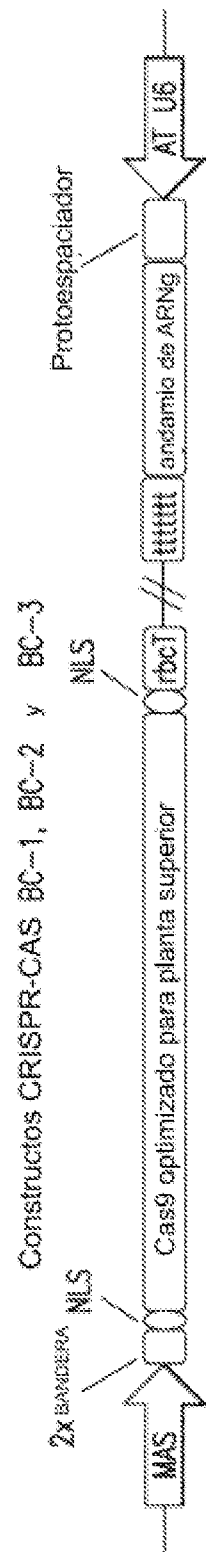


FIG.21A

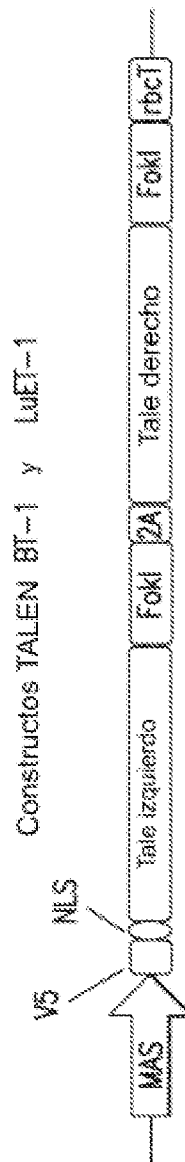


FIG.21B

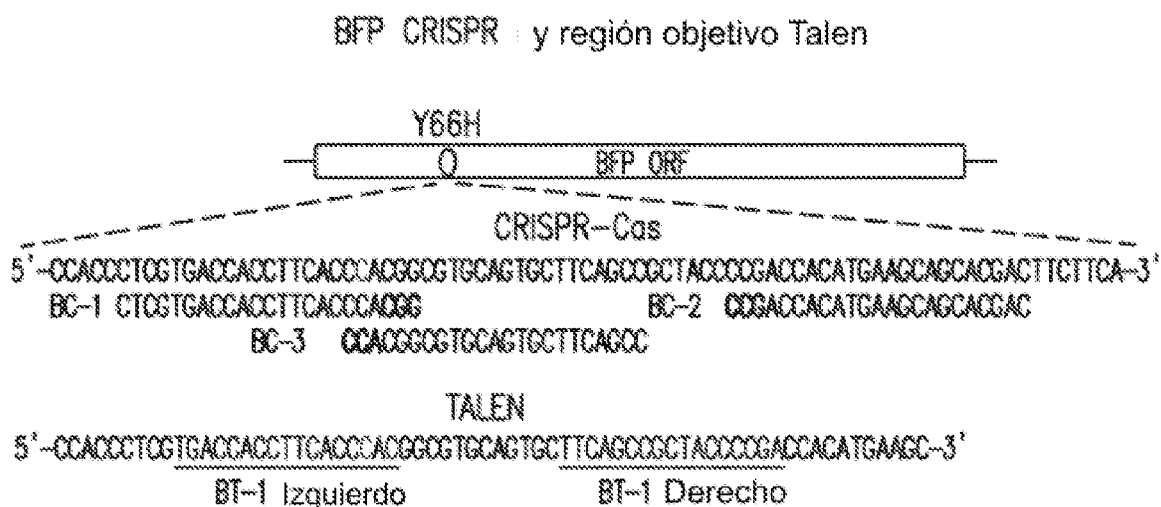


FIG.22A

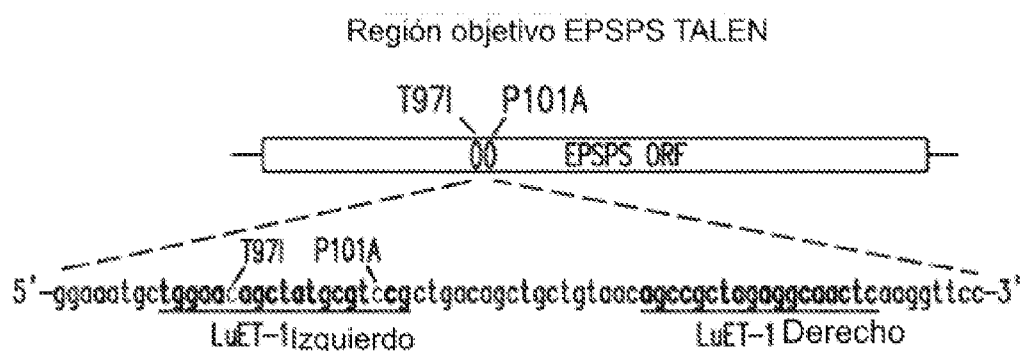


FIG.22B

AMY310767 - Proteína traducida de *Alopecurus myosuroides* plastidal ACCesa
(SEQ ID NO:1)

```

1  MGSTHLPVIG FNASTTPSLG TLRQINSAAA AFQSSSPSRG SKKKSRRVKS
51  IRDDGDGSVP DPAGHGQSIR QGLAGIIDLP KEGASAPDVD TSHGSEDHKA
101 SYQMNGILNE SHNGRHASLS KVEFECTELG GKTPIHSLV ANNGMAAAKF
151 MRSVRTWAND TFGSEKAIQL IAMATPEDMR INAEHTRIDQ OFVEVPQGTN
201 NNNYANVOLI VEIAERTGVS AVWPGWGHAS ENPELPOALT AKGIVFLGPP
251 ASSMNALGDK VGSALIAQAA GYPTLAWSGS HVEIPLELCL DSIPEEMYRK
301 ACVTTADEAV ASCQMIGYPA NIKASWGGGG KGIRKVNNDQ EVKALFKQVQ
351 GEVPGSPIFI MRLASQSRHL EVOLLCDEYG NVAALHSRDC SVQRHHQKII
401 EEGPVTVAPR ETVKELEQAA RRLAKAVGYV GAATVEYLYS METGEYYFLE
451 LNPRLQVEHP VTESIAEVNL PAAQVAVGNG IPLWQIPETR RFYGMONGGG
501 YDIWRKTAAL ATPFNFEVD SQWPKGHCVA VRITSENPDG GFKPTGGKVK
551 EISFKSKPNV WGYFSVKSGG GHEFADSOE GHVFAYGETR SAATTSMSLA
601 LKETQIRGET HTNVDTYVDL LNAPDFRENT IHTSWLDTRI AMRVQAEERP
651 WYISVVGGL YKTITTAET VSEYVSYLK GQIPPKNISL VNSTISLNE
701 ESKYTIETVR SGQGSYRLRL NGSLIEANVQ TLDGGLLMQ LDGNSHVIYA
751 EEEAGETRLI IDGKTCLLQN DHDPSSLAE TPCKLLRFLI ADGAHVDAOV
801 PYAEVEVMKM CMPLLSPAAG VINVLLSEGQ AMQAGDLIAR LDLDPSAVK
851 RAEPFEGSFP EMSLPIAASG QVHKRCAASL NAARMVLGY DHAANKVVDQ
901 LVWCLDTPAL PFLQWEELMS VLATRLPRRL KSELEGKYNE YKLNVDHVKI
951 KQPTTEMLRE TTEENLACVS EKEMVTIERL VDPLMSLLKS YEGGRESHAH
1001 FIVKSLFEY LSVEELFSDG IQSDVIERLR LQYSKDLQKV VDIVLSHGGV
1051 RNKTKLILAL MEKLVYPNPA AYRQQLIRFS SLNHKRYKYL ALKASELLEQ
1101 TKLSELRTSI ARNLSALDMF TEEKADPSLQ DRKLAINESM GDLVTAPLPV
1151 EDALVSLFDC TDQTLQQRVI QTYISRLYQP QLVKDSIQK YQD5GVIALW
1201 EFTEGNHEXR LGAMVILKSL ESVSTAIGAA LKDASHYASS AGNTVHIALL
1251 DADTQNTTE DSGDNDQAOO KMDKLSFVLK QDVVMAOLRA ADVKVVSCIV
1301 QROGAIMPXR RTFLLSEEL CYEEEPILRH VEPPLSALLE LDKLKVKGYN
1351 EMKYTPSRDR QWHIYTLRNT ENPKMLHRVF FRTLVRQPSA GNRFTSDHIT
1401 DVEVGHAEEP LSFTSSSILK SLKIAKEELE LHAIRTGHS MYLCILKEQK
1451 LLDLVPVSGN TVVDVGQDEA TACSLLEKMA LKIHVLGAR MHILSVQWE
1501 VKLKLVSOGP ASGSWRVVT NVTGHTCTVD IYREVEDTES QKLVYHSTAL
1551 SSGPLHGVAL NTSYQPLSVI DLKRCASRNN KTTYCYDFPL TFEAAVQKSW
1601 SNTSSENNOC YVKATELVFA EKNQSWGTP1 IPMORAAGLN DIGMVAWILD
1651 MSTPEFPSCR QIIVIANDIT FRAGSFGPRE DAFPEAVTNL ACEKKLPLTY
1701 LAANSGARIG IADEVKSCFR VQWTDSSPE RGFRTYMTD EDHDIRIGSSV
1751 IAHKMLQDSG ETRWVIDSV GKEDGLGVEN IHGSAIASA YSRAYEFTFT
1801 LTFVTGRTVG IGAYLARLGI RCIQRIDGPI ILTGFSALNK LLGREVYSSH
1851 MQLGGPKIMA TNGVHLTVP DDLEGVSNIL RWLSYVPANI GGPLPITKSL
1901 DPIDRPVAYI PENTCDPRAA ISGIDDSQK WLGGMFOKDS FVETFEQWAK
1951 TVVTGRALG GIPVGVIAVE TQTMQLVPA DPGQPOSHER SVPRAGQWVF
2001 PDSATKTAQA MLDNFREGLP LFI LANWRGF SGGORDLFEG ILQAGSTIVE
2051 NLRTYNQPAF VYIPKAAELR GGAWVIDSK INPDRIEYA ERTAKGNVLE
2101 POGLEIKFR SEELKECMGR LDPELIDLKA RLQGANGLS DGESLOKSTI
2151 ARKKQLPLY TQIAVRFAEL HDTSLRMAAK GVIRKVDWE DSRSFYKRL
2201 RRRLSEDVLA KEIRGVIGK FPHKSAIELI KKWLASEAA AAGSTOWDD
2251 DAFVAWREN ENYKEYIKEL RAQRVSRLLS DVAGSSDLQ ALPQGLSMIL
2301 DKMDPSKRAQ FIEEVMKVLK

```

FIG.23

El producto del gen EPSPS de *E. coli* tiene la siguiente secuencia (Acceso UniProt # P0A6D3, SEQ ID NO:2)

MESLTLOPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVRHMLNALTAL
 GVSYTLSADRTRCEIIGNGGPLHAEGALELFLGNAGTAMRPLAAALCLGSNDIVLTGEPR
 MKERPIGHLVDA LRLGGAKITYLEQENYPPLRLQGGFTGGNVDVDGSSQFLTALLMTA
 PLAPEDTVIRIKGDLVSKPYIDITLNLMTFGVEIENQHYQQFVVKGGQSYQSPGTYLVE
 GDASSASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICWGDDYISCTRGE
 LNAIDMDMNHIPDAAMTIATAALFAKGTTTTLRNIYNWRVKETDRLFAMATELRKVGAEVE
 EGHDIYIRITPPEKLNFAEIATYNDHRMAHCFSLVALSDTPVTILDPKCTAKTFPDYFEQL
 ARISQAA

FIG.24

Proteína	Genbank Acceso #	L82		T97	P101	N111
E. coli	X00557	82		97	101	111*
Arabidopsis thaliana	AF360224	160		179	183	194
Petunia hybrida	M21084.1	155		174	178	189
Brassica napus	X51475.1	155		174	178	189
Zea mays	X63374	84		102	106	117
Oryza sativa	AF413082	150		169	173	184
Arabidopsis thaliana	NM 130093	159		178	182	193

*No hay verdadero ácido homólogo de E. coli

FIG.25

AMY310767 - ADNc de *Alopecurus myosuroides* plastidal ACCasa

```

1  TCGATAAACT TCCTGTTGCA TGTCTCTATC TCTATGGACT AACGGTTCTT ATGTGCATGC
61  ATCTGTCAAG TTTCCAGACC TGGGGTTTAC AATCAGTTTA TGGCAGTCTG TGTTTGAAGA
121  ACACTGCAAC TCTGCTGTCT GTCCAAAGGG AGGACGATGG GATCCACACA TCTGCCATT
181  GTGGGGTTTA ATGCATCCAC AACACCATCG CTATCCACTC TTCGCCAGAT AAATCAGCT
241  GCTGCTGCAT TCCAATCTTC GTCCCTTCA AGSTCATECA AGAAGAAAAG CCGACGTGT
301  AAGTCAATAA GGGATGATGG CGATGGAAGC GTGCCAGACC CTGCAGGCCA TGGCCAGTCT
361  ATTEGECAGG GTCTCGCTGG CATCATCGAC CTCCCAAAGG AGGGCGCATC AGCTCCAGAT
421  GTGGACATTT CACATGGGTC TGAGAGCCAC AAGGCTCTCT ACCAAATGAA TGGGATACTG
481  AATGAATCAC ATAACGGGAG GCACGCTCT CTGTCTAAG TTTATGAATT TTGCACGGAA
541  TTGGGTGGAA AAACACCAAT TCACAGTGTA TTAGTCGCCA ACAATGGAAT GGCACGAGT
601  AAGTTCATGC GGAGTGTCG GACATGGGCT AATGATACAT TTGGGTGAGA GAAGGCGATT
661  CAGTTGATAG CTATGGCAAC TCCGAAGAC ATGAGAATAA ATGCAGAGCA CATTAGAATT
721  GCTGATCAGT TTGTTGAAGT ACCTGGTGGG ACAACAATA ACAACTATGC AAATGTCCAA
781  CTCATAGTGG AGATAGCAGA GAGAACTGGT GTCTCCGCCG TTTGGCTTGG TTGGGGCCAT
841  GCATCTGAGA ATCCTGAACT TCCAGATGCA CTAAGTCAA AAGGAATTGT TTTCTTGGG
901  CCACCAGCAT CATCAATGAA CGCACTAGGC GACAAGGTTG GTTCAGCTCT CATTGCTCAA
961  GCAGCAGGGG TTCCCACTCT TGCTTGGAGT GGATCAGATG TGGAAATTCC ATTAGAATT
1021  TGTTTGGACT CGATACCTGA GGAGATGTAT AGGAAAGCCT GTGTTACAAC CGCTGATGAA
1081  GCAGTTGCAA GTTGTGAGAT GATTGGTTAC CCTGCCATGA TCAAGGCATC CTGGGGTGGT
1141  GGTGGTAAAG GGATTAGAAA GTTAATAAT GATGACGAGG TGAAGCACT GTTTAAGCAA
1201  GTACAGGGTG AAGTTCTTGG CTCCCGATA TTTATCATGA GACTTGATC TCAGAGTCT
1261  CATCTTGAAG TCCAGCTGCT TTGTGATGAA TATGCAATG TAGCAGCACT TCACAGTCT
1321  GATTGCAGTG TGCAACGAG ACACCAAAG ATTATCGAGG AAGGACCACT TACTGTTGCT
1381  CCTCGTGAAA CAGTGAAGA GCTAGAGCAA GCAGCAAGGA GGCTTCTTAA GGCCGTGGT
1441  TACGTGGGTG CTGCTACTGT TGAATATCTC TACAGCATGG AGACTGGTGA ATACTATTTT
1501  CTGGAGCTTA ATCCACGGTT GCAGGTTGAG CACCCAGTCA CCGAGTCGAT AGCTGAAGTA
1561  AATTTGCTG CAGCCCAAGT TGCAGTTGGG ATGGGTATAC CCTTTGGCA GATTCCAGAG
1621  ATCAGACGTT TCTACGGAAT GGACAATGSA GGAGGCTATG ATATTTGGAG GAAACAGCA
1681  GCTCTCGCTA CTCCATTCAA CTTTGATGAA GTAGATTCTC AATGGCCGAA GGSTCATTGT
1741  GTGGCAGTTA GGATAACCAG TGAGAATCCA GATGATGGAT TCAAGCCTAC TGGTGGAAA
1801  GTAAAGGAGA TAAGTTTAA AAGTAAGCCA AATGTCTGGG GATATTTCTC AGTTAAGTCT
1861  GGTGGAGGCA TTCATGAATT TGCGGATCT CAGTTTGGAC ACGTTTGTG CTATGGAGAG
1921  ACTAGATCAG CAGCAATAAC CAGCATGTCT CTTGCACTAA AAGAGATTCA AATTCTGGA
1981  GAAATTGATA CAAACGTTGA TTACACGGTT GATCTCTTGA ATGCCCGAGA CTTCAGAGAA
2041  AACACGATCC ATACCGGTTG GCTGGATAAC AGAATAGCTA TGGTGTTTCA AGCTGAGAGG
2101  CCTCCCTGGT ATATTTCACT GGTGGAGGA GCTCTATATA AAACAATAAC CACCAATGCG
2161  GAGACCGTTT CTGAATATGT TAGCTATCTC ATCAAGGGTC AGATTCCACC AAAGCACATA
2221  TCCCTTGTCC ATTCAACTAT TTCTTTGAAT ATAGAGGAAA GCAATATAC AATTGAGATT
2281  GTGAGGAGTG GACAGGCTAG CTACAGATG AGACTGAATG GATCACTTAT TGAAGCCAAT
2341  GTACAAACAT TATGTGATGG AGGCTTTTTA ATGCAGCTGG ATGGAAATAG CCATGTTATT
2401  TATGCTGAAG AAGAAGCGGG TGGTACACGG CTCTTATTG ATGGAAAAAC ATGCTTGCTA

```

FIG.26A-1

```

2461 CAGAATGACC ATGATCCGTC AAGGTTATTA GCTGAGACAC CCTGCAAAC TCTTCGTTTC
2521 TTGATTGCCG ATGGTGCTCA TGTTGATGCT GATGTACCAT ACGCGGAAGT TGAGGTTATG
2581 AAGATGTGCA TGCCCCTCTT GTCGCCTGCT GCTGGTGTCA TTAATGTTTT GTTGTCTGAG
2641 GGCCAGGCGA TGCAGGCTGG TGATCTTATA GCGAGACTTG ATCTCGATGA CCCTTCTGCT
2701 GTGAAGAGAG CCGAGCCATT TGAAGGATCT TTTCCAGAAA TGAGCCTTCC TATTGCTGCT
2761 TCTGGCCAAG TTCACAAAAG ATGTGCTGCA AGTTTGAACG CTGCTCGAAT GGTCCCTGCA
2821 GGATATGACC ATGCGGCCAA CAAAGTTGTG CAAGATTTGG TATGGTGCCT TGATACACCT
2881 GCTCTTCCCT TCCTACAATG GGAAGAGCTT ATGTCTGTTT TAGCAACTAG ACTTCCAAGA
2941 CGTCTTAAGA GCGAGTTGGA GGGCAAATAC AATGAATACA AGTTAAATGT TGACCATGTG
3001 AAGATCAAGG ATTTCCCTAC CGAGATGCTT AGAGAGACAA TCGAGGAAAA TCTTGCATGT
3061 GTTCCGAGA AGGAAATGGT GACAATTGAG AGGCTTGTTG ACCCTCTGAT GAGCCTGCTG
3121 AAGTCATACG AGGGTGGGAG AGAAAGCCAT GCCCACTTTA TTGTCAAGTC CCTTTTGTAG
3181 GAGTATCTCT CGGTTGAGGA ACTATTCAGT GATGGCATTG AGTCTGACGT GATTGAACGC
3241 CTGCGCTAC AATATAGTAA AGACCTCCAG AAGGTTGTAG ACATTGTTTT GTCTCACCAG
3301 GGTGTGAGAA ACAAACAAA GCTGATACTC GCGCTCATGG AGAAACTGGT CTATCCAAAC
3361 CCTGCTGCCT ACAGAGATCA GTTGATTCGC TTTTCTTCCC TCAACCATAA AAGATATTAT
3421 AAGTTGGCTC TTAAAGCTAG TGAACCTCTT GAACAAACCA AGCTCAGCGA ACTCCGCACA
3481 AGCATTGCAA GGAACCTTTC AGCGCTGGAT ATGTTACCGG AGGAAAAGGC AGATTTCTCC
3541 TTGCAAGACA GAAAATTGGC CATTAAATGAG AGCATGGGAG ATTTAGTCAC TGCCCCACTG
3601 CCAGTTGAAG ATGCACTTGT TTCTTTGTTT GATTGTACTG ATCAAACCTC TCAGCAGAGA
3661 GTGATTCAGA CACATATC TCGATTATAC CAGCCTCAAC TTGTGAAGGA TAGCATCCAG
3721 CTGAAATATC AGGATTCTGG TGTTATTGCT TTATGGGAAT TCACTGAAGG AAATCATGAG
3781 AAGAGATTGG GTGCTATGGT TATCCTGAAG TCACTAGAAT CTGTGTCAAC AGCCATTGGA
3841 GCTGCTCTAA AGGATGCATC ACATTATGCA AGCTCTGCGG GCAACACGGT GCATATTGCT
3901 TTGTTGGATG CTGATACCCA ACTGAATACA ACTGAAGATA GTGGTGATAA TGACCAAGCT
3961 CAAGACAAGA TGGATAAACT TTCTTTTGTA CTGAAACAAG ATGTTGTCAT GGCTGATCTA
4021 CGTGCTGCTG ATGTCAAGGT TGTTAGTTGC ATTGTTCAA GAGATGGAGC AATCATGCCT
4081 ATGCGCCGTA CCTTCCTCTT GTCAGAGGAA AAACCTTGTG ACGAGGAAGA GCCGATTCTT
4141 CGGCATGTGG AGCCTCCACT TTCTGCACTT CTTGAGTTGG ATAAATTGAA AGTGAAAGGA
4201 TACAATGAGA TGAAGTATAC ACCGTCACGT GATCGTCAGT GGCATATATA CACACTTAGA
4261 AATACTGAAA ATCCAAAAT GCTGCACAGG GTATTTTCC GAACACTTGT CAGACAACCC
4321 AGTGCAGGCA ACAGGTTTAC ATCAGACCAT ATCACTGATG TTGAAGTAGG ACACGCAGAG
4381 GAACCTCTTT CATTTACTTC AAGCAGCATA TTAATTCGT TGAAGATTGC TAAAGAAGAA
4441 TTGGAGCTTC ACGCGATCAG GACTGGCCAT TCTCATATGT ACTTGTGCAT ATTGAAAGAG
4501 CAAAAGCTTC TTGACCTTGT TCCTGTTTCA GGGAACTG TTGTGGATGT TGGTCAAGAT
4561 GAAGCTACTG CATGCTCTCT TTTGAAAGAA ATGGCTTTAA AGATACATGA ACTTGTTGGT
4621 GCAAGAATGC ATCATCTTTC TGTATGCCAG TGGGAAGTGA AACTTAAGTT GGTGAGCGAT
4681 GGGCTGCCA GTGGTAGCTG GAGAGTTGTA ACAACCAATG TTAGTGGTCA CACCTGCACT
4741 GTGGATATCT ACCGGGAGGT CGAAGATACA GAATCACAGA AACTAGTATA CCACTCCACC
4801 GCATTGTCAT CTGGTCCTTT GCATGGTGTT GCACTGAATA CTTCGTATCA GCCTTTGAGT
4861 GTTATTGATT TAAACGTTG CTCTGCCAGG AACAACAAA CTACATACTG CTATGATTTT
4921 CCATTGACAT TTGAAGCTGC AGTGCAGAAG TCGTGGTCTA ACATTTCCAG TGAAAACAAC

```

FIG.26A-2

4981 CAATGTTATG TTAAAGCGAC AGAGCTTGTG TTTGCTGAAA AGAATGGGTC GTGGGGCACT
 5041 CCTATAATTC CTATGCAGCG TGCTGCTGGG CTGAATGACA TTGGTATGGT AGCCTGGATC
 5101 TTGGACATGT CCACTCCTGA ATTTCCCAGC GGCAGACAGA TCATTGTTAT CGCAAATGAT
 5161 ATTACATTTA GAGCTGGATC ATTTGGCCCA AGGGAAGATG CATTTTTCGA AGCTGTAACC
 5221 AACCTGGCTT GTGAGAAGAA GCTTCCACTT ATCTACTTGG CTGCAAATC TGGTGCTCGG
 5281 ATTGGCATTG CTGATGAAGT AAAATCTTGC TTCCGTGTTG GATGGACTGA TGATAGCAGC
 5341 CCTGAACGTG GATTTAGGTA CATTTATATG ACTGACGAAG ACCATGATCG TATTGGCTCT
 5401 TCAGTTATAG CACACAAGAT GCAGCTAGAT AGTGGCGAGA TCAGGTGGGT TATTGATTCT
 5461 GTTGTGGGAA AAGAGGATGG ACTAGGTGTG GAGAACATAC ATGGAAGTGC TGCTATTGCC
 5521 AGTGCCTATT CTAGGGCGTA CGAGGAGACA TTTACACTTA CATTCGTTAC TGGACGAACT
 5581 GTTGAATCG GAGCCTATCT TGCTCGACTT GGCATACGGT GCATACAGCG TATTGACCAG
 5641 CCCATTATTT TGACCGGGTT TTCTGCCCTG AACAAGCTTC TTGGGCGGGA GGTGTACAGC
 5701 TCCACATGC AGTTGGGTGG TCCAAAATC ATGGCGACGA ATGGTGTGT CCATCTGACT
 5761 GTTCCAGATG ACCTTGAAGG TGTTTCTAAT ATATTGAGGT GGCTCAGCTA TGTTCTTGCA
 5821 AACATTGGTG GACCTCTTCC TATTACAAAA TCTTTGGACC CAATAGACAG ACCCGTTGCA
 5881 TACATCCCTG AGAATACATG TGATCCTCGT GCAGCCATCA GTGGCATTGA TGACAGCCAA
 5941 GGGAAATGGT TGGGTGGCAT GTTTGACAAA GACAGTTTTG TGGAGACATT TGAAGGATGG
 6001 GCGAAGACAG TAGTTACTGG CAGAGCAAAA CTTGGAGGGA TTCCTGTTGG TGTTATAGCT
 6061 GTGGAGACAC AGACCATGAT GCAGCTCGTC CCCGCTGATC CAGGCCAGCC TGATTCCCAC
 6121 GAGCGGTCTG TTCCTCGTGC TGGGCAAGTT TGGTTTCCAG ATTCTGCTAC CAAGACAGCG
 6181 CAGGCGATGT TGGACTTCAA CCGTGAAGGA TTACCTCTGT TCATACTTGC TAACTGGAGA
 6241 GGCTTCTCTG GAGGGCAAAG AGATCTTTTT GAAGGAATTC TGCAGGCTGG GTCAACAATT
 6301 GTTGAGAACC TTAGGACATA CAATCAGCCT GCCTTTGTAT ATATCCCCAA GGCTGCAGAG
 6361 CTACGTGGAG GAGCCTGGGT CGTGATTGAT AGCAAGATAA ACCCAGATCG CATCGAGTGC
 6421 TATGCTGAGA GGA CTGCAAA GGGTAATGTT CTCGAACCTC AAGGGTTGAT TGAGATCAAG
 6481 TTCAGGTCAG AGGAACTCAA AGAATGCATG GGTAGGCTTG ATCCAGAATT GATAGATCTG
 6541 AAAGCAAGAC TCCAGGGAGC AAATGGAAGC CTATCTGATG GAGAATCCCT TCAGAAGAGC
 6601 ATAGAAGCTC GGAAGAAACA GTTGCTGCCT CTGTACACCC AAATCGCGGT ACGTTTTGCG
 6661 GAATTGCACG ACACTTCCCT TAGAATGGCT GCTAAAGGTG TGATCAGGAA AGTTGTAGAC
 6721 TGGGAAGACT CTCGGTCTTT CTTCTACAAG AGATTACGGA GGAGGCTATC CGAGGACGTT
 6781 CTGGCAAAGG AGATTAGAGG TGTAATTGGT GAGAAGTTTC CTCACAAATC AGCGATCGAG
 6841 CTGATCAAGA AATGGTACTT GGCTTCTGAG GCAGCTGCAG CAGGAAGCAC CGACTGGGAT
 6901 GACGACGATG CTTTTGTCGC CTGGAGGGAG AACCCTGAAA ACTATAAGGA GTATATCAAA
 6961 GAGCTTAGGG CTCAAAGGGT ATCTCGGTTG CTCTCAGATG TTGCAGGCTC CAGTTCGGAT
 7021 TTACAAGCCT TGCCGCAGGG TCTTTCCATG CTA CTAGATA AGATGGATCC CTCTAAGAGA
 7081 GCACAGTTTA TCGAGGAGGT CATGAAGGTC CTGAAATGAT CAAATGATAC CAACACATCC
 7141 AATACAGTAT GTGCATGATA TCTGTTTCTC TTGAAGTACA TATATAGATG GATACAAGGC
 7201 GGCTGTA ACT GATGGTAGCT AATCTGGGCC AACCATTACT TTTGTGA ACT TGCTGGTGGC
 7261 CTTTATTATT CAAGGCACAG CTCGCCTTCG GACCCCTCC GGCTGGTTGA TGATGAGTGT
 7321 AACTGGATGT GTTAGTTCTG CTGCCACAGA ATTCGAGAAG GATAGGGGCA TGCGGGTTTT
 7381 GCCTCCTGTT GGCAAGAACA CTGGTGATTT TGAGTTCTTG TTATGTGGAC TGTGGTAGTC
 7441 TTGTTTCGCT GTAGTTCTGT GATGTTCTAT CTCCTGTAAT TCTAGTCTTG GGAGAGTGAT
 7501 TCAGATGTCC ATTCAATTTT GAACCTGAAT AATAATATGC TTTGTAGGCC TATGCGTACC
 7561 AGTATGTGGA ATAAATGTTT GTTGAGTTA

FIG.26A-3

AMY310767 - Proteína traducida de *Alopecurus myosuroides* plastidal ACCasa

```

1  MGSTHLPVIG FNASTTPSLS TLRQINSAAA AFQSSSPSR SSKKSRRVKS
51  IRDDGGGSVP DPAGHGQSTR QGLAGITDLP KEGASAPDVO ISHGSEDHKA
101 SYQMNGILNE SHNGRHASLS KYVEFCTELG GKTPIHSLV ANNGMAAAKF
151 MRSVRTWAND TFGSEKATQL IANATPEDMR INAHIIRIAD QFVEVPGGTN
201 NNNYANVOLI VEIAERTGVS AVWPGWGHAS ENPELPDALT AKGIVFLGPP
251 ASSMNALGDK VGSALIAQAA QVPTLAWGGS HVEITPLELCL DSIPEEMYRK
301 ACVTTADEAV ASCOMIGYPA MIKASWGGGG KGIRKVNDD EVKALFKQVQ
351 GEVPGSPIFI MRLASQSRHL EVQLLCDEYG NVAALHSRDC SVQRRHQKII
401 EEGPVTAVPR ETVKELEQAA RRLAKAVGYV GAATVEYLYS METGEYFFLE
451 LNPRLQVEHP VTESIAEVNL PAAQVAVGMG IPLWQIPEIR RFGMDNGGG
501 YDIWRKTAAL ATPFNDEVD SQWPKGHQVA VRITSENDD GFKPTGGKVK
551 EISFKSKPNV WGYFSVKS GG IHEFADSQF GHVFAYGETR SAAITSMSLA
601 LKEIQIRGEI HTNVDYTVDL LNAPOFRENT IHTGWLDTRI AMRVDQERPP
651 WYISVVGAL YKTITTNAET VSEYVSYLK GQIPPKHSL VHSTISLNE
701 ESKYTETVR SGQGSYRLRL NGSLIEANVO TLCDGGLLMQ LDGNSHVITYA
751 EEEAGGTRLL IDGKTCLLQN DHDPSRLAE TPCKLLRFLI ADGAHVDAOV
801 PYAEVEVMKM CNPLSPAAG VINULLSEGO AMOAGDLIAR LDLDPSAVK
851 RAEPFEGSFP EMSLPAAAG QVHKRCAASL NAARMVLGY DHAANKVVOO
901 LVMLCDTPAL PFLQWEELMS VLATRLPRRL KSELEGKYNE YKLNVDHVKI
951 KDFPTMLRE TIEENLACVS EKENVTIERL VDPLMSLLKS YEGGESHIAH
1001 FTVKSLFEEY LSVEELFSDG IQSDVIERLR LOYSKDLQKV VOIVLSHOGV
1051 RNKTKLILAL MEKLVYPNPA AYRQQLIRFS SLNHKRYYKL ALKASELLEQ
1101 TKLSELRTSI ARNLSALDMF TEEKADFSLO DRKLATNESM GDLVTAPLPV
1151 EDALVSLFDC TDQTLOQRVI QTYISRLYQP OLVKDSIQLK YDSGVIALW
1201 EFTEGNHEKR LGAMVILKSL ESVSTAIGAA LKQASHYASS AGNTVHIALL
1251 DADTQLNTE DSGDNQQAQD KMDKLSFVLK QDVVMADLRA ADVKVVSCIV
1301 QRDGAINPMR RTFLLSEEKL CYEEEPILRH VEPPLSALLE LDKLKVKGYN
1351 EMKYTPSRDR QWHIYTLRNT ENPKMLHRVF FRTLVRQPSA GNRFTSDHIT
1401 DVEVGHAEEP LSFTSSSILK SLKIAKEELE LHAIRTGSH MYLCILKEOK
1451 LLDLVPVSGN TVVDVGQDEA TACSLLKEMA LKIHVLGAR MHLVSCWE
1501 VKLKLVSOGP ASGSWRVVT NVTGHTCTVD IYREVEDTES QKLVIYSTAL
1551 SSGPLHGVAL NTSYQPLSVI DLKRCARNN KTTYCYDFPL TFEAAVQKSW
1601 SNISSENNQC YVKATELVFA EKNGSWGTP IPMQRAAGLN DIGMVAWILD
1651 MSTPEFPSGR QIIVIANIT FRAGSFGPRE DAFPEAVTNL ACEKKLPLIY
1701 LAANGGARIG IADEVKSCFR VGTDDSSPE RGFYIYMTD EDHDIRIGSY
1751 IAHKMLQDSG EIRWVIDSVV GKEDGLGVEN IHGSAATASA YSRAYEETFT
1801 LTFVTGRTVG IGAYLARLGI RCIORTDQPI ILTGFSALNK LLGREVYSSH
1851 MQLGQPKIMA TNGVHILVTP DDLEGVSNIL RWLSYVPANI GGPLPITKSL
1901 DPIDRPVAYI PENTCDPRAA ISGIDDSQK WLGGMFQKDS FVETFEQWAK
1951 TVVTGRAKLG GIPVGVIAVE TOTMMQLVPA OPGOPDSHER SVPRAGQVWF
2001 PDSATKTAQA MDPNREGLP LFIANWRGF SGGOROLFEG ILOAGSTIVE
2051 NLRTYNQPAF VYIPKAAELR GGAVVVIDSK INPDRIEYA ERTAKGNVLE
2101 PQGLIEIKFR SEELKECMGR LDPELIDLKA RLOGANGSL S DGESLQKSTI
2151 ARKKQLPLY TQIAVRFAEL HDTSLRMAAK GVIRKVVOWE DSRFFYKRL
2201 RRLSEDVLA KEIRGVIGEK FPHKSAIELI KKWYLASEAA AAGSTOWDD
2251 DAFVAVREN ENYKEYIKEL RAQRVSRLS DVAGSSSOLQ ALPOGLSNLL
2301 DKMDPSKRAQ FIEEVMKVLK

```

FIG.26B

ADNc de arroz *Oriza sativa* plastidal ACCasa Os05g22940 (Cypress)

```

1  ATGACATCCA CACATGTGGC GACATTGSSA GTTGGTGECC AGGCACCTCC TCGTCACCAG
61  AAAAAGTCAG CTGGCACTGC ATTTGTATCA TCTGGGTCAT CAAGACCTCC ATACCGAAAG
121 AATGGTCAGC GTA CTGGTC ACTTAGGSA GAAAGCAATG GAGGAGTGTG TGATTCCAAA
181 AAGCTTAACC ACTCTATTGG CCAAGGTCCT GCTGGCATCA TTGACCTGCC AAATGACGCA
241 GCTTCAGAAG TTGATATTTC ACATGGTTCC GAAGATCCCA GGGGGCCTAC GGTCCAGGT
301 TCCTACCAAA TGAATGGGAT TATCAATGAA ACACATAATG GGAGGCATGC TTCAGTCTCC
361 AAGGTTGTTG AGTTTTGTAC GGCACCTGGT GGCAAAACAC CAATTCACAG TGTATTAGTG
421 GCCAACAATG GAATGGCAGC AGCTAAGTTC ATGCGGAGTG TCCGAACATG GGCTAATGAT
481 ACTTTTGGAT CAGAGAAGGC AATTCAGCTG ATAGCTATGG CAACTCCGGA GGATCTGAGG
541 ATAAATGCAG AGCACATCAG AATTGCCGAT CAATTTGTAG AGGTACCTGG TGGAAACAAAC
601 AACAACTAAT ATGCAAAATG CCAACTCATA GTGGAGATAG CAGAGAGAAC AGGTGTTTTCT
661 GCTGTTTGGC CTGTTTGGGG TCATGTCATCT GAGAATCCTG AACTTCCAGA TGGCTGACT
721 GCAAAAGGAA TTGTTTTTCT TGGGCCACCA GCATCATCAA TGCATGCATT AGGAGACAAG
781 GTTGGCTCAG CTCTCATTCG TCAAGCAGCT GGAGTTCCAA CACTTGCTTG GAGTGGATCA
841 CATGTGGAAG TTCTCTGGA GTGTTGCTTG GACTCAATAC CTGATGAGAT GTATAGAAAA
901 GCTGTGTGTA CTACCACAGA GGAAGCAGTT GCAAGTTGTC AGGTGGTTGG TTATCCTGCC
961 ATGATTAAGG CATCTTGGGG TGGTGGTGGT AAAGGAATAA GGAAGGTTCA TAATGATGAT
1021 GAGGTTAGGA CATTATTTAA GCAAGTTCAA GGGGAAGTAC CTGTTTCCC AATATTTATC
1081 ATGAGGCTAG CTGCTCAGAG TCGACATCTT GAAGTTCAGT TGCTTTGTGA TCAATATGTC
1141 AACGTAGCAG CACTTCACAG TCGAGATTGC AGTGTACAAC GGCACACCA AAAGATAATC
1201 GAGGAAGGAC CAGTTACTGT TGCTCCTCST GAGACTGTGA AAGAGCTTGA GCAGGCAGCA
1261 CGGAGGCTTG CTAAAGCTGT GGGTTATGTT GGTGCTGCTA CTGTTGAATA CCTTTACAGC
1321 ATGGAAACTG GTGAATATTA TTTCTGGAA CTTAATCCAC GGTACAGGT TGAGCATCCT
1381 GTCAGTGAAT GGATAGCTGA AGTAAATTTG CCTGCGGCTC AAGTGTCTGT TGGAAATGGT
1441 ATACCCCTTT GGCAGATTCC AGAGATCAGG CGCTTCTACG GAATGAACCA TGGAGGAGGC
1501 TATGACCTTT GGAGGAAAAC AGCAGCTCTA GCGACTCCAT TTAACCTTGA TGAAGTAGAT
1561 TCTAAATGGC CAAAAGGCCA CTGCGTAGCT GTTAGAATAA CTAGCGAGGA TCCAGATGAT
1621 GGGTTTAAGC CTACTGGTGG AAAAGTAAAG GAGATAAGTT TCAAGAGTAA ACCAAATGTT
1681 TGGGCTATT TCTCAGTAAA GTCTGGTGGG GGCATCCATG AATTCGCTGA TTCTCAGTTC
1741 GGACATGTTT TTGCGTATGG AACTACTAGA TCGGCAGCAA TAACTACCAT GGCTCTTGCA
1801 CTAAAGAGAG TTCAAATTCG TGGAGAAATT CATTCAAACG TAGACTACAC AGTTGACCTA
1861 TTAAATGCCT CAGATTTTAG AGAAAATAAG ATTCACTCTG GTTGGCTGGA TACCAGGATA
1921 GCCATGCGTG TTCAAGCTGA GAGGCTTCCA TGGTATATTT CAGTCTTTGG AGGGGCTTTA
1981 TATAAAACAG TAACTGCCAA CACGGCCACT GTTCTGATT ATGTTGGTTA TCTTACCAAG
2041 GGCCAGATTC CACCAAAGCA TATATCCCTT GTCTATACGA CTGTTGCTTT GAATATAGAT
2101 GGGAAAAAAT ATACAATCGA TACTGTGAGG AGTGGACATG GTAGCTACAG ATTGCGAATG
2161 AATGATCAA CGGTTGACGC AAATGTACAA ATATTATGTG ATGTTGGGCT TTTAATGCAG
2221 CTGGATGGAA ACAGCCATGT AATTATGCT GAAGAAGAGG CCAGTGGTAC ACGACTTCTT
2281 ATTGATGGAA AGACATGCAT GTTACAGAA GACCATGACC CATCAAGTT ATTAGCTGAG
2341 ACACCATGCA AACTTCTTCG TTTCTTGGTT GCTGATGGTG CTCATGTTGA TGCTGATGTA

```

FIG.26C-1

2401 CCATATGCGG AAGTTGAGGT TATGAAGATG TGCATGCCCC TCTTATCACC CGCTTCTGGT
 2461 GTCATACATG TTGTAATGTC TGAGGGCCAA GCAATGCAGG CTGGTGATCT TATAGCTAGG
 2521 CTGGATCTTG ATGACCCCTC TGCTGTTAAG AGAGCTGAGC CGTTCGAAGA TACTTTTCCA
 2581 CAAATGGGTC TCCCTATTGC TGCTTCTGGC CAAGTTCACA AATTATGTGC TGCAAGTCTG
 2641 AATGCTTGTC GAATGATCCT TCGGGGTAT GAGCATGATA TTGACAAGGT TGTGCCAGAG
 2701 TTGGTATACT GCCTAGACAC TCCGGAGCTT CCTTCTCTGC AGTGGGAGGA GCTTATGTCT
 2761 GTTTTAGCAA CTAGACTTCC AAGAAATCTT AAAAGTGAGT TGGAGGGCAA ATATGAGGAA
 2821 TACAAAGTAA AATTTGACTC TGGGATAATC AATGATTTCC CTGCCAATAT GCTACGAGTG
 2881 ATAATTGAGG AAAATCTTGC ATGTGGTTCT GAGAAGGAGA AGGCTACAAA TGAGAGGCTT
 2941 GTTGAGCCTC TTATGAGCCT ACTGAAGTCA TATGAGGGTG GGAGAGAAAG TCATGCTCAC
 3001 TTTGTTGTCA AGTCCCTTTT TGAGGAGTAT CTCTATGTTG AAGAAATTGT CAGTGATGGA
 3061 ATTCAGTCTG ATGTGATTGA GCGTCTGCGC CTTCAACATA GTAAAGACCT ACAGAAGGTC
 3121 GTAGACATTG TGTTGTCCCA CCAGAGTGTT AGAAATAAAA CTAAGCTGAT ACTAAACTC
 3181 ATGGAGAGTC TGGTCTATCC AAATCCTGCT GCCTACAGGG ATCAATTGAT TCGCTTTTCT
 3241 TCCCTTAATC ACAAAGCGTA TTACAAGTTG GCACTTAAAG CTAGTGAAC TCTTGAACAA
 3301 ACAAACCTTA GTGAGCTCCG TGCAAGAATA GCAAGGAGCC TTTCAGAGCT GGAGATGTTT
 3361 ACTGAGGAAA GCAAGGGTCT CTCCATGCAT AAGCGAGAAA TTGCCATTAA GGAGAGCATG
 3421 GAAGATTTAG TCACTGCTCC ACTGCCAGTT GAAGATGCGC TCATTTCTTT ATTTGATTGT
 3481 AGTGATACAA CTGTTCAACA GAGAGTGATT GAGACTTATA TAGCTCGATT ATACCAGCCT
 3541 CATCTTGTA AGGACAGTAT CAAATGAAA TGGATAGAAT CGGGTGTTAT TGCTTTATGG
 3601 GAATTTCTTG AAGGGCATT TGTGCAAGA AATGGAGGAG CGGTTCTTGG TGACAAAAGA
 3661 TGGGGTGCCA TGGTCATTGT CAAGTCTCTT GAATCACTTT CAATGGCCAT TAGATTTGCA
 3721 CTAAAGGAGA CATCACACTA CACTAGCTCT GAGGGCAATA TGATGCATAT TGCTTTGTTG
 3781 GGTGCTGATA ATAAGATGCA TATAATTCAA GAAAGTGGTG ATGATGCTGA CAGAATAGCC
 3841 AAACCTCCCT TGATACTAAA GGATAATGTA ACCGATCTGC ATGCCTCTGG TGTGAAAACA
 3901 ATAAGTTTCA TTGTTCAAAG AGATGAAGCA CGGATGACAA TGCCTCGTAC CTTCCTTTGG
 3961 TCTGATGAAA AGCTTTCTTA TGAGGAAGAG CCAATTCTCC GGCATGTGGA ACCTCCTCTT
 4021 TCTGCACTTC TTGAGTTGGA CAAGTTGAAA GTGAAAGGAT ACAATGAAAT GAAGTATACC
 4081 CCATCACGGG ATCGTCAATG GCATATCTAC ACACTTAGAA ATACTGAAAA CCCCCAATG
 4141 TTGCACCGGG TATTTTCCG AACCTTGTC AGGCAACCCA GTGTATCCAA CAAGTTTCTT
 4201 TCGGGCCAGA TTGGTGACAT GGAAGTTGGG AGTGCTGAAG AACCTCTGTC ATTTACATCA
 4261 ACCAGCATAT TAAGATCTTT GATGACTGCT ATAGAGGAAT TGGAGCTTCA CGCAATTAGA
 4321 ACTGGCCATT CACACATGTA TTTGCATGTA TTGAAAGAAC AAAAGCTTCT TGATCTTGT
 4381 CCAGTTTCAG GGAATACAGT TTTGGATGTT GGTCAAGATG AAGCTACTGC ATATTCACCT
 4441 TTAAGAGAAA TGGCTATGAA GATACATGAA CTTGTTGGTG CAAGAATGCA CCATCTTTCT
 4501 GTATGCCAAT GGGAAAGTGA ACTTAAGTTG GACTGCGATG GTCCTGCCAG TGGTACCTGG
 4561 AGGATTGTAA CAACCAATGT TACTAGTCAC ACTTGCACTG TGGATATCTA CCGTGAGATG
 4621 GAAGATAAAG AATCACGGAA GTTAGTATAC CATCCGCCA CTCCGGCGGC TGGTCCTCTG
 4681 CATGGTGTGG CACTGAATAA TCCATATCAG CCTTTGAGTG TCATTGATCT CAAACGCTGT
 4741 TCTGCTAGGA ATAATAGAAC TACATACTGC TATGATTTTC CACTGGCATT TGAACTGCA
 4801 GTGAGGAAGT CATGGTCCTC TAGTACCTCT GGTGCTTCTA AAGGTGTTGA AAATGCCAA
 4861 TGTTATGTTA AAGCTACAGA GTTGGTATTT GCGGACAAAC ATGGGTCATG GGGCACTCCT

FIG.26C-2

```

4921 TTAGTTCAAA TGGACCGGCC TGCTGGGCTC AATGACATTG GTATGGTAGC TTGGACCTTG
4981 AAGATGTCCA CTCCTGAATT TCCTAGTGGT AGGGAGATTA TTGTTGTTGC AAATGATATT
5041 ACGTTCAGAG CTGGATCATT TGGCCCAAGG GAAGATGCAT TTTTGAAGC TGTACCAAC
5101 CTAGCCTGTG AGAAGAAACT TCCTCTTATT TATTTGGCAG CAAATTCTGG TGCTCGAATT
5161 GGCATAGCAG ATGAAGTGAA ATCTTGCTTC CGTGTTGGGT GGTCTGATGA TGGCAGCCCT
5221 GAACGTGGGT TTCAGTACAT TTATCTAAGC GAAGAAGACT ATGCTCGTAT TGGCACTTCT
5281 GTCATAGCAC ATAAGATGCA GCTAGACAGT GGTGAAATTA GGTGGGTTAT TGATTCTGTT
5341 GTGGGCAAGG AAGATGGACT TGGTGTGGAG AATATACATG GAAGTGCTGC TATTGCCAGT
5401 GCTTATTCTA GGGCATATAA GGAGACATTT ACACTTACAT TTGTGACTGG AAGAAGTGT
5461 GGAATAGGAG CTTATCTTGC TCGACTTGGC ATCCGGTGCA TACAGCGTCT TGACCAGCCT
5521 ATTATTCTTA CAGGCTATTC TGCCTGAAC AAGCTTCTTG GGCAGGAAGT GTACAGCTCC
5581 CACATGCAGT TGGGTGGTCC CAAATCATG GCAACTAATG GTGTTGTCCA TCTTACTGTT
5641 TCAGATGACC TTGAAGGCGT TTCTAATATA TTGAGGTGGC TCAGTTATGT TCCTGCCTAC
5701 ATTGGTGGAC CACTTCCAGT AACAACACCG TTGGACCCAC CGGACAGACC TGTTGCATAC
5761 ATTCCTGAGA ACTCGTGTGA TCCTCGAGCG GCTATCCGTG GTGTTGATGA CAGCCAAGGG
5821 AAATGGTTAG GTGGTATGTT TGATAAAGAC AGCTTTGTGG AAACATTTGA AGGTTGGGCT
5881 AAGACAGTGG TTAGTGGCAG AGCAAAGCTT GGTGGAATTC CAGTGGGTGT GATAGCTGTG
5941 GAGACTCAGA CCATGATGCA AACTATCCCT GCTGACCCTG GTCAGCTTGA TTCCCGTGAG
6001 CAATCTGTTC CTCGTGCTGG ACAAGTGTGG TTTCCAGATT CTGCAACCAA GACTGCGCAG
6061 GCATTGCTGG ACTTCAACCG TGAAGGATTA CCTCTGTTCA TCCTCGCTAA CTGGAGAGGC
6121 TTCTCTGGTG GACAAAGAGA TCTTTTGA GGAATTCTTC AGGCTGGCTC GACTATTGTT
6181 GAGAACCTTA GGACATACAA TCAGCCTGCC TTTGTCTACA TTCCCATGGC TGAGAGCTA
6241 CGAGGAGGGG CTTGGGTTGT GGTGATAGC AAGATAAACC CAGACCGCAT TGAGTGCTAT
6301 GCTGAGAGGA CTGCAAAAGG CAATGTTCTG GAACCGCAAG GGTAAATTGA GATCAAGTTC
6361 AGGTCAGAGG AACTCCAGGA TTGCATGAGT CGGCTTGACC CAACATTAAT TGATCTGAAA
6421 GCAAACTCG AAGTAGCAAA TAAAAATGGA AGTGCTGACA CAAAATCGCT TCAAGAAAAT
6481 ATAGAAGCTC GAACAAAACA GTTGATGCCT CTATATACTC AGATTGCGAT ACGGTTTGCT
6541 GAATTGCATG ATACATCCCT CAGAATGGCT GCGAAAGGTG TGATTAAGAA AGTTGTGGAC
6601 TGGGAAGAAT CACGATCTTT CTTCTATAAG AGATTACGGA GGAGGATCTC TGAGGATGTT
6661 CTTGCAAAAG AAATTAGAGC TGTAGCAGGT GAGCAGTTTT CCCACCAACC AGCAATCGAG
6721 CTGATCAAGA AATGGTATTC AGCTTCACAT GCAGCTGAAT GGGATGATGA CGATGCTTTT
6781 GTTGCTTGGA TGGATAACCC TGAAAACCTAC AAGGATTATA TTCAATATCT TAAGGCTCAA
6841 AGAGTATCCC AATCCCTCTC AAGCTTTTCA GATTCCAGCT CAGATTTGCA AGCCCTGCCA
6901 CAGGGTCTTT CCATGTTACT AGATAAGATG GATCCCTCTA GAAGAGCTCA ACTTGTTGAA
6961 GAAATCAGGA AGGTCCTTGG TTGAATCATA TGATGCCAAA ACTATTATTG GAGGCACAAA
7021 TAGCTTGTGG ACCCTGTCGG ATTGTTGGTG AGTGATATT GGATTTGTTA GTTCTGCCAG
7081 ATGAAAGTGC AAGTCTGATG ATTCATGATA CCGTCAGTTG GCAAGAACAC CGGTTAACCT
7141 GAGTGCTTGT TTACAAATGG TCCTTTATGA CAATCGTTGT TTCGCGCTAG TTCCGTGATC
7201 TACTATCATC TGTTAGACGC TGTAATTAGT GAGTCTCCGC GGATCCACAG TATACGGTTG
7261 AGCTGTTGAT TCAATTTTGG ACACGAATAA TATGATTTTG TAGGCATAAA TGCGTCTGTA
7321 TGTGAAATAA ATTGTCTGTT GAGTTAACAC ACAAGATGAC AATATGTTTG TGCTCTACTG
7381 CTATTGTCCA TGAATACTGA TTGCGGAATC AACCACATGC ATTAT

```

FIG.26C-3

Proteína de arroz *Oriza sativa* plastidal ACCasa Os05g22940 (Cypress)

```

1   MTSTHVATLG VQAQAPPRHQ KKSAGTAFVS SGSSRPSYRK NGQRTSLRE
51  ESNGGVSDSK KLNHSTROGL AGIIDLPNDA ASEVDISHGS EDPRGPTVPG
101 SYQMNGIINE THNGRHASVS KVVEFCTALG GKTPIHSLVY ANNGMAAAKF
151 MRSVRTWAND TFGSEKAIQL IAMATPEDLR INAEHIRIAD QFVEVPGGTN
201 NNNYANVQLI VETAERTGVS AVWPGWGHAS ENPELPAALT AKGIVFLGPP
251 ASSMHALGDK VGSALIAQAA GVPTLAWSGS HVEVPLECCL DSIPDEMYRK
301 ACVTTTTEEAV ASCQVVGYPY NIKASWGGGG KGIRKVHNDQ EVRTLFKQVQ
351 GEVPGSPIFI MRLAAOSRHL EVQLLCDQYG NVAALHSRDC SVQRRHQKII
401 EEGPVTVAPR ETVKELEQAA RRLAKAVGYV GAATVEYLYS METGEYYFLE
451 LNPRLQVEHP VTEWIAEVNL PAAQVAVGMG IPLWQIPEIR RFYGMNHGGG
501 YDLWRKTAAL ATPFNDFEVD SKWPKGHCVA VRITSEDPDD GFKPTGGKVK
551 EISFKSKPNV WAYFSVKSGG GIHEFADSQF GHVFAYGTTR SAAITTMALA
601 LKEVQIRGEI HSNVDYTVDL LNASDFRENK IHTGWLDTRI AMRVOAERPP
651 WYISVVGGAAL YKTVTANTAT VSDYVGYLTK GQIPPKHISL VYTTVALNID
701 GKQYITDVR SGHGSYRLRM NGSTVDANVQ ILCDGGLLMQ LDGNSHVIYA
751 EEEASGTRLI IDGKTCMLQN DHDPSKLLAE TPCKLLRFLV ADGAHVDAV
801 PYAEVEVMKM CMPLLSPASG VINVMSEGO AMQAGDLIAR LDLDPSAVK
851 RAEPFEDTFP QMGLPIAASG QVHKLCAASL NACRMILAGY EHDIDKVVPE
901 LVYCLDTPEL PFLQWEELMS VLATRLPRNL KSELEGKYEE YKVKFDSGII
951 NDFPANMLRV IIEENLACGS EKEKATNERL VEPLMSLLKS YEGGRESHAN
1001 FVVKSLFEY LYVEELFSDG IQSDVIERLR LQHSKDLQKV VDIVLSHQSV
1051 RNICKLILKL MESLVYPNPA AYRDOLIRFS SLNHKAYYKL ALKASELLEQ
1101 TKLSELRARI ARSLSELEMF TEESKGLSMH KREIATKESM EDLVTAPLPV
1151 EDALISLFDC SDTTVQORVI ETYIARLYQP HLVKDSIKMK WIESGVIALW
1201 EFPEGHFDAR NGGAVLGDKR WGAMVIVKSL ESLSMATRFA LKETSHYTSS
1251 EGNMHIALL GADNMHIHQ ESGODADRIA KLPLILKDNV TDLHSGVKVT
1301 ISFTVQRDEA RMTMRRFTLW SDEKLSYEEE PILRHVEPPL SALLELDKLL
1351 VKGYNEMKYT PSRDQRWHYI TLRNTENPKM LHRVFFRTL V ROPSVSNKFS
1401 SGOIGDMEVG SAEPLSFTS TSILRSLMTA IEELELHAIR TGHSHMYLHV
1451 LKEQKLLDLV PVSQNTVLDV GQDEATAYSL LKEMANKIHE LVGARMHLS
1501 VCQWEVKKLK DCDGPASGTW RIVTTNVTSH TCTVDIYREM EDKESRKL VY
1551 HPATPAAGPL HGVALNNPYQ PLSVIDLKRC SARNNRTTYC YDFPLAFETA
1601 VRKSWSSSTS GASKGVENAO CYVKATELVF ADKHGSGWTP LVQMDRPA GL
1651 NDIGMVAWTL KMSTPEFPSG RELIIVANDI TFRAGSFGPR EDAFFEAVTN
1701 LACEKKLPLI YLAANSRARI GIADSVKSCF RVGWSDDGSP ERGFQYIYLS
1751 EEDYARIGTS VIANKMQLDS GEIRWVIDSV VGKEDGLGVE NTHGSAATAS
1801 AYSRAYKETF TLTFVTGRV GIGAYLARLG IRCIQRLDQ IILTGYSA LN
1851 KLLGREVYSS HMQLGGPKIM ATNGVVHLTV SDDLEGVSNI LRWLSYVPAY
1901 IGGPLPVTP LDPPDRPVAY IPENSCDPA AIRGVDSQK KWLGGMFQD
1951 SFVETFEQWA KTVVTGRKL GGIPVGVIAV ETQTMNQITP ADPGQLDSRE
2001 QSVPRAGQVW FPD SATKTAQ ALLDFNREGL PLFILANWRG FSGGQDRLFE
2051 GILQAGSTIV ENLRTYNQPA FVYIPMAEL RGGAWVVVDS KINPDRIECY
2101 AERTAKGNVL EPQGLIEIKF RSEELQDCMS RLDPTLIDLK AKLEVANKNG
2151 SADTKSLOEN IEARTKQLMP LYTQIAIRFA ELHDTSLRMA AKGVIKKVVD
2201 WEESRSFFYK RLRRRISESV LAKETRAVAG EQFSHOPATE LIKKWYSASH
2251 AAEWDDDAF VAWMDNPENY KOYIOYLKAO RVSQSLSSLS DSSSDLQALP
2301 QGLSMLLDKM DPSRRAQLVE EIRKVLG

```

FIG.26D

ADN genómico de arroz de Oriza sativa plasmid ACCasa Os05g22940 - AP008211
fragmento complementado inverso

```

1 TCATTCTTAT ATATTTTCAT CTGTCAGATT TCACACATCT GGGGATTTAT CTTCTCTTTG
61 TATGGCACTA CACATTTGAG AAACCGTGCA ATTCTACTGT TTGGTCAGCA GGACAACAAT
121 GACATCCACA CATGTGSCGA CATTGGGAGT TGGTGCCAG GCACCTCCTC GTCACCAGAA
181 AAAGTCAGCT GGCACGSCAT TTGTATCATC TGGGTCATCA AGACCTCAT ACCGAAAGAA
241 TGGTCAGCGT ACTCGGTCAC TTAGGGAAGA AAGCAATGGA GGAGTGTCTG ATTCACAAAA
301 GCTTAACCAC TCTATTGCCC AAGGTGACCA CTAGCTACTT TACATATGCT ATAATTTGTG
361 CCAAACATAA ACATGCAATG GCTGCTATTA TTTAAACGTT AATGTTGAAA TAGCTGCTAT
421 AGGATACAGC AAAAATATAT AATTGACTGG GCAAGATGCA ACAATTGTTT TTCACTAAAG
481 TTAGTTATCT TTTGCTGTAA AAGACAACCTG TTTTTCATAT AAAATGGTAT TAATAACCTT
541 GTAATATTCA ATGCAACATG TTCTCAAGTA AAAAAAACA TTGCTGGTT GTATAAGCAA
601 ATGTTGCGTT GTAGACATCT TATTAAACCT TTTTGTGATA TCTATTACCG TAGGGAACAG
661 GGGAGCTGTT TAAATCTGTT ATCATAGAGT AATATGAGAA AAGTGGATTG TGCGACTTTG
721 GCATGTATAC CTGCTCAATT TCAAATATAT GTCTATGTGC AGGTCTTGCT GGCATCATTG
781 ACCTCCAAA TGACGAGCT TCAGAAGTTG ATATTTTACA GTAAGGACTT TATATTTTAT
841 AATAATTATT ATATAATTTT CTGACATGTT TTGAGAACCT CAAAACATGT GATTGCACCT
901 TCCTTTTTTA TGTCTGGTTC AGAAACTGAT AAGTTTTGAC AGTGTTTAGG ATGGATCTTT
961 GATGGGCACA GTGCTTTCTA ATGTTTTCAT TTTTGAAAGT AATGTTTATG GAAGAATAT
1021 CTGATTAAAT TTATACTTTA TCTTTACAAA AGTCAAATGC GTTCTGTATC AATTGCGGTT
1081 TGTAATATGG CAAGAACATG CTTTCAGAAAT TTGTTTATAC AATGCTTTCT TTCTATTATT
1141 ATGTAGACAA AATACCTAAT ACTTTGTTCA CCTTTTATAG TGGACACCTC TCACAGCTTT
1201 TTCAGTAAGT GATGCAATTT TGTACATTTG TAAGATGTGT TCCAGAAACC TTTTCCTCTG
1261 CAATTCATAT GTACCCACTC AAAGTGGTAT CACCAAGAT CTCCATCTGA TTGAAAAAAA
1321 GCTGCGTGAA GATGCTTTAT TTATGCTAAC CATACTGAT TTATACTGTT TTATAGTACA
1381 ATGCTTATTT ATGCTAACCA TACATAATTT TATTCTGTTT TCTAGTACAT TATTTGTGCC
1441 CCTGACCATA AATGATCCTT TCTTTTACAG TGGTTCCGAA GATCCAGGG GGCCTACGGT
1501 CCCAGGTTCC TACCAAATGA ATGGGATTAT CAATGAAACA CATAATGGA GGCATGCTTC
1561 AGTCTCCAAG GTTGTGAGT TTTGTACGSC ACTTGSTGGC AAAACACCAA TTCACAGTGT
1621 ATTAGTGGCC AACAATGEAA TGGCAGCAGC TAAGTTCATG CGGAGTGTCC GAACATGGGC
1681 TAATGATACT TTTGGATCAG AGAAGGCAAT TCAGCTGATA GCTATGGCAA CTCGGAGGA
1741 TCTGAGGATA AATGCAGAGC ACATCAGAAAT TGCCGATCAA TTTGTAGAGG TACCTGGTGG
1801 AACAAACAAC AACAACTATG CAAATGTCCA ACTCATAGTG GAGGTTAGTT CAGCTCATCC
1861 CTCAACACAA CATTTTGGTT TCTATTTAAG TTAGGAAAAA ATCTCTACGA CCTCCAATT
1921 TCTGAACATC CAATTTTCAC CATCAACTGC AATCACAGAT AGCAGAGAGA ACAGGTGTTT
1981 CTGCTGTTTG GCTSGTTGG GGTGATGSCAT CTGAGAATCC TGAACCTCCA GATGCGCTGA
2041 CTGCAAAAGG AATTGTTTTT CTTGGGCCAC CAGCATCATC AATGCATGCA TTAGGAGACA
2101 AGGTTGGCTC AGCTCTCATT GCTCAAGCAG CTGGAGTTCC AACACTTGCT TGGAGTGGAT
2161 CACATGTGAG CCTTGTCTTC TCTTTTATAG CTTATCATCT TATCTTTTCG GTGATGCATT
2221 ATCCCAATGA CACTAAACCA TAGGTGGAAG TTCCTCTGGA GTGTTGCTTG GACTCAATAC
2281 CTGATGAGAT GTATAGAAAA GCTTGTGTTA CTACCACAGA GGAAGCAGTT GCAAGTTGTC
2341 AGGTGGTTGG TTATCTGCC ATGATTAAGG CATCTTGGGG TGCTGGTGGT AAAGGAATAA

```

FIG.26E-1

```

2401 GGAAGGTTTG TTCTTCTTGT AGTTATCAAG AGATTGTTTG GATTGCAAGT GTTTAGTGCC
2461 CATAGTTAAC TCTGGTCTTT CTAACATGAG TAACTCAACT TTCTTGCAAG TTCATAATGA
2521 TGATGAGGTT AGGACATTAT TTAAGCAAGT TCAAGGCGAA GTACCTGGTT CCCCATATT
2581 TATCATGAGG CTAGCTGCTC AGGTGGGGCC TTTTATGGAA GTTACACCTT TTCCCTTAAT
2641 GTTGAGTTAT TCCGGAGTTA TTATGGTTAT GTTCTGTATG TTTGATCTGT AAATTATTGA
2701 AATTCACCTC CATTGGTTCT CCAGATTAGC AGACCTACAA TTCTACATAT GGTTTATACT
2761 TTATAAATAC TAGGATTTAG GGATCTTCAT ATAGTTTATA CATGGTATTT AGATTTCACT
2821 TGTAACCCTA TTGAAGACAT CCTGATTGTT GTCTTATGTA GAGTCGACAT CTTGAAGTTC
2881 AGTTGCTTTG TGATCAATAT GGCAACGTAG CAGCACTTCA CAGTCGAGAT TGCAGTGTAC
2941 AACGGCGACA CCAAAAGGTC TGCTGTCTCA GTTAAATCAC CCCTCTGAAT GATCTACTTC
3001 TTGCCTGCTG CGTTGGTCAG AGGAATAATG GTTGTATTCT ACTGAACAGA TAATCGAGGA
3061 AGGACCAGTT ACTGTTGCTC CTCGTGAGAC TGTGAAAGAG CTTGAGCAGG CAGCACGGAG
3121 GCTTGCTAAA GCTGTGGGTT ATGTTGGTGC TGCTACTGTT GAATACCTTT ACAGCATGGA
3181 AACTGGTGAA TATTATTTTC TGGAACTTAA TCCACGGCTA CAGGTCGGCT CCTTTGACAT
3241 TCTTCAGGAA TTAATTTCTG TTGACCACAT GATTTACATT GTCAAATGGT CTCACAGGTT
3301 GAGCATCCTG TCACTGAGTG GATAGCTGAA GTAAATTTGC CTGCGGCTCA AGTTGCTGTT
3361 GGAATGGGTA TACCCCTTTG GCAGATTCCA GGTAATGCTT CTTCATTTAG TTCCTGCTCT
3421 TTGTTAATTG AATGAGCTCT TATACAGACC ATGAGACACA TTCTACTGTT AATTCATAGT
3481 ATCCCCTGAC TTGTTAGTGT TAGAGATACA GAGATGTATC ACAAATTCAT TGTATCTCCT
3541 CAAGGACTGT AAAAATCCTA TAATTAATTT TCTGAAAATT TGTTCTTTTA AGCAGAAAAA
3601 AAATCTCTAA ATTATCTCCC TGTATACAGA GATCAGGCGC TTCTACGGAA TGAACCATGG
3661 AGGAGGCTAT GACCTTTGGA GGAAAACAGC AGCTCTAGCG ACTCCATTTA ACTTTGATGA
3721 AGTAGATTCT AAATGGCCAA AAGGCCACTG CGTAGCTGTT AGAATAACTA GCGAGGATCC
3781 AGATGATGGG TTTAAGCCTA CTGGTGGAAG AGTAAAGGTG CGGTTTCCTG ATGTTAGGTG
3841 TATGAATTGA ACACATTGCT ATATTGCAGC TAGTGAAATG ACTGGATCAT GGTTCTCTTA
3901 TTTTCAGGAG ATAAGTTTCA AGAGTAAACC AAATGTTTGG GCCTATTTCT CAGTAAAGGT
3961 AGTCCTCAAT ATTGTTGCAC TGCCACATTA TTTGAGTTGT CCTAACAAAT GTGCTGCAAT
4021 TGTTAGTTTT CAACTATTTG TTGTTCTGTT TGGTTGACTG GTACCTCTC TTTGCAGTCT
4081 GGTGGAGGCA TCCATGAATT CGCTGATTCT CAGTTCGGTA TGTAAAGTTA AAAGAGTAAT
4141 ATTGTCTTTG CTATTTATGT TTGTCCTCAC TTTTAAAAGA TATTGCCTTC CATTACAGGA
4201 CATGTTTTTG CGTATGGAAC TACTAGATCG GCAGCAATAA CTACCATGGC TCTTGCACTA
4261 AAAGAGGTTT AAATTCGTGG AGAAATTCAT TCAAACGTAG ACTACACAGT TGACCTATTA
4321 AATGTAAGGA CTAATATCT GCTTATTGAA CCTTGCTTTT TGGTTCCCTA ATGCCATTTT
4381 AGTCTGGCTA CTGAAGAACT TATCCATCAT GCCATTTCTG TTATCTTAAA TTCAGGCCTC
4441 AGATTTTAGA GAAAATAAGA TTCATACTGG TTGGCTGGAT ACCAGGATAG CCATGCGTGT
4501 TCAAGCTGAG AGGCCTCCAT GGTATATTTT AGTCGTTGGA GGGGCTTTAT ATGTAAGACA
4561 AACTATGCCA CTCATTAGCA TTTATGTGAA GCAAATGCGG AAAACATGAT CAATATGTCG
4621 TCTTATTTAA ATTTATTTAT TTTTGTGCTG CAGAAAACAG TAACTGCCAA CACGGCCACT
4681 GTTTCTGATT ATGTTGGTTA TCTTACCAAG GGCCAGATTC CACCAAAGGT ACTATTCTGT
4741 TTTTTCAGGA TATGAATGCT GTTTGAATGT GAAAACCATT GACCATAAAT CCTTGTTTGC
4801 AGCATATATC CCTTGTCTAT ACGACTGTTG CTTTGAATAT AGATGGGAAA AAATATACAG
4861 TAAGTGTGAC ATCTTAATG GGGAACTTA ATTTGTTGTA AATAATCAAT ATCATATTGA

```

FIG.26E-2

```

4921 CTCGTGTATG CTGCATCATA GATCGATACT GTGAGGAGTG GACATGGTAG CTACAGATTG
4981 CGAATGAATG GATCAACGGT TGACGCAAAT GTACAAATAT TATGTGATGG TGGGCTTTTA
5041 ATGCAGGTAA TATCTTCTTC CTAGTTAAAG AAGATATATC TTGTTCAAAG AATTCTGATT
5101 ATTGATCTTT TAATGTTTTT AGCTGGATGG AAACAGCCAT GTAATTTATG CTGAAGAAGA
5161 GGCCAGTGGT ACACGACTTC TTATTGATGG AAAGACATGC ATGTTACAGG TAATGATAGC
5221 CTTGTTCTTT TTAGTTCTAG TCACGGTGTT TGCTTGCTAT TTGTTGTATC TATTTAATGC
5281 ATTCACTAAT TACTATATTA GTTTGCATCA TCAAGTTAAA ATGGAAC TTC TTTCTTGCAG
5341 AATGACCATG ACCCATCAAA GTTATTAGCT GAGACACCAT GCAAAC TTCT TCGTTTCTTG
5401 GTTGCTGATG GTGCTCATGT TGATGCTGAT GTACCATATG CGGAAGTTGA GGTTATGAAG
5461 ATGTGCATGC CCCTCTTATC ACCCGCTTCT GGTGTCATAC ATGTTGTAAT GTCTGAGGGC
5521 CAAGCAATGC AGGTACATTC CTACATTCCA TTCATTGTGC TGTGCTGACA TGAACATTTT
5581 AAGTAAATAC CTGTAAC TTG TTTATTATTC TAGGCTGGTG ATCTTATAGC TAGGCTGGAT
5641 CTTGATGACC CTTCTGCTGT TAAGAGAGCT GAGCCGTTTC AAGATACTTT TCCACAAATG
5701 GGTCTCCCTA TTGCTGCTTC TGGCCAGTT CACAAATTAT GTGCTGCAAG TCTGAATGCT
5761 TGTGCAATGA TCCTTGCGGG GTATGAGCAT GATATTGACA AGGTAAACAT CATGTCCTCT
5821 TGTTTTTTCT TTTGTTTATC ATGCATTCTT ATGTTTCATCA TGTCTCTGG CAAATCTAGA
5881 TTCCGCTGTC GTTTCACACA GATTTTCTC ATTCTCATAA TGGTGCCAAA CATAAATATG
5941 CTGCTATATT CATCAATGTT TTCACTCGAT TTCTAATTTT GCTTTTGAGT TTTAAACTTT
6001 AGTACAATCC ATATCTAATC TCCTTTGGCA ACAGTGAATC CATTATATAT ATTTTATTA
6061 AACTGCTTTC TTTTTCAGGT TGTGCCAGAG TTGGTATACT GCCTAGACAC TCCGGAGCTT
6121 CCTTTCCTGC AGTGGGAGGA GCTTATGTCT GTTTTAGCAA CTAGACTTCC AAGAAATCTT
6181 AAAAGTGAGG TATATTATGG TTGACAAGAT AGCTAGTCTC ATGCTCTAAG GACTTGTACA
6241 TTTGCGCACA TAGGTTAATT TTCCATATCA AGTTCTAATG TACGATATAA AAGTAGTACT
6301 GGCCTAAAAC AGTATTGGTG GTTGACTATC TTTGTTGTGT AAGATCAAGT ATTTCTTTTT
6361 CATGCTTAGT TTGTCAATAC TTCACATTTA TCACTGACTT GTCGAGCTAA ATGAGATTTT
6421 ATTTGATTTT TGTGCTCCAT TATTTTGTG TATATATATA TATATTTAAC TATGACTATA
6481 TGTTATGCCT CAAACGTTTC AAACCTTTTC AGTTGGAGGG CAAATATGAG GAATACAAAG
6541 TAAATTTGA CTCTGGGATA ATCAATGATT TCCCTGCCAA TATGCTACGA GTGATAATTG
6601 AGGTCAGTTA TTCAATTTGT TGTGATAATC ACTGCCTTAA CTGTTGCTTC TTTTAACAAG
6661 CGGTTTTATA GGAAAATCTT GCATGTGGTT CTGAGAAGGA GAAGGCTACA AATGAGAGGC
6721 TTGTTGAGCC TCTTATGAGC CTA CTGAAAGT CATATGAGGG TGGGAGAGAA AGTCATGCTC
6781 ACTTTGTTGT CAAGTCCCTT TTTGAGGAGT ATCTCTATGT TGAAGAATTG TTCAGTGATG
6841 GAATTCAGGT TAACTTACCT ATTCGCATTA AACAAATCAT CAGTTGTTTT ATGATAAAGT
6901 CAAAATGTTT ATATTTCCCA TTCTTCTGTG GATCAAATAT ATCACGGACA TGATATAGTT
6961 TCCTTAGGCT ATATAATGGT TCTTCATCAA ATAATATTGC AGGAAACAGT ATAGCAAAC
7021 ATTTGTATAT ACTCGAGATG GAAATTGTTA GAAACATCAT TGA CTAATC TGTCTTTTGT
7081 TACGCTGTTT TTGTAGTCTG ATGTGATTGA GCGTCTGCGC CTTCAACATA GTAAAGACCT
7141 ACAGAAGGTC GTAGACATTG TGTTGTCCCA CCAGGTAAAT TTCTTCATGG TCTGATGACT
7201 TCACTGCGAA TGGTTACTGA ACTGTCTTCT TGTTCTGACA ATGTGACTTT TCTTTGTAGA
7261 GTGTTAGAAA TAAACTAAG CTGATACTAA AACTCATGGA GAGTCTGGTC TATCCAAATC
7321 CTGCTGCCTA CAGGGATCAA TTGATTGCT TTTCTTCCCT TAATCACAAA GCGTATTACA
7381 AGGTGACCAG GATAAACATA AATAACGTG AATTTTTCAT TGACCTTTTC TTCTGACATC

```

FIG.26E-3

```

7441 TGAATCTGAT GAATTTCTTG CATATTAATA CAGTTGGCAC TTAAAGCTAG TGAACCTTCTT
7501 GAACAAACAA AACTTAGTGA GCTCCGTGCA AGAATAGCAA GGAGCCTTTC AGAGCTGGAG
7561 ATGTTTACTG AGGAAAGCAA GGGTCTCTCC ATGCATAAGC GAGAAATTGC CATTAAAGGAG
7621 AGCATGGAAG ATTTAGTCAC TGCTCCACTG CCAGTTGAAG ATGCGCTCAT TTCTTTATTT
7681 GATTGTAGTG ATACAACGTG TCAACAGAGA GTGATTGAGA CTTATATAGC TCGATTATAC
7741 CAGGTATGAG AAGAAAGACC TTTTGAAATT ATTTATATTA ACATATCCTA GTAAAACAGC
7801 ATGCTCATCA TTTCTTAAAA AAAGTTTACA GCACCTGATG TTTGGTTACT GACCGCATCA
7861 TTAAATAAAA GTTACTTGTT GTGGAGAGAT GTATTTTGA ACTTGTGGCA CATGCAGTAA
7921 CATGCTACTG CTCGATATGT TTGCTAACTT GACAACAATA TTTTTCAGCC TCATCTTGTA
7981 AAGGACAGTA TCAAAATGAA ATGGATAGAA TCGGGTGTTA TTGCTTTATG GGAATTTCTT
8041 GAAGGGCATT TTGATGCAAG AAATGGAGGA GCGGTTCTTG GTGACAAAAG ATGGGGTGCC
8101 ATGGTCATTG TCAAGTCTCT TGAATCACTT TCAATGGCCA TTAGATTTGC ACTAAAGGAG
8161 ACATCACACT ACACTAGCTC TGAGGGCAAT ATGATGCATA TTGCTTTGTT GGGTGCTGAT
8221 AATAAGATGC ATATAATTCA AGAAAGGTAT GTTCATATGC TATGTTGGTG CTGAAATAGT
8281 TATATATGTA GTTAGCTGGT GGAGTTCTGG TAATTAACCT ATCCCATTTG TCAGTGGTGA
8341 TGATGCTGAC AGAATAGCCA AACTTCCCTT GATACTAAAG GATAATGTAA CCGATCTGCA
8401 TGCCTCTGGT GTGAAAACAA TAAGTTTCAT TGTTCAAAGA GATGAAGCAC GGATGACAAT
8461 GCGTCGTACC TTCCTTTGGT CTGATGAAAA GCTTTCTTAT GAGGAAGAGC CAATTCTCCG
8521 GCATGTGGAA CCTCCTCTTT CTGCACTTCT TGAGTTGGTA CGTGATATCA TCAAAATGAT
8581 AATGTTTTGG TATGGCATTG ATTATCTTCT ATGCTCTTTG TATTTATTCA GCCTATTGTG
8641 GATACAGGAC AAGTTGAAAG TGAAAGGATA CAATGAAATG AAGTATACCC CATCACGGGA
8701 TCGTCAATGG CATATCTACA CACTTAGAAA TACTGAAAAC CCCAAAATGT TGCACCGGGT
8761 ATTTTTCCGA ACCCTTGTC AACCACCCAG TGTATCCAAC AAGTTTTCTT CGGGCCAGAT
8821 TGGTGACATG GAAGTTGGGA GTGCTGAAGA ACCTCTGTCA TTTACATCAA CCAGCATATT
8881 AAGATCTTTG ATGACTGCTA TAGAGGAATT GGAGCTTCAC GCAATTAGAA CTGGCCATTC
8941 ACACATGTAT TTGCATGTAT TGAAAGAACA AAAGCTTCTT GATCTTGTTT CAGTTTCAGG
9001 GTAAGTGCGC ATATTTCTTT TTGGGAACAT ATGCTTGCTT ATGAGGTTGG TCTTCTCAAT
9061 GATCTTCTTA TCTTACTCAG GAATACAGTT TTGGATGTTG GTCAAGATGA AGCTACTGCA
9121 TATTCACCTT TAAAAGAAAT GGCTATGAAG ATACATGAAC TTGTTGGTGC AAGAATGCAC
9181 CATCTTTCTG TATGCCAATG GGAAGTGAAA CTTAAGTTGG ACTGCGATGG TCCTGCCAGT
9241 GGTACCTGGA GGATTGTAAC AACCATGTT ACTAGTCACA CTTGCACTGT GGATGTAAGT
9301 TTAATCCTCT AGCATTTTGT TTTCTTTGGA AAAGCATGTG ATTTTAAGCC GGCTGGTCTT
9361 CATACCCAGA CCTAGTGATC TTTATATAGT GTAGACATTT TTCTAACTGC TTTTAATTGT
9421 TTTAGATCTA CCGTGAGATG GAAGATAAAG AATCACGGAA GTTAGTATAC CATCCCGCCA
9481 CTCCGGCGGC TGGTCCTCTG CATGGTGTGG CACTGAATAA TCCATATCAG CCTTTGAGTG
9541 TCATTGATCT CAAACGCTGT TCTGCTAGGA ATAATAGAAC TACATACTGC TATGATTTTC
9601 CACTGGTGAG TTGACTGCTC CTTATATTC AATGCATTAC CATAGCAAAT TCATATTCGT
9661 TCATGTTGTC AAAATAAGCC GATGAAAATT CAAAACGTGA GGCATTTGAA ACTGCAGTGA
9721 GGAAGTCATG GTCCTCTAGT ACCTCTGGTG CTTCTAAAGG TGTTGAAAAT GCCCAATGTT
9781 ATGTTAAAGC TACAGAGTTG GTATTTGCGG ACAAACATGG GTCATGGGGC ACTCCTTTAG
9841 TTCAAATGGA CCGGCCTGCT GGGCTCAATG ACATTGGTAT GGTAGCTTGG ACCTTGAAGA
9901 TGTCCACTCC TGAATTTCTT AGTGGTAGGG AGATTATTGT TGTTGCAAAT GATATTACGT

```

FIG.26E-4

```

9961 TCAGAGCTGG ATCATTGGC CCAAGGGAAG ATGCATTTT TGAAGCTGTT ACCAACCTAG
10021 CCTGTGAGAA GAAACTTCCT CTTATTTATT TGGCAGCAAA TTCTGGTGCT CGAATTGGCA
10081 TAGCAGATGA AGTGAAATCT TGCTCCGTG TTGGGTGGTC TGATGATGGC AGCCCTGAAC
10141 GTGGGTTTCA GTACATTTAT CTAAGCGAAG AAGACTATGC TCGTATTGGC ACTTCTGTCA
10201 TAGCACATAA GATGCAGCTA GACAGTGGTG AAATTAGGTG GGTTATTGAT TCTGTTGTGG
10261 GCAAGGAAGA TGGACTTGGT GTGGAGAATA TACATGGAAG TGCTGCTATT GCCAGTGCTT
10321 ATTCTAGGGC ATATAAGGAG ACATTTACAC TTACATTTGT GACTGGAAGA ACTGTTGGAA
10381 TAGGAGCTTA TCTTGCTCGA CTTGGCATCC GGTGCATACA GCGTCTTGAC CAGCCTATTA
10441 TTCTTACAGG CTATTCTGCA CTGAACAAGC TTCTTGGGCG GGAAGTGTAC AGCTCCACA
10501 TGCAGTTGGG TGGTCCCAA ATCATGGCAA CTAATGGTGT TGTCCATCTT ACTGTTTCAG
10561 ATGACCTTGA AGGCGTTTCT AATATATTGA GGTGGCTCAG TTATGTTCTT GCCTACATTG
10621 GTGGACCACT TCCAGTAACA ACACCGTTGG ACCCACCAGA CAGACCTGTT GCATACATTC
10681 CTGAGAACTC GTGTGATCCT CGAGCGGCTA TCCGTGGTGT TGATGACAGC CAAGGGAAT
10741 GGTTAGGTGG TATGTTTGAT AAAGACAGCT TTGTGGAAC ATTTGAAGGT TGGGCTAAGA
10801 CAGTGGTTAC TGGCAGAGCA AAGCTTGGTG GAATTCCAGT GGGTGTGATA GCTGTGGAGA
10861 CTCAGACCAT GATGCAAAT ATCCCTGCTG ACCCTGGTCA GCTTGATTCC CGTGAGCAAT
10921 CTGTTCTCG TGCTGGACAA GTGTGGTTTC CAGATTCTGC AACCAAGACT GCGCAGGCAT
10981 TGCTGGACTT CAACCGTGA GGATTACCTC TGTTATCCT CGCTAACTGG AGAGGCTTCT
11041 CTGGTGGACA AAGAGATCTT TTTGAAGGAA TTCTTCAGGC TGGCTCGACT ATTGTTGAGA
11101 ACCTTAGGAC ATACAATCAG CCTGCCTTTG TCTACATTCC CATGGCTGCA GAGCTACGAG
11161 GAGGGGCTTG GGTGTGGTT GATAGCAAGA TAAACCAGA CCGCATTGAG TGCTATGCTG
11221 AGAGGACTGC AAAAGGCAAT GTTCTGGAAC CGCAAGGTT AATTGAGATC AAGTTCAGGT
11281 CAGAGGAACT CCAGGATTGC ATGAGTCGGC TTGACCCAAC ATTAATTGAT CTGAAAGCAA
11341 AACTCGAAGT AGCAAATAAA AATGGAAGTG CTGACACAAA ATCGCTTCAA GAAAATATAG
11401 AAGCTCGAAC AAAACAGTTG ATGCCTCTAT AACTCAGAT TGCATACGG TTTGCTGAAT
11461 TGCATGATAC ATCCCTCAGA ATGGCTGCGA AAGGTGTGAT TAAGAAAGTT GTGGACTGGG
11521 AAGAATCAG ATCTTTCTC TATAAGAGAT TACGGAGGAG GATCTCTGAG GATGTTCTTG
11581 CAAAAGAAAT TAGAGCTGTA GCAGGTGAGC AGTTTTCCA CCAACCAGCA ATCGAGCTGA
11641 TCAAGAAATG GTATTAGCT TCACATGCAG CTGAATGGGA TGATGACGAT GCTTTTGTG
11701 CTTGGATGGA TAACCCTGAA AACTACAAGG ATTATATTCA ATATCTTAAG GCTCAAAGAG
11761 TATCCCAATC CCTCTCAAGT CTTTCAGATT CCAGCTCAGA TTTGCAAGCC CTGCCACAGG
11821 GTCTTTCCAT GTTACTAGAT AAGGTAATTA GCTTACTGAT GCTTATATAA ATTCTTTTTT
11881 ATTACATATG GCTGGAGAAC TATCTAATCA AATAATGATT ATAATTCAA TCGTTCTTTT
11941 TATGCCATTA TGATCTTCTG AAATTTCTT CTTTGGACAC TTATTAGAT GGATCCCTCT
12001 AGAAGAGCTC AACTTGTTGA AGAAATCAGG AAGGTCCTG GTTGAATCAT ATGATGCCAA
12061 AACTATTATT GGAGGCACAA ATAGCTTGTG GACCCTGTCG GATTGTTGGT GAGTGTATAT
12121 TGGATTTGTT AGTTCTGCCA GATGAAAGTG CAAGTCTGAT GATTCATGAT ACCGTCAGTT
12181 GGCAAGAACA CCGGTTAACC TGAGTGCTTG TTTACAAATG GTCCTTTATG ACAATCGTTG
12241 TTTCGCGCTA GTTCCGTGAT CTAATATCAT CTGTTAGACG CTGTAATTAG TGAGTCTCCG
12301 CGGATCCACA GTATACGGT GAGCTGTTGA TTCAATTTT GACACGAATA ATATGATTTT
12361 GTAGGCATAA ATGCGTCTGT ATGTGAAATA AATTGTCTGT TGAGTTAACA CACAAGATGA
12421 CAATATGTTT GTGCTCTACT GCTATTGTCC ATGAATACTG ATTGCGGAAT CAACCACATG
12481 CATTAT

```

FIG.26E-5

EAZ33685 - Proteína traducida de arroz japonico *Oriza sativa* plastidal
ACCasa Os05G22940

```

1  MTSTHVATLG VQAQAPPRHQ KKSAGTAFVS SGSSRPSYRK NGQRTSLRE
51  ESNGGVSDSK KLNHSIROGL AGIIDLPNDA ASEVDISHGS EDPRGPTVPG
101 SYQNGIINE THNGRHASVS KVVEFCTALG GKTPINSVLV ANNGMAAAKF
151 MRSVRTWAND TFGSEKAIOL IAMATPEDLR INAEHRTAD QFVEVPGGTN
201 NNNYANVOLI VETAEARTGS AVWPGMGHAS ENPELPAALT AKGIVFLGPP
251 ASSMHALGDK VGSALIAQAA GVPTLAWSSS HVEVPLECCL DSIPDEMYRK
301 ACVTTTEEAV ASCQVVGYP NIKASWGGGG KGIRKVHNDQ EVRTLKQVQ
351 GEVPGSPIFI MRLAASRHL EVQLLCDQYG NVAALHSRDC SVQRHKKII
401 EEGPVTVAPR ETVKELEQAA RRLAKAVGYV GAATVEYLYS METGEYFFLE
451 LNPRLQVEHP VTEWIAEVNL PAAQVAVGMG IPLWQIPETR RPYGMHGGG
501 YDLWRKTAAL ATPNFDEVD SKWPKGHCHA VRITSEDPDD GPKPTGGKVK
551 EISFKSKPNV WAFVSVKSGG GIHEFADSQF GNVFAYGTTR SAAITTMALA
601 LKEVOIRGEI HSNVDYTVOL LNASDFRENK IHTQWLDTRI AMRVQAEERP
651 WYISVGGAL YKTVTANTAT VSDYVGYLTK GQIPPKHISL VYTTVALNID
701 GKKYTIOTVR SGHGSYRLRM NGSTVDANVQ ILCGGGLLMQ LGGNSHVIYA
751 EEEASGTRLL TDGKTCMLQN DHOPSKLLAE TPCKLLRFLV ADGAHVDAOV
801 PYAEVEVMKM CMPLLSPASG VINVMSEGG AMOAGDLIAR LDLDPSAYK
851 RAEPFEDTFP QKGLPIAASG QVHKLCAASL NACRMILAGY EHDIDKVPE
901 LVYCLDTPEL PFLQWEELMS VLATRLPRNL KSELEGKYEE YKVKFDSGII
951 NDFPANMLRV ITEENLAGS EKEKATNERL VEPLMSLLKS YEGGRESHAH
1001 FVVKSLFEFY LYVEELFSDG IQSDVIERLR LQMSKDLQKV VOIVLSHOSV
1051 RNKTKLILKL NESLVYPNPA AYRDQLRFS SLNHKAYYKL ALKASELLEQ
1101 TKLSELRARI ARSLSELMF TEESKGLSMH KRETAIKESH EDLVTAPLPV
1151 EDALISLFDC SDTTVQQRVI ETYIARLYQP HLVKDSIKMK WIESGVIALW
1201 EFPEGHFDAR NGGAVLGDKR WGANVIVKSL ESLMAIRFA LKETSHYTSS
1251 EGNMHIALL GADNMHIIQ ESGDDADRIA KLPLILKQNV TDLHASGVKT
1301 ISFLVORDEA RMTMRRTFLW SDEKLSYEEE PILRHVEPPL SALLELDKIK
1351 VKGYNEMKYT PSRDQRWHLY TLRNTENPKM LHRVFFRTL V RQPSVSNKFS
1401 SGOIGOMEVG SAEELPSFTS TSILRSLMTA IEELELHAIR TGHSHMYLHV
1451 LKEQKLLDLV PVSQNTVLDV GODEATAYSL LKEMANKIHE LVGARMHILS
1501 VCGWEVKLKL DCDGPASGTW RIVTTNVTSH TCTVDIYREM EDKESRKL VY
1551 HPATPAAGPL HGVALNPNYQ PLSVIDLKRC SARNNRTTYC YDFPLAFETA
1601 VRKSWSSSTS GASKGVENAO CYVKATELVQ MORLAGLNDI GNVAVTLKMS
1651 TPEFLSGREI IVVANDITFR AGSFGPREDA FFEAVTNLAC EKKLPLIYLA
1701 ANSGARIGIA DEVKSCFRVG WSDGSPERG FOYTYLSEED YARIGTSVIA
1751 HKMQLDSGEI RWVIDSVVGK EDGLQVENIH GSAAIASAYS RAYKETFTLT
1801 FVTGRTVGIG AYLARLGIRC IQRLDQPIIL TGYSALNKL GREVYSSHMQ
1851 LGGPKIMATN GVVHLTVSDO LEGVSNILRW LSYVPAYIGG PLPVTTPLDP
1901 PDRPVAYIPE NSCDPRAAIR GVDDSGKWL GGMFDKDSFV ETFEGWAKTV
1951 VTGRAKLGSI PVGVIAVETQ THMQTIPADP GOLDSREQSV PRAGQVWFPD
2001 SATKTAQALL DFNREGPLPF ILANWRGFSG GORDLFEGIL QAGSTIVENL
2051 RTYNQPAFVY IPMAAELRGG AWVVVDSKIN PORIECYAER TAKGNVLEPO
2101 GLIEIKFRSE ELQDCMSRLD PTLIDLKAKL EVANKNGSAD TKSLQENIEA
2151 RTKQLMPLYT QIAIRFAELH DTSLRMAAKG VIKKVVDWEE SRSFFYKRLR
2201 RRISEDVLAK EIRAVAGEOF SHOPAIELIK KWYSASHAAE WODDAFVAW
2251 MDNPNENKDY IOYLKQQRVS QSLSSLSQSS SOLQALPOGL SMLLDKMDPS
2301 RRAQLVEETR KVLG

```

FIG.26F

B9FK36 – Proteína de Oriza sativa ACCasa

10	20	30	40	50	60
MTSTHVATLG	VGADAPPRHQ	KKSAGTAFVS	SGSSRPSYRK	NGQRTSLRE	ESNGGVSDSK
70	80	90	100	110	120
KLNHSIRQGL	AGIIDLPNDA	ASEVDISHGS	EDPRGPTVPG	SYQNGIINE	THNGRHASVS
130	140	150	160	170	180
KVVEFCTALG	GKTIHSLV	ANNQMAAKF	MRSVRTWAND	TFGSEKAIQL	IAMATPEDLR
190	200	210	220	230	240
INAEHIRIAD	QFVEVPGGTN	NNNYANVQLI	VEIAERTGVS	AVWPGWGHAS	ENPELPDALT
250	260	270	280	290	300
AKGIVFLGPP	ASSMHALGDK	VGSALIAQAA	GVPTLAWSGS	HVEVPLECCL	DSIPDEMYRK
310	320	330	340	350	360
ACVTTTEEAV	ASQVVGYP	MIKASWGGGG	KGIRKVINDD	EVRTLFGQVQ	GEVPGSPIFI
370	380	390	400	410	420
MRLAAQSRHL	EVQLLCDQYG	NVAALHSRDC	SVQRRHOKII	EEGPVTVAPR	ETVKELEQAA
430	440	450	460	470	480
RRLAKAVGYV	GAATVEYLYS	METGEYYFLE	LNPRLOVENP	VTEWIAEVNL	PAAQVAVGMG
490	500	510	520	530	540
IPLWQIPEIR	RFYGMNHGGG	YDLWRKTAAL	ATPFNFDEVD	SKWPKGHCVA	VRITSEDPDD
550	560	570	580	590	600
GFKPTGSKVK	EISFKSKPNV	WAYFSVKSGG	GIHEFADSQF	GHVFAYGTTR	SAAITTMALA
610	620	630	640	650	660
LKEVQIRGEI	HSNVDYTVOL	LNASDFRENK	INTGWLDTRI	AMRVOAERPP	WYISVVGGL
670	680	690	700	710	720
YKTVTANTAT	VSDYVGYLTK	GQIPPKHISL	VYTTVALNID	GKKYTIDTVR	SGHGSYRLRM
730	740	750	760	770	780
NGSTVDANVQ	ILCDGGLIMQ	LDGNSHVIYA	EEEASGTRLL	IDGKTCMLON	DHDPKLLAE

FIG.26G-1

ES 2 974 625 T3

790	800	810	820	830	840
TPCKLLRFLV	ADGAHVDA DV	PYAEVEVMKM	CMPLLSPASG	VIHVVMSEGO	AMQAGDLIAR
850	860	870	880	890	900
LDLDDPSAVK	RAEPFEDTFP	QMGLPIAASG	QVHKLCAASL	NACRMILAGY	EHDIDKVVPE
910	920	930	940	950	960
LVYCLDTPEL	PFLQWEELMS	VLATRLPRNL	KSELEGKYEE	YKVKFDSGII	NDFPANMLRV
970	980	990	1000	1010	1020
IIEENLACGS	EKEKATNERL	VEPLMSLLKS	YEGGRESHAH	FVVKSLFEEY	LYVEELFSDG
1030	1040	1050	1060	1070	1080
IQSDVIERLR	LQHSKDLQKV	VDIVLSHQSV	RNKTCLILKL	MESLVYPNPA	AYRDQLIRFS
1090	1100	1110	1120	1130	1140
SLNHKAYYKL	ALKASELLEQ	TKLSELRARI	ARSLSELEMF	TEESKGLSMH	KREIAIKESM
1150	1160	1170	1180	1190	1200
EDLVTAPLPV	EDALISLFDC	SOTTVQQRVI	ETYIARLYQP	HLVKDSIKMK	WIESGVIALW
1210	1220	1230	1240	1250	1260
EFPEGHFDAR	NGGAVLGDKR	WGAMVIVKSL	ESLSMAIRFA	LKETSHYTSS	EGNMMHIAL
1270	1280	1290	1300	1310	1320
GADNKMHIQ	ESGDDADRIA	KLPLILKDNV	TDLHASGVKT	ISFIVQRDEA	RMTMRRTFLW
1330	1340	1350	1360	1370	1380
SDEKLSYEEE	PILRHVEPPL	SALLELDKLG	VKGYNEMKYT	PSRDROWHIY	TLRNTENPKM
1390	1400	1410	1420	1430	1440
LHRVFFRTL	V RQPSVSNKFS	SGQIGDMEVG	SAEEPLSFTS	TSILRSLMTA	IEEELHAIR
1450	1460	1470	1480	1490	1500
TGSHMYLHV	LKEQKLLDLV	PVSGNTVLDV	GQDEATAYSL	LKEMAMKIHE	LVGARMHLS
1510	1520	1530	1540	1550	1560
VCQWEVKLKL	DCDGPASGTW	RIVTTNVTSH	TCTVDIYREM	EDKESRKL	VY HPATPAAGPL
1570	1580	1590	1600	1610	1620
HGVALNNPYQ	PLSVIDLKRC	SARNNRTTYC	YDFPLAFETA	VRKSWSSSTS	GASKGVENAQ

FIG.26G-2

1630	1640	1650	1660	1670	1680
CYVKATELVF	ADKHGSWGTP	LVQMDRPAGL	NDIGMVAWTL	KMSTPEFPSG	REIIVVANDI
1690	1700	1710	1720	1730	1740
TFRAGSFGPR	EDAFFEAVTN	LACEKKLPLI	YLAANS GARI	GIADDEVKSCF	RVGWSDDGSP
1750	1760	1770	1780	1790	1800
ERGFQYIYLS	EEDYARIGTS	VIAHKMQLDS	GEIRWVIDSV	VGKEDGLGVE	NIHGSAAIAS
1810	1820	1830	1840	1850	1860
AYSRAYKETF	TLTFVTGRTV	GIGAYLARLG	IRCIQRDQ P	IILTGYSALN	KLLGREVYSS
1870	1880	1890	1900	1910	1920
HMQLGGPKIM	ATNGVVHLTV	SDDLEGVSNI	LRWLSYVPAY	IGGPLPV TTP	LDPPDRPVAY
1930	1940	1950	1960	1970	1980
IPENSCDPRA	AIRGVDDSQG	KWLGGMFDKD	SFVETFEGWA	KTVVTGRAKL	GGIPVGVIAV
1990	2000	2010	2020	2030	2040
ETQTMQTIP	ADPGQLDSRE	QSVPRAGQVW	FPDSATKTAQ	ALLDFNREGL	PLFILANWRG
2050	2060	2070	2080	2090	2100
FSGGQRDLFE	GILQAGSTIV	ENLR TYNQPA	FVYIPMAAEL	RGGAWVVVDS	KINPDRIECY
2110	2120	2130	2140	2150	2160
AERTAKGNVL	EPQGLIEIKF	RSEELQDCMS	RLDPTLIDLK	AKLEVANKNG	SADTKSLQEN
2170	2180	2190	2200	2210	2220
IEARTKQLMP	LYTQIAIRFA	ELHDTSLRMA	AKGV IKKVVD	WEESRSFFYK	RLRRRISE DV
2230	2240	2250	2260	2270	2280
LAKEIRAVAG	EQFSHQPAIE	LIKKWYSASH	AAEWDDDDAF	VAWMDNPENY	KDYIQYLKAQ
2290	2300	2310	2320		
RVSQSLSSLS	DSSSDLQALP	QGLSMLLDKM	DPSRRAQLVE	EIRKVLG	

FIG.26G-3