

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 585 157**

51 Int. Cl.:

**C12N 15/82** (2006.01)

**C12N 9/22** (2006.01)

**C12N 15/90** (2006.01)

**A01H 5/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.09.2008 E 08833882 (7)**

97 Fecha y número de publicación de la concesión europea: **18.05.2016 EP 2205749**

54 Título: **Proteínas de dedo de cinc de diseño que se dirigen a los genes de 5-enolpiruvil shikimato-3-fosfato sintasa**

30 Prioridad:

**27.09.2007 US 995557 P**

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

**04.10.2016**

73 Titular/es:

**DOW AGROSCIENCES LLC (50.0%)  
9330 Zionsville Road  
Indianapolis IN 46268-1054, US y  
SANGAMO BIOSCIENCES, INC. (50.0%)**

72 Inventor/es:

**GUPTA, MANJU;  
PALTA, ASHA, M.;  
NOVAK, STEPHEN;  
URNOV, FYODOR y  
GOPALAN, SUNITA**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 585 157 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

Proteínas de dedo de cinc de diseño que se dirigen a los genes de 5-enolpiruvil shikimato-3-fosfato sintasa

### Campo técnico

5 La divulgación se refiere en general a los campos de la ingeniería genómica, la integración cromosómica dirigida y la expresión de proteínas en plantas. En particular, la presente divulgación se refiere a proteínas de dedo de cinc de diseño que se dirigen a genes de 5-enolpiruvil shikimato-3-fosfato sintasa (EPSPS) y a métodos para usar dichas proteínas de dedo de cinc en la modulación de la expresión génica, la inactivación génica y la modificación génica dirigida. Más particularmente, la divulgación se refiere a nucleasas de dedo de cinc de diseño para la escisión y alteración dirigida de genes de EPSPS.

### 10 Antecedentes

Un área principal de interés en la agricultura, especialmente a la luz de la determinación de las secuencias nucleotídicas completas de una serie de genomas de plantas, es la regulación dirigida de la expresión génica y la alteración de las secuencias génicas. En particular, la capacidad para modular la expresión génica o para modificar secuencias endógenas de plantas podrían facilitar numerosas aplicaciones, tales como, por ejemplo, la optimización de características del cultivo que afectan al valor nutricional, el rendimiento, la tolerancia al estrés, la resistencia a patógenos, la cantidad de aceite y la resistencia a agentes agroquímicos y/o a la adaptación de las plantas para su uso como fábricas biológicas para la producción de compuestos farmacéuticos o agentes químicos industriales.

20 Las proteínas de dedo de cinc (ZFP) se han usado ventajosamente para modular de manera selectiva la expresión génica y para la alteración dirigida de secuencias génicas en plantas (véanse, por ejemplo, las Patentes de Estados Unidos n.º 7.262.054, 7.235.354, 7.220.719, 7.001.768 y 6.534.261; y el n.º de Serie de Estados Unidos 60/874.911). Las proteínas de dedo de cinc (ZFP) son proteínas que se unen a ADN, ARN y/o proteínas, de una manera específica de secuencia, por virtud de un dominio estabilizado por metales conocido como dedo de cinc. Véase, por ejemplo, Miller et al. (1985) EMBO J. 4:1609-1614; Rhodes et al. (1993) Sci. Amer. 268(2):56-65; y Klug (1999) J. Mol. Biol. 293:215-218. Las ZFP se encuentran comúnmente en factores de transcripción y hasta la fecha, se han identificado más de 10.000 secuencias de dedos de cinc en varios miles de factores de transcripción conocidos o presupuestos.

30 La regulación y alteración de dianas génicas seleccionadas puede lograrse, en teoría, mediante el diseño de ZFP con una especificidad de secuencia de ADN predeterminada que tienen actividades biológicas deseadas. Los dominios de dedo de cinc se han combinado, por ejemplo, en proteínas de fusión con dominios reguladores para producir factores de transcripción de dedos de cinc modificados para controlar la regulación génica (véase, por ejemplo, la Patente de Estados Unidos n.º 6.534.261). También se han combinado los dominios de dedo de cinc con dominios de escisión de nucleasa para producir nucleasas de dedo de cinc (ZFN) para el direccionamiento específico de un corte bicatenario a la región de un genoma donde se desea la modificación (por ejemplo, eliminación, mutación, recombinación de homólogos o inserción de una secuencia exógena) (véanse, por ejemplo, 35 las Publicaciones de Solicitud de Patente de Estados Unidos n.º 2007/0134796 y 2005/0064474). Las ZFP modificadas facilitan en gran medida la inserción de secuencias exógenas o la modificación de secuencias endógenas en sitios diana específicos en plantas y posibilitan la alteración dirigida de genomas de plantas con mayor eficacia que con los métodos convencionales (véanse, por ejemplo, las Patentes de Estados Unidos n.º 7.262.054, 7.235.354, 7.220.719, 7.001.768 y 6.534.261).

40 Sin embargo, la duplicación genómica es común entre plantas y sigue habiendo una necesidad de composiciones y métodos para la alteración dirigida de dichos genes parálogos en genomas de plantas y para la modulación de la expresión de genes parálogos en plantas.

### Compendio

45 La presente invención se define por las reivindicaciones adjuntas. La presente descripción proporciona composiciones y métodos para modular la expresión y para la alteración dirigida de uno o más genes parálogos (por ejemplo, genes EPSPS) en células vegetales. Las células vegetales pueden ser de especies de plantas monocotiledóneas (monocot) o dicotiledóneas (dicots) y también incluyen células cultivadas, células en una planta en cualquier etapa del desarrollo y células vegetales que se han extraído de una planta completa y cuyas células (o sus descendientes) se devolverán a la planta. Las células vegetales pueden contener una o más secuencias génicas homólogas o parálogas, pudiendo dirigirse a cualquier número de ellas o a todas ellas para la modificación mediante 50 los métodos descritos en esta memoria.

En un aspecto, la presente memoria describe una proteína de dedo de cinc (ZFP) que se une a una región genómica diana de EPSPS de interés, en donde la ZFP comprende uno o más dominios de unión de dedo de cinc modificados, en donde la proteína de dedo de cinc comprende las regiones de hélice de reconocimiento tal como se muestran en una sola fila de la tabla A. En determinadas realizaciones, el gen EPSPS usado como diana por la ZFP comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia que tiene una identidad de secuencia de al menos aproximadamente el 80-100 % con las mismas, incluyendo cualquier 55

- 5 porcentaje de identidad entre estos intervalos, tal como una identidad del 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % con las mismas. En determinadas realizaciones, la ZFP es una proteína de fusión que comprende uno o más dominios reguladores. En una realización, se seleccionan uno o más dominios reguladores del grupo que consiste en un represor transcripcional, una endonucleasa, una metil transferasa, una histona desacetilasa, un activador transcripcional y una histona acetiltransferasa.
- 10 En una realización, la ZFP es una nucleasa de dedo de cinc (ZFN) que escinde una región genómica diana de EPSPS de interés, en donde la ZFN comprende uno o más dominios de unión de dedos de cinc modificados y un dominio de escisión de nucleasa. En determinadas realizaciones, la ZFN comprende un polipéptido de fusión que comprende un dominio de unión de dedo de cinc modificado que tiene especificidad por una secuencia génica de EPSPS y un dominio de escisión y/o uno o más polipéptidos de fusión que comprenden un dominio de unión de dedo de cinc modificado y un semidominio de escisión. En determinadas realizaciones, los dominios de dedos de cinc comprenden una secuencia seleccionada del grupo que consiste en proteínas de dedo de cinc que comprenden los dominios de reconocimiento mostrados en la tabla A. Los dominios de escisión y los semidominios de escisión pueden obtener, por ejemplo, a partir de diversas endonucleasas de restricción y/o endonucleasas *homíng* (endonucleasas de asentamiento). En una realización, los semidominios de escisión proceden de una endonucleasa de restricción de tipo IIS (por ejemplo, Fok I). La ZFN puede escindir específicamente una secuencia génica de EPSPS particular. Como alternativa, la ZFN puede escindir dos o más secuencias génicas de EPSPS homólogas, que pueden incluir secuencias génicas parálogas u ortólogas de EPSPS.
- 15 En determinadas realizaciones, el gen de EPSPS diana de ZFN comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia que tiene una identidad de secuencia de al menos el 80-100 % con las mismas, incluyendo cualquier porcentaje de identidad dentro de estos intervalos, tal como una identidad de secuencia del 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % con las mismas.
- 20 La ZFN puede unirse a y/o escindir un gen EPSPS dentro de la región codificante del gen o en una secuencia no codificante dentro de o adyacente al gen, tal como, por ejemplo, una secuencia líder, una secuencia no traducida o intrón o dentro de una región no transcrita, ya sea cadena arriba o cadena abajo de la región codificante. En determinadas realizaciones, la ZFN se une a y/o escinde una secuencia codificante o una secuencia reguladora del gen EPSPS. En determinadas realizaciones, la ZFN se une a y escinde un gen EPSPS dentro de una región que consiste en una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14.
- 25 En otro aspecto, se describen en la presente memoria composiciones que comprenden una o más ZFP, que pueden incluir una o más ZFN. Las células vegetales pueden contener un único gen EPSPS o múltiples genes EPSPS parálogos. Por lo tanto, las composiciones pueden comprender una o más ZFP que se dirigen a uno o más genes EPSPS en una célula vegetal, por ejemplo, a 1, 2, 3, 4, 5 o hasta cualquier número de parálogos de EPSPS o a todos los parálogos de EPSPS presentes en una célula vegetal. En una realización, la composición comprende una o más ZFP que se dirigen a todos los genes parálogos de EPSPS en una célula vegetal. En otra realización, la composición comprende una ZFP que se dirige específicamente a un gen parálogo de EPSPS en una célula vegetal. Por ejemplo, la composición puede comprender una ZFN que se une específicamente a y escinde un gen parálogo de EPSPS particular en una célula vegetal o múltiples ZFN que se unen a y escinden dos o más genes parálogos de EPSPS en una célula vegetal. Además, las composiciones pueden contener ZFP no nucleasas que alteran la regulación transcripcional de uno o más genes parálogos de EPSPS.
- 30 En otro aspecto, se describe en la presente memoria un polinucleótido que codifica una o más ZFP descritas en la presente memoria. En una realización, el polinucleótido codifica al menos una ZFN. Los polinucleótidos ejemplares comprenden una secuencia nucleotídica que codifica cualquiera de las proteínas de dedo de cinc como se muestran en la tabla A.
- 35 En otro aspecto, se describe en la presente memoria un vector de expresión de ZFP que comprende un polinucleótido que codifica una o más ZFP descritas en la presente memoria, unidas operativamente a un promotor. En una realización, una o más de las ZFP son ZFN.
- 40 En otro aspecto, se describe en la presente memoria una célula hospedadora vegetal que comprende uno o más vectores de expresión de ZFP. La célula hospedadora vegetal puede transformarse de manera estable o transfectarse de manera transitoria o una combinación de ambos con uno o más vectores de expresión de ZFP. En una realización, los uno o más vectores de expresión de ZFP expresan una o más ZFN en la célula hospedadora vegetal.
- 45 En otro aspecto, se describe en la presente memoria un método para escindir uno o más genes parálogos en una célula vegetal, comprendiendo el método: (a) introducir, en la célula vegetal, uno o más vectores de expresión que codifican una o más ZFN como se definen anteriormente que se unen a un sitio diana en los uno o más genes parálogos en condiciones tales que la ZFN (o las ZFN) se expresa (o expresan) y los uno o más genes parálogos y en donde el sitio diana es un gen EPSPS. En una realización, se escinde un gen parálogo de EPSPS particular en una célula vegetal. En otra realización, se escinde más de un parálogo de EPSPS, por ejemplo, se escinden 2, 3, 4, 5 o hasta cualquier número de parálogos de EPSPS o todos los parálogos de EPSPS presentes en una célula

vegetal.

En otro aspecto, se describe en la presente memoria un vector donante que comprende las secuencias primera y segunda de ADN, en donde (i) la primera secuencia es homóloga de una tercera secuencia y la segunda secuencia es homóloga de una cuarta secuencia; y (ii) las secuencias tercera y cuarta son secuencias de ADN cromosómico de EPSPS. En determinadas realizaciones, los extremos próximos de las secuencias tercera y cuarta están separados por al menos 1 par de nucleótidos. En determinadas realizaciones, las secuencias de ADN de EPSPS pertenecen a un gen de EPSPS que comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia que tiene una identidad de secuencia del 80-100 % con las mismas, incluyendo cualquier porcentaje de identidad entre estos intervalos, tal como una identidad de secuencia del 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % con las mismas.

En determinadas realizaciones, al menos una de las secuencias primera y segunda en el vector donante tiene una longitud de 100 nucleótidos o menos. Además, cualquiera de los vectores descritos en la presente memoria puede comprender además una quinta secuencia, en donde la quinta secuencia: (a) está interpuesta entre las secuencias primera y segunda; y (b) es una secuencia exógena. En determinadas realizaciones, la quinta secuencia tiene un tamaño de al menos 1 par de bases, pero puede ser tan grande como de 22 pares de kilobases o más.

Los vectores donantes (por ejemplo, la quinta secuencia) también pueden comprender secuencias que codifican una proteína o partes de una proteína. En determinadas realizaciones, la secuencia codificante de proteína codifica un marcador de selección (por ejemplo, proteína fluorescente verde (GFP),  $\beta$ -glucuronidasa (GUS), fosfinotricina N-acetil transferasa (PAT, BAR), neomicina fosfotransferasa, higromicina fosfotransferasa,  $\beta$ -lactamasa, catecol dioxigenasa,  $\alpha$ -amilasa, tirosinasa,  $\beta$ -galactosidasa, luciferasa, aecuatorina, EPSP sintasa, nitrilasa, acetolactato sintasa (ALS), dihidrofolato reductasa (DHFR), dalapon deshalogenasa y antranilato sintasa). En otras realizaciones, la secuencia codificante de proteína (por ejemplo, la quinta secuencia) codifica una proteína o una parte de proteína, por ejemplo, una secuencia que es homóloga de secuencias cromosómicas.

En otras realizaciones más, los vectores donantes (por ejemplo, la quinta secuencia) comprenden una o más secuencias reguladoras transcripcionales. Por ejemplo, un vector donante puede comprender una o más secuencias reguladoras transcripcionales que aumentan o reducen la expresión del gen parálogo (por ejemplo, EPSPS). En determinadas realizaciones, el vector donante comprende una o más secuencias de direccionamiento a proteínas que potencian o reducen el transporte de proteínas.

En otras realizaciones adicionales, los vectores donantes (por ejemplo, la quinta secuencia) pueden comprender un homólogo de tipo silvestre de una secuencia cromosómica mutante (por ejemplo, EPSPS) o, como alternativa, un homólogo mutante de una secuencia cromosómica de tipo silvestre (por ejemplo, EPSPS). En determinadas realizaciones, la secuencia cromosómica mutante comprende una o más mutaciones seleccionadas del grupo que consiste en una mutación puntual, una sustitución, una eliminación y una inserción. En una realización, el vector donante comprende una secuencia cromosómica de EPSPS mutante que aumenta la tolerancia de una planta al herbicida glifosato.

En cualquiera de los vectores donantes descritos en la presente memoria, la primera secuencia puede tener una homología de al menos el 35 % con la tercera secuencia. De manera similar, en cualquiera de los vectores descritos en la presente memoria, la segunda secuencia puede tener una homología de al menos el 35 % con la cuarta secuencia. En algunas realizaciones, la primera secuencia tiene una homología de al menos el 35 % al 50 %, al menos 50 % al 70 %, al menos el 70 % al 80 %, al menos 80 % al 85 %, al menos el 85 % al 90 %, al menos el 90 % al 95 %, al menos el 95 %, 96 %, 97 %, 98 %, 99 % o 100 % con la tercera secuencia. En algunas realizaciones, la segunda secuencia tiene una homología de al menos el 35 % al 50 %, al menos 50 % al 70 %, al menos el 70 % al 80 %, al menos 80 % al 85 %, al menos el 85 % al 90 %, al menos el 90 % al 95 %, al menos el 95 %, 96 %, 97 %, 98 %, 99 % o 100 % con la cuarta secuencia.

En otro aspecto más, se describe en la presente memoria un método para introducir una secuencia exógena en el genoma de una célula vegetal, comprendiendo el método las etapas de: (a) poner en contacto la célula con cualquiera de los vectores donantes descritos en la presente memoria; y (b) expresar una o más nucleasas de dedo de cinc en la célula, en donde las una o más nucleasas de dedo de cinc escinden ADN cromosómico en entre 0,4 y 3 pares de kilobases de una de las secuencias tercera y cuarta; de tal forma que la escisión del ADN cromosómico en la etapa (b) estimula la incorporación del vector donante en el genoma mediante recombinación de homólogos. En determinadas realizaciones, las una o más nucleasas son fusiones entre el dominio de escisión de una endonucleasa de restricción de tipo IIS y un dominio de unión de dedo de cinc modificado.

En determinadas realizaciones, las nucleasas de dedo de cinc escinden un gen EPSPS que comprende una secuencia seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia que tiene una identidad de secuencia de al menos aproximadamente el 80-100 % con las mismas, incluyendo cualquier porcentaje de identidad dentro de estos intervalos, tal como una identidad de secuencia del 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % con las mismas.

En otro aspecto, se describe en la presente memoria un método para expresar el producto de una secuencia de ácido nucleico exógena en una célula vegetal, comprendiendo el método las etapas de: (a) poner en contacto la célula con un vector donante que comprende una secuencia de ácido nucleico exógena; y (b) expresar una nucleasa de dedo de cinc (ZFN) en la célula, en donde la ZFN escinde uno o más genes parálogos (por ejemplo, uno o más genes de EPSPS) en ADN cromosómico dentro de 3 pares de kilobases de una de las secuencias tercera o cuarta. La escisión del ADN cromosómico en la etapa (b) da como resultado la incorporación del vector donante en el genoma mediante recombinación de homólogos y la expresión del producto de la secuencia de ácido nucleico exógena.

En determinadas realizaciones, la nucleasa de dedo de cinc escinde un gen EPSPS que comprende una secuencia seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia que tiene una identidad de secuencia de al menos aproximadamente el 80-100 % con las mismas, incluyendo cualquier porcentaje de identidad dentro de estos intervalos, tal como una identidad de secuencia del 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % con las mismas.

En otro aspecto, se describe en la presente memoria un método para la recombinación de homólogos intramolecular en el genoma de una célula vegetal, comprendiendo el método las etapas de: (a) proporcionar un segmento de ADN que comprende una secuencia de un gen diana y que comprende además una primera secuencia que es homóloga de una segunda secuencia; y (b) poner en contacto dicho segmento de ADN con la ZFN como se describe en la presente memoria, en donde la ZFN escinde el segmento de ADN en una secuencia de gen diana estimulando de este modo la recombinación de homólogos intramolecular. En determinadas realizaciones, el segmento de ADN es endógeno para la célula. En otras realizaciones, el segmento de ADN es exógeno para la célula. En determinadas realizaciones, el gen diana es único para la célula. En otras realizaciones, el gen diana es un gen parálogo. En cualquiera de estos métodos, el gen diana puede comprender un gen EPSPS único o parálogo y la ZFN comprende cualquiera de las secuencias mostradas en la tabla A. En determinadas realizaciones, puede producirse recombinación de homólogos en un cromosoma. En una realización, se elimina del cromosoma el ADN entre las secuencias primera y segunda. En una realización, las secuencias eliminadas del cromosoma pueden codificar la totalidad o parte del gen diana. En otra realización, las secuencias eliminadas del cromosoma pueden codificar la totalidad o parte de un marcador de selección, por ejemplo, proteína fluorescente verde (GFP),  $\beta$ -glucuronidasa (GUS), fosfinotricina N-acetil transferasa (PAT, BAR), neomicina fosfotransferasa, higromicina fosfotransferasa,  $\beta$ -lactamasa, catecol dioxigenasa,  $\alpha$ -amilasa, tirosinasa,  $\beta$ -galactosidasa, luciferasa, aecuorina, EPSP sintasa, nitrilasa, acetolactato sintasa (ALS), dihidrofolato reductasa (DHFR), dalapon deshalogenasa y antranilato sintasa.

En determinadas realizaciones, se reemplaza el ADN eliminado con una secuencia exógena, comprendiendo el método además introducir un polinucleótido en la célula, en donde el polinucleótido comprende (i) secuencias cuarta y quinta, en donde la cuarta secuencia es homóloga de secuencias no eliminadas en proximidad a la primera secuencia y la quinta secuencia es homóloga de secuencias no eliminadas en proximidad a la segunda secuencia; y (ii) la secuencia exógena.

En determinadas realizaciones, el ADN eliminado se reemplaza con una secuencia génica que puede comprender un homólogo mutante de una secuencia génica de tipo silvestre. En determinadas realizaciones, la secuencia génica mutante comprende una o más mutaciones seleccionadas del grupo que consiste en una mutación puntual, una sustitución, una eliminación y una inserción. En una realización, el ADN eliminado se reemplaza con una secuencia génica de EPSPS, por ejemplo, una secuencia génica de EPSPS que comprende una mutación que aumenta la tolerancia de una planta al herbicida glifosato.

En otra realización, la secuencia exógena puede ser un marcador de selección, por ejemplo, proteína fluorescente verde (GFP),  $\beta$ -glucuronidasa (GUS), fosfinotricina N-acetil transferasa (PAT, BAR), neomicina fosfotransferasa, higromicina fosfotransferasa,  $\beta$ -lactamasa, catecol dioxigenasa,  $\alpha$ -amilasa, tirosinasa,  $\beta$ -galactosidasa, luciferasa, aecuorina, EPSP sintasa, nitrilasa, acetolactato sintasa (ALS), dihidrofolato reductasa (DHFR), dalapon deshalogenasa y antranilato sintasa.

En otra realización, se describe en la presente memoria un método para eliminar una secuencia génica del genoma de una célula vegetal, comprendiendo el método (a) proporcionar una célula vegetal que comprende una secuencia génica; y (b) expresar primeras y segundas nucleasas de dedo de cinc (ZFN, como se han descrito anteriormente) en la célula, en donde la primera ZFN escinde en un primer sitio de escisión y la segunda ZFN escinde en un segundo sitio de escisión, en donde la secuencia génica se localiza entre el primer sitio de escisión y el segundo sitio de escisión, en donde la escisión de los sitios de escisión primero y segundo da como resultado la eliminación de la secuencia génica. En determinadas realizaciones, la secuencia génica es un gen EPSPS. El tamaño de la eliminación en la secuencia génica viene determinado por la distancia entre los sitios de escisión primero y segundo. Por consiguiente, pueden obtenerse eliminaciones de cualquier tamaño, en cualquier región genómica de interés. Pueden obtenerse eliminaciones de 25, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.000 pares de nucleótidos o de cualquier valor de pares de nucleótidos integrado dentro de este intervalo. Además, pueden obtenerse eliminaciones de una secuencia de cualquier valor integral de pares de nucleótidos mayores de 1.000 pares de nucleótidos usando los métodos y composiciones descritas en la presente memoria. En una realización, los sitios de escisión primero y segundo están separados por al menos 100 nucleótidos. En una realización, se elimina el gen EPSPS completo. En otra realización, se elimina una parte del gen EPSPS. En una realización, se elimina la

secuencia génica de EPSPS de una célula vegetal transgénica. La secuencia génica de EPSPS puede ser una secuencia endógena o exógena.

5 En otro aspecto, se describe en la presente memoria un método para modular la regulación de un gen de planta, comprendiendo el método (a) proporcionar una célula vegetal que comprende una secuencia génica diana; y (b) expresar una ZFP como se ha descrito anteriormente en la célula, en donde la ZFP se une a una secuencia reguladora del gen diana, modulando de este modo la regulación del gen diana, en donde la secuencia génica es un gen EPSPS. La unión de la ZFP a la secuencia reguladora puede aumentar o reducir la transcripción del gen diana de EPSPS. En determinadas realizaciones, la ZFP también aumenta o reduce la tolerancia de una planta al herbicida glifosato.

10 En un aspecto adicional más, también se proporciona una célula vegetal transgénica obtenida según uno cualquiera de los métodos descritos en la presente memoria.

En otro aspecto, se proporciona en la presente memoria una planta que comprende una célula vegetal transgénica obtenida como se ha descrito en la presente memoria.

15 En cualquiera de los métodos descritos en la presente memoria, puede usarse la modificación de las secuencias génicas diana de planta (por ejemplo, secuencias reguladoras transcripcionales o secuencias codificantes de EPSPS) para aumentar o reducir la tolerancia de una planta al herbicida glifosato.

Por lo tanto, la presente descripción abarca las siguientes realizaciones numeradas:

1. Una proteína de dedo de cinc (ZFP) que se une a una región genómica diana de EPSPS de interés, comprendiendo dicha ZFP uno o más dominios de unión de dedo de cinc modificados.
- 20 2. La ZFP de la realización 1, en donde la región genómica diana se encuentra en una célula de una planta dicotiledónea.
3. La ZFP de la realización 2, en donde la región genómica diana se encuentra en una célula de una planta de colza.
4. La ZFP de la realización 2, en donde la región genómica diana se encuentra en una célula de una planta de *Brassica*.
- 25 5. La ZFP de la realización 1, en donde la región genómica diana de EPSPS de interés pertenece a un gen EPSPS que comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia idéntica en al menos un 95 % con las mismas.
6. La ZFP de la realización 1, en donde la ZFP es una proteína de fusión que comprende uno o más dominios funcionales.
- 30 7. La ZFP de la realización 6, que comprende uno o más dominios funcionales seleccionados del grupo que consiste en un represor transcripcional, una endonucleasa, una metil transferasa, una histona desacetilasa, un activador transcripcional y una histona acetiltransferasa.
8. La ZFP de cualquiera de las realizaciones 1-7, en donde la ZFP se une a una secuencia reguladora transcripcional del gen EPSPS.
- 35 9. La ZFP de cualquiera de las realizaciones 1-7, en donde la ZFP se une arriba de un sitio de inicio de la transcripción del gen EPSPS.
10. La ZFP de cualquiera de las realizaciones 1-7, en donde la ZFP se une adyacente a un sitio de inicio de la transcripción del gen EPSPS.
11. La ZFP de cualquiera de las realizaciones 1-10, en donde la ZFP aumenta la transcripción del gen EPSPS.
- 40 12. La ZFP de cualquiera de las realizaciones 1-10, en donde la ZFP reduce la transcripción del gen EPSPS.
13. La ZFP de la realización 1, en donde la ZFP es una nucleasa de dedo de cinc (ZFN) que escinde la región genómica diana de EPSPS de interés, comprendiendo dicha ZFN uno o más dominios de unión de dedo de cinc modificados y un dominio de escisión de nucleasa.
14. La ZFN de la realización 13, en donde el dominio de escisión comprende dos semidominios de escisión.
- 45 15. La ZFN de la realización 14, en donde los semidominios de escisión proceden de la misma nucleasa.
16. La ZFN de la realización 15, en donde los semidominios de escisión proceden de una endonucleasa de restricción de tipo IIS.
17. La ZFN de la realización 16, en donde la endonucleasa de restricción de tipo IIS es Fok I.

18. La ZFN de la realización 13, en donde la región genómica diana de EPSPS de interés pertenece a un gen EPSPS que comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia idéntica en al menos un 95 % con las mismas.
19. La ZFN de la realización 13, en donde la ZFN se une a una secuencia en la región codificante de un gen EPSPS.
- 5 20. La ZFN de la realización 13, en donde la ZFN se une a una secuencia en la región no codificante de un gen EPSPS.
21. La ZFN de la realización 20, en donde la ZFN se une a una secuencia reguladora del gen EPSPS.
22. La ZFN de la realización 13, en donde la ZFN escinde una o más secuencias génicas parálogas u ortólogas de EPSPS.
- 10 23. La ZFN de la realización 13, en donde la ZFN escinde específicamente una secuencia génica paróloga u ortóloga de EPSPS.
24. La ZFN de la realización 13 que comprende un dominio de unión de dedo de cinc que comprende una secuencia como la mostrada en la tabla A.
- 15 25. La ZFN de la realización 13, en donde la ZFN se une a y escinde un gen EPSPS dentro de una región que consiste en una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14.
26. La ZFN de la realización 13, en donde la ZFN comprende:
- (a) una primera proteína de fusión que comprende un primer dominio de unión de dedo de cinc y un primer semidominio de escisión, en donde el primer dominio de unión de dedo de cinc se une a una primera secuencia nucleotídica. y
- 20 (b) una segunda proteína de fusión que comprende un segundo dominio de unión de dedo de cinc y un segundo semidominio de escisión, en donde el segundo dominio de unión de dedo de cinc se une a una segunda secuencia nucleotídica.
27. La ZFN de la realización 26, en donde la segunda secuencia nucleotídica se ubica a entre 2 y 50 nucleótidos respecto de la primera secuencia nucleotídica.
- 25 28. La ZFN de la realización 26, en donde la escisión se produce entre las secuencias nucleotídicas primera y segunda.
29. Una composición que comprende una o más proteínas de dedo de cinc (ZFP) según cualquiera de las realizaciones 1-28.
30. La composición de la realización 29, en donde una o más de las ZFP son nucleasas de dedo de cinc (ZFN).
- 30 31. La composición de la realización 29, que comprende una o más ZFP que se dirigen a uno o más genes de EPSPS en una célula vegetal.
32. La composición de la realización 29, que comprende dos o más ZFP que en combinación se dirigen a todos los genes parálogos de EPSPS en una célula vegetal.
- 35 33. La composición de la realización 30, que comprende una ZFN que se une específicamente a y escinde un gen parálogo de EPSPS en una célula vegetal.
34. La composición de la realización 30, que comprende dos o más ZFN que se unen a y escinden dos o más genes parálogos de EPSPS en una célula vegetal.
35. La composición de la realización 30, que comprende una o más ZFN que se unen a y escinden todos los genes parálogos de EPSPS en una célula vegetal.
- 40 36. Un polinucleótido que codifica una o más proteínas de dedo de cinc (ZFP) según cualquiera de las realizaciones 1 a 28.
37. El polinucleótido de la realización 36 que comprende una secuencia nucleotídica que codifica una proteína de dedo de cinc como se muestra en la tabla A.
- 45 38. Un vector de expresión de ZFP que comprende el polinucleótido de cualquiera de las realizaciones 36 o 37 unido operablemente a un promotor.
39. Una célula hospedadora vegetal que comprende uno o más vectores de expresión de ZFP según la realización 38.

40. La célula hospedadora vegetal de la realización 39, en donde la célula se transfecta de manera estable con uno o más vectores de expresión de ZFP.
41. La célula hospedadora vegetal de la realización 39, en donde la célula se transfecta de manera transitoria con uno o más vectores de expresión de ZFP.
- 5 42. Un método para escindir uno o más genes de EPSPS en una célula vegetal, comprendiendo el método:
- (a) transfectar la célula vegetal con uno o más vectores de expresión de ZFP que codifican una o más ZFN según la realización 10; y
  - (b) expresar las una o más ZFN en la célula, en donde las ZFN escinden a uno o más genes EPSPS.
- 10 43. El método de la realización 42, en donde se transfecta de manera estable al menos un vector de expresión de ZFP en la célula vegetal.
44. El método de la realización 42, en donde se transfecta de manera transitoria al menos un vector de expresión de ZFP en la célula vegetal.
45. El método de las realizaciones 42-44, en donde los al menos dos vectores de expresión de ZFP se transfectan en la célula.
- 15 46. El método de la realización 45, en donde los al menos dos vectores de expresión de ZFP se transfectan conjuntamente en la célula.
47. El método de la realización 45, en donde los al menos dos vectores de expresión de ZFP se transfectan secuencialmente en la célula.
- 20 48. El método de cualquiera de las realizaciones 42 a 47, en donde se escinden todos los genes parálogos de EPSPS en la célula vegetal.
49. El método de cualquiera de las realizaciones 42 a 47, en donde se escinde un gen parálogo de EPSPS en la célula vegetal.
50. El método de cualquiera de las realizaciones 42 a 47, en donde se escinden al menos dos genes parálogos de EPSPS en la célula vegetal.
- 25 51. Un vector donante que comprende las secuencias de ADN primera y segunda;
- en donde la primera secuencia es homóloga de una tercera secuencia y la segunda secuencia es homóloga de una cuarta secuencia; y
- en donde las secuencias tercera y cuarta son secuencias de ADN cromosómico de EPSPS.
- 30 52. El vector donante de la realización 51, en donde los extremos próximos de las secuencias tercera y cuarta son contiguos.
53. El vector donante de la realización 51, en donde los extremos próximos de las secuencias tercera y cuarta están separados por al menos 1 par de nucleótidos.
- 35 54. El vector de cualquiera de las realizaciones 51 a 53, en donde las secuencias cromosómicas de ADN de EPSPS pertenecen a un gen EPSPS que comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia idéntica en al menos un 95 % con las mismas.
55. El vector de cualquiera de las realizaciones 51 a 54, en donde las secuencias tercera y cuarta son secuencias exógenas.
56. El vector de cualquiera de las realizaciones 51 a 54, en donde las secuencias tercera y cuarta son secuencias endógenas.
- 40 57. El vector de cualquiera de las realizaciones 51 a 56, en donde al menos una de las secuencias primera o segunda tiene una longitud de 100 nucleótidos o menos.
58. El vector de cualquiera de las realizaciones 51 a 57, que además comprende una quinta secuencia, en donde la quinta secuencia:
- (a) está interpuesta entre las secuencias primera y segunda; y
- 45 (b) es una secuencia de ácido nucleico exógena.

59. El vector de la realización 58, en donde la quinta secuencia tiene un tamaño de al menos 1 par de bases.
60. El vector de la realización 58 o 59, en donde la quinta secuencia comprende secuencias que codifican un marcador de selección.
- 5 61. El vector de la realización 60, en donde el marcador de selección se selecciona del grupo que consiste en proteína fluorescente verde (GFP),  $\beta$ -glucuronidasa (GUS), fosfinotricin N-acetil transferasa (PAT, BAR), neomicina fosfotransferasa, higromicina fosfotransferasa,  $\beta$ -lactamasa, catecol dioxigenasa,  $\alpha$ -amilasa, tirosinasa,  $\beta$ -galactosidasa, luciferasa, aecuatorina, EPSP sintasa, nitrilasa, acetolactato sintasa (ALS), dihidrofolato reductasa (DHFR), dalapon deshalogenasa y antranilato sintasa.
- 10 62. El vector de la realización 58, en donde la quinta secuencia comprende secuencias que codifican una proteína distinta de un marcador de selección.
63. El vector de cualquiera de las realizaciones 58 a 62, en donde la quinta secuencia comprende una o más secuencias reguladoras transcripcionales.
64. El vector de cualquiera de las realizaciones 58 a 63, en donde la quinta secuencia comprende una o más secuencias que potencian o reducen el direccionamiento de la proteína.
- 15 65. El vector de la realización 63, en donde las una o más secuencias reguladoras transcripcionales aumentan la expresión de EPSPS.
66. El vector de la realización 63, en donde las una o más secuencias reguladoras transcripcionales reducen la expresión de EPSPS.
- 20 67. El vector de cualquiera de las realizaciones 58 a 66, en donde la quinta secuencia comprende una o más secuencias que codifican una parte de una proteína o un ARN pequeño interferente o un micro ARN.
68. El vector de la realización 67, en donde las secuencias que codifican la parte de la proteína comprenden secuencias homólogas a secuencias cromosómicas de EPSPS.
69. El vector de la realización 58, en donde la quinta secuencia comprende un homólogo de tipo silvestre de una secuencia cromosómica de EPSPS mutante.
- 25 70. El vector de la realización 58, en donde la quinta secuencia comprende un homólogo de tipo silvestre de una secuencia cromosómica de EPSPS.
71. El vector de la realización 70, en donde la secuencia cromosómica de EPSPS aumenta la tolerancia de una planta al herbicida glifosato.
- 30 72. El vector de cualquiera de las realizaciones 51 a 71, en donde la primera secuencia tiene una homología de al menos el 35 % con la tercera secuencia.
73. El vector de cualquiera de las realizaciones 51 a 72, en donde la segunda secuencia tiene una homología de al menos el 35 % con la cuarta secuencia.
74. Un método para introducir una secuencia de ácido nucleico exógena en el genoma de una célula vegetal, comprendiendo el método las etapas de:
- 35 (a) poner en contacto la célula con un vector donante según cualquiera de las realizaciones 51 a 73; y
- (b) expresar una nucleasa de dedo de cinc (ZFN) en la célula, en donde la ZFN escinde un gen EPSPS en ADN cromosómico entre 3 pares de kilobases de cualquiera de las secuencias tercera y cuarta;
- de tal forma que la escisión del ADN cromosómico en la etapa (b) estimula la incorporación del vector donante en el genoma mediante recombinación de homólogos.
- 40 75. El método de la realización 74, en donde el gen EPSPS comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia idéntica al menos idéntica al 95 % con las mismas.
76. Un método para expresar el producto de una secuencia de ácido nucleico exógena en una célula vegetal, comprendiendo el método las etapas de:
- (a) poner en contacto la célula con el vector donante de la realización 58-73; y
- 45 (b) expresar una nucleasa de dedo de cinc (ZFN) en la célula, en donde la ZFN escinde un gen EPSPS en ADN cromosómico entre 3 pares de kilobases de cualquiera de las secuencias tercera y cuarta;
- de tal forma que la escisión del ADN cromosómico en la etapa (b) da como resultado la incorporación del vector

donante en el genoma mediante recombinación de homólogos y la expresión del producto de la secuencia de ácido nucleico exógena.

77. El método de la realización 76, en donde el gen EPSPS comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia idéntica al menos idéntica al 95 % con las mismas.

5 78. Una célula vegetal transgénica obtenida según el método de cualquiera de las realizaciones 74 o 75.

79. Una planta que comprende una célula vegetal transgénica según la realización 78.

80. Un método para la recombinación de homólogos intramolecular en el genoma de una célula vegetal, comprendiendo el método las etapas de:

10 (a) proporcionar un segmento de ADN que comprende un gen de EPSPS y que comprende además una primera secuencia que es homóloga de una segunda secuencia; y

(b) poner en contacto dicho segmento de ADN con la ZFN de cualquiera de las realizaciones 14 a 28, en donde la ZFN escinde el ADN en una secuencia génica de EPSPS estimulando de este modo la recombinación de homólogos intramolecular.

81. El método de la realización 80, en donde el segmento de ADN es endógeno para la célula.

15 82. El método de la realización 80 u 81, en donde la recombinación de homólogos se produce en un cromosoma.

83. El método de la realización 82, en donde se elimina del cromosoma el ADN entre las secuencias primera y segunda.

84. El método de cualquiera de las realizaciones 80 a 83, en donde el gen EPSPS es único en el genoma.

20 85. El método de cualquiera de las realizaciones 80 a 83, en donde están presentes uno o más parálogos del gen EPSPS en el genoma.

86. El método de cualquiera de las realizaciones 80 a 85, en donde la ZFN comprende un par de proteínas de fusión, en donde cada proteína de fusión es una fusión entre el dominio de escisión de una endonucleasa de restricción de tipo IIS y un dominio de unión de dedo de cinc modificado.

25 87. El método de cualquiera de las realizaciones 80 a 86, en donde la segunda secuencia se encuentra a al menos 100 pares de bases respecto de la primera secuencia.

88. El método de cualquiera de las realizaciones 80 a 86, en donde la secuencia génica de EPSPS se encuentra a al menos 100 pares de bases respecto de la primera o la segunda secuencia.

89. El método de cualquiera de las realizaciones 80 a 86, en donde la secuencia génica de EPSPS se encuentra entre las secuencias primera y segunda.

30 90. El método de cualquiera de las realizaciones 80 a 89, en donde una de las secuencias primera o segunda es exógena para el organismo.

91. El método de cualquiera de las realizaciones 80 a 90, en donde ambas de las secuencias primera y segunda son exógenas para el organismo.

35 92. El método de la realización 83, en donde las secuencias eliminadas del cromosoma codifican la totalidad o parte de un gen EPSPS.

93. El método de la realización 83, en donde las secuencias eliminadas del cromosoma codifican la totalidad o parte de un marcador de selección.

40 94. El método de la realización 93, en donde el marcador de selección se selecciona del grupo que consiste en proteína fluorescente verde (GFP),  $\beta$ -glucuronidasa (GUS), fosfinotricin N-acetil transferasa (PAT, BAR), neomicina fosfotransferasa, higromicina fosfotransferasa,  $\beta$ -lactamasa, catecol dioxigenasa,  $\alpha$ -amilasa, tirosinasa,  $\beta$ -galactosidasa, luciferasa, aecuatorina, EPSP sintasa, nitrilasa, acetolactato sintasa (ALS), dihidrofolato reductasa (DHFR), dalapon deshalogenasa y antranilato sintasa.

95. El método de la realización 83, en donde el ADN eliminado se reemplaza por una secuencia exógena, comprendiendo el método:

45 introducir un polinucleótido en la célula, en donde el polinucleótido comprende:

(a) las secuencias cuarta y quinta, en donde la cuarta secuencia es homóloga de secuencias no eliminadas en la proximidad de la primera secuencia y la quinta secuencia es homóloga de secuencias no eliminadas en la

proximidad de la segunda secuencia; y

(b) la secuencia exógena.

96. El método de la realización 95, en donde la secuencia exógena es un marcador de selección.

5 97. El método de la realización 96, en donde el marcador de selección se selecciona del grupo que consiste en proteína fluorescente verde (GFP),  $\beta$ -glucuronidasa (GUS), fosfinotricin N-acetil transferasa (PAT, BAR), neomicina fosfotransferasa, higromicina fosfotransferasa,  $\beta$ -lactamasa, catecol dioxigenasa,  $\alpha$ -amilasa, tirosinasa,  $\beta$ -galactosidasa, luciferasa, aecuatorina, EPSP sintasa, nitrilasa, acetolactato sintasa (ALS), dihidrofolato reductasa (DHFR), dalapon deshalogenasa y antranilato sintasa.

98. El método de la realización 95, en donde la secuencia génica de EPSPS es una secuencia exógena.

10 99. El método de la realización 98, en donde la secuencia génica de EPSPS comprende una mutación.

100. El método de la realización 99, en donde la mutación aumenta la tolerancia de una planta al herbicida glifosato.

101. Un método para eliminar una secuencia génica de EPSPS del genoma de una célula vegetal, comprendiendo el método:

(a) proporcionar una célula vegetal que comprende una secuencia génica de EPSPS; y

15 (b) expresar primeras y segundas nucleasas de dedo de cinc (ZFN) en la célula, en donde la primera ZFN escinde en un primer sitio de escisión y la segunda ZFN escinde en un segundo sitio de escisión, en donde la secuencia génica de EPSPS se localiza entre el primer sitio de escisión y el segundo sitio de escisión, en donde la escisión de los sitios de escisión primero y segundo da como resultado la eliminación de la secuencia génica de EPSPS.

20 102. El método de la realización 101, en donde la secuencia génica de EPSPS se elimina mediante unión de extremos no homólogos de los sitios de escisión primero y segundo.

103. El método de la realización 101 o 102, en donde los sitios de escisión primero y segundo están separados por al menos 100 nucleótidos.

104. El método de la realización 101 o 102, en donde la célula vegetal es una célula vegetal transgénica.

25 105. El método de la realización 101 o 102, en donde la secuencia génica de EPSPS es una secuencia exógena.

106. El método de la realización 101 o 102, en donde la secuencia génica de EPSPS es una secuencia endógena.

107. Un método para modular la regulación de un gen EPSPS, comprendiendo el método:

(a) proporcionar una célula vegetal que comprende una secuencia génica de EPSPS; y

30 (b) expresar una ZFP en la célula, en donde dicha ZFP se une a un sitio diana en el gen EPSPS, de este modo modulando la regulación del gen EPSPS.

108. El método de la realización 107, en donde el sitio diana es una secuencia reguladora del gen EPSPS.

109. El método de la realización 107, en donde el sitio diana es se encuentra cadena arriba de un sitio de inicio de la transcripción del gen EPSPS.

35 110. El método de la realización 107, en donde el sitio diana es se encuentra adyacente a un sitio de inicio de la transcripción del gen EPSPS.

111. El método de la realización 107, en donde el sitio diana es se encuentra cadena abajo de un sitio de inicio de la transcripción del gen EPSPS.

112. El método de la realización 107, en donde la ZFP aumenta la transcripción del gen EPSPS.

113. El método de la realización 112, en donde la ZFP aumenta la tolerancia de una planta al herbicida glifosato.

40 114. El método de la realización 107, en donde la ZFP reduce la transcripción del gen EPSPS.

115. El método de la realización 114, en donde la ZFP reduce la tolerancia de una planta al herbicida glifosato.

#### Breve descripción de los dibujos

La **figura 1** muestra una transferencia de Southern que proporciona una estimación del número de genes EPSPS en los genomas de la variedad Nex710 de *B. napus*, *B. rapa* y *B. oleracea*. (Los marcadores patrón son los marcadores

de amplio espectro de ADN analíticos de Promega).

Las **figuras 2A-2E** muestran una representación esquemática de la estrategia de clonación usada para generar construcciones de expresión de ZFN. Se usó una estrategia por etapas: Los genes individuales que codifican la ZFN se clonaron en vectores y pVAX-C2A-NLSop2-EGFP-FokMono (figura 2A) y pVAX-N2A-NLSop2-EGFP-FokMono (figura 2B) para crear un casete de proteína dual (figura 2C). Este casete se ligó en pDAB3731 para generar un plásmido final (figura 2D) para expresar el heterodímero de la ZFN. El casete de ZFN se transfirió entonces a un vector binario con la tecnología Gateway para crear una construcción (figura 2E) para la transformación de *B. napus* mediada por *Agrobacterium*. ZFN d2 = 10654-CH3-v2; ZFN rb2 = 10657-CH3-v2.

Las **figuras 3A-3D** muestran la amplificación específica de parálogo de los genes EPSPS. Las figuras 3A-3D representan los ensayos PCR específicos de los parálogos A, B, C y D, respectivamente. Los carriles 1-6 contenían los siguientes ADN. Carril 1: Control de PCR sin ADN; carril 2: ADN de la variedad Nex710 de *B. napus* (10 ng/μl); los carriles 3, 4, 5 y 6 contenían ADN amplificado de los parálogos D, C, B y A (1570 pb) como controles positivos a una concentración de 1 ng/μl. Los productos PCR se ejecutaron en un E-GEL 96 al 2 % (Invitrogen, Carlsbad, CA) y se visualizaron usando el sistema de documentación de gel GEL DOC 2000 (Bio-Rad, Hercules, CA). La imagen se capturó y se analizó usando el programa informático QUANTITY ONE (Bio-Rad, Hercules, CA) y se procesó adicionalmente usando el programa informático E-EDITOR (Invitrogen, Carlsbad, CA). Los tamaños de los fragmentos se muestran en pares de bases (pb).

La **figura 4** muestra los sitios de unión y escisión de las ZFN para los parálogos de EPSPS de *B. napus*. Se necesitan dos proteínas ZFN para llevar a cabo la escisión bicatenaria (bc) del ADN. Cadena arriba del sitio de escisión, indicado con una flecha hacia abajo, se unió una proteína (10657 o 10658) a los nucleótidos subrayados, donde otra proteína (10654) se unió cadena abajo a la secuencia subrayada de la manera que se muestra. Únicamente cuando ambas proteínas estaban unidas a sus respectivos sitios se produjo la escisión. Las diferencias menores de secuencia (subrayadas) en los sitios de unión de una o ambas ZFN de un par entre los 5 parálogos (como se muestra más adelante) proporcionaron especificidad de secuencia y dieron como resultado la escisión bicatenaria de los parálogos.

La **figura 5** muestra una eliminación mediada por ZFN en el parálogo D de EPSPS. Se produjo una eliminación de 2 pb como resultado de la reparación de unión de extremos no homólogos (NHEJ) de una rotura de ADN bicatenaria mediada por ZFN presente en el parálogo D de EPSPS de *B. napus*. La diana de escisión para la ZFN pDAB7151 fue CAGTT, que corresponde a la eliminación de GT de 2 pb. Parte inferior: secuencia de tipo silvestre predicha. Parte superior: alineamiento de 26 secuencias de clones del parálogo D con el programa informático SEQUENCHER que muestra la eliminación de 2 pb. Estas secuencias se obtuvieron a partir de la secuenciación de cebador directo e inverso de 13 clones.

La **figura 6** muestra roturas bicatenarias mediadas por ZFN que dan como resultado NHEJ en el parálogo D de EPSPS de *B. napus*. Se muestra el alineamiento de múltiples eliminaciones de NHEJ con respecto al sitio de escisión previsto (parte superior) en el ADN de tipo silvestre. Los números en los paréntesis en el lado de la derecha muestran el número de moléculas idénticas observadas en el alineamiento.

Las **figuras 7A-7B** muestran roturas bicatenarias mediadas por ZFN que dan como resultado NHEJ en los parálogos C y D de EPSPS de *B. napus*. Se muestra el alineamiento de múltiples eliminaciones de NHEJ respecto del sitio de escisión predicho (parte superior) en las muestras de tipo silvestre y tratadas. Los números de muestra corresponden a aquellos mostrados en la tabla 5.

Las **figuras 8A-8B** muestran roturas bicatenarias mediadas por ZFN que dan como resultado NHEJ en los parálogos A y B de EPSPS de *B. napus*. Se muestra el alineamiento de múltiples eliminaciones de NHEJ respecto del sitio de escisión predicho (parte superior) en las muestras de tipo silvestre y transgénicas. Los números de muestra corresponden a aquellos mostrados en la tabla 6.

La **figura 9** (SEQ ID NO: 10) muestra la secuencia nucleotídica de la secuencia del parálogo A de EPSPS de *B. napus*.

La **figura 10** (SEQ ID NO: 11) muestra la secuencia nucleotídica de la secuencia del parálogo B de EPSPS de *B. napus*.

Las **figuras 11A-11B** (SEQ ID NO: 12) muestran la secuencia nucleotídica del parálogo C de EPSPS de *B. napus*.

Las **figuras 12A-12B** (SEQ ID NO: 13) muestran la secuencia nucleotídica del parálogo D de EPSPS de *B. napus*.

Las **figuras 13A-13B** (SEQ ID NO: 14) muestran la secuencia nucleotídica del parálogo E de EPSPS de *B. napus*.

La **figura 14**, paneles A a E son gráficas que ilustran la actividad de corrección génica de ZFN de EPSPS ejemplares (véase también el ejemplo 3) en células indicadoras 293 de riñón. La fig. 14A muestra la corrección génica usando el par de ZFN 10654 y 10658. La fig. 14B muestra la actividad de corrección génica usando el par de ZFN 10654 y 10658. La fig. 14C muestra la actividad de corrección génica usando el par de ZFN 9875 y 10275. La

fig. 14D muestra la actividad de corrección génica para los pares de ZFN 10740/10741 y 10749/10742. La fig. 14E muestra la actividad de corrección génica para los pares de ZFN indicados debajo de cada barra.

### Descripción detallada

5 En la presente memoria se describen composiciones y métodos útiles para la modulación de la expresión y la escisión dirigida y alteración de genes en plantas, en particular genes parálogos en plantas. La regulación de un gen parálogo puede modularse, por ejemplo, usando factores de transcripción de ZFP modificados o modificando regiones reguladoras génicas. Los genes pueden alterarse, por ejemplo, mediante escisión dirigida seguida de recombinación de homólogos intracromosómica o mediante escisión dirigida seguido de recombinación de homólogos entre un polinucleótido exógeno (que comprende una o más regiones de homología con la secuencia nucleotídica del gen) y una secuencia genómica. Un ejemplo de un gen parálogo en plantas es el gen EPSPS.

10 Las secuencias genómicas incluyen aquellas presentes en cromosomas, episomas, genomas organulares (por ejemplo, mitocondrias, cloroplastos), cromosomas artificiales y cualquier otro tipo de ácido nucleico presente en una célula, tal como, por ejemplo, secuencias amplificadas, cromosomas diminutos dobles y los genomas de bacterias y virus endógenos o infecciosos. Las secuencias genómicas pueden ser normales (es decir, de tipo silvestre) o mutantes; las secuencias mutantes pueden comprender, por ejemplo, inserciones, eliminaciones, translocaciones, reordenamientos y/o mutaciones puntuales. Una secuencia genómica también puede comprender uno de una serie de alelos diferentes.

15 Las composiciones descritas en la presente memoria comprenden una o más ZFP que comprenden dominios de unión de dedo de cinc modificados, polinucleótidos que codifican estos polipéptidos y combinaciones de ZFP y polinucleótidos que codifican ZFP. Un dominio de unión de dedo de cinc puede comprender uno o más dedos de cinc (por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9 o más dedos de cinc) y pueden estar modificados para que se unan a cualquier secuencia genómica de EPSPS.

20 Las ZFP, como se describen en la presente memoria, pueden usarse para regular la expresión génica de EPSPS, ya sea mediante la activación o represión de la transcripción génica. Pueden construirse ZFP que comprenden fusiones de dominios de dedo de cinc unidas a dominios reguladores para crear factores de transcripción quiméricos que activan o reprimen la transcripción. También pueden usarse las ZFP para la escisión dirigida de una región genómica de EPSPS de interés mediante la unión de dominios de dedo de cinc con dominios de escisión de nucleasa (o semidominios de escisión) para producir nucleasas de dedo de cinc. Por lo tanto, mediante la identificación de una región genómica de EPSPS de interés en la que se desea la regulación, escisión o recombinación génica, se puede, según los métodos descritos en la presente memoria, construir una proteína de 25 dedo de cinc que comprende una o más proteínas de fusión que comprenden uno o más dominios reguladores y/o dominios de escisión (o semidominios de escisión) unidos a un dominio de dedo de cinc modificado para que reconozca una secuencia génica de EPSPS en esa región genómica. La presencia de dicha ZFP que comprende una proteína (o proteínas) de fusión en una célula dará como resultado la unión de la proteína (o las proteínas) de fusión a su (o sus) sitio (o sitios) diana y a la regulación o escisión alterada dentro de o próxima a la región genómica. Además, si se escinde una región genómica de EPSPS y también está presente un polinucleótido exógeno homólogo a esa región genómica de EPSPS en la célula, se produce la recombinación de homólogos a una 30 alta tasa entre la región genómica de EPSPS y el polinucleótido exógeno.

35 Las células vegetales pueden contener una o más secuencias génicas homólogas o parálogas de EPSPS, pudiendo dirigirse a cualquier número de ellas o a todas ellas para la modificación mediante los métodos descritos en esta memoria. Por lo tanto, las composiciones descritas en la presente memoria pueden dirigirse a uno o más genes EPSPS en una célula vegetal, por ejemplo, a 1, 2, 3, 4, 5 o hasta cualquier número de parálogos de EPSPS o a todos los parálogos de EPSPS presentes en una célula vegetal. Algunas ZFP pueden unirse específicamente a un gen parálogo de EPSPS particular en una célula vegetal. Otras ZFP pueden unirse a múltiples genes parálogos de EPSPS en una célula vegetal. Por lo tanto, pueden combinarse una o más ZFP o vectores de expresión que 40 codifican ZFP de diferentes especificidades para dirigirse a los genes de EPSPS de interés deseados en una planta.

### General

45 La puesta en práctica de los métodos, así como la preparación y uso de las composiciones descritas en la presente memoria emplean, a menos que se indique lo contrario, técnicas convencionales de biología molecular, bioquímica, estructura y análisis de cromatina, química computacional, cultivo celular, ADN recombinante y campos relacionados dentro de las capacidades de la técnica. Estas técnicas se explican detalladamente en la bibliografía. Véase, por ejemplo, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, Segunda edición, Cold Spring Harbor Laboratory Press, 1989 y tercera edición, 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, Nueva York, 1987 y actualizaciones periódicas; la serie METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, tercera edición, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman y A. P. Wolffe, eds.), Academic Press, San Diego, 1999; y METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

## Definiciones

Los términos "ácido nucleico," "polinucleótido" y "oligonucleótido" se usan de manera intercambiable y se refieren a un polímero de desoxirribonucleótidos o ribonucleótidos, en conformación lineal o circular y en forma ya sea monocatenaria o bicatenaria. Para los fines de la presente descripción, estos términos no deben entenderse como limitantes respecto de la longitud de un polímero. Los términos pueden abarcar análogos conocidos de nucleótidos naturales, así como nucleótidos que se modifican en los restos de la base, el azúcar y/o de fosfato (por ejemplo, armazones de fosforotioato). En general, un análogo de un nucleótido particular tiene la misma especificidad de emparejamiento de bases; es decir, un análogo de A se emparejará con una base de T.

Los términos "polipéptido", "péptido" y "proteína" se usan de manera intercambiable para referirse a un polímero de restos de aminoácidos. El término también se aplica a polímeros de aminoácidos en los que uno o más aminoácidos son análogos químicos o derivados modificados de un aminoácido de origen natural.

"Unión" se refiere a una interacción no covalente, específica de secuencia entre macromoléculas (por ejemplo, entre una proteína y un ácido nucleico). No todos los componentes de una interacción de unión tienen por qué ser específicos de secuencia (por ejemplo, contactos con restos fosfato en un armazón de ADN), en la medida en la que la interacción sea específica de secuencia en su conjunto. Dichas interacciones se caracterizan generalmente por una constante de disociación ( $K_d$ ) de  $10^{-6} \text{ M}^{-1}$  o menor. "Afinidad" se refiere a la fuerza de la unión: una afinidad de unión aumentada se correlaciona con una menor  $K_d$ .

Una "proteína de unión" es una proteína que es capaz de unirse de manera no covalente a otra molécula. Una proteína de unión puede unirse, por ejemplo, a una molécula de ADN (una proteína de unión a ADN), una molécula de ARN (una proteína de unión a ARN) y/o una molécula de proteína (una proteína de unión a proteínas). En el caso de una proteína de unión a proteínas, puede unirse a sí misma (para formar homodímeros, homotrímeros, etc.) y/o puede unirse a una o más moléculas de una proteína o proteínas diferentes. Una proteína de unión puede tener más de una actividad de unión. Por ejemplo, las proteínas de dedo de cinc tienen actividad de unión a ADN, de unión a ARN y de unión a proteínas.

Una "proteína de unión a ADN de dedo de cinc" (o su dominio de unión) es una proteína o un dominio dentro de una proteína mayor, que se une a ADN de una manera específica de secuencia mediante uno o más dedos de cinc, que son regiones de secuencia de aminoácidos dentro del dominio de unión cuya estructura se estabiliza mediante la coordinación de un ión de cinc. La expresión proteína de unión a ADN de dedo de cinc se abrevia normalmente a proteína de dedo de cinc o ZFP.

Los dominios de unión de dedo de cinc pueden "modificarse" para que se unan a una secuencia nucleotídica predeterminada (por ejemplo, una secuencia génica de EPSPS). Los ejemplos de métodos para modificar proteínas de dedo de cinc son el diseño y la selección. Una proteína de dedo de cinc diseñada es una proteína de origen no natural cuyo diseño/composición es el resultado principalmente de criterios racionales. Los criterios racionales para el diseño incluyen reglas de aplicación de sustitución y algoritmos informáticos para procesar información en una base de datos que almacena información de diseños y datos de unión de ZFP existentes. Véanse, por ejemplo, las Patentes de Estados Unidos 6.140.081; 6.453.242; 6.534.261; y 6.785.613; véanse también los documentos WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 y WO 03/016496; y las Patentes de Estados Unidos 6.746.838; 6.866.997; y 7.030.215. Por lo tanto, una proteína de dedo de cinc "modificada" o una proteína de dedo de cinc "de origen no natural" es una en la que uno o más de los dominios de unión a ADN del componente de dedo de cinc (hélices de reconocimiento) son de origen no natural y se han modificado para que se unan a un sitio diana preseleccionado.

Una proteína de dedo de cinc "seleccionada" es una proteína que no se encuentra en la naturaleza cuya producción es el resultado principalmente de un proceso empírico tal como presentación en fagos, trampa de interacción o selección híbrida. Véanse, por ejemplo, los documentos US 5.789.538; US 5.925.523; US 6.007.988; US 6.013.453; US 6.200.759; US 6.733.970; US RE39.229; y los documentos WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970, WO 01/88197 y WO 02/099084.

El término "secuencia" se refiere a una secuencia nucleotídica de cualquier longitud, que puede ser ADN o ARN; puede ser lineal, circular o ramificada y puede ser monocatenaria o bicatenaria. La expresión "secuencia donante" se refiere a una secuencia nucleotídica que se inserta en un genoma. Una secuencia donante puede ser de cualquier longitud, por ejemplo, de entre 2 y 25.000 nucleótidos de longitud (o cualquier valor entero intermedio o superior), preferentemente de entre 100 y 5.000 nucleótidos de longitud (o cualquier número entero intermedio), más preferentemente de entre 200 y 2.500 nucleótidos de longitud.

Una "secuencia homóloga" se refiere a una primera secuencia que comparte un grado de identidad de secuencia con una segunda secuencia y cuya secuencia puede ser idéntica a aquella de la segunda secuencia. Una "secuencia homóloga, no idéntica" se refiere a una primera secuencia que comparte un grado de identidad de secuencia con una segunda secuencia, pero cuya secuencia no es idéntica a aquella de la segunda secuencia. Por ejemplo, un polinucleótido que comprende la secuencia de tipo silvestre de un gen mutante es homólogo y no idéntico a la secuencia del gen mutante. En determinadas realizaciones, el grado de homología entre las dos

secuencias es suficiente para permitir la recombinación de homólogos entre las mismas, utilizando mecanismos celulares normales. Dos secuencias homólogas no idénticas pueden tener cualquier longitud y su grado de no homología puede ser tan pequeño como de un solo nucleótido (por ejemplo, para la corrección de una mutación genómica puntual mediante recombinación de homólogos dirigida) o tan grande como de 10 o más kilobases (por ejemplo, para la inserción de un gen en un sitio predeterminado en un cromosoma). Dos polinucleótidos que comprenden las secuencias homólogas no idénticas no tienen que tener necesariamente la misma longitud. Por ejemplo, puede usarse un polinucleótido exógeno (es decir, polinucleótido donante) de entre 20 y 10.000 nucleótidos o partes de nucleótidos.

Las técnicas para determinar la identidad de secuencia de ácidos nucleicos y aminoácidos se conocen en la técnica. Típicamente, dichas técnicas incluyen la determinación de la secuencia de nucleótido del ARNm para un gen y/o determinar la secuencia de aminoácidos codificada por la misma y comparar estas secuencias con una segunda secuencia de nucleótidos o de aminoácidos. También pueden determinarse y compararse de este modo secuencias genómicas. En general, la identidad se refiere a una correspondencia exacta de nucleótido a nucleótido o de aminoácido a aminoácido de dos secuencias de polinucleótidos o polipéptidos, respectivamente. Pueden compararse dos o más secuencias (de polinucleótidos o aminoácidos) determinando su porcentaje de identidad. El porcentaje de identidad de dos secuencias, ya sean secuencias de ácidos nucleicos o de aminoácidos, es el número de coincidencias exactas entre dos secuencias alineadas dividido entre la longitud de las secuencias más cortas y multiplicado por 100. Un alineamiento aproximado para secuencias de ácidos nucleicos se proporciona mediante el algoritmo de homología local de Smith y Waterman, *Advances in Applied Mathematics* 2:482-489 (1981). Este algoritmo puede aplicarse a secuencias de aminoácidos mediante el uso de la matriz de puntuación desarrollada por Dayhoff, *Atlas of Protein Sequences and Structure*, M.O. Dayhoff ed., 5 supl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., EE.UU. y normalizado por Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). Una implementación ejemplar de este algoritmo para determinar el porcentaje de identidad de una secuencia se proporciona por el Genetics Computer Group (Madison, WI) en la aplicación de la utilidad "BestFit". Los parámetros por defecto para este método se describen en el manual del paquete del programa Wisconsin Sequence Analysis Package, versión 8 (1995) (disponible a través del Genetics Computer Group, Madison, WI). Un método preferido para determinar el porcentaje de identidad en el contexto de la presente descripción es usar el paquete de programas MPSRCH con derechos de autor para la Universidad de Edimburgo, desarrollado por John F. Collins y Shane S. Sturrok y distribuido por IntelliGenetics, Inc. (Mountain View, CA). Puede emplearse el algoritmo Smith-Waterman a partir de este conjunto de paquetes, en donde se usan parámetros por defecto para la tabla de puntuación (por ejemplo, penalización por hueco abierto 12, penalización por extensión de hueco de uno y un hueco de seis). Para la generación de datos, el valor "Coincidencia" refleja la identidad de secuencia. Se conocen generalmente en la técnica otros programas adecuados para calcular el porcentaje de identidad o similitud entre secuencias, por ejemplo, otro programa de alineamiento es BLAST, usado con parámetros por defecto. Por ejemplo, pueden usarse BLASTN y BLASTP usando los siguientes parámetros por defecto: código genético = estándar; filtro = ninguno; hebra = ambas; corte = 60; esperado = 10; Matriz = BLOSUM62; Descripciónes = 50 secuencias; clasificar por = ALTA PUNTUACIÓN; Bases de datos = no redundantes, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Los detalles de estos programas pueden encontrarse en Internet. Con respecto a las secuencias descritas en la presente memoria, el intervalo de grados de identidad de secuencia deseado es de aproximadamente el 35 % al 100 % y cualquier valor entero intermedio. Típicamente, los porcentajes de identidad entre secuencias son una identidad de secuencia de al menos el 35 %-40 %; 40 %-45 %; 45 %-50 %; 50 %-60 %; 60 %-70 %; 70-75 %, preferentemente del 80-82 %, más preferentemente del 85-90 %, incluso más preferentemente del 92 %, aún más preferentemente del 95 % y lo más preferentemente del 98 %.

Como alternativa, puede determinarse el grado de similitud de secuencia entre polinucleótidos mediante hibridación de los polinucleótidos en condiciones que permiten la formación de dúplex estables entre regiones homólogas, seguido de digestión con nucleasas específicas de cadenas sencillas y determinación del tamaño de los fragmentos digeridos. Dos ácidos nucleicos o dos secuencias polipeptídicas son sustancialmente homólogas entre sí cuando las secuencias muestran al menos aproximadamente un 70 %-75 %, preferentemente un 80 %-82 %, más preferentemente un 85 %-90 %, incluso más preferentemente un 92 %, aún más preferentemente un 95 % y lo más preferentemente un 98 % de identidad de secuencia a lo largo de una longitud definida de las moléculas, según se determina usando los métodos anteriores. Como se usa en la presente memoria, sustancialmente homólogo se refiere también a secuencias que muestran identidad completa con un ADN o secuencia polipeptídica específica. Pueden identificarse secuencias de ADN que son sustancialmente homólogas en un experimento de hibridación de Southern en, por ejemplo, condiciones rigurosas, tal como se definen para ese sistema particular. La definición de las condiciones de hibridación adecuadas se encuentra dentro de las capacidades de la técnica. Véase, por ejemplo, Sambrook et al., anteriormente citado; *Nucleic Acid Hybridization: A Practical Approach*, editores B.D. Hames y S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

Puede determinarse la hibridación selectiva de dos fragmentos de ácido nucleico del modo siguiente. El grado de identidad de secuencia entre dos moléculas de ácido nucleico afecta a la eficacia y a la fuerza de los eventos de hibridación entre dichas moléculas. Una secuencia de ácido nucleico parcialmente idéntica inhibirá al menos parcialmente la hibridación de una secuencia completamente idéntica a una molécula diana. La inhibición de la hibridación de la secuencia completamente idéntica puede evaluarse usando ensayos de hibridación que se conocen bien en la técnica (por ejemplo, transferencia de Southern (ADN), transferencia de Northern (ARN),

hibridación en solución o similares, véase Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Segunda edición, (1989) Cold Spring Harbor, N.Y.). Dichos ensayos pueden llevarse a cabo usando diversos grados de selectividad, por ejemplo, usando condiciones que varían de baja a alta rigurosidad. Si se emplean las condiciones de baja rigurosidad, puede evaluarse la ausencia de unión no específica usando una sonda secundaria que carece incluso de un grado parcial de identidad de secuencia (por ejemplo, una sonda que tiene menos de aproximadamente un 30 % de identidad de secuencia con la molécula diana), de tal forma que, en ausencia de sucesos de unión no específica, la sonda secundaria no se unirá a la diana.

Cuando se utiliza un sistema de detección basado en hibridación, se selecciona una sonda de ácido nucleico que es complementaria a una secuencia de ácido nucleico de referencia y después mediante la selección de condiciones adecuadas, se hibridan la sonda y la secuencia de referencia o se unen entre sí para formar una molécula dúplex. Una molécula de ácido nucleico que es capaz de hibridar de manera selectiva con una secuencia de referencia en condiciones de hibridación moderadamente rigurosas hibrida típicamente en condiciones que permiten la detección de una secuencia de ácido nucleico diana de al menos 10-14 nucleótidos de longitud que tiene una identidad de secuencia de al menos el 70 % con la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación rigurosas permiten típicamente la detección de secuencias de ácido nucleico diana de al menos 10-14 nucleótidos de longitud que tienen una identidad de secuencia mayor del 90-95 % respecto de la secuencia de la sonda de ácido nucleico seleccionada. Las condiciones de hibridación útiles para la hibridación de la sonda/secuencia de referencia, donde la sonda y la secuencia de referencia tienen un grado de identidad específico, pueden determinarse tal como se conoce en la técnica (véase, por ejemplo, *Nucleic Acid Hybridization: A Practical Approach*, editores B.D. Hames y S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

Las condiciones de hibridación se conocen bien por los expertos en la materia. La rigurosidad de la hibridación se refiere al grado en que las condiciones de hibridación desalientan la formación de híbridos que contienen nucleótidos desparejados, correlacionándose una mayor rigurosidad con una menor tolerancia de híbridos desparejados. Los factores que afectan a la rigurosidad de la hibridación se conocen bien por los expertos en la materia e incluyen temperatura, pH, fuerza iónica y concentración de disolventes orgánicos, tales como, por ejemplo, formamida y dimetilsulfóxido. Como es sabido para los expertos en la materia, la rigurosidad de la hibridación aumenta por mayores temperaturas, menor fuerza iónica y menores concentraciones de disolvente.

Con respecto a las condiciones de rigurosidad para la hibridación, se sabe en la técnica que pueden emplearse numerosas condiciones equivalentes para establecer una rigurosidad particular variando, por ejemplo, los siguientes factores: los parámetros de longitud y naturaleza de las secuencias, la composición de bases de las diversas secuencias, las concentraciones de sales y otros componentes de la solución, la presencia o ausencia de agentes bloqueantes en las soluciones de hibridación (por ejemplo, sulfato de dextrano y polietilenglicol), la temperatura de la reacción de hibridación y el tiempo, así como diversas condiciones de lavado. La selección de un conjunto de condiciones de hibridación particular se efectúa siguiendo métodos estándar en la técnica (véase, por ejemplo, Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Segunda edición, (1989) Cold Spring Harbor, N.Y.).

"Recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos. Para los fines de la presente descripción, "recombinación de homólogos (HR)" se refiere a la forma especializada de dicho intercambio que tiene lugar, por ejemplo, durante la reparación de roturas bicatenarias en células. Este proceso requiere de homología de secuencia de nucleótidos, usa una molécula "donante" para el molde de reparación de una molécula "diana" (es decir, aquella que sufre la rotura bicatenaria) y se conoce de manera diversa como "conversión génica sin sobrecruzamiento" o "conversión génica de tracto corto", debido a que da lugar a la transferencia de información genética del donante a la diana. Sin desear quedar ligados a cualquier teoría particular, dicha transferencia puede implicar la corrección de desparejamiento de heterodúplex de ADN que se forma entre la diana rota y el donante y/o "hibridación de hebra dependiente de síntesis", en la que se usa el donante para volver a sintetizar información genética que formará parte de la diana y/o procesos relacionados. Dicha HR especializada da como resultado a menudo una alteración de la secuencia de la molécula diana de tal forma que se incorpora una parte o la totalidad de la secuencia del polinucleótido donante en el polinucleótido diana.

"Escisión" se refiere a la rotura del armazón covalente de una molécula de ADN. La escisión puede iniciarse por una diversidad de métodos, incluyendo la hidrólisis enzimática o química de un enlace fosfodiéster. Son posibles la escisión tanto monocatenaria como bicatenaria y la escisión bicatenaria puede suceder como resultado de dos sucesos de escisión monocatenarios. La escisión de ADN puede dar como resultado la producción de extremos romos o de extremos escalonados. En determinadas realizaciones, se usan polipéptidos de fusión para la escisión dirigida de ADN bicatenario.

Un "dominio de escisión" comprende una o más secuencias polipeptídicas que comprenden actividad catalítica para la escisión de ADN. Un dominio de escisión puede estar contenido en una sola cadena polipeptídica o la actividad de escisión puede ser el resultado de la asociación de dos (o más) polipéptidos.

Un "semidominio de escisión" es una secuencia polipeptídica que conjuntamente con un segundo polipéptido (ya sea idéntico o diferente) forma un complejo que tiene actividad de escisión (preferentemente actividad de escisión bicatenaria).

La "cromatina" es la estructura de nucleoproteína que comprende el genoma celular. La cromatina celular comprende ácido nucleico, principalmente ADN y proteínas, que incluyen histonas y proteínas cromosómicas no histonas. La mayoría de la cromatina celular eucariota existe en forma de nucleosomas, en donde un núcleo de nucleosoma comprende aproximadamente 150 pares de bases de ADN asociadas con un octámero que comprende dos de cada una de las histonas H2A, H2B, H3 y H4; y ADN enlazador (de longitud variable dependiendo del organismo) que se extiende entre los núcleos del nucleosoma. Generalmente se asocia una molécula de histona H1 con el ADN enlazador. Para los fines de la presente descripción, se entiende que el término "cromatina" abarca todos los tipos de nucleoproteína celular, tanto procariota como eucariota. La cromatina celular incluye cromatina tanto cromosómica como episómica.

Un "cromosoma" es un complejo de cromatina que comprende la totalidad o parte del genoma de una célula. A menudo, el genoma de una célula se caracteriza por su cariotipo, que es la colección de todos los cromosomas que comprenden el genoma de la célula. El genoma de una célula puede comprender uno o más cromosomas.

Un "episoma" es un complejo de ácido nucleico y nucleoproteína u otra estructura que comprenda un ácido nucleico que no forma parte del cariotipo cromosómico de una célula. Los ejemplos de episomas incluyen plásmidos y determinados genomas víricos.

Una "región accesible" es un sitio en la cromatina celular donde puede unirse una molécula exógena que reconoce al sitio diana para un sitio diana presente en el ácido nucleico. Sin desear quedar ligados a cualquier teoría particular, se cree que una región accesible es una que no está empaquetada en una estructura de nucleosoma. A menudo puede detectarse la estructura distinta de una región accesible según su sensibilidad a sondas químicas y enzimáticas, por ejemplo, nucleasas.

Un "sitio diana" o una "secuencia diana" es una secuencia de ácido nucleico que define una parte de un ácido nucleico a la que se unirá una molécula de unión, siempre que existan condiciones suficientes para la unión. Por ejemplo, la secuencia 5'-GAATTC-3' es un sitio diana para la endonucleasa de restricción Eco RI.

Una molécula "exógena" es una molécula que normalmente no está presente en una célula pero que puede introducirse en una célula mediante uno o más métodos genéticos, bioquímicos u otros. La "presencia normal en la célula" se determina respecto del estado de desarrollo particular y las condiciones ambientales de la célula. Por lo tanto, por ejemplo, una molécula que está presente únicamente durante el desarrollo embrionario del músculo es una molécula exógena con respecto a una célula muscular adulta. De manera similar, una molécula inducida por choque térmico es una molécula exógena respecto de una célula a la que nos e ha sometido a choque térmico. Una molécula exógena puede comprender, por ejemplo, una versión funcional de una molécula endógena disfuncional o una versión disfuncional de una molécula endógena que esté funcionando normalmente.

Una molécula exógena puede ser, entre otras cosas, una molécula pequeña, tal como generada mediante un proceso de química combinatoria o una macromolécula, tal como una proteína, ácido nucleico, carbohidrato, lípido, glucoproteína, lipoproteína, polisacárido y un derivado modificado de las moléculas anteriores o cualquier complejo que comprende una o más de las moléculas anteriores. Los ácidos nucleicos incluyen ADN y ARN, pueden ser mono o bicatenarios; pueden ser lineales, ramificados o circulares; y pueden tener cualquier longitud. Los ácidos nucleicos incluyen aquellos capaces de formar dúplex, así como ácidos nucleicos formadores de tripletes. Véanse, por ejemplo, las Patentes de Estados Unidos n.º 5.176.996 y 5.422.251. Las proteínas incluyen proteínas de unión a ADN, factores de transcripción, factores remodeladores de la cromatina, proteínas de unión a ADN metilado, polimerasas, metilasas, desmetilasas, acetilasas, desacetilasas, cinasas, fosfatasa, recombinasas, ligasas, topoisomerasas, girasas y helicinas.

Una molécula exógena puede ser el mismo tipo de molécula que una molécula endógena, por ejemplo, una proteína o ácido nucleico exógeno. Por ejemplo, un ácido nucleico exógeno puede comprender un genoma vírico infeccioso, una hebra T de *Agrobacterium tumefaciens*, un plásmido o episoma introducido en una célula o un cromosoma que no está normalmente presente en la célula. Los métodos para la introducción de moléculas exógenas en células se conocen por los expertos en la materia e incluyen transferencia mediada por lípidos (es decir, liposomas, incluyendo lípidos neutros y catiónicos), electroporación, inyección directa, fusión celular, bombardeo de partículas, co-precipitación de fosfato de calcio, transferencia mediada por DEAE-dextrano y transferencia mediada por vector vírico. La molécula exógena puede ser una molécula no vegetal, por ejemplo, un anticuerpo de mamífero (por ejemplo, humano o humanizado).

Por el contrario, una molécula "endógena" es una que está normalmente presente en una célula particular en un estado de desarrollo particular en condiciones ambientales particulares. Por ejemplo, un ácido nucleico endógeno puede comprender un cromosoma, el genoma de una mitocondria, cloroplasto u otro orgánulo o un ácido nucleico episómico de origen natural. Las moléculas exógenas adicionales pueden incluir proteínas, por ejemplo, factores de transcripción y enzimas.

Una molécula de "fusión" es una molécula en la que se unen dos o más subunidades de moléculas, preferentemente de manera covalente. Las subunidades de moléculas pueden ser del mismo tipo químico de molécula o pueden ser moléculas de diferente tipo químico. Los ejemplos del primer tipo de molécula de fusión incluyen proteínas de fusión

(por ejemplo, una ZFN que comprende una fusión entre un dominio de unión a ADN de ZFP y un dominio de escisión) y ácidos nucleicos de fusión (por ejemplo, un ácido nucleico que codifica la proteína de escisión descrita anteriormente). Los ejemplos del segundo tipo de molécula de fusión incluyen una fusión entre un ácido nucleico formador de tripletes y un polipéptido y una fusión entre una molécula de unión al surco menor y un ácido nucleico.

- 5 La expresión de una proteína de fusión en una célula puede ser el resultado del suministro de la proteína de fusión a la célula o del suministro de un polinucleótido que codifica la proteína de fusión a una célula, en donde el polinucleótido se transcribe y el transcrito se traduce para generar la proteína de fusión. También pueden estar implicados el corte y empalme en *trans*, la escisión polipeptídica y la ligadura polipeptídica en la expresión de una proteína en la célula. Los métodos para el suministro de polinucleótidos y polipéptidos a células se representan en  
10 otras partes de la presente memoria.

Un "gen", para los fines de la presente descripción, incluye una región de ADN que codifica un producto génico (véase más abajo), así como todas las regiones de ADN que regulan la producción del producto génico, estén o no adyacentes secuencias reguladoras a las secuencias codificantes y/o transcritas. Por consiguiente, un gen incluye, pero no necesariamente se limita a, secuencias promotoras, terminadores, secuencias reguladoras traduccionales, tales como sitios de unión a ribosomas y sitios de entrada a ribosomas internos, potenciadores, silenciadores, aislantes, elementos de límite, orígenes de replicación, sitios de unión a matriz y locus de región de control.  
15

La "expresión génica" se refiere a la conversión de la información, contenida en un gen, en un producto génico. Un producto génico puede ser el producto transcripcional directo de un gen, (por ejemplo, ARNm, ARNt, ARNr, ARN antisentido, ribozima, ARNhc, micro ARN, ARN estructural o cualquier otro tipo de ARN) o una proteína producida por la traducción de un ARNm. Los productos génicos también incluyen ARN que se modifican mediante procesos tales como taponado, poliadenilación, metilación y edición y proteínas modificadas mediante, por ejemplo, metilación, acetilación, fosforilación, ubiquitinación, ADP-ribosilación, miristilación y glucosilación.  
20

La "modulación" de la expresión génica se refiere a un cambio en la actividad de un gen. La modulación de la expresión puede incluir la activación génica y la represión génica.

- 25 Las células "vegetales" incluyen células de plantas monocotiledóneas (monocot) o dicotiledóneas (dicot). Los ejemplos de monocot incluyen plantas de cereales, tales como maíz, arroz, cebada, avena, trigo, sorgo, centeno, caña de azúcar, piña, cebolla, plátano y coco. Los ejemplos de dicot incluyen tabaco, tomate, girasol, algodón, remolacha azucarera, patata, lechuga, melón, soja, colza y alfalfa. Las células vegetales pueden ser de cualquier parte de la planta y/o de cualquier etapa del desarrollo de la planta.

- 30 Una "región de interés" es cualquier región de la cromatina celular, tal como, por ejemplo, un gen o una secuencia no codificante dentro de o adyacente a un gen en la que es deseable unir una molécula exógena (por ejemplo, una región genómica de EPSPS de interés incluye una región dentro de o adyacente a un gen de EPSPS). La unión puede ser con fines de escisión de ADN dirigida y/o de recombinación dirigida. Una región de interés puede estar presente en un cromosoma, un episoma, un genoma organular (por ejemplo, mitocondrial, de cloroplasto) o un  
35 genoma vírico infeccioso, por ejemplo. Una región de interés puede estar dentro de la región codificante de un gen, dentro de regiones no codificantes transcritas, por ejemplo, secuencias líder, secuencias no codificantes o intrones o dentro de regiones no traducidas, ya sea cadena arriba o cadena abajo de la región codificante. Una región de interés puede ser tan pequeña como de un solo par de nucleótidos o de hasta 25.000 pares de nucleótidos de longitud o cualquier valor intermedio de pares de nucleótidos.

- 40 Las expresiones "unión operativa" y "unido operativamente" (o "unido operablemente") se usan de manera intercambiable en referencia a una yuxtaposición de dos o más componentes (tales como elementos de secuencia, en los que los componentes se disponen de tal modo que ambos componentes funcionan de manera normal y permiten la posibilidad de que al menos uno de los componentes pueda mediar una función que se ejerce por al menos uno de los otros componentes. A modo ilustrativo, una secuencia reguladora transcripcional, tal como un  
45 promotor, se une operablemente a una secuencia codificante si la secuencia reguladora transcripcional controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o ausencia de uno o más factores reguladores de la transcripción. Una secuencia reguladora generalmente se une operablemente en *cis* con una secuencia codificante, pero no tiene que estar necesariamente adyacente a esta. Por ejemplo, un potenciador es una secuencia reguladora transcripcional que se une operablemente a una secuencia codificante, aunque no esté  
50 contiguo.

Con respecto a los polipéptidos de fusión, la expresión "unido operativamente" puede referirse al hecho de que cada uno de los componentes lleva a cabo la misma función enlazado al otro componente, como en caso de que no estuviesen enlazados de ese modo. Por ejemplo, con respecto a un polipéptido de fusión en el que se fusiona un dominio de unión a ADN de ZFP a un dominio de escisión, el dominio de unión a ADN de ZFP y el dominio de escisión se encuentran en unión operativa si, en el polipéptido de fusión, la parte de dominio de unión a ADN de ZFP es capaz de unirse a su sitio diana y/o su sitio de unión, mientras que el dominio de escisión es capaz de escindir al ADN en las proximidades del sitio diana.  
55

Un "fragmento funcional" de una proteína, polipéptido o ácido nucleico es una proteína, polipéptido o ácido nucleico cuya secuencia no es idéntica a la proteína, polipéptido o ácido nucleico de longitud completa, aunque retiene la misma función que la proteína, polipéptido o ácido nucleico de longitud completa. Un fragmento funcional puede poseer más, menos o el mismo número de restos que la molécula nativa correspondiente y/o puede contener una o más sustituciones de aminoácidos o nucleótidos. Los métodos para determinar la función de un ácido nucleico (por ejemplo, función codificante, capacidad para hibridar con otro ácido nucleico) se conocen bien en la técnica. De manera similar, se conocen bien métodos para determinar la función de una proteína. Por ejemplo, puede determinarse la función de unión a ADN de un polipéptido, por ejemplo, mediante ensayos de unión a filtro, de cambio en la movilidad electroforética o de inmunoprecipitación. La escisión de ADN puede ensayarse mediante electroforesis en gel. Véase *Ausubel et al.*, anteriormente citado. Puede determinarse la capacidad de una proteína para interactuar con otra proteína, por ejemplo, mediante ensayos de co-inmunoprecipitación, de dos híbridos o de complementación, tanto genética como bioquímica. Véase, por ejemplo, Fields et al. (1989) *Nature* 340:245-246; la Patente de Estados Unidos n.º 5.585.245 y el documento PCT WO 98/44350.

### Sitios diana

Los métodos y composiciones descritos incluyen ZFP que comprenden proteínas de fusión que comprenden un dominio regulador o dominio de escisión (o un semidominio de escisión) y un dominio de dedo de cinc, en el que el dominio de dedo de cinc, mediante la unión a una secuencia en la cromatina celular (por ejemplo, un sitio génico diana o sitio de unión a EPSPS), dirige la actividad del dominio regulador o del dominio de escisión (o semidominio de escisión) a las proximidades de la secuencia y de este modo, modula la transcripción o induce la escisión en las proximidades de la secuencia diana. Tal como se expone en otras partes de la presente descripción, un dominio de dedo de cinc puede modificarse para que se una a prácticamente cualquier secuencia deseada. Por consiguiente, después de identificar una región de interés que contiene una secuencia en la que se desea la regulación, escisión o recombinación génica, pueden modificarse dos o más dominios de unión de dedo de cinc para que se unan a una o más secuencias en la región de interés.

La selección de una región genómica de EPSPS de interés en la cromatina celular para su unión por parte de un dominio de dedo de cinc (por ejemplo, un sitio diana) puede lograrse, por ejemplo, según los métodos descritos en la Patente de Estados Unidos n.º 6.453.242 (17 de septiembre de 2002) del mismo propietario que la presente, que también describe métodos para diseñar ZFP para que se unan a una secuencia seleccionada. Será evidente para los expertos en la materia que también puede usarse una simple inspección visual de una secuencia nucleotídica para la selección de un sitio diana. Por consiguiente, puede usarse cualquier medio para la selección del sitio diana en los métodos reivindicados.

Los sitios diana están compuestos generalmente de una diversidad de subsitios diana adyacentes. Un subsitio diana se refiere a la secuencia (normalmente un triplete de nucleótidos o un cuadruplete de nucleótidos que pueden solaparse con un cuadruplete adyacente) unido por un dedo de cinc individual. Véase, por ejemplo, el documento WO 02/077227. Si la hebra con la que la proteína de dedo de cinc efectúa la mayoría de contactos se denomina la "hebra de reconocimiento principal" o la "hebra de contacto primaria" de la hebra diana, algunas proteínas de dedo de cinc se unen a un triplete de tres bases en la hebra diana y a una cuarta base en la hebra no diana. Un sitio diana tiene generalmente una longitud de al menos 9 nucleótidos y, por consiguiente, se une a un dominio de unión de dedo de cinc que comprende al menos tres dedos de cinc. Sin embargo, también es posible la unión, por ejemplo, de un dominio de unión de 4 dedos a un sitio diana de 12 nucleótidos, de un dominio de unión de 5 dedos a un sitio diana de 15 nucleótidos o de un dominio de unión de 6 dedos a un sitio diana de 18 nucleótidos. Tal como será evidente, también es posible la unión de dominios de unión mayores (por ejemplo, de 7, 8, 9 y más dedos) a sitios diana más largos.

No es necesario que un sitio diana sea un múltiplo de tres nucleótidos. Por ejemplo, en los casos donde se producen interacciones cruzadas entre hebras (véanse, por ejemplo, la Patente de Estados Unidos 6.453.242 y el documento WO 02/077227), uno o más de los dedos de cinc individuales de un dominio de unión de múltiples dedos pueden unirse a cuadrupletes de sub-sitios solapantes. Como resultado, una proteína de tres dedos puede unirse a una secuencia de 10 nucleótidos, en donde el décimo nucleótido forma parte de un cuadruplete unido por un dedo terminal, una proteína de cuatro dedos puede unirse a una secuencia de 13 nucleótidos, en donde el decimotercer nucleótido forma parte de un cuadruplete unido por un dedo terminal, etc.

La longitud y la naturaleza de las secuencias enlazadoras de aminoácidos entre dedos de cinc individuales en un dominio de unión de múltiples dedos también afecta a la unión a una secuencia diana. Por ejemplo, la presencia de lo que se denomina un "enlazador no canónico", "enlazador largo" o "enlazador estructurado" entre los dedos de cinc adyacentes en un dominio de unión de múltiples dedos puede permitir que esos dedos se unan a subsitios que no se encuentran inmediatamente adyacentes. Los ejemplos de dichos enlazadores se describen, por ejemplo, en la Patente de Estados Unidos n.º 6.479.626 y el documento WO 01/53480. Por consiguiente, pueden separarse entre sí uno o más subsitios en un sitio diana para un dominio de unión de dedo de cinc por 1, 2, 3, 4, 5 o más nucleótidos. Para proporcionar solo un ejemplo, puede unirse un dominio de unión de cuatro dedos a un sitio diana de 13 nucleótidos que comprende, en secuencia, dos subsitios contiguos de 3 nucleótidos, un nucleótido intermedio y dos subsitios de triplete contiguos. Véase también la Solicitud de Estados Unidos n.º 61/130.099 para composiciones y métodos para enlazar nucleasas artificiales para que se unan a sitios diana separados por

diferentes números de nucleótidos. La distancia entre secuencias (por ejemplo, sitios diana) se refiere al número de nucleótidos o de pares de nucleótidos intermedios entre las dos secuencias, medidos a partir de los extremos de las secuencias más próximas entre sí.

5 En determinadas realizaciones, se diseñan ZFP con función de factor de transcripción. Para la función de factor de transcripción, generalmente lo único que se necesita es una unión sencilla y una proximidad suficiente al promotor. El posicionamiento exacto en relación al promotor, la orientación y dentro de unos límites la distancia no tienen gran importancia. Esta característica permite una flexibilidad considerable a la hora de seleccionar sitios diana para construir factores de transcripción artificiales. Por lo tanto, el sitio diana reconocido por la ZFP puede ser cualquier sitio adecuado en el gen diana que permitirá la activación o la represión de la expresión génica por una ZFP, opcionalmente unida a un dominio regulador. Los sitios diana preferidos incluyen regiones adyacentes a, cadena abajo o cadena arriba del sitio de inicio de la transcripción. Además, sitios diana que se localizan en regiones potenciadoras, sitios represores, sitios de pausa de ARN polimerasa y sitios regulatorios específicos (por ejemplo, sitios SP-1, elementos de respuesta a hipoxia, elementos de reconocimiento de receptor nuclear, sitios de unión de p53), sitios en el ADNc que codifican una región o en una región codificante de marcador de secuencia expresada (EST).

15 En otras realizaciones, se diseñan ZFP con actividad de nucleasa. La expresión de una ZFN que comprende una proteína de fusión que comprende un dominio de unión de dedo de cinc y un dominio de escisión (o de dos proteínas, comprendiendo cada una un dominio de dedo de cinc y un semidominio de escisión), en una célula, efectúa la escisión en la proximidad de la secuencia diana. En determinadas realizaciones, la escisión depende de la unión de dos moléculas de dominio de dedo de cinc/semidominio de escisión a sitios diana separados. Los dos sitios diana pueden encontrarse en hebras de ADN opuestas o, como alternativa, ambos sitios diana pueden encontrarse en la misma hebra de ADN.

#### **Dominios de unión de dedo de cinc**

25 Un dominio de unión de dedo de cinc comprende uno o más dedos de cinc. Miller et al. (1985) EMBO J. 4:1609-1614; Rhodes (1993) Scientific American Feb.:56-65; Patente de Estados Unidos n.º 6.453.242. Típicamente, un dominio de dedo de cinc tiene aproximadamente 30 aminoácidos de longitud. Los estudios estructurales han demostrado que cada dominio (motivo) de dedo de cinc contiene dos beta láminas (que se mantienen en un giro beta que contiene los dos restos de cisteína invariantes) y una alfa hélice (que contiene los dos restos de histidina invariantes), que se mantienen en una conformación particular mediante la coordinación de un átomo de cinc por las dos cisteínas y las dos histidinas.

30 Los dedos de cinc incluyen tanto dedos de cinc canónicos de  $C_2H_2$  (es decir, aquellos en los que el ión cinc se coordina por dos restos de cisteína y dos restos de histidina) y dedos de cinc no canónicos, tales como, por ejemplo, dedos de cinc de  $C_3H$  (aquellos en los que el ión cinc se coordina por tres restos de cisteína y un resto de histidina) y dedos de cinc de  $C_4$  (aquellos en los que el ión cinc se coordina por cuatro restos de cisteína). Véase también el documento WO 02/057293.

35 Los dominios de unión de dedo de cinc pueden modificarse para que se unan a una secuencia de elección. Véase, por ejemplo, Beerli et al. (2002) Nature Biotechnol. 20:135-141; Pabo et al. (2001) Ann. Rev. Biochem. 70:313-340; Isalan et al. (2001) Nature Biotechnol. 19:656-660; Segal et al. (2001) Curr. Opin. Biotechnol. 12:632-637; Choo et al. (2000) Curr. Opin. Struct. Biol. 10:411-416. Un dominio de unión de dedo de cinc modificado puede tener una nueva especificidad de unión en comparación con una proteína de dedo de cinc de origen natural. Los métodos de modificación incluyen el diseño racional y diversos tipos de selección. El diseño racional incluye, por ejemplo, el uso de bases de datos que incluyen secuencias de tripletes (o cuadrupletes) de nucleótidos y secuencias de aminoácidos de dedo de cinc individuales, en las que cada secuencia de triplete o cuadruplete de nucleótidos se asocia con una o más secuencias de aminoácidos de dedos de cinc que se unen a la secuencia de triplete o cuadruplete particular. Véanse, por ejemplo, las Patentes de Estados Unidos 6.453.242 y 6.534.261 del mismo propietario que la presente. Se divulgan métodos de diseño adicionales, por ejemplo, en las Patentes de Estados Unidos 6.746.838; 6.785.613; 6.866.997; y 7.030.215.

40 Los métodos de selección ejemplares, incluyendo presentación en fagos y sistemas de dos híbridos, se describen en las Patentes de Estados Unidos 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.410.248; 6.140.466; 6.200.759; y 6.242.568; así como los documentos WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y GB 2.338.237.

45 La potenciación de la especificidad de unión para los dominios de unión de dedo de cinc se ha descrito, por ejemplo, en la Patente de Estados Unidos n.º 6.794.136 del mismo propietario que la presente.

50 Ya que un dedo de cinc individual se une a una secuencia de tres nucleótidos (es decir, un triplete) (o a una secuencia de cuatro nucleótidos que puede solaparse, en un nucleótido, con el sitio de unión a cuatro nucleótidos de un dedo de cinc adyacente), la longitud de una secuencia para cuya unión se modifica un dominio de dedo de cinc (por ejemplo, una secuencia diana) determinará el número de dedos de cinc en un dominio de unión de dedo de cinc modificado. Por ejemplo, para las ZFP en las que los motivos de dedo no se unen a subsitios solapantes, una secuencia diana de seis nucleótidos se une a un dominio de unión de dos dedos; una secuencia diana de nueve

nucleótidos se une a un dominio de unión de tres dedos, etc. Tal como se indica en la presente memoria, los sitios de unión para dedos de cinc individuales (es decir, subsitios) en un sitio diana no se encuentran necesariamente contiguos, pero pueden separarse por uno o varios nucleótidos, dependiendo de la longitud y la naturaleza de las secuencias de aminoácidos entre los dedos de cinc (es decir, los enlazadores entre dedos) en un dominio de unión de múltiples dedos.

En un dominio de unión de dedo de cinc de múltiples dedos, los dedos de cinc adyacentes pueden estar separados por secuencias enlazadoras de aminoácidos de aproximadamente 5 aminoácidos (denominados enlazadores entre dedos "canónicos") o, como alternativa, por uno o más enlazadores no canónicos. Véanse, por ejemplo, las Patentes de Estados Unidos n.º 6.453.242 y 6.534.261 del mismo propietario que la presente. Para los dominios de unión de dedo de cinc modificados que comprenden más de tres dedos, puede preferirse la inserción de enlazadores entre dedos más largos ("no canónicos") ya que pueden aumentar la afinidad y/o especificidad de la unión por el dominio de unión. Véanse, por ejemplo, la Patente de Estados Unidos n.º 6.479.626 y el documento WO 01/53480. Por consiguiente, los dominios de unión de dedo de cinc de múltiples dedos también pueden caracterizarse respecto de la presencia y localización de enlazadores entre dedos no canónicos. Por ejemplo, un dominio de unión de dedo de cinc de seis dedos que comprende tres dedos (unidos por dos enlazadores entre dedos canónicos), un enlazador canónico y tres dedos (unidos por dos enlazadores entre dedos canónicos) se indica como una configuración 2x3. De manera similar, un dominio de unión que comprende dos dedos (con un enlazador canónico entre los mismos), un enlazador largo y dos dedos adicionales (unidos por un enlazador canónico) se indica como una proteína 2x2. Una proteína que comprende tres unidades de dos dedos (en cada una de las cuales los dos dedos están unidos por un enlazador canónico) y en la que cada unidad de dos dedos está unida a la unidad adyacente de dos dedos por un enlazador largo, se cita como una proteína 3x2.

La presencia de un enlazador entre dedos largo o no canónico entre dos dedos de cinc adyacentes en un dominio de unión de múltiples dedos a menudo permite que los dos dedos se unan a subsitios que no se encuentran inmediatamente contiguos en la secuencia diana. Por consiguiente, puede haber huecos de uno o más nucleótidos entre los subsitios en un sitio diana; es decir, un sitio diana puede contener uno o más nucleótidos que no entran en contacto con un dedo de cinc. Por ejemplo, un dominio de unión de dedo de cinc de 2x2 puede unirse a dos secuencias de seis nucleótidos separadas por un nucleótido, es decir, se une a un sitio diana de 13 nucleótidos. Véase también Moore et al. (2001a) Proc. Natl. Acad. Sci. USA 98:1432-1436; Moore et al. (2001b) Proc. Natl. Acad. Sci. USA 98:1437-1441 y el documento WO 01/53480.

Tal como se ha mencionado anteriormente, un subsitio diana es una secuencia de tres o cuatro nucleótidos que se une a un solo dedo de cinc. Para determinados fines, una unidad de dos dedos se denomina como un módulo de unión. Puede obtenerse un módulo de unión, por ejemplo, seleccionando dos dedos adyacentes en el contexto de una proteína de múltiples dedos (generalmente tres dedos) que se unen a una secuencia diana de seis nucleótidos particular. Como alternativa, pueden construirse los módulos mediante el ensamblaje de los dedos de cinc individuales. Véanse también los documentos WO 98/53057 y WO 01/53480.

Puede diseñarse un dominio de dedo de cinc para que se una a una o más secuencias genómicas diana de EPSPS homólogas (por ejemplo, ortólogas o parálogas). Por ejemplo, puede diseñarse un dominio de unión de dedo de cinc para que se una específicamente a una secuencia diana de EPSPS única. Como alternativa, puede diseñarse un dominio de unión de dedo de cinc para que se una a múltiples secuencias genómicas de EPSPS ortólogas o parálogas.

En una realización, se describe en la presente memoria un dominio de dedo de cinc que comprende una secuencia de aminoácidos tal como se muestra en una sola fila de la tabla A. En otra realización, la divulgación proporciona un polinucleótido que codifica un dominio de unión de dedo de cinc, en donde el dominio de unión de dedo de cinc comprende una secuencia de aminoácidos como se muestra en una sola fila de la tabla A.

#### 45 **Dominios reguladores**

Las ZFP descritas en la presente memoria pueden asociarse opcionalmente con dominios reguladores para la modulación de la expresión génica. La ZFP puede asociarse covalentemente o no covalentemente con uno o más dominios reguladores, como alternativa a dos o más dominios reguladores, siendo los dos o más dominios dos copias del mismo dominio o dos dominios diferentes. Los dominios reguladores pueden unirse covalentemente a la ZFP, por ejemplo, mediante un aminoácido enlazador, como parte de una proteína de fusión. Las ZFP también pueden asociarse con un dominio regulador mediante un dominio de dimerización no covalente, por ejemplo, una cremallera de leucina, un dominio N-terminal de la proteína STAT o una proteína de unión FK506 (véase, por ejemplo, O'Shea, Science 254: 539 (1991), Barahmand-Pour et al, Curr. Top. Microbiol. Immunol. 211:121-128 (1996); Klemm et al., Annu. Rev. Immunol. 16:569-592 (1998); Klemm et al., Annu. Rev. Immunol. 16:569-592 (1998); Ho et al., Nature 382:822-826 (1996); y Pomeranz et al., Biochem. 37:965 (1998)). El dominio regulador puede asociarse con la ZFP en cualquier posición adecuada, incluyendo el extremo C o N-terminal de la ZFP.

Los dominios reguladores comunes para su adición a la ZFP incluye, por ejemplo, dominios efectores de factores de transcripción (activadores, represores, co-activadores, co-represores), silenciadores, receptores de hormonas nucleares, factores de transcripción de oncogenes (por ejemplo, miembros de las familias de myc, jun, fos, myb,

max, mad, rel, ets, bcl, myb, mos, etc.); enzimas de reparación de ADN y sus factores y modificadores asociados; enzimas de reordenamiento de ADN y sus factores y modificadores asociados; proteínas asociadas a cromatina y sus modificadores (por ejemplo, cinasas, acetilasas y desacetilasas); y enzimas modificadoras de ADN (por ejemplo, metiltransferasas, topoisomerasas, helicasas, ligasas, cinasas, fosfatasas, polimerasas, endonucleasas) y sus factores y modificadores asociados.

Los polipéptidos de factor de transcripción a partir de los cuales se puede obtener un dominio regulador incluyen aquellos implicados en la transcripción regulada y basal. Dichos polipéptidos incluyen factores de transcripción, sus dominios efectores, coactivadores, silenciadores, receptores de hormonas nucleares (véase, por ejemplo, Goodrich et al., *Cell* 84:825-30 (1996) para una revisión de proteínas y elementos de ácido nucleico implicados en la transcripción; los factores de transcripción en general se revisan en Barnes y Adcock, *Clin. Exp. Allergy* 25 Suppl. 2:46-9 (1995) y Roeder, *Methods Enzymol.* 273:165-71(1996)). Se conocen bases de datos dedicadas a factores de transcripción (véase, por ejemplo, *Science* 269:630 (1995)). Los factores de transcripción de receptores de hormonas nucleares se describen en, por ejemplo, Rosen et al., *J. Med. Chem.* 38:4855-74 (1995). La familia de factores de transcripción C/EBP se revisa en Wedel et al., *Immunobiology* 193:171-85 (1995). Los coactivadores y co-represores que median la regulación de la transcripción por receptores de hormonas nucleares se revisan en, por ejemplo, Meier, *Eur. J. Endocrinol.* 134(2): 158-9 (1996); Kaiser et al., *Trends Biochem. Sci.* 21:342-5 (1996); y Uitley et al., *Nature* 394:498-502 (1998)). Los factores de transcripción GATA, que están implicados en la regulación de la hematopoyesis, se describen en, por ejemplo, Simon, *Nat. Genet.* 11:9-11 (1995); Weiss et al., *Exp. Hematol.* 23:99-107. La proteína de unión a la caja de TATA (TBP) y sus polipéptidos TAP asociados (que incluyen TAF30, TAF55, TAF80, TAF 10, TAFI 50 y TAF250) se describen en Goodrich y Tjian, *Curr. Opin. Cell Biol.* 6:403-9 (1994) y Hurley, *Curr. Opin. Struct. Biol.* 6:69-75 (1996). La familia de factores de transcripción STAT se revisa en, por ejemplo, Barahmand-Pour et al., *Curr. Top. Microbiol. Immunol.* 211:121-8 (1996). Los factores de transcripción implicados en enfermedades se revisan en Aso et al., *J. Clin. Invest.* 97:1561-9 (1996).

En una realización, se usa el dominio de represión KRAB de la proteína KOX-1 humana como represor transcripcional (Thiesen et al., *New Biologist* 2:363-374 (1990); Margolin et al., *PNAS* 91:4509-4513 (1994); Pengue et al., *Nucl. Acids Res.* 22:2908-2914 (1994); Witzgall et al., *PNAS* 91:4514-4518 (1994)). En otra realización, se usa KAP-1, un co-represor de KRAB, con KRAB (Friedman et al., *Genes Dev.* 10:2067-2078 (1996)). Como alternativa, puede usarse KAP-1 solo con una ZFP. Otros factores de transcripción y dominios de factor de transcripción que actúan como represores transcripcionales incluyen MAD (véase, por ejemplo, Sommer et al., *J. Biol. Chem.* 273:6632-6642 (1998); Gupta et al., *Oncogene* 16:1149-1159 (1998); Queva et al., *Oncogene* 16:967-977 (1998); Larsson et al., *Oncogene* 15:737-748 (1997); Laherty et al., *Cell* 89:349-356 (1997); y Cultraro et al., *Mol. Cell. Biol.* 17:2353-2359 (1997)); FKHR (forkhead en el gen de rabdosarcoma; Ginsberg et al., *Cancer Res.* 15:3542-3546 (1998); Epstein et al., *Mol. Cell. Biol.* 18:4118-4130 (1998)); EGR-1 (producto génico 1 de respuesta a crecimiento temprano; Yan et al., *PNAS* 95:8298-8303 (1998); y Liu et al., *Cancer Gene Ther.* 5:3-28 (1998)); el dominio represor del factor represor ets2 (ERD; Sgouras et al., *EMBO J* 14:4781-4793 ((1995)); y el dominio de interacción smSIN3 de MAD (SID; Ayer et al., *Mol. Cell. Biol.* 16:5772-5781 (1996)).

En una realización, se usa el dominio de activación VP16 de HSV como activador transcripcional (véase, por ejemplo, Hagmann et al., *J. Virol.* 71:5952-5962 (1997)). Otros factores de transcripción que podrían proporcionar dominios de activación incluyen el dominio de activación de VP64 (Seipel et al., *EMBO J.* 11:4961-4968 (1996)); receptores de hormonas nucleares (véase, por ejemplo, Torchia et al., *Curr. Opin. Cell. Biol.* 10:373-383 (1998)); la subunidad p65 del factor nuclear kappa B (Bitko y Barik, *J. Virol.* 72:5610-5618 (1998) y Doyle y Hunt, *Neuroreport* 8:2937-2942 (1997)); y EGR-1 (producto génico 1 de respuesta a crecimiento temprano; Yan et al., *PNAS* 95:8298-8303 (1998); y Liu et al., *Cancer Gene Ther.* 5:3-28 (1998)).

También son útiles las cinasas, fosfatasas y otras proteínas que modifican a péptidos implicados en la regulación génica como dominios reguladores para las ZFP. A menudo dichos modificadores están implicados en la activación o desactivación de la transcripción mediada por, por ejemplo, hormonas. Las cinasas implicadas en la regulación transcripcional se revisan en Davis, *Mol. Reprod. Dev.* 42:459-67 (1995), Jackson et al., *Adv. Second Messenger Phosphoprotein Res.* 28:279-86 (1993) y Boulikas, *Crit. Rev. Eukaryot. Gene Expr.* 5:1-77 (1995), mientras que las fosfatasas se revisan en, por ejemplo, Schonthal y Semin, *Cancer Biol.* 6:239-48 (1995). Las tirosina cinasas nucleares se describen en Wang, *Trends Biochem. Sci.* 19:373-6 (1994).

Tal como se ha descrito, también pueden obtenerse dominios útiles a partir de productos génicos de oncogenes (por ejemplo, los miembros de la familia myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos) y sus factores y modificadores asociados. Los oncogenes se describen en, por ejemplo, Cooper, *Oncogenes*, 2<sup>a</sup> ed., The Jones and Bartlett Series in Biology, Boston, Mass., Jones and Bartlett Publishers, 1995. Los factores de transcripción ets se revisan en Waslyk et al., *Eur. J. Biochem.* 211:7-18 (1993) y Crepieux et al., *Crit. Rev. Oncog.* 5:615-38 (1994). Los oncogenes Myc se revisan en, por ejemplo, Ryan et al., *Biochem. J.* 314:713-21 (1996). Los factores de transcripción jun y fos se describen en, por ejemplo, *The Fos and Jun Families of Transcription Factors*, Angel y Herrlich, eds. (1994). El oncogén max se revisa en Hurlin et al., *Cold Spring Harb. Symp. Quant. Biol.* 59:109-16. La familia génica myb se revisa en Kanei-Ishii et al., *Curr. Top. Microbiol. Immunol.* 211:89-98 (1996). La familia de mos se revisa en Yew et al., *Curr. Opin. Genet. Dev.* 3:19-25 (1993).

Las ZFP pueden incluir dominios reguladores obtenidos a partir de enzimas de reparación de ADN y sus factores y modificadores asociados. Los sistemas de reparación de ADN se revisan en, por ejemplo, Vos, *Curr. Opin. Cell Biol.* 4:385-95 (1992); Sancar, *Ann. Rev. Genet.* 29:69-105 (1995); Lehmann, *Genet. Eng.* 17:1-19 (1995); y Wood, *Ann. Rev. Biochem.* 65:135-67 (1996). Las enzimas de reordenamiento de ADN y sus factores y modificadores asociados también pueden usarse como dominios reguladores (véase, por ejemplo, Gangloff et al., *Experientia* 50:261-9 (1994); Sadowski, *FASEB J.* 7:760-7 (1993)).

De manera similar, los dominios reguladores pueden proceder de enzimas modificadoras de ADN (por ejemplo, ADN metiltransferasas, topoisomerasas, helicasas, ligasas, cinasas, fosfatasa, polimerasas) y sus factores y modificadores asociados. Las helicasas se revisan en Matson et al., *Bioessays*, 16:13-22 (1994) y las metiltransferasas se describen en Cheng, *Curr. Opin. Struct. Biol.* 5:4-10 (1995). Las proteínas asociadas a cromatina y sus modificadores (por ejemplo, cinasas, acetilasas y desacetilasas), tales como histona desacetilasa (Wolffe, *Science* 272:371-2 (1996)) también son dominios útiles para la adición de la ZFP elegida. En una realización preferida, el dominio regulador es una ADN metil transferasa que actúa como represor transcripcional (véase, por ejemplo, Van den Wyngaert et al., *FEBS Lett.* 426:283-289 (1998); Flynn et al., *J. Mol. Biol.* 279:101-116 (1998); Okano et al., *Nucleic Acids Res.* 26:2536-2540 (1998); y Zardo y Caiafa, *J. Biol. Chem.* 273:16517-16520 (1998)). En otra realización preferida, se usan endonucleasas, tales como Fok1 como represores transcripcionales, que actúan mediante escisión génica (véanse, por ejemplo, los documentos WO95/09233; y PCT/US94/01201).

También pueden usarse factores que controlan la estructura, el movimiento y la localización de la cromatina y el ADN y sus factores y modificadores asociados; factores procedentes de microbios (por ejemplo, procariontes, eucariotes y virus) y factores que se asocian o los modifican para obtener proteínas quiméricas. En una realización, se usan recombinasas e integrasas como dominios reguladores. En una realización, se usa histona acetiltransferasa como activador transcripcional (véase, por ejemplo, Jin y Scotto, *Mol. Cell. Biol.* 18:4377-4384 (1998); Wolffe, *Science* 272:371-372 (1996); Taunton et al., *Science* 272:408-411 (1996); y Hassig et al., *PNAS* 95:3519-3524 (1998)). En otra realización, se usa histona desacetilasa como represor transcripcional (véase, por ejemplo, Jin y Scotto, *Mol. Cell. Biol.* 18:4377-4384 (1998); Syntichaki y Thireos, *J. Biol. Chem.* 273:24414-24419 (1998); Sakaguchi et al., *Genes Dev.* 12:2831-2841 (1998); y Martinez et al., *J. Biol. Chem.* 273:23781-23785 (1998)).

Pueden incluirse dominios enlazadores entre dominios polipeptídicos, por ejemplo, entre dos ZFP o entre una ZFP y un dominio regulador. Dichos enlazadores son típicamente secuencias polipeptídicas, tales como secuencias de poligly de entre aproximadamente 5 y 200 aminoácidos. Los enlazadores preferidos son típicamente subsecuencias de aminoácidos flexibles que se sintetizan como parte de una proteína de fusión recombinante. Véase, por ejemplo, la Patente de Estados Unidos n.º 6.534.261; Liu et al., *Proc. Nat. Acad. Sci. USA*, 95:5525-5530 (1997); Pomerantz et al., *Proc. Nat. Acad. Sci. USA* 92:9752-9756 (1995); Kim et al., *Proc. Nat. Acad. Sci. USA* 93:1156-1160 (1996).

Como alternativa, pueden diseñarse racionalmente enlazadores flexibles usando un programa informático capaz de modelar tanto los sitios de unión a ADN como los péptidos en sí (Desjarlais y Berg, *Proc. Nat. Acad. Sci. USA* 90:2256-2260 (1993), *Proc. Nat. Acad. Sci. USA* 91:11099-11103 (1994) o mediante métodos de presentación en fagos.

En otras realizaciones, se usa un enlazador químico para conectar secuencias de dominio producidas de manera sintética o recombinante. Dichos enlazadores flexibles son conocidos por los expertos en la materia. Por ejemplo, están disponibles enlazadores de poli(etilenglicol) a través de Shearwater Polymers, Inc. Huntsville, Ala. Estos enlazadores tienen opcionalmente enlaces amida, enlaces sulfhidrido o enlaces heterofuncionales. Además del enlace covalente de las ZFP a dominios reguladores, pueden usarse métodos no covalentes para producir moléculas con ZFP asociadas con dominios reguladores

### Dominios de escisión

La porción del dominio de escisión de las proteínas de fusión descritas en la presente memoria puede obtenerse a partir de cualquier endonucleasa o exonucleasa. Las endonucleasas a partir de las cuales puede proceder un dominio de escisión incluyen endonucleasas de restricción y endonucleasas *homing*. Véase, por ejemplo, 2002-2003 Catalogue, New England Biolabs, Beverly, MA; y Belfort et al. (1997) *Nucleic Acids Res.* 25:3379-3388. Se conocen enzimas adicionales que escinden al ADN (por ejemplo, la nucleasa S1; la nucleasa del frijol mungo; la DNasa I pancreática; la nucleasa mitocondrial; la endonucleasa HO de levadura; véase también Linn et al. (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993). Puede usarse una o más de estas enzimas (o fragmentos funcionales de las mismas) como fuente de dominios de escisión y de semidominios de escisión.

De manera similar, un semidominio de escisión (por ejemplo, proteínas de fusión que contienen un dominio de unión de dedo de cinc y un semidominio de escisión) puede proceder de cualquier nucleasa o parte de la misma, tal como se ha indicado anteriormente, que requiere de dimerización para su actividad de escisión. En general, se necesitan dos proteínas de fusión para la escisión si las proteínas de fusión comprenden semidominios de escisión. Como alternativa, puede usarse una sola proteína que comprenda dos semidominios de escisión. Los dos semidominios de escisión pueden proceder de la misma endonucleasa (o fragmentos funcionales de la misma) o cada semidominio de escisión puede proceder de una endonucleasa diferente (o fragmentos funcionales de la misma). Además, los sitios diana para las dos proteínas de fusión están dispuestos preferentemente, con respecto al otro, de tal forma

que la unión de las dos proteínas de fusión a sus sitios diana respectivos coloca a los semidominios de escisión en una orientación espacial respecto del otro que permite a los semidominios de escisión formar un dominio de escisión funcional, por ejemplo, mediante dimerización. Por lo tanto, en determinadas realizaciones, los extremos próximos de los sitios diana están separados por 5-8 nucleótidos o por 15-18 nucleótidos. Sin embargo, puede intervenir cualquier número intermedio de nucleótidos o de pares de nucleótidos entre dos sitios diana (por ejemplo, de 2 a 50 nucleótidos o más). En general, el punto de escisión se encuentra entre los sitios diana.

Las endonucleasas de restricción (enzimas de restricción) están presentes en muchas especies y son capaces de unirse a ADN de manera específica de secuencia (en un sitio de reconocimiento) y de escindir el ADN en o próximo al sitio de unión. Determinadas enzimas de restricción (por ejemplo, del tipo IIS) escinden al ADN en sitios que se eliminan del sitio de reconocimiento y tienen dominios de unión y de escisión separables. Por ejemplo, la enzima de tipo IIS *Fok I* cataliza la escisión bicatenaria del ADN, a 9 nucleótidos desde su sitio de reconocimiento en una hebra y a 13 nucleótidos desde su sitio de reconocimiento en la otra. Véanse, por ejemplo, las Patentes de Estados Unidos 5.356.802; 5.436.150 y 5.487.994; así como Li et al. (1992) Proc. Natl. Acad. Sci. USA 89:4275-4279; Li et al. (1993) Proc. Natl. Acad. Sci. USA 90:2764-2768; Kim et al. (1994a) Proc. Natl. Acad. Sci. USA 91:883-887; Kim et al. (1994b) J. Biol. Chem. 269:31,978-31,982. Por lo tanto, en una realización, las proteínas de fusión comprenden el dominio de escisión (o el semidominio de escisión) de al menos una enzima de restricción del tipo IIS y uno o más dominios de unión de dedo de cinc, que pueden estar o no modificados.

Una enzima de restricción de tipo IIS ejemplar, cuyo dominio de escisión es separable del dominio de unión, es *Fok I*. Esta enzima particular está activa en forma de dímero (Bitinaite et al. (1998) Proc. Natl. Acad. Sci. USA 95: 10.570-10.575). Por consiguiente, para los fines de la presente descripción, la porción de la enzima *Fok I* usada en las proteínas de fusión descritas se considera un semidominio de escisión. Por lo tanto, para la escisión bicatenaria dirigida y/o el reemplazo dirigido de secuencias celulares usando fusiones de dedo de cinc-*Fok I*, pueden usarse dos proteínas de fusión, comprendiendo cada una un semidominio de escisión de *FokI*, para reconstituir un dominio de escisión catalíticamente activo. Como alternativa, también puede usarse una sola molécula polipeptídica que contiene un dominio de unión de dedo de cinc y dos semidominios de escisión de *Fok I*. Los parámetros para la escisión dirigida y para la alteración de secuencias dirigida usando fusiones de dedo de cinc-*Fok I* se proporcionan en otras partes de la presente descripción.

Un dominio de escisión o un semidominio de escisión puede ser cualquier parte de una proteína que retenga la actividad de escisión o que retenga la capacidad para multimerizar (por ejemplo, dimerizar) para formar un dominio de escisión funcional.

Las enzimas de restricción de tipo IIS ejemplares se listan en la tabla 1. Las enzimas de restricción adicionales también contienen dominios de unión y de escisión separables y estas están contempladas por la presente descripción. Véase, por ejemplo, Roberts et al. (2003) Nucleic Acids Res. 31:418-420.

**Tabla 1: Algunas enzimas de restricción de tipo IIS**

Aar I	BsrB I	SspD5 I
Ace III	BsrD I	Sth132 I
Aci I	BstF5 I	Sts I
Alo I	Btr I	TspDT I
Bae I	Bts I	TspGW I
Bbr7 I	Cdi I	Tth111 II
Bbv I	CjeP I	UbaP I
Bbv II	Drd II	Bsa I
BbvC I	Eci I	BsmB I
Bcc I	Eco31 I	
Bce83 I	Eco57 I	
BceA I	Eco57M I	
Bcef I	Esp3 I	
Bcg I	Fau I	
BciV I	Fin I	

Bfi I	Fok I
Bin I	Gdi II
Bmg I	Gsu I
Bpu10 I	Hga I
BsaX I	Hin4 II
Bsb I	Hph I
BscA I	Ksp632 I
BscG I	Mbo II
BseR I	Mly I
BseY I	Mme I
Bsi I	Mnl I
Bsm I	Pfl1108 I
BsmA I	Ple I
BsmF I	Ppi I
Bsp24 I	Psr I
BspG I	RleA I
BspM I	Sap I
BspNC I	SfaN I
Bsr I	Sim I

**Proteínas de fusión de dedo de cinc**

Los métodos para el diseño y construcción de proteínas de fusión (y de polinucleótidos que codifican a las mismas) son conocidos para los expertos en la materia. Por ejemplo, se describen métodos para el diseño y construcción de proteínas de fusión que comprenden dominios de dedo de cinc y dominios reguladores o de escisión (o semidominios de escisión) y polinucleótidos que codifican dichas proteínas de fusión en las Patentes de Estados Unidos 6.453.242 y 6.534.261 de cesión común y en las Publicaciones de Patente de Estados Unidos 2007/0134796 y 2005/0064474.

En determinadas realizaciones, se construyen polinucleótidos que codifican las proteínas de fusión. Estos polinucleótidos pueden insertarse en un vector y puede introducirse el vector en una célula (véase más adelante para una descripción adicional en referencia a vectores y métodos para introducir polinucleótidos en células).

En determinadas realizaciones de los métodos descritos en la presente memoria, una nucleasa de dedo de cinc comprende una proteína de fusión que comprende un dominio de unión de dedo de cinc y un semidominio de escisión de la enzima de restricción *Fok I* y dos de dichas proteínas de fusión se expresan en una célula. La expresión de dos proteínas de fusión en una célula puede ser el resultado del suministro de las dos proteínas a la célula; del suministro de una proteína y un ácido nucleico que codifica una de las proteínas a la célula; del suministro de dos ácidos nucleicos, codificando cada uno una de las proteínas, a la célula; o mediante el suministro de un solo ácido nucleico, que codifica ambas proteínas, a la célula. En realizaciones adicionales, una proteína de fusión comprende una sola cadena polipeptídica que comprende dos semidominios de escisión y un dominio de unión de dedo de cinc. En este caso, se expresa una sola proteína de fusión en una célula y sin desear quedar ligados a la teoría, se cree que escinde el ADN como resultado de la formación de un dímero intramolecular de los semidominios de escisión.

En determinadas realizaciones, los componentes de las proteínas de fusión (por ejemplo, fusiones de ZFP-*Fok I*) están dispuestos de tal modo que el dominio de dedo de cinc se encuentra más próximo al extremo amino terminal de la proteína de fusión y el semidominio de escisión se encuentra más próximo al extremo carboxi terminal. Esto imita a la orientación relativa del dominio de escisión en los dominios de escisión de dimerización de origen natural, tales como aquellos procedentes de la enzima *Fok I*, en la que el dominio de unión a ADN está más próximo al extremo amino terminal y el semidominio de escisión está más próximo al extremo carboxi terminal. En estas realizaciones, se provoca la dimerización de los semidominios de escisión para formar una nucleasa funcional

mediante la unión de las proteínas de fusión a sitios en hebras de ADN opuestas, encontrándose los extremos 5' de los sitios de unión próximos entre sí.

5 En realizaciones adicionales, los componentes de las proteínas de fusión (por ejemplo, fusiones de ZFP-*Fok I*) están dispuestos de tal modo que el semidominio de unión se encuentra más próximo al extremo amino terminal de la proteína de fusión y el dominio de dedo de cinc se encuentra más próximo al extremo carboxi terminal. En estas realizaciones, se provoca la dimerización de los semidominios de escisión para formar una nucleasa funcional mediante la unión de las proteínas de fusión a sitios en hebras de ADN opuestas, encontrándose los extremos 3' de los sitios de unión próximos entre sí.

10 En otras realizaciones adicionales, una primera proteína de fusión contiene el semidominio de escisión más próximo al extremo amino terminal de la proteína de fusión y el dominio de dedo de cinc más próximo al extremo carboxi terminal y una segunda proteína de fusión está dispuesta de tal modo que el dominio de dedo de cinc está más próximo al extremo amino terminal de la proteína de fusión y el semidominio de escisión está más próximo al extremo carboxi terminal. En estas realizaciones, ambas proteínas de fusión se unen a la misma hebra de ADN, encontrándose el sitio de unión de la primera proteína de fusión que contiene el dominio de dedo de cinc más próxima al extremo carboxi terminal situado en el lado 5' del sitio de unión de la segunda proteína de fusión que contiene el dominio de dedo de cinc más próximo al extremo amino terminal.

15 En determinadas realizaciones de las proteínas de fusión, la secuencia de aminoácidos entre el dominio de dedo de cinc y el dominio de escisión (o el semidominio de escisión) se denomina "enlazador ZC". El enlazador ZC debe distinguirse de los enlazadores entre los dedos discutidos anteriormente. Véanse, por ejemplo, las Publicaciones de Patente de Estados Unidos 20050064474A1 y 20030232410 y la Publicación Internacional de Patente WO05/084190, para detalles acerca de la obtención de enlazadores ZC que optimizan la escisión.

20 En una realización, la descripción proporciona una ZFN que comprende una proteína de dedo de cinc que tiene las secuencias de aminoácido de hélice de reconocimiento mostradas en una sola fila de la tabla A. En otra realización, se proporciona en la presente memoria un vector de expresión de ZFP que comprende una secuencia nucleotídica que codifica una ZFP que tiene las hélices de reconocimiento mostradas en una sola fila de la tabla A.

### Regulación de la expresión génica

25 Pueden usarse una diversidad de ensayos para determinar si una ZFP modula la expresión génica. Puede evaluarse la actividad de una ZFP concreta usando una diversidad de ensayos inmunohistoquímicos con anticuerpos *in vitro* e *in vivo*, midiendo, *por ejemplo*, los niveles de proteína o de ARNm, los niveles de producto, la actividad enzimática; la activación o represión transcripcional de un gen indicador usando, por ejemplo, inmunoensayos (por ejemplo, ELISA y ensayos inmunohistoquímicos con anticuerpos), ensayos de hibridación (por ejemplo, protección de RNasa, transferencias de Northern, hibridación *in situ*, estudios de matriz de oligonucleótidos), ensayos colorimétricos, ensayos de amplificación, ensayos de actividad enzimática, ensayos fenotípicos y similares.

30 Típicamente, las ZFP se ensayan en primer lugar respecto de su actividad *in vitro* usando ensayos ELISA y después usando células de riñón. La ZFP a menudo se ensaya en primer lugar usando un sistema de expresión transitoria con un gen indicador y después se ensaya la regulación del gen endógeno diana en células y en plantas completas, tanto *in vivo* como *ex vivo*. La ZFP puede expresarse de manera recombinante en una célula, expresarse de manera recombinante en células trasplantadas a una planta o expresarse de manera recombinante en una planta transgénica, así como suministrarse en forma de una proteína a una planta o célula usando vehículos de suministro descritos más adelante. Las células pueden estar inmovilizadas, encontrarse en solución, inyectarse en una planta o ser de origen natural en una planta transgénica o no transgénica.

35 La modulación de la expresión génica se ensaya usando uno de los ensayos *in vitro* o *in vivo* descritos en la presente memoria. Las muestras o los ensayos se tratan con una ZFP y se comparan con muestras de control sin el compuesto de ensayo para examinar el alcance de la modulación. Para la regulación de la expresión génica endógena, la ZFP tiene típicamente una  $K_d$  de 200 nM o menos, más preferentemente de 100 nM o menos, más preferentemente de 50 nM, lo más preferentemente de 25 nM o menos.

40 Los efectos de las ZFP pueden medirse examinando cualquiera de los parámetros descritos anteriormente. Puede usarse cualquier cambio en la expresión génica, fenotípico o fisiológico para evaluar la influencia de una. Cuando se determinan las consecuencias funcionales usando células o plantas intactas, también se puede medir una diversidad de efectos, tales como el crecimiento de la planta, los cambios transcripcionales en marcadores genéticos conocidos y no caracterizados (por ejemplo, transferencias de Northern o estudios de matrices de oligonucleótidos), cambios en el metabolismo celular, tales como cambios en el crecimiento celular o el pH y cambios en los segundos mensajeros intracelulares, tales como GMPc.

45 Los ensayos preferidos para la regulación de ZFP de la expresión génica endógena pueden llevarse a cabo *in vitro*. En un formato de ensayo *in vitro* preferido, se mide la regulación de ZFP de la expresión génica endógena en células cultivadas mediante el examen de la producción de proteínas usando un ensayo ELISA. Se compara la muestra de ensayo con células de control tratadas con un vector vacío o con una ZFP no relacionada que se dirige a otro gen.

En otra realización, se determina la regulación de ZFP de la expresión génica endógena *in vitro* midiendo el nivel de expresión de ARNm del gen diana. El nivel de expresión génica se mide usando amplificación, por ejemplo, usando PCR, LCR o ensayos de hibridación, por ejemplo, hibridación de Northern, protección de RNasa, transferencia por puntos. En una realización, se usa la protección de RNasa. Se detecta el nivel de proteína o de ARNm usando agentes de detección marcados directa o indirectamente, por ejemplo, ácidos nucleicos marcados fluorescente o radiactivamente, anticuerpos marcados radiactiva o enzimáticamente y similares, tal como se describe en la presente memoria.

Como alternativa, puede diseñarse un sistema de gen indicador usando el promotor del gen diana unido operablemente a un gen indicador, tal como luciferasa, proteína verde fluorescente, CAT o  $\beta$ -gal. La construcción indicadora se transfecta conjuntamente de manera típica en una célula cultivada. Después del tratamiento con la ZFP elegida, se mide la cantidad de transcripción, traducción o actividad del gen indicador según técnicas estándar conocidas por los expertos en la materia.

También se usan plantas transgénicas y no transgénicas como una realización preferida para examinar la regulación de la expresión génica endógena *in vivo*. Las plantas transgénicas pueden expresar de manera estable la ZFP elegida. Como alternativa, pueden usarse plantas que expresan de manera transitoria la ZFP elegida o a las que se ha suministrado la ZFP en un vehículo de suministro. Se ensaya la regulación de la expresión génica endógena usando uno cualquiera de los ensayos descritos en la presente memoria.

### Métodos para la escisión dirigida

Los métodos y composiciones descritos pueden usarse para escindir el ADN en una región de interés en la cromatina celular (por ejemplo, en un sitio deseado o predeterminado en un genoma dentro de o adyacente a un gen EPSPS, ya sea mutante o de tipo silvestre). Para dicha escisión dirigida de ADN, se modifica un dominio de unión de dedo de cinc para que se una a un sitio diana en o próximo al sitio de escisión predeterminado y se expresa en una célula una proteína de fusión que comprende el dominio de unión de dedo de cinc modificado y un dominio de escisión. Tras la unión de la parte de dedo de cinc de la proteína de fusión al sitio diana, se escinde el ADN próximo al sitio de unión mediante el dominio de escisión. El sitio de escisión exacto puede depender de la longitud del enlazador ZC.

Como alternativa, se expresan en una célula dos proteínas de fusión, comprendiendo cada una un dominio de unión de dedo de cinc y un semidominio de escisión y se unen a sitios diana que están yuxtapuestos de tal modo que se reconstituye un dominio de escisión funcional y el ADN se escinde en las proximidades de los sitios diana. En una realización, la escisión se produce entre los sitios diana de los dos dominios de unión de dedo de cinc. Uno o ambos dominios de unión de dedo de cinc pueden estar modificados.

Para la escisión dirigida usando un polipéptido de fusión de dominio de escisión de dedo de cinc-dominio de escisión, el sitio de unión puede abarcar el sitio de escisión o el extremo próximo del sitio de unión puede encontrarse a 1, 2, 3, 4, 5, 6, 10, 25, 50 o más nucleótidos (o cualquier número intermedio entre 1 y 50 nucleótidos) del sitio de escisión. La localización exacta del sitio de unión con respecto al sitio de escisión dependerá del dominio de escisión particular y de la longitud del enlazador ZC. Para los métodos en los que se usan dos polipéptidos de fusión, comprendiendo cada uno un dominio de dedo de cinc y un semidominio de escisión, los sitios de unión generalmente se encabalgan al sitio de escisión. Por lo tanto, el extremo próximo del primer sitio de unión puede encontrarse a 1, 2, 3, 4, 5, 6, 10, 25 o más nucleótidos (o a cualquier valor intermedio entre 1 y 50 nucleótidos) a un lado del sitio de escisión y el extremo próximo del segundo sitio de unión puede encontrarse a 1, 2, 3, 4, 5, 6, 10, 25 o más nucleótidos (o a cualquier valor intermedio entre 1 y 50 nucleótidos) al otro lado del sitio de escisión. Los expertos en la materia conocen métodos para mapear sitios de escisión *in vitro* e *in vivo*.

Por lo tanto, los métodos descritos en la presente memoria pueden usar un dominio de unión de dedo de cinc modificado fusionado a un dominio de escisión. En estos casos, se modifica el dominio de unión para que se una a una secuencia diana, en o próximo a donde se desea la escisión. La proteína de fusión o un polinucleótido que la codifica, se introduce en una célula vegetal. Una vez introducida o expresada en la célula, la proteína de fusión se une a la secuencia diana y escinde en o en la proximidad de la secuencia diana. El sitio de escisión exacto depende de la naturaleza del dominio de escisión y/o la presencia y/o naturaleza de las secuencias enlazadores entre los dominios de unión y de escisión. En los casos donde se usan dos proteínas de fusión, comprendiendo cada una un semidominio de escisión, la distancia entre los extremos próximos de los sitios de unión puede ser de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 25 o más nucleótidos (o a cualquier valor intermedio entre 1 y 50 nucleótidos). Los niveles óptimos de escisión también pueden depender tanto de la distancia entre los sitios de unión de las dos proteínas de fusión (véase, por ejemplo, Smith et al. (2000) *Nucleic Acids Res.* 28:3361-3369; Bibikova et al. (2001) *Mol. Cell. Biol.* 21:289-297) como de la longitud del enlazador ZC en cada proteína de fusión. Véanse también la Publicación de Patente de Estados Unidos 20050064474A1 y las Publicaciones Internacionales de Patente WO05/084190, WO05/014791 y WO03/080809.

En determinadas realizaciones, el dominio de escisión comprende dos semidominios de escisión, formando parte los dos de un solo polipéptido que comprende un dominio de escisión, un primer semidominio de escisión y un segundo semidominio de escisión. Los semidominios de escisión pueden tener la misma secuencia de aminoácidos o

diferentes secuencias de aminoácidos, en tanto que funcionen para escindir el ADN.

Los semidominios de escisión también pueden proporcionarse en moléculas separadas. Por ejemplo, pueden introducirse en una célula dos polipéptidos de fusión, en donde cada polipéptido comprende un dominio de unión y un semidominio de escisión. Los semidominios de escisión pueden tener la misma secuencia de aminoácidos o diferentes secuencias de aminoácidos, en tanto que funcionen para escindir el ADN. Además, los dominios de unión se unen a secuencias diana que están dispuestas típicamente de tal forma que, tras la unión de los polipéptidos de fusión, los dos semidominios de escisión se presentan en una orientación espacial respecto del otro que permite la reconstitución de un dominio de escisión (por ejemplo, mediante dimerización de los semidominios), de este modo posicionando a los semidominios en relación al otro para que se forme un dominio de escisión funcional, dando como resultado la escisión de la cromatina celular en una región de interés. En general, la escisión por parte del dominio de escisión reconstituido se produce en un sitio ubicado entre las dos secuencias diana. Puede modificarse una o las dos proteínas para que se unan a su sitio diana.

Las dos proteínas de fusión pueden unirse en la región de interés con la misma polaridad u opuesta y sus sitios de unión (es decir, sitios diana) pueden separarse por una serie de nucleótidos, por ejemplo, de 0 a 200 nucleótidos o a cualquier valor intermedio entre estos. En determinadas realizaciones, los sitios de unión para dos proteínas de fusión, comprendiendo cada una un dominio de unión de cinco y un semidominio de escisión, pueden localizarse separados por entre 5 y 18 nucleótidos, por ejemplo, separados por 5-8 nucleótidos o separados por 15-18 nucleótidos o separados por 6 nucleótidos o separados por 16 nucleótidos, medido desde el extremo de cada sitio de unión más próximo al otro sitio de unión y la escisión se produce entre los sitios de unión.

El sitio en el que se escinde el ADN se encuentra generalmente entre los sitios de unión para las dos proteínas de fusión. La escisión bicatenaria del ADN a menudo es el resultado de dos escisiones monocatenarias o "cortes", separados por 1, 2, 3, 4, 5, 6 o más nucleótidos, (por ejemplo, la escisión de ADN bicatenario por *Fok I* nativa es el resultado de roturas monocatenarias separadas por 4 nucleótidos). Por lo tanto, la escisión no se produce necesariamente en sitios exactamente opuestos en cada hebra de ADN. Además, la estructura de las proteínas de fusión y la distancia entre los sitios diana puede influenciar si la escisión tiene lugar adyacente a un solo par de nucleótidos o si la escisión se produce en varios sitios. Sin embargo, para muchas aplicaciones, incluyendo la recombinación dirigida y la mutagénesis dirigida (véase más adelante) la escisión dentro de un intervalo de nucleótidos es generalmente suficiente y no es necesaria la escisión entre un par de bases particular.

Tal como se ha indicado anteriormente, las proteínas de fusión pueden introducirse en forma de polipéptidos o de polinucleótidos. Por ejemplo, pueden introducirse en una célula dos polinucleótidos, comprendiendo cada uno secuencias que codifican uno de los polipéptidos anteriormente mencionados y cuando se expresan los polipéptidos y cada uno se une a su secuencia diana, se produce la escisión en o próxima a la secuencia diana. Como alternativa, se introduce en una célula un solo polinucleótido que comprende secuencias que codifican ambos polipéptidos de fusión. Los polinucleótidos pueden ser ADN, ARN o cualquier forma modificada o análogo de ADN y/o ARN.

Para potenciar la especificidad de escisión, también pueden emplearse composiciones adicionales en los métodos descritos en la presente memoria. Por ejemplo, los semidominios de escisión individuales pueden mostrar una actividad de escisión bicatenaria limitada. En los métodos en los que se introducen en la célula dos proteínas de fusión, conteniendo cada una un dominio de tres dedos de cinc y un semidominio de escisión, las dos proteínas especifican un sitio diana de 9 nucleótidos. Aunque es probable que la secuencia agregada de 18 nucleótidos sea única en un genoma de mamífero, cualquier sitio diana de 9 nucleótidos dado aparece, de media, aproximadamente 23.000 veces en el genoma humano. Por lo tanto, puede producirse la unión no específica, debida a la unión de sitio específico de un solo semidominio. Por consiguiente, los métodos descritos en la presente memoria contemplan el uso de un mutante negativo dominante de un semidominio de escisión tal como *Fok I* (o un ácido nucleico que codifica al mismo) que se expresa en una célula junto con las dos proteínas de fusión. El mutante negativo dominante es capaz de dimerizarse, pero es incapaz de escindir y también bloquea la actividad de escisión de un semidominio con el que se dimeriza. Al proporcionar el mutante negativo dominante en exceso molar respecto de las proteínas de fusión, solo las regiones en las que ambas proteínas de fusión están unidas tendrán una concentración local de semidominios de escisión funcionales suficiente para que se produzca la dimerización y la escisión. En los sitios donde solo una de las proteínas de fusión está unida, su semidominio de escisión forma un dímero con el semidominio mutante negativo funcional y no se produce la escisión inespecífica no deseada.

Se han identificado tres restos de aminoácidos catalíticos en el semidominio de escisión de *Fok I*: Asp 450, Asp 467 y Lys 469. Bitinaite et al. (1998) Proc. Natl. Acad. Sci. USA 95: 10.570-10.575. Por lo tanto, puede usarse una o más mutaciones en uno de estos restos para generar una mutación negativa dominante. Además, se conocen muchos de los restos de aminoácido catalíticos de otras endonucleasas de tipo IIS y/o pueden determinarse, por ejemplo, mediante el alineamiento con secuencias de *Fok I* y/o mediante la generación y ensayo de mutantes para actividad catalítica.

**Mutaciones del dominio de dimerización en el semidominio de escisión**

Los métodos para la escisión dirigida que implican el uso de fusiones entre una ZFP y un semidominio de escisión (tales como, por ejemplo, una fusión ZFP/*FokI*) requieren del uso de dos de dichas moléculas de fusión, cada una dirigida generalmente a una secuencia diana distinta. Las secuencias diana para las dos proteínas de fusión pueden seleccionarse de tal modo que la escisión dirigida se dirige a un sitio único en un genoma, tal como se ha discutido anteriormente. Una fuente potencial de especificidad de escisión reducida podría ser el resultado de la homodimerización de una de las dos fusiones de ZFP/semidominio de escisión. Esto podría suceder, por ejemplo, debido a la presencia, en un genoma, de repeticiones invertidas de las secuencias diana para una de las fos fusiones ZFP/semidominio de escisión, ubicadas de modo que permiten que dos copias de la misma fusión se unan con una orientación y un espaciado que permitan la formación de un dímero funcional.

Una estrategia para reducir la probabilidad de este tipo de escisión aberrante en secuencias distintas del sitio diana previsto implica la generación de variantes del semidominio de escisión que minimizan o previenen la homodimerización. Preferentemente, se alteran uno o más aminoácidos en la región del semidominio implicada en su dimerización. En la estructura cristalina del dímero de proteína de *FokI*, se comunica que la estructura de los semidominios de escisión es similar a la disposición de los semidominios de escisión durante la escisión del ADN por *FokI*. Wah et al. (1998) Proc. Natl. Acad. Sci. USA 95:10564-10569. Esta estructura indica que los restos de aminoácido en las posiciones 483 y 487 desempeñan un papel clave en la dimerización de los semidominios de escisión de *FokI*. La estructura también indica que los restos de aminoácidos en las posiciones 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537 y 538 están lo suficientemente próximos a la interfaz de dimerización como para influenciar la dimerización. Por consiguiente, es probable que las alteraciones de la secuencia de aminoácidos en una o más de las posiciones anteriormente mencionadas alteren las propiedades de dimerización del semidominio de escisión. Dichos cambios pueden introducirse, por ejemplo, construyendo una biblioteca que contiene (o que codifica) diferentes restos de aminoácidos en estas posiciones y seleccionando variantes con las propiedades deseadas o diseñando racionalmente mutantes individuales. Además de prevenir la homodimerización, también es posible que algunas de estas mutaciones puedan aumentar la eficacia de escisión más allá de la obtenida con dos semidominios de escisión de tipo silvestre.

Por consiguiente, la alteración de un semidominio de escisión de *FokI* en cualquier resto de aminoácido que afecta a la dimerización puede usarse para prevenir que un par de fusiones de ZFP/*FokI* sufran homodimerización que pueda dar lugar a la escisión en secuencias no deseadas. Por lo tanto, para la escisión dirigida usando un par de fusiones de ZFP/*FokI*, una o ambas proteínas de fusión pueden comprender una o más alteraciones de aminoácidos que inhiben la auto-dimerización pero que permiten que se produzca la heterodimerización de las dos proteínas de fusión, de tal forma que se produce la escisión en el sitio diana deseada. En determinadas realizaciones, las alteraciones están presentes en ambas proteínas de fusión y las alteraciones tienen efectos aditivos; es decir, se minimiza o suprime la homodimerización de cualquier fusión que dé lugar a una escisión aberrante, mientras que se facilita la heterodimerización de las dos proteínas de fusión en comparación con aquella obtenida con los semidominios de escisión de tipo silvestre.

**Métodos para la alteración dirigida de secuencias genómicas parálogos y recombinación dirigida**

También se describen en la presente memoria métodos para reemplazar una secuencia genómica de EPSPS, por ejemplo, de uno o más genes parálogos (por ejemplo, una región genómica diana de EPSPS en la cromatina celular) con una secuencia homóloga no idéntica (es decir, recombinación dirigida. Los intentos previos para reemplazar secuencias concretas han implicado poner en contacto una célula con un polinucleótido que comprende secuencias que portan homología con una región cromosómica (es decir, un ADN donante) seguido de la selección de células en las que la molécula de ADN donante ha sufrido recombinación de homólogos en el genoma. La tasa de éxito de estos métodos es baja debido a la baja eficacia de la recombinación de homólogos y a una alta frecuencia de inserción no específica del ADN donante en regiones del genoma distintas del sitio diana.

La presente descripción proporciona métodos para la alteración de secuencias dirigida caracterizada por una mayor eficacia de recombinación dirigida y una menor frecuencia de sucesos de inserción no específica. Los métodos implican la producción y uso de dominios de unión de dedo de cinc que se unen en o próximos a secuencias génicas de EPSPS parálogas, fusionados a dominios de escisión (o semidominios de escisión) para producir una o más roturas bicatenarias dirigidas en el ADN celular. Debido a que las roturas bicatenarias en el ADN celular estimulan los mecanismos de reparación celular varios cientos de veces en la proximidad del sitio de escisión, dicha escisión dirigida permite la alteración o reemplazo (mediante reparación dirigida por homología) de las secuencias génicas de EPSPS en prácticamente cualquier sitio en el genoma.

Los métodos descritos en la presente memoria son aplicables a cualquier secuencia génica paróloga de EPSPS de cualquier organismo o especie. En determinadas realizaciones, la región genómica de EPSPS diana que se altera pertenece a un gen EPSPS que comprende una secuencia seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia que tiene una identidad de secuencia de al menos aproximadamente el 80-100 % con las mismas, incluyendo cualquier porcentaje de identidad dentro de estos intervalos, tal como una identidad de secuencia del 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % con las mismas.

Además de las moléculas de fusión descritas en la presente memoria, el reemplazo dirigido de una secuencia genómica seleccionada también requiere de la introducción de la secuencia de reemplazo (o donante). La secuencia donante puede introducirse en una célula antes de, concurrentemente con o posteriormente a la expresión de la proteína (o las proteínas) de fusión. El polinucleótido donante contiene homología suficiente con una secuencia genómica de EPSPS como para soportar la recombinación de homólogos (o la reparación dirigida por homología) entre esta y la secuencia genómica de EPSPS con la que porta homología. Aproximadamente 25, 50, 100, 200, 500, 750, 1.000, 1.500, 2.000 nucleótidos o más de homología de secuencia entre una secuencia donante y una genómica (o cualquier valor intermedio entre 10 y 2.000 nucleótidos o más) soportarán la recombinación de homólogos entre los mismos. La longitud de las secuencias donante puede variar de 10 a 5.000 nucleótidos (o cualquier valor intermedio de nucleótidos entre los mismos) o mayor. Será fácilmente evidente que la secuencia donante típicamente no es idéntica a la secuencia genómica a la que reemplaza. Por ejemplo, la secuencia del polinucleótido donante puede contener uno o más cambios de una sola base, inserciones, eliminaciones, inversiones o reordenamientos con respecto a la secuencia genómica, en la medida en que esté presente una homología suficiente con las secuencias cromosómicas. Como alternativa, una secuencia donante puede contener una secuencia no homóloga flanqueada por dos regiones de homología. Además, las secuencias donantes pueden comprender una molécula de vector que contiene secuencias que no son homólogas a la región de interés en la cromatina celular. En general, las regiones de homólogas de una secuencia donante tendrán una identidad de secuencia de al menos el 50 % con una secuencia genómica con la que se desea la recombinación. En determinadas realizaciones, hay presente una identidad de secuencia del 60 %, 70 %, 80 %, 90 %, 95 %, 98 %, 99 % o 99,9 %. Puede estar presente cualquier valor de identidad de secuencia de entre 1 % y el 100 %, dependiendo de la longitud del polinucleótido donante.

Una molécula donante puede contener varias regiones de homología discontinuas para la cromatina celular. Por ejemplo, para la inserción dirigida de secuencias no presentes normalmente en una región de interés, dichas secuencias pueden estar presentes en una molécula de ácido nucleico donante y flanqueadas por regiones de homología para una secuencia génica en la región de interés.

Para simplificar los ensayos (por ejemplo, hibridación, PCR, digestión con enzima de restricción) para determinar la inserción exitosa de la secuencia donante, pueden estar presentes determinadas diferencias de secuencia en la secuencia donante en comparación con la secuencia genómica de EPSPS. Preferentemente, si se localiza en una región codificante, dichas diferencias de secuencia de nucleótidos no cambiarán la secuencia de aminoácidos o producirán cambios de aminoácido silenciosos (es decir, cambios que no afectan a la estructura o función de la proteína). El polinucleótido donante puede contener opcionalmente cambios en las secuencias correspondientes a los sitios de unión del dominio de dedo de cinc en la región de interés para prevenir la escisión de las secuencias donantes que se han introducido en la cromatina celular mediante recombinación de homólogos.

El polinucleótido donante puede ser ADN o ARN, monocatenario o bicatenario y puede introducirse en una célula en forma lineal o circular. Si se introducen en forma lineal, pueden protegerse los extremos de la secuencia donante (por ejemplo, de la degradación exonucleolítica) mediante métodos conocidos por los expertos en la materia. Por ejemplo, se añaden uno o más dideoxinucleótidos al extremo 3' de una molécula lineal y/o se ligan oligonucleótidos auto-complementarios a uno o ambos extremos. Véase, por ejemplo, Chang et al. (1987) Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889. Los métodos adicionales para proteger polinucleótidos exógenos de la degradación incluyen la adición de grupos amino terminales y el uso de engarces internucleótido tales como, por ejemplo, fosforotioatos, fosforamidatos y restos de O-metil ribosa o desoxirribosa.

Puede introducirse un polinucleótido en una célula como parte de una molécula de vector que tenga secuencias adicionales tales como, por ejemplo, orígenes de replicación, promotores y genes que codifican resistencia a antibióticos. Además, pueden introducirse los polinucleótidos donantes en forma de ácido nucleico desnudo, en forma de ácido nucleico complejado con un agente, tal como un lisosoma o poloxámero o pueden suministrarse mediante bacterias o virus (por ejemplo, *Agrobacterium*, *Rhizobium* sp. NGR234, *Sinorhizobium meliloti*, *Mesorhizobium loti*, virus del mosaico del tabaco, virus X de la patata, virus del mosaico de la coliflor y virus del mosaico de la vena de cassava. Véase, por ejemplo, Chung et al. (2006) Trends Plant Sci. 11(1):1-4.

Sin quedar ligados a una teoría, parece que la presencia de una rotura bicatenaria en una secuencia celular, acoplada a la presencia de una molécula de ADN exógena que tiene homología con una región adyacente a o que rodea a la rotura activa mecanismos celulares que reparan la rotura mediante la transferencia de información de secuencia desde la molécula donante a la secuencia celular (por ejemplo, genómica o cromosómica; es decir, mediante un proceso de reparación dirigida por homología, también conocido como "conversión génica". Los métodos de los solicitantes combinan ventajosamente las potentes capacidades de direccionamiento de las ZFP modificadas con un dominio de escisión (o semidominio de escisión) para dirigirse específicamente a genes parálogos, tales como genes de EPSPS de tal forma que la escisión de la secuencia diana produce una rotura bicatenaria en la región del genoma donde se desea la inserción de las secuencias exógenas.

Para la alteración de una secuencia cromosómica, no es necesario que se copie la secuencia completa del donante en el cromosoma, en tanto que se copie lo suficiente de la secuencia donante para efectuar la alteración de secuencia deseada.

La eficacia de la inserción de secuencias donantes mediante recombinación de homólogos es inversamente proporcional a la distancia, en el ADN celular, entre la rotura bicatenaria y el sitio donde se desea la recombinación. En otras palabras, se observan mayores eficacias de recombinación cuando la rotura bicatenaria está más próxima al sitio en el que se desea la recombinación. En los casos donde no se predetermina un sitio de recombinación preciso (es decir, el suceso de recombinación deseado puede producirse a lo largo de un intervalo de la secuencia genómica), se seleccionan la longitud y la secuencia del ácido nucleico donante, junto con el sitio (o los sitios) de escisión, para obtener el suceso de recombinación deseado. En los casos en los que se diseña el suceso deseado para cambiar la secuencia de un solo par de nucleótidos en una secuencia genómica, la cromatina celular se escinde a 10.000 nucleótidos desde cualquiera de los dos extremos de ese par de nucleótidos. En determinadas realizaciones, la escisión tiene lugar a 1.000, 500, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5 o 2 nucleótidos o cualquier número intermedio entre 2 y 1.000 nucleótidos, en cualquiera de los dos lados del par de nucleótidos cuya secuencia va a cambiarse.

Tal como se ha detallado anteriormente, los sitios de unión para dos proteínas de fusión, comprendiendo cada una un dominio de unión de dedo de cinc y un semidominio de escisión, pueden localizarse a 5-8 o a 15-18 nucleótidos de distancia, medidos desde el extremo de cada sitio de unión más próximo al otro sitio de unión y la escisión se produce entre los sitios de unión. Que la escisión se produzca en un solo sitio o en múltiples sitios dentro de los sitios de unión es algo inmaterial, ya que las secuencias genómicas escindidas se reemplazan por las secuencias donantes. Por lo tanto, para una alteración eficaz de la secuencia de un solo par de nucleótidos mediante recombinación dirigida, el punto medio de la región entre los sitios de unión se encuentra a 10.000 nucleótidos de distancia de ese par de nucleótidos, preferentemente a 1.000 nucleótidos o 500 nucleótidos o 200 nucleótidos o 100 nucleótidos o 50 nucleótidos o 20 nucleótidos o 10 nucleótidos o 5 nucleótidos o 2 nucleótidos o un nucleótido o en el par de nucleótidos de interés.

En determinadas realizaciones, un cromosoma homólogo puede servir de polinucleótido donante. Por lo tanto, por ejemplo, la corrección de una mutación en un heterocigoto puede lograrse modificando proteínas de fusión que se unen a y escinden la secuencia mutante en un cromosoma pero que no escinden a la secuencia de tipo silvestre en el cromosoma homólogo. La rotura bicatenaria en el cromosoma portador de la mutación estimula un proceso de "conversión génica" basada en homología en el que la secuencia de tipo silvestre del cromosoma homólogo se copia en el cromosoma escindido, restaurando de este modo dos copias de la secuencia de tipo silvestre.

También se proporcionan métodos y composiciones que pueden potenciar los niveles de recombinación dirigida que incluyen el uso de fusiones adicionales de ZFP-dominio funcional para activar la expresión de genes implicados en la recombinación de homólogos, tales como, por ejemplo, miembros del grupo de epistasias de RAD52 (por ejemplo, *Rad50*, *Rad51*, *Rad51B*, *Rad51C*, *Rad51D*, *Rad52*, *Rad54*, *Rad54B*, *Mre11*, *XRCC2*, *XRCC3*), genes cuyos productos interactúan con los productos génicos anteriormente mencionados (por ejemplo, *BRCA1*, *BRCA2*) y/o genes en el complejo de NBS1. Véase, por ejemplo, Boyko et al. (2006) *Plant Physiology* 141:488-497 y LaFarge et al. (2003) *Nucleic Acids Res* 31(4): 1148-1155. De manera similar, pueden usarse fusiones de ZFP-dominio funcional, en combinación con los métodos y composiciones descritos en la presente memoria, para reprimir la expresión de genes implicados en la unión de extremos no homólogos (por ejemplo, *Ku70/80*, *XRCC4*, poli(ADP ribosa) polimerasa, ADN ligasa 4). Véase, por ejemplo, Riha et al. (2002) *EMBO* 21:2819-2826; Freisner et al. (2003) *Plant J.* 34:427-440; Chen et al. (1994) *European Journal of Biochemistry* 224:135-142. Los métodos para la activación y represión de la expresión génica usando fusiones entre un dominio de unión de dedo de cinc y un dominio funcional se describen en, por ejemplo, las Patentes de Estados Unidos de cesión común 6.534.261; 6.824.978 y 6.933.113. Los métodos de represión adicionales incluyen el uso de oligonucleótidos antisentido y/o ARN pequeño interferente (ARNpi o iARN) o ARNhc dirigidos a la secuencia del gen que se va a reprimir.

Como alternativa a o además de activar la expresión de los productos génicos implicados en la recombinación de homólogos, la fusión de estas proteínas (o fragmentos funcionales de las mismas) con un dominio de unión de dedo de cinc dirigido a la región genómica de interés (por ejemplo, EPSPS), puede usarse para reclutar estas proteínas (proteínas de recombinación) a la región de interés, aumentando de este modo su concentración local y estimulando de este modo los procesos de recombinación de homólogos. Como alternativa, un polipéptido implicado en la recombinación de homólogos (o un fragmento funcional del mismo) como se ha descrito anteriormente puede formar parte de una proteína de fusión triple que comprende un dominio de unión de dedo de cinc, un dominio de escisión (o un semidominio de escisión) y la proteína de recombinación (o un fragmento funcional de la misma). Las proteínas adicionales implicadas en la conversión génica y en el remodelado de la cromatina relacionado con la recombinación que pueden usarse en los métodos y composiciones anteriormente mencionados, incluyen histona acetiltransferasas (por ejemplo, *Esalp*, *Tip60*), histona metiltransferasas (por ejemplo, *Dot1p*), histona cinasas e histona fosfatasa. Véase también Bhat et al. (1999) *Plant J.* 33:455-469

Se logran aumentos adicionales en la eficacia de la recombinación dirigida en células que comprenden una molécula de fusión de dedo de cinc/nucleasa y una molécula de ADN donante mediante el bloqueo de las células en la fase G<sub>2</sub> del ciclo celular, cuando los procesos de reparación dirigida por homología se encuentran en su activación máxima. Dicha detención puede lograrse mediante una serie de modos. Por ejemplo, puede tratarse a las células con, por ejemplo, fármacos, compuestos y/o moléculas pequeñas que influyen la progresión del ciclo celular para detener a las células en la fase G<sub>2</sub>. Las moléculas de este tipo ejemplares incluyen compuestos que afectan a la polimerización de microtúbulos (por ejemplo, vinblastina, nocodazol, Taxol), compuestos que interactúan con el ADN

(por ejemplo, dicloruro de *cis*-platino(II) diamina, Cisplatino, doxorubicina) y/o compuestos que afectan a la síntesis de ADN (por ejemplo, timidina, hidroxiurea, L-mimosina, etopósido, 5-fluorouracilo). Se logran aumentos adicionales en la eficacia de recombinación mediante el uso de inhibidores de histona desacetilasa (HDAC) (por ejemplo, butirato de sodio, trichostatina A) que alteran la estructura de la cromatina para producir ADN genómico más accesible a la maquinaria de recombinación celular.

Los métodos adicionales para la detención del ciclo celular incluyen la sobreexpresión de proteínas que inhiben la actividad de las cinasas CDK del ciclo celular, por ejemplo, introduciendo un ADNc que codifica la proteína en la célula o mediante la introducción en la célula de una ZFP modificada que activa la expresión del gen que codifica la proteína. La detención del ciclo celular también se logra inhibiendo la actividad de ciclinas y CDK, por ejemplo, usando métodos de iARN (por ejemplo, la Patente de Estados Unidos n.º 6.506.559) o introduciendo en la célula una ZFP modificada que reprime la expresión de uno o más genes implicados en la progresión del ciclo celular, tales como, por ejemplo, genes de ciclinas y/o CDK. Véase, por ejemplo, la Patente de Estados Unidos de cesión común n.º 6.534.261 para métodos para la síntesis de proteínas de dedo de cinc modificadas para la regulación de la expresión génica.

Como alternativa, en determinados casos, la escisión dirigida se lleva a cabo en ausencia de un polinucleótido donante (preferentemente en fase S o G<sub>2</sub>) y la recombinación se produce entre cromosomas homólogos.

### Vectores de expresión

Puede clonarse un ácido nucleico que codifica una o más ZFP en un vector para la transformación en células procariontas o eucariotas para su replicación y/o expresión. Los vectores pueden ser vectores procariontas, por ejemplo, plásmidos o vectores lanzadera, vectores de insecto o vectores eucariotas. También puede clonarse un ácido nucleico que codifica una ZFP en un vector de expresión para su administración a una célula vegetal.

Para expresar las ZFP, las secuencias que codifican las ZFP se subclonan típicamente en un vector de expresión que contiene un promotor para dirigir la transcripción. Los promotores bacterianos y eucariotas adecuados se conocen bien en la técnica y se describen, por ejemplo, en Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2ª ed. 1989; 3ª ed., 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); y *Current Protocols in Molecular Biology* (Ausubel et al., anteriormente citado. Los sistemas de expresión bacterianos para expresar la ZFP se encuentran disponibles en, por ejemplo, *E. coli*, *Bacillus sp.* y *Salmonella* (Palva et al., *Gene* 22:229-235 (1983)). Se encuentran comercialmente disponibles kits para dichos sistemas de expresión. Los sistemas de expresión eucariotas para células de mamífero, levaduras y células de insectos se conocen bien por los expertos en la materia y también están disponibles comercialmente.

El promotor usado para dirigir la expresión de un ácido nucleico codificante de una ZFP depende de la aplicación particular. Por ejemplo, se usa típicamente un promotor constitutivo fuerte adaptado a la célula hospedadora para la expresión y purificación de las ZFP.

Por el contrario, cuando se administra una ZFP *in vivo* para la regulación de un gen vegetal (véase, la sección de "Suministro de ácido nucleico a células vegetales" más adelante), se usa un promotor constitutivo o uno inducible, dependiendo del uso concreto de la ZFP. Los ejemplos de promotores vegetales incluyen secuencias promotoras procedentes de ubiquitina-3 (ubi-3) de *A. thaliana* (Callis, et al., 1990, *J. Biol. Chem.* 265:12486-12493); manopina sintasa ( $\Delta$ mas) de *A. tumifaciens* (Petolino et al., Patente de Estados Unidos n.º 6.730.824); y/o virus del mosaico de la vena de Cassava (CsVMV) (Verdaguer et al., 1996, *Plant Molecular Biology* 31:1129-1139). Véanse también los ejemplos.

Además del promotor, el vector de expresión contiene típicamente una unidad de transcripción o un casete de expresión que contiene todos los elementos adicionales necesarios para la expresión del ácido nucleico en las células hospedadoras, ya sean procariontas o eucariotas. Un casete de expresión típico contiene, por lo tanto, un promotor unido operablemente, por ejemplo, a una secuencia de ácido nucleico que codifica la ZFP y las señales necesarias, por ejemplo, para la poliadenilación eficaz del transcrito, la terminación transcripcional, los sitios de unión a ribosoma o la terminación de la traducción. Los elementos adicionales del casete pueden incluir, por ejemplo, potenciadores y señales de corte y empalme heterólogas.

El vector de expresión particular usado para transportar la información genética en la célula se selecciona respecto del uso previsto de la ZFP, por ejemplo, expresión en plantas, animales, bacterias, hongos, protozoos, etc. (véanse los vectores de expresión descritos más adelante). Los vectores de expresión bacteriana y animal estándar se conocen por los expertos en la materia y se describen en detalle, por ejemplo, en la Publicación de Patente de Estados Unidos 20050064474A1 y en las Publicaciones Internacionales de Patente WO05/084190, WO05/014791 y WO03/080809.

Pueden usarse métodos de transfección estándar para producir líneas celulares bacterianas, de mamífero, de levadura o de insecto que expresan grandes cantidades de proteína, que puede purificarse usando técnicas estándar (véase, por ejemplo, Colley et al., *J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification*, en *Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). La transformación de células eucariotas o procariontas se lleva a cabo según técnicas estándar (véase, por ejemplo, Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss y

Curtiss, *Methods in Enzymology* 101:347-362 (Wu et al., eds., 1983).

Puede usarse cualquiera de los procedimientos bien conocidos para introducir secuencias nucleotídicas exógenas en dichas células hospedadoras. Estos incluyen el uso de transfección por fosfato de calcio, polibreno, fusión de protoplastos, electroporación, métodos ultrasónicos (por ejemplo, sonoporación), liposomas, microinyección, ADN desnudo, vectores plasmídicos, vectores víricos, tanto episómicos como integrativos y cualquier otro de los métodos bien conocidos para introducir ADN genómico clonado, ADNc, ADN u otro material genético exógeno en una célula hospedadora (véase, por ejemplo, Sambrook *et al.*, anteriormente citado). Solo es necesario que el procedimiento de modificación genética empleado sea capaz de introducir con éxito al menos un gen en la célula hospedadora capaz de expresar la proteína elegida.

#### 10 **Suministro de ácido nucleico a células vegetales**

Como se ha indicado anteriormente, pueden introducirse construcciones de ADN en (por ejemplo, en el genoma de) un hospedador vegetal deseado mediante una diversidad de técnicas convencionales. Para revisiones de dichas técnicas véase, por ejemplo, Weissbach y Weissbach *Methods for Plant Molecular Biology* (1988, Academic Press, N.Y.) Sección VIII, págs. 421-463; y Grierson y Corey, *Plant Molecular Biology* (1988, 2ª Ed.), Blackie, Londres, Caps. 7-9.

Por ejemplo, la construcción de ADN puede introducirse directamente en el ADN genómico de la célula vegetal usando técnicas tales como electroporación y microinyección de protoplastos de células vegetales o pueden introducirse las construcciones de ADN directamente al tejido vegetal usando métodos biolísticos, tales como bombardeo con partículas de ADN (véase, por ejemplo, Klein et al (1987) *Nature* 327:70-73). Como alternativa, pueden combinarse las construcciones de ADN con regiones flanqueantes de ADN-T adecuadas e introducirse en un vector hospedador convencional de *Agrobacterium tumefaciens*. Las técnicas de transformación mediadas por *Agrobacterium tumefaciens*, incluyendo el desmontaje y uso de vectores binarios, está bien descrita en la bibliografía científica. Véase, por ejemplo, Horsch et al (1984) *Science* 233:496-498 y Fraley et al (1983) *Proc. Nat'l. Acad. Sci. USA* 80:4803.

Además, puede lograrse la transferencia génica usando bacterias no de *Agrobacterium* o virus, tales como *Rhizobium sp.* NGR234, *Sinorhizobium meliloti*, *Mesorhizobium loti*, virus X de la patata, virus del mosaico de la coliflor y virus del mosaico de la vena de la cassava y/o el virus del mosaico del tabaco. Véase, por ejemplo, Chung et al. (2006) *Trends Plant Sci.* 11(1):1-4.

Las funciones de virulencia del hospedador de *Agrobacterium tumefaciens* dirigirá la inserción de la construcción y el marcador adyacente en el ADN de la célula vegetal cuando se infecta la célula por la bacteria usando el vector de ADN T binario (Bevan (1984) *Nuc. Acid Res.* 12:8711-8721) o el procedimiento de cultivo conjunto (Horsch et al (1985) *Science* 227:1229-1231). En general, se usa el sistema de transformación de *Agrobacterium* para modificar plantas dicotiledóneas (Bevan et al (1982) *Ann. Rev. Genet* 16:357-384; Rogers et al (1986) *Methods Enzymol.* 118:627-641). El sistema de transformación de *Agrobacterium* también puede usarse para transformar, así como para transferir, ADN a plantas y células vegetales monocotiledóneas. Véase la Patente de Estados Unidos n.º 5.591.616; Hernalsteen et al (1984) *EMBO J*3:3039-3041; Hooykass-Van Slogteren et al (1984) *Nature* 311:763-764; Grimsley et al (1987) *Nature* 325:1677-179; Boulton et al (1989) *Plant Mol. Biol.* 12:31-40.; y Gould et al (1991) *Plant Physiol.* 95:426-434.

Los métodos de transferencia génica y transformación alternativos incluyen transformación de protoplastos mediante captación mediada por calcio, polietilenglicol (PEG) o electroporación de ADN desnudo (véase Paszkowski et al. (1984) *EMBO J*3:2717-2722, Potrykus et al. (1985) *Molec. Gen. Genet.* 199:169-177; Fromm et al. (1985) *Proc. Nat. Acad. Sci. USA* 82:5824-5828; y Shimamoto (1989) *Nature* 338:274-276) y electroporación de tejidos vegetales (D'Halluin et al. (1992) *Plant Cell* 4:1495-1505). Los métodos adicionales para la transformación de células vegetales incluyen microinyección, captación de ADN mediada por carburo de silicio (Kaepler et al. (1990) *Plant Cell Reporter* 9:415-418) y bombardeo con micropartículas (véase Klein et al. (1988) *Proc. Nat. Acad. Sci. USA* 85:4305-4309; y Gordon-Kamm et al. (1990) *Plant Cell* 2:603-618).

Los métodos y composiciones descritos pueden usarse para insertar secuencias exógenas en una ubicación predeterminada en un genoma de célula vegetal. Esto es útil en tanto que la expresión de un transgén introducido en un genoma vegetal depende de manera crítica de su sitio de integración. Por consiguiente, los genes que codifican, por ejemplo, nutrientes, antibióticos o moléculas terapéuticas pueden insertarse, mediante recombinación dirigida, en regiones de un genoma vegetal favorables para su expresión.

Las células vegetales transformadas que se producen mediante cualquiera de las técnicas de transformación anteriores pueden cultivarse para regenerar una planta completa que posee el genotipo transformado y por lo tanto el fenotipo deseado. Dichas técnicas de regeneración se basan en la manipulación de determinadas fitohormonas en un medio de cultivo tisular, basándose típicamente en un marcador de biocida y/o herbicida que se ha introducido junto con las secuencias nucleotídicas. La regeneración de plantas a partir de protoplastos cultivados se describe en Evans, et al., "Protoplasts Isolation and Culture" en *Handbook of Plant Cell Culture*, págs. 124-176, Macmillian Publishing Company, Nueva York, 1983; y Binding, *Regeneration of Plants, Plant Protoplasts*, págs. 21-73, CRC

Press, Boca Raton, 1985. La regeneración también puede obtenerse a partir de callos, explantes, órganos, pólenes o embriones vegetales o partes de los mismos. Dichas técnicas de regeneración se describen de manera general en Klee et al (1987) Ann. Rev. of Plant Phys. 38:467-486.

5 Los ácidos nucleicos introducidos en una célula vegetal pueden usarse para conferir rasgos deseados esencialmente en cualquier planta. Pueden modificarse una gran variedad de plantas y sistemas de células vegetales respecto de las características fisiológicas y agronómicas descritas en la presente memoria usando las construcciones de ácido nucleico de la presente descripción y los diversos métodos de transformación. En realizaciones preferidas, las plantas y células vegetales diana para su modificación incluyen plantas monocotiledóneas y dicotiledóneas, tales como cultivos, incluyendo cultivos de grano (por ejemplo, trigo, maíz, arroz, mijo, cebada), cultivos frutales (por ejemplo, tomate, manzana, pera, fresa, naranja), cultivos de forraje (por ejemplo, alfalfa), cultivos de vegetales de raíz (por ejemplo, zanahoria, patata, remolacha azucarera, batata), cultivos de vegetales foliares (por ejemplo, lechuga, espinaca); plantas de flor (por ejemplo, petunia, rosa, crisantemo), coníferas y pinos (por ejemplo, pino, abeto, picea); plantas usadas en fitorremediación (por ejemplo, plantas acumuladoras de metales pesados); cultivos oleaginosos (por ejemplo, girasol, colza) y plantas usadas con fines experimentales (por ejemplo, *Arabidopsis*). Por lo tanto, los métodos y composiciones descritos son útiles en una gran variedad de plantas, incluyendo especies de los géneros *Asparagus*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Cucurbita*, *Daucus*, *Erigeron*, *Glycine*, *Gossypium*, *Hordeum*, *Lactuca*, *Lolium*, *Lycopersicon*, *Malus*, *Manihot*, *Nicotiana*, *Orychophragmus*, *Oryza*, *Persea*, *Phaseolus*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Secale*, *Solanum*, *Sorghum*, *Triticum*, *Vitis*, *Vigna* y *Zea*.

20 Un experto en la materia reconocerá que después de que se incorpore de manera estable el casete de expresión en plantas transgénicas y se confirme que está operativo, puede introducirse en otras plantas mediante cruzamiento sexual. Puede usarse cualquiera de una serie de técnicas de fitomejoramiento convencionales, dependiendo de la especie que se vaya a cruzar.

25 Una célula vegetal, callo, tejido o planta transformada puede identificarse y aislarse mediante selección o exploración del material vegetal modificado respecto de rasgos codificados por los genes marcadores presentes en el ADN de transformación. Por ejemplo, la selección puede llevarse a cabo cultivando el material vegetal modificado en medios que contienen una cantidad inhibidora del antibiótico o el herbicida para el que la construcción génica de transformación confiere resistencia. Además, también pueden identificarse plantas y células vegetales transformadas mediante exploración respecto de las actividades de cualquier gen marcador visible (por ejemplo, los genes de  $\beta$ -glucuronidasa, luciferasa, B o C1) que pueden estar presentes en las construcciones de ácido nucleico recombinantes. Dichas metodologías de selección y exploración se conocen bien por los expertos en la materia.

30 También pueden usarse métodos físicos y bioquímicos para identificar plantas o células vegetales transformantes que contienen las construcciones génicas insertadas. Estos métodos incluyen: 1) análisis de Southern o amplificación PCR para detectar y determinar la estructura de la inserción de ADN recombinante; 2) transferencia de Northern, protección de RNasa S1, extensión del cebador o amplificación PCR de transcriptasa inversa para detectar y examinar transcritos de ARN de las construcciones génicas; 3) ensayos enzimáticos para detectar la actividad de enzimas o ribozimas, en donde dichos genes están codificados por la construcción génica; 4) electroforesis de proteínas en gel, técnicas de transferencia de Western, inmunoprecipitación o inmunoensayos ligados a enzimas, donde los productos de la construcción génica son proteínas. También pueden usarse técnicas adicionales, tales como hibridación *in situ*, tinción enzimática e inmunotinción para detectar la presencia o la expresión de la construcción recombinante en órganos y tejidos vegetales específicos. Los métodos para efectuar todos estos ensayos se conocen bien por los expertos en la materia.

45 Los efectos de la manipulación génica usando los métodos divulgados en la presente memoria pueden observarse mediante, por ejemplo, transferencias de Northern del ARN (por ejemplo, ARNm) aislado de los tejidos de interés. Típicamente, si ha aumentado la cantidad de ARNm, puede asumirse que el correspondiente gen endógeno se está expresando a una tasa mayor que antes. Pueden usarse otros métodos para medir la actividad génica y/o de CYP74B. Pueden usarse diferentes tipos de ensayos enzimáticos, dependiendo del sustrato usado y del método para detectar el aumento o reducción de un producto o subproducto de reacción. Además, pueden medirse los niveles de y/o la proteína CYP74B expresada de manera inmunoquímica, es decir, ELISA, RIA, EIA y otros ensayos basados en anticuerpos bien conocidos para los expertos en la materia, tales como mediante ensayos de detección electroforéticos (ya sea con tinción o de transferencia de Western). El transgén puede expresarse de manera selectiva en algunos tejidos de la planta o en algunos estados de desarrollo o puede expresarse el transgén en sustancialmente todos los tejidos vegetales, sustancialmente a lo largo de su ciclo vital completo. Sin embargo, también puede aplicarse cualquier modo de expresión combinatoria.

55 La presente descripción también abarca semillas de las plantas transgénicas descritas anteriormente en donde la semilla tiene el transgén o la construcción génica. La presente descripción abarca además la descendencia, clones, líneas celulares o células de las plantas transgénicas descritas anteriormente en donde dicha descendencia, clon, línea celular o célula tiene el transgén o construcción génica.

60 Las ZFP y los vectores de expresión que codifican las ZFP pueden administrarse directamente a la planta para la regulación génica, la escisión dirigida y/o la recombinación. En determinadas realizaciones, la planta contiene

múltiples genes diana parálogos. Se sabe que las plantas pueden contener múltiples genes parálogos, por ejemplo, *B. napus* tiene 5 genes EPSPS (SEQ ID NO: 10-14), que pueden usarse como diana por una o más ZFP (véanse los ejemplos). Por lo tanto, pueden administrarse una o más ZFP o vectores de expresión que codifican ZFP diferentes a una planta para dirigirse a uno o más genes EPSPS en la planta. Por ejemplo, pueden usarse como diana 1, 2, 3, 4, 5 o hasta cualquier número de parálogos (por ejemplo, parálogos de EPSPS) o todos los parálogos (por ejemplo, parálogos de EPSPS) presentes en una planta.

En determinadas realizaciones, el gen de EPSPS comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia que tiene una identidad de secuencia de al menos el 80-100 % con las mismas, incluyendo cualquier porcentaje de identidad dentro de estos intervalos, tal como una identidad de secuencia del 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 % con las mismas.

La administración de cantidades eficaces es mediante cualquiera de las rutas usadas normalmente para introducir ZFP en contacto íntimo con la célula vegetal que va a tratarse. Las ZFP se administran de cualquier modo adecuado, preferentemente con vehículos farmacéuticamente aceptables. Los métodos adecuados para aplicar dichos moduladores están disponibles y se conocen bien por los expertos en la materia y, aunque puede usarse más de una ruta para administrar una composición particular, una ruta concreta puede a menudo proporcionar una reacción más inmediata y más eficaz que otra ruta.

También pueden usarse vehículos y se determinan en parte por la composición particular que se esté administrando, así como por el método particular usado para administrar la composición. Por consiguiente, hay una gran variedad de formulaciones adecuadas de composiciones farmacéuticas que están disponibles (véase, por ejemplo, Remington's Pharmaceutical Sciences, 17<sup>a</sup> ed. 1985).

### Aplicaciones

Un área de particular interés en la agricultura es la mejora genética de plantas para conferir resistencia a herbicidas. Muchos herbicidas actúan inhibiendo una enzima vegetal o proteína clave necesaria para el crecimiento. Por ejemplo, el herbicida glifosato destruye las plantas inhibiendo la actividad de la enzima 5-enolpiruvil-3-fosfoshikimato sintasa (EPSPS), que sintetiza aminoácidos aromáticos. Se han producido plantas tolerantes al glifosato insertando transgenes de EPSPS en genomas de plantas, mediante sobreexpresión de EPSPS y mediante mutación selectiva de EPSPS para producir mutantes resistentes al glifosato (véanse, por ejemplo, las Patentes de Estados Unidos n.º 5.312.910 y 6.987.213; y Papanikou et al. (2004) *Planta* 218(4):589-598).

Por ejemplo, los métodos y composiciones descritos pueden usarse para modular la expresión y para la alteración dirigida de genes de EPSPS. En un aspecto, se usan ZFP modificadas para regular positiva o negativamente la expresión de EPSPS en una planta. Las ZFP se asocian opcionalmente con dominios reguladores para la modulación de la expresión génica, que pueden asociarse covalente o no covalentemente y activar o reprimir la transcripción génica de EPSPS. Dichas ZFP pueden usarse para aumentar o reducir la producción de la enzima EPSPS, controlar la biosíntesis de aminoácidos aromáticos en plantas o aumentar o reducir la tolerancia de la planta al herbicida glifosato, por ejemplo, para producir cultivos resistentes al herbicida glifosato, aumentar el rendimiento de los cultivos o revertir la resistencia al glifosato en hierbas o plantas silvestres.

Las composiciones que comprenden una o más ZFP o los polinucleótidos que las codifican pueden administrarse a una célula vegetal. En una realización, se administran a una célula al menos dos ZFP que reconocen a la misma secuencia diana de un gen EPSPS o una secuencia diana diferente o polinucleótidos que codifican dichas ZFP. La primera ZFP se asocia opcionalmente con la segunda ZFP, ya sea covalente o no covalentemente. Puede usarse el reconocimiento de sitios diana adyacentes por ZFP asociadas o individuales para producir una unión cooperativa de las ZFP, dando como resultado afinidades que son mayores que las afinidades de las ZFP cuando se unen individualmente a sus respectivos sitios diana.

Para la represión de la expresión génica, se reduce la expresión del gen típicamente en aproximadamente un 20 % (es decir, un 80 % de expresión no modulada por ZFP), más preferentemente en aproximadamente un 50 % (es decir, un 50 % de expresión no modulada por ZFP), más preferentemente en aproximadamente un 75-100 % (es decir, de un 25 % a un 0 % de expresión no modulada por ZFP). Para la activación de la expresión génica, la expresión se activa típicamente en aproximadamente 1,5 veces (es decir, un 150 % de expresión no modulada por ZFP), preferentemente 2 veces (es decir, un 200 % de expresión no modulada por ZFP), más preferentemente 5-10 veces (es decir, un 500-1000 % de expresión no modulada por ZFP), hasta al menos 100 veces o más.

La expresión de activadores y represores de ZFP modificados también puede controlarse mediante sistemas reguladores de molécula pequeña, tales como los sistemas regulados por tet y el sistema RU-486 (véase, por ejemplo, Gossen y Bujard, *PNAS* 89:5547 (1992); Oligino et al., *Gene Ther.* 5:491-496 (1998); Wang et al., *Gene Ther.* 4:432-441 (1997); Neering et al, *Blood* 88:1147-1155 (1996); y Rendahl et al., *Nat. Biotechnol.* 16:757-761 (1998)). Estos sistemas reguladores confieren control de molécula pequeña en la expresión de los activadores y represores de ZFP e imponen un nivel de control adicional en los genes diana de interés (por ejemplo, EPSPS).

En otro aspecto, se usan ZFN para inducir mutaciones en una secuencia génica de EPSPS, por ejemplo, escindiendo en dos sitios y eliminando secuencias entre medias, mediante la escisión en un solo sitio seguida de la

- unión de extremos no homólogos, escindiendo en uno o dos sitios con inserción de una secuencia exógena entre los cortes y/o escindiendo en un sitio para eliminar uno o dos o unos pocos nucleótidos. También puede usarse la escisión dirigida para crear *knockouts* génicos (por ejemplo, para genómica funcional o validación de la diana) y para facilitar la inserción dirigida de una secuencia en un genoma (es decir, un *knock-in* génico); por ejemplo, con fines de modificación celular o sobreexpresión de proteínas. La inserción puede ser mediante reemplazos de secuencias cromosomas mediante recombinación de homólogos o mediante integración dirigida, en la que una nueva secuencia (es decir, una secuencia no presente en la región de interés), flanqueada por secuencias homólogas a la región de interés en el cromosoma, se inserta en un sitio diana predeterminado. También pueden usarse los mismos métodos para reemplazar una secuencia génica de EPSPS de tipo silvestre por una secuencia génica de EPSPS mutante o para convertir un alelo en un alelo diferente. Las composiciones y métodos descritos en la presente memoria también pueden usarse para generar líneas de plantas que tienen ZFP inducibles y/o ZFN integradas de manera estable en el genoma. Por consiguiente, las secuencias integradas de manera estable que codifican las proteínas que contienen dedo de cinc pueden expresarse tras su introducción adecuada para lograr el efecto deseado en la planta a lo largo de múltiples generaciones de la planta y en cualquier estado de desarrollo de la planta.
- Además, puede usarse la escisión dirigida de patógenos vegetales infecciosos o integrados para tratar infecciones patógenas en un hospedador vegetal, por ejemplo, escindiendo el genoma del patógeno de tal forma que se reduce o elimina su patogenicidad. Además, puede usarse la escisión dirigida de genes que codifican receptores para virus de plantas para bloquear la expresión de dichos receptores, de este modo previniendo la infección vírica y/o la diseminación vírica en la planta.
- Los ejemplos de patógenos incluyen virus de plantas, tales como *Alfamoviruses*, *Alphacryptoviruses*, *Badnaviruses*, *Betacryptoviruses*, *Bigeminiviruses*, *Bromoviruses*, *Bymoviruses*, *Capilloviruses*, *Carlaviruses*, *Carmoviruses*, *Caulimoviruses*, *Closteroviruses*, *Comoviruses*, *Cucumoviruses*, *Cytorhabdoviruses*, *Dianthoviruses*, *Enamoviruses*, *Fabaviruses*, *Fijiviruses*, *Furoviruses*, *Hordeiviruses*, *Hybrigeminiviruses*, *Idaeoviruses*, *Ilarviruses*, *Ipomoviruses*, *Luteoviruses*, *Machlomoviruses*, *Macluraviruses*, *Marafiviruses*, *Monogeminiviruses*, *Nanaviruses*, *Necroviruses*, *Nepoviruses*, *Nucleorhabdoviruses*, *Oryzaviruses*, *Ourmiaviruses*, *Phytoreoviruses*, *Potexviruses*, *Potyvirus*, *Rymoviruses*, ARN satélite, *satelliviruses*, *Sequiviruses*, *Sobemoviruses*, *Tenuiviruses*, *Tobamoviruses*, *Tobravirus*, *Tombusviruses*, *Tospoviruses*, *Trichoviruses*, *Tymoviruses*, *Umbraviruses*, *Varicosaviruses* y *Waikaviruses*; patógenos fúngicos, tales como el tizón (por ejemplo, *Ustilaginales*), royas (*Uredinales*), ergóticos (*Claviceps pupurea*) y añublos; mohos (*Oomycetes*) tales como *Phytophthora infestans* (tizón de la patata); patógenos bacterianos, tales como *Erwinia* (por ejemplo, *E. herbicola*), *Pseudomonas* (por ejemplo, *P. aeruginosa*, *P. syringae*, *P. fluorescence* y *P. putida*), *Ralstonia* (por ejemplo, *R. solanacearum*), *Agrobacterium* y *Xanthomonas*; lombrices (*Nematoda*); y *Phytopmyxa* (*Polymyxa* y *Plasmodiophora*).
- Los métodos descritos para la recombinación dirigida pueden usarse para reemplazar una o más secuencias genómicas de EPSPS con secuencias homólogas no idénticas. Por ejemplo, puede reemplazarse una secuencia genómica mutante por una secuencia de tipo silvestre o como alternativa, puede reemplazarse una secuencia genómica de tipo silvestre por una secuencia mutante para, por ejemplo, producir cultivos resistentes al herbicida glifosato, aumentar el rendimiento de los cultivos, revertir la resistencia al glifosato en hierbas o plantas silvestres, etc. De un modo similar, puede reemplazarse un alelo de un gen por un alelo diferente usando los métodos de recombinación dirigida descritos en la presente memoria.
- En muchos de estos casos, una región genómica de EPSPS de interés comprende una mutación y el polinucleótido donante comprende la secuencia de tipo silvestre correspondiente. De manera similar, puede reemplazarse una secuencia genómica de tipo silvestre por una secuencia mutante, en caso de que esto sea deseable. Por ejemplo, puede revertirse o reducirse la resistencia al glifosato en una planta reemplazando un gen de EPSPS mutado o exógeno por un gen de tipo silvestre, eliminando un gen de EPSPS exógeno, mutando un gen de EPSPS para reducir la resistencia al glifosato o reemplazando las secuencias de control de un gen de EPSPS por secuencias que soportan un menor nivel de expresión de EPSPS. Como alternativa, puede introducirse una mutación en un gen EPSPS que confiere resistencia al glifosato en una planta bien mutando el gen EPSPS para producir una enzima EPSPS tolerante al glifosato o reemplazando las secuencias de control del gen de EPSPS por secuencias que aumentan el nivel de expresión de EPSPS. Las modificaciones génicas de EPSPS y las enzimas EPSPS mutantes que aumentan la tolerancia al herbicida glifosato se conocen en la técnica (véanse, por ejemplo, las Patentes de Estados Unidos n.º 7.238.508, 7.214.535, 7.141.722, 7.045.684, 6.803.501, 6.750.377, 6.248.876, 6.225.114, 6.040.497, 5.866.775, 5.804.425, 5.776.760, 5.633.435, 5.627.061, 5.554.798, 5.463.175, 5.312.910, 5.310.667, 5.188.642, 5.145.783, 4.971.908 y 4.940.835 y el documento WO 00/66748).
- También pueden usarse la escisión dirigida y la recombinación dirigida para alterar secuencias no codificantes (por ejemplo, secuencias reguladoras, tales como promotores, potenciadores, iniciadores, terminadores, sitios de corte y empalme) para alterar los niveles de expresión de un producto génico de EPSPS. Dichos métodos pueden usarse, por ejemplo, para aumentar la expresión de una variante de EPSPS tolerante al glifosato en un cultivo.
- Puede lograrse la inactivación de un gen EPSPS, por ejemplo, mediante un solo suceso de escisión, mediante escisión seguida de unión de extremos no homólogos, mediante escisión en dos sitios seguida de unión para eliminar la secuencia entre los dos sitios de escisión, mediante recombinación dirigida de un codón de sentido perdido o sin sentido en la región codificante o mediante la recombinación dirigida de una secuencia irrelevante (por

ejemplo, una secuencia de "relleno") en el gen o su región reguladora para alterar el gen o la región reguladora.

La modificación dirigida de la estructura de la cromatina, tal como se describe en el documento WO 01/83793 de cesión común, puede usarse para facilitar la unión de proteínas de fusión a la cromatina celular.

5 En realizaciones adicionales, pueden usarse una o más fusiones entre un dominio de unión de dedo de cinc y una recombinasa (o un fragmento funcional de la misma) además de o en lugar de fusiones de dedo de cinc-dominio de escisión descritas en la presente memoria para facilitar la recombinación dirigida. Véanse, por ejemplo, la Patente de Estados Unidos n.º 6.534.261 y Akopian et al. (2003) Proc. Natl. Acad. Sci. USA 100:8688-8691.

10 En realizaciones adicionales, los métodos y composiciones descritos se usan para proporcionar fusiones de dominios de unión de ZFP con dominios de activación o represión transcripcional que necesitan dimerización (ya sea homodimerización o heterodimerización) para su actividad. En estos casos, un polipéptido de fusión comprende un dominio de unión de dedo de cinc y un monómero de dominio funcional (por ejemplo, un monómero de un dominio de activación o represión transcripcional dimérico). La unión de dos de dichos polipéptidos de fusión a sitios diana situados de manera adecuada permite la dimerización para reconstituir un dominio funcional de activación o represión de la transcripción.

15 Además, tal como se ha descrito anteriormente, los métodos y composiciones expuestos en la presente memoria pueden usarse para la integración dirigida de secuencias exógenas en una región de interés en el genoma de una célula (por ejemplo, una secuencia reguladora o codificante de un gen de EPSPS), por ejemplo, en el que la escisión potencia la inserción mediante mecanismos dependientes de homología (por ejemplo, inserción de una secuencia donante que comprende una secuencia exógena junto con una o más secuencias que son idénticas u homólogas pero no idénticas con una secuencia genómica predeterminada (es decir, un sitio diana)).

20 Tal como se ha indicado anteriormente, en determinadas realizaciones, la integración dirigida mediante mecanismos tanto dependientes de homología como no dependientes de homología implica la inserción de una secuencia exógena entre los extremos generados por la escisión. La secuencia exógena insertada puede ser de cualquier longitud, por ejemplo, una secuencia "parche" relativamente corta de entre 1 y 50 nucleótidos de longitud (por ejemplo, una secuencia de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45 o 50 nucleótidos).

25 En los casos en los que la integración dirigida es dependiente de homología, una secuencia de ácido nucleico o donante comprende una secuencia exógena junto con una o más secuencias que son bien idénticas u homólogas, pero no idénticas con una secuencia genómica predeterminada (es decir, un sitio diana). En determinadas realizaciones, dos de las secuencias idénticas o dos de las secuencias homólogas, pero no idénticas (o una de cada) están presentes, flanqueando a la secuencia exógena. Una secuencia exógena (o ácido nucleico exógeno o polinucleótido exógeno) es uno que contiene una secuencia nucleotídica que normalmente no está presente en la región de interés.

35 Las secuencias exógenas ejemplares incluyen ADNc, secuencias promotoras, secuencias potenciadoras, marcadores epitópicos, genes marcadores, sitios de reconocimiento de enzimas de escisión y diversos tipos de construcciones de expresión. Los genes marcadores incluyen secuencias que codifican proteínas que median la resistencia química o a antibióticos (por ejemplo, resistencia a ampicilina, resistencia a kanamicina, resistencia a G418, resistencia a higromicina B, resistencia a puromicina, resistencia a herbicida), secuencias que codifican proteínas coloreadas o fluorescentes o luminiscentes (por ejemplo, proteína verde fluorescente, proteína verde fluorescente mejorada, proteína fluorescente roja, luciferasa) y proteínas que median el crecimiento celular y/o la amplificación génica potenciada (por ejemplo, dihidrofolato reductasa). Los marcadores epitópicos incluyen, por ejemplo, una o más copias de FLAG, His, myc, Tap, HA o cualquier secuencia de aminoácidos detectable.

40 Las construcciones de expresión de proteínas incluyen ADNc y secuencias de control transcripcional en unión operativa con secuencias de ADNc. Las secuencias de control transcripcional incluyen promotores, potenciadores y aisladores. Las secuencias reguladoras transcripcionales y traduccionales adicionales que pueden incluirse en las construcciones de expresión incluyen, por ejemplo, sitios de entrada a ribosoma interno, secuencias que codifican péptidos 2A y señales de poliadenilación. Una construcción de expresión de proteína ejemplar es una construcción de expresión de anticuerpo que comprende una secuencia que codifica una cadena pesada de anticuerpo y una secuencia que codifica una cadena ligera de anticuerpo, estando cada secuencia unida operativamente a un promotor (siendo los promotores iguales o diferentes) y estando una de las dos o ambas secuencias opcionalmente unidas operablemente a un potenciador (y, en el caso de que ambas secuencias codificantes estén unidas a potenciadores, los potenciadores son iguales o diferentes).

45 Los sitios de reconocimiento de la enzima de escisión incluyen, por ejemplo, secuencias reconocidas por endonucleasas de restricción, endonucleasas *homing* y/o meganucleasas. La integración dirigida de un sitio de reconocimiento de una enzima de escisión (ya sea mediante mecanismos dependientes de homología o independientes de homología) es útil para generar células cuyo genoma contiene únicamente un solo sitio que puede escindirse por una enzima particular. La puesta en contacto de dichas células con una enzima que reconoce y escinde en el sitio único facilita la posterior integración dirigida de secuencias exógenas (ya sea mediante

mecanismos dependientes de homología como independientes de homología) y/o mutagénesis dirigida en el sitio que se escinde.

Las endonucleasas *homing* ejemplares incluyen I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-Scel, I-Scell, I-Ppol, I-ScellI, I-Crel, I-TevI, I-TevII e I-TevIII. Se conocen sus secuencias de reconocimiento. Véanse también la Patente de Estados Unidos n.º 6.833.252, la Patente de Estados Unidos n.º 5.420.032; Belfort et al. (1997) Nucleic Acids Res. 25:3379-3388; Dujon et al. (1989) Gene 82:115-118; Perler et al. (1994) Nucleic Acids Res. 22, 1125-1127; Jasin (1996) Trends Genet. 12:224-228; Gimble et al. (1996) J. Mol. Biol. 263:163-180; Argast et al. (1998) J. Mol. Biol. 280:345-353 y el catálogo de New England Biolabs.

Aunque la especificidad de escisión de la mayoría de endonucleasas *homing* no es absoluta con respecto a sus sitios de reconocimiento, los sitios tienen una longitud suficiente como para que pueda obtenerse un solo suceso de escisión por genoma de tamaño de mamífero expresando una endonucleasa *homing* en una célula que contiene una sola copia de su sitio de reconocimiento. También se ha comunicado que pueden modificarse enzimas de escisión para que se unan a sitios diana no naturales. Véanse, por ejemplo, Chevalier et al. (2002) Molec. Cell 10:895-905; Epinat et al. (2003) Nucleic Acids Res. 31:2952-2962; Ashworth et al. (2006) Nature 441:656-659.

Los métodos anteriores para obtener la recombinación e integración dirigida usando endonucleasas *homing* padecían del problema de que la inserción dirigida del sitio de reconocimiento es extremadamente ineficaz, requiriendo de una exploración laboriosa para identificar células que contenían el sitio de reconocimiento insertado en la ubicación deseada. Los presentes métodos superan estos problemas permitiendo una integración dirigida altamente eficaz (ya sea dependiente de homología o independiente de homología) de un sitio de reconocimiento para una enzima escisora de ADN.

En determinadas realizaciones, se usa integración dirigida para insertar una construcción de expresión de ARN, por ejemplo, secuencias responsables de la expresión regulada de micro ARN, ARNhc o ARNpi. También pueden incorporarse promotores, potenciadores y secuencias reguladoras de la transcripción adicionales en una construcción de expresión de ARN.

En realizaciones en las que se produce la integración dirigida mediante un mecanismo dependiente de homología, la secuencia donante contiene homología suficiente, en las regiones que flanquean a la secuencia exógena, para soportar la reparación dirigida por homología de una rotura bicatenaria en una secuencia genómica, insertando de este modo la secuencia exógena en el sitio genómico diana. Por lo tanto, el ácido nucleico donante puede ser de cualquier tamaño suficiente para soportar la integración de la secuencia exógena mediante mecanismos de reparación dependientes de homología (por ejemplo, recombinación de homólogos). Sin desear quedar ligados a cualquier teoría particular, se cree que las regiones de homología que flanquean a la secuencia exógena proporcionan a los extremos rotos del cromosoma un molde para volver a sintetizar la información genética en el sitio de rotura bicatenaria.

Puede usarse la integración dirigida de secuencias exógenas, tal como se describe en la presente memoria, para generar células y líneas celulares para expresión de proteínas. Véase, por ejemplo, la Publicación de Solicitud de Patente de Estados Unidos n.º 2006/0063231 de cesión común. Para una expresión óptima de una o más proteínas codificadas por secuencias exógenas integradas en un genoma, el sitio de integración cromosómica debe ser compatible con la transcripción a alto nivel de las secuencias integradas, preferentemente en un amplio intervalo de tipos celulares y estados del desarrollo. Sin embargo, se ha observado que la transcripción de secuencias integradas varía dependiendo del sitio de integración debido a, entre otras cosas, la estructura de la cromatina del genoma en el sitio de integración. Por lo tanto, son deseables sitios diana genómicos que soporten la transcripción de alto nivel de secuencias integradas. En determinadas realizaciones, también será deseable que la integración de secuencias exógenas no dé como resultado la activación ectópica de uno o más genes celulares (por ejemplo, oncogenes). Por otro lado, en el caso de integración de secuencias promotoras y/o potenciadoras, puede desearse la expresión ectópica.

Para determinadas realizaciones, es deseable que un sitio de integración no esté presente en un gen esencial (por ejemplo, un gen esencial para la viabilidad celular) de tal forma que la inactivación de dicho gen esencial no sea el resultado de la integración de la secuencia exógena. Por otro lado, si la intención es desactivar la función génica (es decir, crear un "*knock-out*" génico) la integración dirigida de una secuencia exógena para alterar un gen endógeno es un método eficaz. En estos casos, la secuencia exógena puede ser cualquier secuencia capaz de bloquear la transcripción del gen endógeno o de generar un producto de traducción no funcional, por ejemplo, un parche corto de secuencia de aminoácidos, que es opcionalmente detectable (véase más arriba). En determinadas realizaciones, las secuencias exógenas pueden comprender un gen marcador (descrito anteriormente), que permite la selección de células que han sufrido integración dirigida.

Los sitios diana genómicos adicionales que soportan la transcripción de alto nivel de las secuencias integradas pueden identificarse como regiones de cromatina abierta o "regiones accesibles", tal como se describe, por ejemplo, en las Publicaciones de Patente de Estados Unidos de cesión común 2002/0064802 (30 de mayo de 2002) y 2002/0081603 (27 de junio de 2002).

La presencia de una rotura bicatenaria en una secuencia genómica facilita no solo la integración dependiente de homología de secuencias exógenas (es decir, recombinación de homólogos) sino también la integración independiente de homología de secuencias exógenas en el genoma en el sitio de la rotura bicatenaria. Por consiguiente, las composiciones y métodos descritos en la presente memoria pueden usarse para la escisión dirigida de una secuencia genómica, seguidos de integración no dependiente de homología de una secuencia exógena en o próxima al sitio de escisión diana. Por ejemplo, puede ponerse en contacto una célula con una o más proteínas de fusión de ZFP-dominio de escisión (o semidominio de escisión) modificadas para escindir en una región de interés en un genoma tal como se describe en la presente memoria (o uno o más polinucleótidos que codifican dichas proteínas de fusión) y un polinucleótido que comprende una secuencia exógena que caree de homología con la región de interés, para obtener una célula en la que la totalidad o una parte de la secuencia exógena se integra en la región de interés.

Los métodos de integración dirigida (es decir, inserción de una secuencia exógena en un genoma) tanto dependientes como independientes de homología descritos en la presente memoria pueden usarse para una serie de fines. Estos incluyen la inserción de un gen o secuencia de ADNc en el genoma de una célula para permitir la expresión de los productos de transcripción y/o traducción del gen o del ADNc por parte de la célula. Para las situaciones en las que puede producirse una enfermedad o patología a causa de una o una diversidad de mutaciones (por ejemplo, múltiples mutaciones puntuales diseminadas a lo largo de la secuencia del gen), es particularmente eficaz la integración dirigida (ya sea dependiente de homología o independiente de homología) de una copia de ADNc del gen de tipo silvestre. Por ejemplo, dicho ADNc de tipo silvestre se inserta en una secuencia líder no traducida o en el primer exón de un gen cadena arriba de todas las mutaciones conocidas. En determinados integrantes, en los que se conserva la fase de lectura traduccional, el resultado es que se expresa el ADNc de tipo silvestre y su expresión se regula mediante las secuencias reguladoras transcripcionales endógenas adecuadas. En realizaciones adicionales, dichas secuencias de ADNc integradas pueden incluir señales de terminación transcripcional (y/o traduccional) dispuestas cadena abajo del ADNc de tipo silvestre y cadena arriba del gen endógeno mutante. De este modo, se expresa una copia de tipo silvestre del gen causante de enfermedad y no se expresa el gen endógeno mutante. En otras realizaciones, se inserta una porción del ADNc de tipo silvestre en la región adecuada de un gen (por ejemplo, un gen en el que están agrupadas las mutaciones causantes de enfermedad).

### Ejemplos

A continuación, hay ejemplos de realizaciones específicas para llevar a cabo la presente descripción. Los ejemplos se proporcionan solo con fines ilustrativos.

Se han aplicado esfuerzos para asegurar la precisión respecto de los números usados (por ejemplo, cantidades, temperaturas, etc.) pero por supuesto debe permitirse cierto grado de error y desviación experimental.

#### Ejemplo 1: Identificación de secuencia diana en *B. napus*

##### A. Identificación de secuencia

Se seleccionaron secuencias de ADN para genes de colza nativos de función conocida como dianas para la edición genómica usando nucleasas de dedo de cinc modificadas. Las secuencias de estos genes, citadas como gene de 5-enolpiruvil shikimato-3-fosfato sintasa (EPSPS), procedían de *Brassica napus* L, Nex710. La enzima EPSPS es la sexta enzima de la ruta de shikimato, que es esencial para la síntesis de aminoácidos aromáticos y muchos metabolitos aromáticos en plantas (Bentley (1990) Crit. Rev. Biochem. Mol Biol. 25:307-384). Cataliza la transferencia del resto enolpiruvil del piruvato de fosfoenol (PEP) al 5-hidroxilo de shikimato 3-fosfato (S3P). Ya que *B. napus* es una especie anfiploide resultante de la combinación de os conjuntos de cromosomas de *B. rapa* ( $2n = 20$ , AA) y *B. oleracea* ( $2n = 18$ , CC) (Morinaga, 1934; U, 1935), se espera que pueda haber más de un gen de EPSPS en esta especie.

##### B. Aislamiento de ADN

Se plantaron semillas de la variedad Nex710 de *B. napus* en el invernadero. Se recogieron muestras en el 13<sup>er</sup> día después de la siembra, se congelaron inmediatamente en nitrógeno líquido y se almacenaron a -80 °C hasta su uso.

El ADN genómico se aisló usando precipitación de bromuro de cetil-trimetilamonio (CTAB) o el kit de extracción PLANT DNEASY para el aislamiento del ADN de planta (Qiagen, Valencia, CA). Para el procedimiento usando CTAB, se trituró 1 g de tejido foliar (grupos de 6 plantas) en nitrógeno líquido. El ADN se aisló tal como se describe por Permingeat et al. (Plant Mol. Biol. Repr. (1998) 16:1-6) salvo que se modificó el tampón de extracción. El tampón de extracción modificado contenía Tris-HCl 100 mM (pH 8,0), NaCl 2 M, EDTA 25 mM, CTAB al 2,5 % (Sigma n.º de catálogo H-5882) y polivinilpirrolidona-40 (PVP-40) al 1,5 %. El ADN genómico total se aisló con el kit de extracción PLANT DNEASY (Qiagen, Valencia, CA) según las recomendaciones del fabricante con una modificación. Se añadió PVP-40 al tampón AP1 (Qiagen) a una concentración final del 1 %. Cuando se iba a digerir el ADN con enzimas de restricción, se purificó adicionalmente mediante dos etapas de precipitación de polietilenglicol (PEG, PM 8.000) del modo siguiente. Se añadió un volumen igual de NaCl 1,2 M/PEG al 13 % al ADN y se incubó sobre hielo durante 2 horas. Entonces se centrifugaron las muestras a 5.000xg durante 10 minutos, se

desechó el sobrenadante y se lavó el sedimento con etanol al 70 %. El etanol se retiró completamente mediante liofilización y se resuspendió el sedimento de ADN en tampón EB (Qiagen).

Entonces se midió el ADN usando tinción de ácidos nucleicos fluorescente PICOGREEN para cuantificar el ADN bicatenario según las instrucciones del comercializador (Molecular Probes, Eugene, OR) y mediante lecturas de absorbancia a 260 y 280 nm. Se comprobó la calidad del ADN corriendo muestras de ADN en un gel de agarosa al 0,8 % usando tampón de Tris-acetato-EDTA (TAE) (Sambrook et al. (1989) Gel electrophoresis of DNA, Molecular Cloning. Cold Spring Harbor Laboratory Press, pág. 6.7).

### C. Estimación del número de copias del gen EPSPS mediante análisis de Southern

Se llevó a cabo una estimación del número de copias del gen de EPSPS mediante análisis de Southern antes de la amplificación, clonación y secuenciación génica del ADN genómico de *B. napus*. Se seleccionaron enzimas de restricción que cortan en ADN genómico una vez en el gen para digerir el ADN genómico (Gasser y Klee (1990) Nucleic Acid Research 18:2821) y una segunda vez en secuencias genómicas flanqueantes, de tal forma que se creó un fragmento de ADN genómico de tamaño único para cada uno de los genes de EPSPS con la hibridación de una sonda para el EPSPS. La mayoría de las enzimas de restricción que se seleccionaron (*Pvu* II, *Nde* I, *Bsr* BI, *Bsa* I, *Bcl* I, *Bsm* I, *Afl* II) cortan bien hacia el extremo 5' o hacia la mitad del gen salvo *Bcl* I, que corta en el extremo 3' del gen, donde se hibrida la sonda (véase más adelante).

Se digirieron muestras de ADN (5 µg cada una para Nex710, 4 µg cada una para *B. rapa* y 3 µg cada una para *B. oleracea*) durante toda la noche con 30 unidades de cada enzima de restricción, *Pvu* II, *Nde* I, *Bsr* BI, *Bsa* I, *Bcl* I, *Bsm* I y *Afl* II, por separado en tubos Eppendorf según las instrucciones del fabricante (New England BioLabs). Entonces se sometió a las muestras de ADN digeridas a una precipitación de etanol y se liofilizaron los sedimentos.

Los sedimentos secos se resuspendieron en tampón de carga 2x, se cargaron en un gel de agarosa al 0,85 % y se sometieron a electroforesis en tampón de Tris-acetato 0,4x a pH 8,0 (Sambrook et al. (1989) Gel electrophoresis of DNA, Molecular Cloning, Cold Spring Harbor Laboratory Press, pág. 6.7). Entonces se tiñó el gen con bromuro de etidio y se visualizaron las bandas de ADN mediante UV. Posteriormente se transfirió el ADN a una membrana de transferencia de hibridación GENESCREEN PLUS (DuPont NEN, Boston, MA, EE.UU.) en tampón de pirofosfato de Na 25 mM (Murray et al. (1992) Plant Mol. Biol. Repr. 10:173-177). La hibridación previa se llevó a cabo durante un mínimo de 2 horas a 65 °C en tampón de hibridación SIGMA PERFECT HYB PLUS (Sigma, St. Louis, MO). La hibridación se llevó a cabo en el tampón durante toda la noche después de añadir una sonda radiactiva (véase más adelante). Se usó un horno de hibridación (Robbins Scientific Corp, Sunnyvale, CA, EE.UU.) para las etapas tanto de pre-hibridación como de hibridación. La membrana se lavó en una dilución de factor 20 del tampón de lavado que comprendía fosfato de sodio 200 mM, pH 7,8, pirofosfato de sodio 50 mM, EDTA 10 mM y SDS al 2 % (Murray et al., anteriormente citado). Se usó un enjuagado inicial de 5 minutos seguido de dos lavados de 15 cada uno. Entonces se expuso la transferencia a una pantalla de obtención de imágenes fosforescentes a temperatura ambiente durante 12 horas antes de escanearla en un dispositivo de obtención de imágenes fosforescentes BIORAD PERSONAL FX (Bio-Rad, Hercules, CA).

La sonda de EPSPS para la hibridación de transferencia de Southern se generó mediante PCR usando *B. napus* var. Nex710 como molde de ADN genómico. Se diseñaron cebadores a partir de la secuencia del exón 8 basados en la secuencia de ADN genómico publicada de *B. napus* (Gasser y Klee, anteriormente citado) con el programa informático VECTORNTI (Invitrogen, Carlsbad, CA) y se sintetizaron a medida por MWG BIOTECH, Inc. (High Pint, NC, EE.UU.). Las secuencias de los cebadores en orientación directa e inversa fueron TTGGAGCTACAGTGAAGAAGGTT (SEQ ID NO: 1) y CGATTGCATCTCACTCAGTTCATTA (SEQ ID NO: 2), respectivamente. Las reacciones PCR contenían 5 µl de tampón PCR HOT START 10X (Qiagen, Valencia, CA, EE.UU.), 2 µl de MgCl<sub>2</sub> 25 mM, 4 µl de mezcla de nucleótidos 10 mM, 1 µl de cada cebador (20 µM), 1,5 unidades de ADN polimerasa Taq de HOT START (Qiagen, Valencia, CA), 5 µl de ADN molde de Nex710 y agua estéril en un volumen total de 50 µl. La amplificación se ejecutó en un instrumento de PCR en tiempo real ICYCLER IQ (Bio-Rad, Hercules, CA) usando los siguientes parámetros: desnaturalización inicial durante 15 minutos a 95 °C seguido de 35 ciclos de 30 segundos a 95 °C, hibridación a 55,5 °C y 52,9 °C durante 30 segundos y 30 segundos a 72 °C. Se purificó un producto de PCR de 350 pares de bases con un kit de eliminación de nucleótidos QIAQUICK (Qiagen, Valencia, CA). Se verificó el tamaño y la integridad del ADN mediante electroforesis en un gel de agarosa E-GEL al 2,0 % (Invitrogen, Carlsbad, CA). Se determinó la cantidad de fragmentos usando el reactivo de cuantificación de ADN PICOGREEN (Invitrogen, Carlsbad, CA). Las sondas de ADN se marcaron usando perlas de marcaje de ADN READY-TO-GO (-dCTP) (Amersham Biosciences, Piscataway, NJ).

Los análisis de transferencia de Southern mostraron múltiples bandas específicas de EPSPS de *B. napus* (cuatro o más) y potencialmente un número igual de genes (figura 1). El ADN de *B. rapa* y *B. oleracea* se hibridó en menos bandas y sus números y posiciones no se sumaron a los patrones de *B. napus*, lo que indica que la diversidad de secuencia surgió en los genomas parentales y de *B. napus* debido a la anfidiploidía. Las bandas menores pudieron deberse a la hibridación cruzada con otros genes de *B. napus* con una homología de secuencia limitada.

D. Amplificación génica y análisis de secuencia

En el presente estudio, se usó la secuencia de ADNc de EPSPS de *B. rapa* (n.º de referencia de GenBank AY512663, SEQ ID NO: 3) para efectuar búsquedas en la base de datos TIGR *Brassica napus*EST (disponible en Internet en [tigrblast.tigr.org/tgi/](http://tigrblast.tigr.org/tgi/)) usando algoritmos BLAST. Se identificaron dos secuencias, TC1307 (parcial y no anotada) y TC1308 (EPSPS de longitud completa). La secuencia TC1307 era una secuencia génica de EPSPS. Se usaron las secuencias AY512663 y TC1307 para diseñar múltiples oligonucleótidos cortos para su uso como cebadores de PCR, incluyendo los siguientes nucleótidos de orientación directa:

5'- ATGGCGCAAGCTAGCAGAATCTGCC-3'(SEQ ID NO: 4)

5'-ATGGCGCAAGCTAGCAGAATC-3' (SEQ ID NO: 5)

10 5'-CCAGCAGCAGCGTGGAGCTTATCAGATA-3' (SEQ ID NO: 6) y los siguientes oligonucleótidos de orientación inversa:

5'- GGCCAAAACACTGATTCAACGATTGC -3' (SEQ ID NO: 7)

5'-CGTTGCCACCAGCAGCAGTA-3' (SEQ ID NO: 8)

5'-GATGGTCCAGTCACAGTCACACTGTTCTCTGT-3' (SEQ ID NO: 9).

15 Todos los cebadores oligonucleotídicos se sintetizaron y adquirieron a través de Integrated DNA Technologies (IDT, Coralville, IA).

Para los análisis basados en la PCR, la amplificación de ADN se llevó a cabo en una mezcla de reacción de la PCR que contenía 2,5 µl de tampón II LA PCR 10X (más Mg<sup>2+</sup>) (Takara Bio Inc., Otsu, Shiga, Japón), 0,7 µl de MgCl<sub>2</sub> 25 mM, 4 µl de mezcla de nucleótidos 10 mM, 0,5 µl de cada cebador (20 µM), 1,25 unidades de polimerasa Taq de TAKARA LA (Takara Bio Inc.), 1 µl de ADN molde de la variedad Nex710 de *B. napus* (3-10 ng de ADN) y agua estéril hasta un volumen total de 25 µl. La amplificación se llevó en un termociclador MJ (Bio-Rad, Hercules, CA) o un sistema PCR GENEAMP 9700 (Applied Biosystems, Foster City, CA) usando los siguientes parámetros: desnaturalización inicial durante 1 minuto a 94 °C seguido de 30 ciclos de 20 segundos a 94 °C, 30 segundos a 59 °C y 2 minutos a 72 °C. Se verificaron el tamaño y la integridad de los productos de la PCR mediante electroforesis.

25 Cuando se llevó a cabo la PCR con los cebadores SR130 y SR131 (SEQ ID NO: 4 y 7, respectivamente), se amplificaron fragmentos de ADN con un tamaño de 2,6 kb - 3 kb. Estos fragmentos se clonaron directamente en el vector pCR2.1 (Invitrogen, Carlsbad, CA) usando el kit de clonación TA de Invitrogen (Carlsbad, CA) según las recomendaciones del fabricante. Se secuenciaron los fragmentos clonados en DAS con el kit de secuenciación de colorante CEQ (Beckman Coulter, Fullerton, CA) según las recomendaciones del fabricante o se contrataron los servicios de secuenciación a Cogenics (anteriormente Lark Technologies, Inc. Houston, TX). El análisis de secuencia de múltiples clones reveló 3 fragmentos génicos distintos. Estos fragmentos génicos se denominaron parálogos C, D y E de EPSPS (SEQ ID NO: 12-14).

35 Para identificar otras variantes de los genes que pudieran existir en el genoma de *B. napus*, se ejecutó la PCR con un gradiente de temperatura en las mismas condiciones para la PCR descritas anteriormente. La amplificación se llevó a cabo en el instrumento de PCR en tiempo real ICYCLER IQ (Bio-Rad, Hercules, CA) usando los siguientes parámetros: desnaturalización inicial durante 1 minuto a 94 °C seguido de 30 ciclos de 20 segundos a 94 °C, 30 segundos a gradientes de temperatura entre 40 °C a 60 °C y 4 minutos a 72 °C. Se llevó a cabo una extensión final de 30 minutos a 72 °C seguido de una retención indefinida a 4 °C. En estas condiciones, se produjo una banda específica correspondiente a ADN amplificado de aproximadamente 2,5 kb a 52,5 °C. El ADN amplificado se clonó en el vector pCR2.1 y se secuenció tal como se ha descrito anteriormente. El análisis de secuencia de múltiples clones indicó claramente que este producto de la PCR representó un gen diferente, que se identificó como el parálogo B de EPSPS (SEQ ID NO: 11).

45 Se usaron los cebadores 1307F (SEQ ID NO: 5) y 1307R (SEQ ID NO: 8), correspondientes al fragmento TC1307, en las reacciones PCR de la siguiente composición: 5 µl de tampón PCR HOT START 10X (Qiagen, Valencia, CA), 3 µl de MgCl<sub>2</sub> 25 mM, 4 µl de mezcla de nucleótidos 10 mM, 1µl de cada cebador (20 µM), 1,5 unidades de ADN polimerasa Taq de HOT START (Qiagen, Valencia, CA), 5 µl (20 ng) de ADN molde de Nex710 y agua estéril en un volumen total de 50 µl. La amplificación se ejecutó en un instrumento de PCR ICYCLER IQ (Bio-Rad, Hercules, CA) usando los siguientes parámetros: desnaturalización inicial durante 15 minutos a 95 °C seguido de 35 ciclos de 30 segundos a 95 °C, 30 segundos a una temperatura de 40 °C a 60 °C y 1 minuto a 72 °C. Se llevó a cabo una extensión final de 10 minutos a 72 °C seguido de una retención indefinida a 4 °C. En estas condiciones, se produjo una banda de ADN amplificado de 700 pb a 41,4 °C. Este fragmento se clonó en el vector TOPO PCR2.1 y se secuenció tal como se ha descrito anteriormente. El alineamiento de múltiples clones dio como resultado una secuencia de 669 pb, que se identificó como el parálogo A de EPSPS (SEQ ID NO: 10).

55 Se llevaron a cabo reacciones PCR adicionales con los cebadores 4°\_Gene\_F2 (SEQ ID NO: 6) y EPSP\_ADNc\_R9 (SEQ ID NO: 9) para amplificar la secuencia más larga del parálogo A. La mezcla de reacción de la PCR tenía la

5 siguiente composición: 10,0 µl de reactivos ACCUPRIME SUPERMIX II para amplificación PCR (Invitrogen, Carlsbad, CA), 0,5 µl de cada cebador (20 µM), 3 µl de ADN molde de Nex710 y agua estéril hasta un volumen total de 20 µl. La amplificación se ejecutó usando los siguientes parámetros: 95 °C durante 3 minutos, seguido de 10 ciclos de 95 °C durante 30 segundos, 73 °C (-0,5 °C/ciclo) durante 30 segundos y 68 °C durante 3 minutos, seguido de 30 ciclos de 95 °C durante 30 segundos, 68 °C durante 30 segundos y 68 °C durante 3 minutos. Esto fue seguido de una extensión final a 68 °C durante 30 minutos. Se clonó un fragmento amplificado de aproximadamente 2 kb en el vector TOPO PCR2.1 y después se secuenció tal como se ha descrito previamente. El alineamiento de múltiples clones dio como resultado la secuencia de 1571 pb del parálogo A.

10 Una comparación del ADNc de *B. rapa* (SEQ ID NO: 3) y el ADN genómico de *B. napus* (SEQ ID NO: 14) mostró la presencia de 8 exones y 7 intrones en el gen de EPSPS. Basándose en esta comparación, el alineamiento de todos los 5 parálogos genómicos aislados de ADN de *B. napus* var. Nex710 indicaron que existían pequeñas diferencias, tales como polimorfismos de un solo nucleótido (SNP), entre los genes en las regiones codificantes predichas, mientras que las secuencias intrónicas variaron significativamente más a nivel de nucleótidos. En general, había una homología de secuencia del 84 % o más entre los 5 parálogos de EPSPS (véase la tabla 2).

15 **Tabla 2.** Homología de secuencia (%) entre los parálogos A-E de EPSPS (SEQ ID NO: 10-14)

Parálogos de EPSPS	Parálogo E	Parálogo A	Parálogo B	Parálogo C
Parálogo D	98	88	84	97
Parálogo E		88	84	95
Parálogo A			92	87
Parálogo B				84

20 Estas diferencias entre los 5 parálogos se anotaron debido a que destacan regiones de las secuencias que pueden discriminarse mediante una proteína de unión a ADN dependiente de secuencia, tal como una proteína de dedo de cinc. Es deseable diseñar un dominio de unión a ADN de dedo de cinc que se una a una secuencia génica y no a otra, incluso si la secuencia es elevadamente similar. Se seleccionaron secuencias prácticamente de longitud completa para cuatro de los parálogos B, C, D y E (SEQ ID NO: 11-14) y una secuencia génica parcial para el parálogo A de 1575 kb (SEQ ID NO: 10) como dianas para el diseño de nucleasas de dedo de cinc, tal como se describe a continuación.

**Ejemplo 2: Diseño de dominios de unión a ADB de EPSPS de dedo de cinc**

25 Usando sitios diana identificados a partir de los parálogos A-E de EPSPS de *B. napus* (ejemplo 1, figs. 9-13), se seleccionaron hélices de reconocimiento de EPSPS de dedo de cinc. Las hélices de reconocimiento para los diseños representativos de EPSPS de dedo de cinc se muestran a continuación en la tabla A.

Tabla A: Diseños de dedo de cinc de EPSPS

Nombre de ZFN	F1	F2	F3	F4	F5	F6
10654 EPSPS	QSGDLTR (SEQ ID NO: 91)	RSDTLST (SEQ ID NO: 92)	RNDNRIT (SEQ ID NO: 93)	QSSDLR (SEQ ID NO: 94)	QSSDLTR (SEQ ID NO: 95)	ninguno
10658* EPSPS	DRSNLSR (SEQ ID NO: 96)	RSDALAR (SEQ ID NO: 97)	QNAHRKT (SEQ ID NO: 98)	RSDHLSE (SEQ ID NO: 99)	NSRNRKT (SEQ ID NO: 100)	ninguno
9875 EPSPS	QSSDLR (SEQ ID NO: 94)	RSDHLR (SEQ ID NO: 101)	QSSDLRR (SEQ ID NO: 102)	QSGNLAR (SEQ ID NO: 103)	QSGNLAR (SEQ ID NO: 103)	ninguno
10275 EPSPS	RSDVLSQ (SEQ ID NO: 104)	RNANRKK (SEQ ID NO: 105)	RSDNLST (SEQ ID NO: 106)	RNDARIT (SEQ ID NO: 107)	RSDNLST (SEQ ID NO: 106)	DNSSRIT (SEQ ID NO: 108)
10740 EPSPS	RSDVLSE (SEQ ID NO: 109)	TSGHLR (SEQ ID NO: 110)	RSDDLK (SEQ ID NO: 111)	DSSARKK (SEQ ID NO: 112)	ninguno	ninguno

ES 2 585 157 T3

Nombre de ZFN	F1	F2	F3	F4	F5	F6
10741 EPSPS	TSGNLTR (SEQ ID NO: 113)	TSGSLTR (SEQ ID NO: 114)	RSDHLST (SEQ ID NO: 115)	QSANRTK (SEQ ID NO: 116)	ninguno	ninguno
10742 EPSPS	TSGNLTR (SEQ ID NO: 113)	TSGSLTR (SEQ ID NO: 114)	RSDHLSQ (SEQ ID NO: 117)	TSSNRIT (SEQ ID NO: 118)	ninguno	ninguno
9876 EPSPS	QSSDLRR (SEQ ID NO: 102)	RSDHLRSR (SEQ ID NO: 101)	QSSDLRR (SEQ ID NO: 102)	DRSALSR (SEQ ID NO: 119)	QSGNLAR (SEQ ID NO: 103)	ninguno
9882 EPSPS	QSSDLRR (SEQ ID NO: 102)	RSDHLST (SEQ ID NO: 115)	HSDTRKK (SEQ ID NO: 120)	QSGNLAR (SEQ ID NO: 103)	QSGNLAR (SEQ ID NO: 103)	ninguno
11038 EPSPS	QSGNLAR (SEQ ID NO: 103)	TSGSLTR (SEQ ID NO: 114)	RSDHLST (SEQ ID NO: 115)	QSANRTK (SEQ ID NO: 116)	ninguno	ninguno
11039 EPSPS	QSGNLAR (SEQ ID NO: 103)	TSGSLTR (SEQ ID NO: 114)	RSDHLSQ (SEQ ID NO: 117)	TSSNRIT (SEQ ID NO: 118)	ninguno	ninguno
10744 EPSPS	RSDDLSE (SEQ ID NO: 121)	TNSNRKR (SEQ ID NO: 122)	RSDLSA (SEQ ID NO: 123)	TSANLSR (SEQ ID NO: 124)	ninguno	ninguno
10743 EPSPS	RREDLIT (SEQ ID NO: 125)	TSSNLSR (SEQ ID NO: 126)	RSDTLSE (SEQ ID NO: 127)	QANRKT (SEQ ID NO: 128)	ninguno	ninguno
10745 EPSPS	RSDTLSE (SEQ ID NO: 127)	TSANLSR (SEQ ID NO: 124)	RSDLSA (SEQ ID NO: 123)	TSANLSR (SEQ ID NO: 124)	ninguno	ninguno
9892 EPSPS	RSDNLSA (SEQ ID NO: 129)	QNRDRKN (SEQ ID NO: 130)	QSGDLTR (SEQ ID NO: 91)	RSDALAR (SEQ ID NO: 97)	RSDNLRE (SEQ ID NO: 131)	ninguno
9895 EPSPS	RSAALAR (SEQ ID NO: 132)	RSDDLTR (SEQ ID NO: 133)	QSGDLTR (SEQ ID NO: 91)	RSDTLSQ (SEQ ID NO: 134)	QSGSLTR (SEQ ID NO: 135)	ninguno
9896 EPSPS	RSDALAR (SEQ ID NO: 97)	RSDDLTR (SEQ ID NO: 133)	QSGDLTR (SEQ ID NO: 91)	RSDTLSQ (SEQ ID NO: 134)	QSGSLTR (SEQ ID NO: 135)	ninguno
10657* EPSPS	DRSNLSR (SEQ ID NO: 96)	RSDALAR (SEQ ID NO: 97)	QNAHRKT (SEQ ID NO: 98)	RSDHLSE (SEQ ID NO: 99)	NSRNRKT (SEQ ID NO: 100)	ninguno
12385 EPSPS	QSGDLTR (SEQ ID NO: 91)	RSDTLST (SEQ ID NO: 92)	RNDNRIT (SEQ ID NO: 93)	QSSDLRSR (SEQ ID NO: 94)	LLTTLKA (SEQ ID NO: 136)	ninguno
12202* EPSPS	DRSNLSR (SEQ ID NO: 96)	RSDALAR (SEQ ID NO: 97)	QNAHRKT (SEQ ID NO: 98)	RSDHLSE (SEQ ID NO: 99)	KNFNLHQ (SEQ ID NO: 137)	ninguno
14318* EPSPS	DRSNLSR (SEQ ID NO: 96)	RSDALAR (SEQ ID NO: 97)	QNAHRKT (SEQ ID NO: 98)	RSDHLSE (SEQ ID NO: 99)	KNFNLHQ (SEQ ID NO: 137)	ninguno

Nombre de ZFN	F1	F2	F3	F4	F5	F6
14320* EPSPS	DRSNLSR (SEQ ID NO: 96)	RSDALAR (SEQ ID NO: 97)	QNAHRKT (SEQ ID NO: 98)	RSDHLSE (SEQ ID NO: 99)	KNFNLHQ (SEQ ID NO: 137)	ninguno
13969 EPSPS	DRSNLSR (SEQ ID NO: 96)	RSDALAR (SEQ ID NO: 97)	TSTGLLI (SEQ ID NO: 138)	RSDHLSE (SEQ ID NO: 99)	KNFNLHQ (SEQ ID NO: 137)	ninguno
12540 EPSPS	DRSNLSR (SEQ ID NO: 96)	RSDALAR (SEQ ID NO: 97)	VSHTRLD (SEQ ID NO: 139)	RSDHLSE (SEQ ID NO: 99)	NSRNRKT (SEQ ID NO: 100)	ninguno
12352 EPSPS	QSGDLTR (SEQ ID NO: 91)	RSDTLST (SEQ ID NO: 92)	TRYKLMS (SEQ ID NO: 140)	QSSDLR (SEQ ID NO: 94)	QSSDLTR (SEQ ID NO: 95)	ninguno
11034 EPSPS	RSDVLSE (SEQ ID NO: 109)	TSGHLSR (SEQ ID NO: 110)	RSDDLK (SEQ ID NO: 111)	DSSARKK (SEQ ID NO: 112)	ninguno	ninguno
11036 EPSPS	TSGNLTR (SEQ ID NO: 113)	TSGSLTR (SEQ ID NO: 114)	RSDHLST (SEQ ID NO: 115)	QSANRTK (SEQ ID NO: 116)	ninguno	ninguno
11037 EPSPS	TSGNLTR (SEQ ID NO: 113)	TSGSLTR (SEQ ID NO: 114)	RSDHLSQ (SEQ ID NO: 117)	TSSNRIT (SEQ ID NO: 118)	ninguno	ninguno

(\*Nota- las ZFN 10657 y 10658, las 12202, 14318 y 14320, las 10740 y 11034, las 10741 y 11036 y las 10742 y 11037 difieren entre sí por mutaciones no localizadas en las hélices de reconocimiento).

Los sitios diana de los diseños de dedo de cinc se muestran a continuación en la tabla B. Las ZFP 10654 y 10658 se diseñaron para sitios en los parálogos C y D; las ZFP 9875 y 10275 se diseñaron para sitios diana en el parálogo D; y las ZFP 10740, 10741 y 10742 se diseñaron para su unión a sitios diana en los parálogos A y B.

Tabla B: Sitios diana de dedos de cinc de EPSPS

Nombre de ZFN	Sitio diana (5' a 3')
10654 EPSPS	ttACTGCTgCAGGTGGCAac (SEQ ID NO: 141)
10658 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO: 142)
9875 EPSPS	ttGAAGAAGCTGGGGCTta (SEQ ID NO: 143)
10275 EPSPS	gcATCAAGcATGTAGTTGATGtc (SEQ ID NO: 144)
10740 EPSPS	aaATCTCGGGTCTGat (SEQ ID NO: 145)
10741 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO: 146)
10742 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO: 146)
9876 EPSPS	ttGAAGAAGCTGGGGCTta (SEQ ID NO: 143)
9882 EPSPS	ttGAAGAAGCTGGGGCTta (SEQ ID NO: 143)
11038 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO: 146)
11039 EPSPS	tcTAATGGGTTGAAgca (SEQ ID NO: 146)
10744 EPSPS	gaGATTTGGATCCGgg (SEQ ID NO: 147)
10743 EPSPS	tcCAACCGGATTCTtc (SEQ ID NO: 148)

10745 EPSPS	gaGATTTGGATCCGgg (SEQ ID NO: 147)
9892 EPSPS	tgCAGGTGGCAaCGCAAGgat (SEQ ID NO: 149)
9895 EPSPS	caGTAACGGCAGCGGTGag (SEQ ID NO: 150)
9896 EPSPS	caGTAACGGCAGCGGTGag (SEQ ID NO: 150)
10657 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO: 142)
12385 EPSPS	ttACTGCTgCAGGTGGCAac (SEQ ID NO: 141)
12202 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO: 142)
14318 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO: 142)
14320 EPSPS	ggCAGCGGTGAGTGGACgc (SEQ ID NO: 142)
13969 EPSPS	GCAGCGGTGAGTGGACG (SEQ ID NO: 151)
12540 EPSPS	tgCAGCTGTAAGTGGACgc (SEQ ID NO: 152)
12352 EPSPS	ttACTGCTgCTGGTGGCAac (SEQ ID NO: 153)
11034 EPSPS	aaATCTCGGGTCTGat (SEQ ID NO: 145)
11036 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO: 146)
11037 EPSPS	tcTAATGGGTTGAAgc (SEQ ID NO: 146)

Los diseños de EPSPS se incorporaron en vectores de expresión de dedo de cinc que codificaban una proteína que tenía al menos un dedo con una estructura CCHC. Véase la Solicitud de Patente de Estados Unidos con n.º de serie 60/874.911. En particular, el último dedo en cada proteína tenía un armazón de CCHC. Entonces se fusionaron las secuencias codificantes de dedo de cinc al dominio de nucleasa de la enzima de restricción de tipo IIS *FokI* (aminoácidos 384-579 de la secuencia de Wah et al. (1998) Proc. Natl. Acad. Sci. USA 95:10564-10569 a través de un enlazador ZC de cuatro aminoácidos) para formar ZFN de EPSPS. Se ensayaron diversas ZFN respecto de su actividad biológica y/o toxicidad tal como se describe en la Solicitud de Patente de Estados Unidos n.º 60/995.566.

### Ejemplo 3: Validación funcional de ZFN específicas de EPSPS en células HEK293

Se ensayó la capacidad de las ZFN de EPSPS tal como se describen en la presente memoria para facilitar la recombinación de homólogos en el sistema de GFP descrito en Urnov (2005) Nature 435(7042):646-51 y en la Publicación de Patente de Estados Unidos n.º 20050064474 (por ejemplo, los ejemplos 6-11). En resumen, se generaron células indicadoras HEK 293 que portaban la región génica de EPSP de interés del modo siguiente. Se amplificó la región génica de EPSP de interés mediante la PCR y posteriormente se clonó en pcDNA4TO- GFP. Se transfectoron células HEK 293 con el plásmido anterior y posteriormente se seleccionaron a las 48 horas después de la transfección, en presencia de 400µg/ml de zeocina.

Entonces se ensayaron agrupaciones de clones estables obtenidos con las ZFN dirigidas hacia la región específica de interés del gen de EPSP en las líneas celulares indicadoras generadas anteriormente del modo siguiente. Se sembraron las líneas celulares indicadoras a 350.000 células/pocillo en una placa de 12 pocillos en medio de 1 ml de DMEM, FBS al 10 % (sin PSG) y se transfectoron con 50 o 100 ng de cada ZFN y 500 ng del donante de GFP sin promotor (Urnov (2005) Nature) en 500.000 células indicadoras, usando 2 µl de lipofectamina 2000 (Invitrogen) por muestra, según el protocolo de lipofectamina 2000 de Invitrogen. Las transfecciones para cada par de ZFN se llevaron a cabo en triplicados. Un día después de la transfección, se añadió 1 ml de medio DMEM con 1,5 µl de vinblastina a una concentración final de 0,2 µM a 1 mM a cada pocillo y se retiró 72 horas después de la transfección. Las células se ensayaron respecto de la expresión de GFP 5 días después de la transfección midiendo 40.000 por transfección en el analizador FACS Guava benchtop. Los resultados ilustrativos se muestran en la fig. 14, paneles A a E.

### Ejemplo 4: Una ZFN puede escindir dos o más parálogos de EPSPS en *B. napus var. Nex710*

Para evaluar la funcionalidad de las proteínas de nucleasa de dedo de cinc en células vegetales, se utilizaron métodos para la expresión de dichas proteínas en células vegetales vivas. El ADN que codifica proteínas nucleasas de dedo de cinc puede suministrarse a células vegetales en condiciones donde el ADN no se incorpora al genoma de la célula vegetal. Por lo tanto, la molécula de ADN se mantiene de manera transitoria en las células vegetales y actúa como molde para la expresión génica. Como alternativa, puede suministrarse el ADN que codifica las

proteínas nucleasas de dedo de cinc en condiciones que permiten que el ADN se incorpore en el genoma de la célula vegetal, dando como resultado la transgénesis de los genes que codifican la nucleasa de dedo de cinc de tal forma que la molécula de ADN se mantiene de manera estable en las células vegetales y actúa como molde para la expresión génica. Un experto en la materia puede utilizar expresión transitoria o transgénica de ADN que codifican nucleasas de dedo de cinc para evaluar la funcionalidad de estas proteínas en células vegetales vivas.

A. Diseño del vector

Se construyeron vectores plasmídicos para la expresión de proteínas ZFN en células de *B. napus*. Para optimizar la expresión y la estequiometría relativa de las 2 proteínas distintas necesarias para formar un heterodímero de nucleasa de dedo de cinc funcional, se adoptó una estrategia de expresión que dio como resultado la inserción de las fases abiertas de lectura de ambos monómeros de ZFN en un solo vector, dirigidas por un solo promotor. Esta estrategia exploró la funcionalidad de una secuencia 2A (Mattion, et al. (1996) J. Virol. 70, 8124-8127) procedente del virus *Thesoa assigna*, una señal de localización nuclear (NLS) del virus SMV (PKKKRKV (SEQ ID NO: 15); Kalderon et al. (1984a) Nature 311: 33-38; Kalderon et al. (1984b) Cell 39: 499-509) o una NLS del gen *opaque-2* (*op-2*; Maddaloni et al. (1989) Nucleic Acids Research 17:7532; Van Eenennaam et al. (2004) Metabolic Engineering 6:101-108) y un promotor procedente del virus del mosaico de la vena de la cassava o CsVMV (véase la tabla 3).

**Tabla 3.** Descripción de pares de ZFN y elementos de expresión presentes en diversas construcciones que se usaron para la transformación de *B. napus*.

N.S.	Par de ZFN	Número de construcción	Tipo de construcción	Casetes génicos
1	10654-v2/10657-v2	pDAB7147	Binaria	CsVMV/par de ZFN/AtuORF23// AtUbi 10/PabAtuORF1
2	10654-v3/10658-v3	pDAB7150	Binaria	CsVMV/par de ZFN/AtuORF23// AtUbi10/Pat/AtuORF1
3	10654-v2/10657-v2	pDAB7151	No binaria	CsVMV/par de ZFN/AtuORF23
4	10654-v3/10658-v3	pDAB7154	No binaria	CsVMV/par de ZFN/AtuORF23
5	10740-v2/10741-v2	pDAB7185	Binaria	RB7 MAR//CsVMV//par de ZFN//AtuORF23//Orf del gen ipt interrumpido (onc 4)// 4OCS delta mas 2'//PAT//AtuORF1
6	10740-v2/10742-v2	pDAB7186	Binaria	RB7 MAR//CsVMV//par de ZFN//AtuORF23//Orf del gen ipt interrumpido (onc 4)// 4OCS delta mas 2'//PAT//AtuORF1

CsVMV = Secuencia promotora y líder del virus del mosaico de la vena de la Cassava de 517 pb (Verdaguer et al. (1998) Plant Mol. Biol. 37:1055-1067); AtuORF23 =3' UTR del ORF23 de *Agrobacterium tumefaciens*; AtUbi10=promotor del gen 10 de ubiquitina de *Arabidopsis thaliana*; Pat = gen de fosfotricin acetil transferasa de *Streptomyces viridochromogenes*. Es un gen reconstruido de lo que se comunicó en la Patente de Estados Unidos 5633434); AtuORF1 = 3' UTR del ORF1 de *Agrobacterium tumefaciens* (n.º de registro de Genebank X00493, NC\_002377); RB7 MAR = región de unión a matriz del tabaco; 4OCS delta mas 2'= un promotor de manopina sintasa modificado que contiene elementos OCS 4X para potenciar la expresión; Orf del gen ipt interrumpido (onc 4) = gen *ipt* de *Agrobacterium tumefaciens* alterado (se usó la ID de secuencia de Genebank ATTMPTI para el diseño).

Se ideó una estrategia de clonación modular por etapas para desarrollar estos vectores de expresión para cualquier par dado de genes que codifican ZFN seleccionados del archivo de la biblioteca o sintetizados desde cero. En primer lugar, se modificó un vector pVAX (véase, por ejemplo, la Publicación de Patente de Estados Unidos 2005-0267061) para que abarcara el dominio de expresión N-terminal, tal como se muestra en las figuras 2A-2E. Las características de este plásmido modificado (pVAX-N2A-NLSop2-EGFP-FokMono) (figure 2A) incluyen un segmento rediseñado y sintetizado que codifica una NLS y un segmento rediseñado y sintetizado que codifica el dominio de nucleasa de *FokI* utilizando el sesgo de codones de dicot. Además, una sola inserción de nucleótidos (C) cadena abajo del único sitio de *Xho I* creó un sitio de *Sac I* adicional por conveniencia para la clonación.

En segundo lugar, también se modificó un vector pVAX (véase, por ejemplo, la Publicación de Patente de Estados

Unidos 2005-0267061) para abarcar el dominio de expresión C-terminal. Las características de este plásmido modificado (pVAX-C2A-NLSop2-EGFP-FokMono) (figura 2B) incluyeron un segmento rediseñado y sintetizado que codifica una NLS y un segmento rediseñado y sintetizado que codifica el dominio de nucleasa de *FokI* utilizando el sesgo de codones de dicot. Además, se introdujo la secuencia de 2A del virus *Thosea asigna* virus (EGRGSLTTCGDVEENPGP, SEQ ID NO: 16) en el extremo N-terminal del ORF de ZFN con el fin de la unión posterior de los dos dominios codificantes de la proteína.

Se clonaron los casetes génicos que codificaban los ORF de proteínas de dedo de cinc individuales en el vector N2A o C2A mediante ligamiento usando las enzimas de restricción *Kpn I* y *BamH I* para crear extremos compatibles. A continuación, se insertó el fragmento *Bgl II/Xho I* del vector C2A en el vector N2A usando los mismos sitios de restricción, produciendo una construcción intermedia que contenía un casete que incluía 2 dominios codificantes de ZFN flanqueados por los sitios de restricción de *Nco I* y *Sac I* (figura 2C).

Finalmente, se cortó el casete *Nco I/Sac I* de esta construcción intermedia (figura 2C), que contenía ambos genes de ZFN, usando esas enzimas de restricción y se ligó en un armazón de plásmido pDAB3731. Los plásmidos resultantes, tales como pDAB7151 (figura 2D), incluyeron los genes de ZFN más las secuencias promotoras y terminadoras relevantes, más marcadores de selección para el mantenimiento del plásmido (tabla 2). Las secuencias se confirmaron mediante digestión de enzima de restricción y secuenciación. En esta construcción, el casete de expresión de ZFN (incluyendo los elementos promotores y terminadores) está flanqueado por sitios attL para una manipulación conveniente usando el sistema GATEWAY de Invitrogen (Carlsbad, CA). Cada una de las construcciones de ZFN generadas usando esta estrategia de clonación se transformó en células de *E. coli* DH5 $\alpha$  (Invitrogen, Carlsbad, CA) y posteriormente se mantuvieron en la selección adecuada.

Para la transformación de plantas mediada por *Agrobacterium*, se clonaron los casetes de ZFN en una construcción binaria usando la reacción GATEWAY LR CLONASE (Invitrogen, n.º de cat. 11791-019). La construcción binaria resultante (figura 2E) se confirmó mediante digestión de enzima de restricción y después se transformó en la cepa Z707s de *Agrobacterium tumefaciens*. Se confirmaron las colonias que contenían el gen mediante digestión con enzima de restricción y reacción de secuenciación.

#### B. Sistemas de expresión transitoria y estable

Se generaron preparaciones de plásmidos de construcciones de expresión de ZFN, tales como pDAB 7151, tal como se ilustra en la figura 2, a partir de cultivos de 2 l de *E. coli* cultivadas en medio LB que contenía antibióticos usando un kit GIGAPREP sin endonucleasas (Qiagen, Valencia, CA) según las recomendaciones del fabricante. El ADN plasmídico se suministró directamente en células de hipocotilo de *B. napus* usando una diversidad de métodos.

En un ejemplo de suministro transitorio de ZFN, se sometieron segmentos de hipocotilo de colza a suministro de ADN mediante transformación transitoria mediada por whiskers de segmentos de hipocotilo. Las semillas de *B. napus*, var. Nex710 se esterilizaron superficialmente con CLOROX al 10 % (v/v) (hipoclorito de sodio al 5,25 %) durante 10 minutos y se enjuagó 3 veces con agua destilada estéril. Posteriormente, se germinaron las semillas en medio MS de concentración ½ (sales basales con vitaminas 1/2 MS, sacarosa al 1 %, agar al 0,8 %, pH 5,8) contenidas en Phytatrays con 25 semillas por Phytatray. Las semillas se colocaron en una sala de cultivo para su germinación durante 5 días a 23 °C con un fotoperiodo de 16 horas de luz, 8 horas de oscuridad. En el día 5, se extirparon asépticamente segmentos de hipocotilo de 3 mm de longitud y se colocaron en agua estéril para evitar el secado mientras se cortaban segmentos adicionales. Se desecharon las secciones del tallo y la raíz. Las secciones se colocaron horizontalmente sobre un trozo de papel de filtro estéril, reposando sobre la superficie del medio MSK1D1 (sales basales con vitaminas MS, cinetina 1 mg/l, ácido 2,4-diclorofenoxiacético [2,4-D] 1 mg/ml, sacarosa 30 g/l, TC-agar 7 g/l, pH 5,8). Los segmentos se cultivaron durante 3 días a 23 °C y 16 horas de luz.

En el día del tratamiento de whiskers, se colocaron 300 segmentos de hipocotilo parcialmente encallados en una botella Sorvall junto con 30 ml de 'medio alto osmótico' (sales MS con vitamina B5, 4,42 mg/l de 2,4-D y sacarosa al 12 %) durante un pretratamiento de una hora a temperatura ambiente. Este pretratamiento es un medio para plasmizar parcialmente el tejido en un intento para mejorar el daño celular cuando se penetró la pared celular durante el posterior tratamiento de whiskers. Posteriormente, se añadieron 8,1 ml de solución de whisker de carburo de silicio Silar SO-9 al 5 % (Advanced Composite Materials, LLC Greer, SC) y 170  $\mu$ g del ADN de plásmido de ZFN no binario (tabla 2) preparado como se ha descrito anteriormente a la botella Sorval. Entonces se agitó vigorosamente la botella durante 30 segundos en un mezclador de pinturas (Red Devil Equipment Co., Minneapolis, MN) en el que el conjunto de pinza para la lata de pintura se ajustó para sujetar una botella Sorval. Después de la agitación, se añadieron inmediatamente 100 ml de 'medio alto osmótico' a la botella que después se dejó recuperar durante 20 minutos a temperatura ambiente. Entonces se recogieron los segmentos vertiendo los contenidos de la botella a través de una malla de alambre estéril de tamaño adecuado para separar los segmentos de los whiskers y el contenido líquido de la botella. Finalmente, se volvieron a colocar los segmentos en una placa reciente de medio MSK1D1 con papel de filtro. Se tomaron muestras de aproximadamente 100 mg para el análisis de expresión transitoria en los días 1, 2, 3 y 7 después del tratamiento de.

En otro ejemplo, el sistema de suministro transitorio usó transformación mediada por polietilenglicol (PEG) de protoplastos de hipocotilo. Los protoplastos se prepararon a partir de tejidos de hipocotilo de plántulas de *B. napus*,

var. Nex710, usando los métodos descritos por Sun et al. (Can. J. Bot. (1998) 76: 530-541) con modificaciones. Las semillas se esterilizaron superficialmente y se germinaron durante 7 días en medio ½ MS para colza tal como se ha descrito anteriormente. Se recogió un gramo de hipocotilos para cada tratamiento. Los hipocotilos se cortaron en finas secciones de  $\leq 1$  mm de tamaño y se colocaron en medio MS9m (manitol al 9 %, MES 5 mM, arginina 10 mM, polivinilpirrolidona-40 (PVP-40) al 0,3 %) contenidos en placas de Petri de 100 mm. Después de colocar todas las secciones de hipocotilo en la placa de Petri, se retiró el medio líquido con una pipeta y se reemplazó con 6 ml de solución enzimática (MS9m que contenía Macroenzima-R10 al 0,1 % (Yakult Honsha Co. Ltd, Tokio, Japón), celulasa-R10 al 1 % (Yakult Honsha Co. Ltd, Tokio, Japón), pectinasa al 1 % (Sigma Chemical Co.).

Los tejidos se colocaron en la oscuridad a 25 °C durante 16 horas con agitación suave en un agitador rotatorio a 40 rpm para digerir las paredes celulares. Después de la incubación, la solución de enzima-protoplasto se filtró en condiciones asépticas a través de un colador celular de 100  $\mu$ m (Sigma Chemical Co.) colocada sobre un tubo de centrifugación desechable de 50 ml. La solución se centrifugó a 50  $\times$  g durante 5 minutos. Después de desechar el sobrenadante, se resuspendió el sedimento de protoplastos en 4 ml de medio MS9m. La suspensión de protoplastos se dispuso suavemente en capas sobre 4,5 ml de MS con solución de sacarosa 0,5 M en un tubo de centrifuga de 15 ml y se centrifugó a 50  $\times$  g. Los protoplastos se extrajeron con una micropipeta de una banda espesa ubicada en la interfase y se lavaron con 5 ml de medio MS9m mediante centrifugación a 50  $\times$  g durante 5 minutos.

Para el tratamiento del ADN, se resuspendió el sedimento de protoplastos en 200  $\mu$ l de solución de Mg-manitol hasta una concentración final de  $1 \times 10^5$  protoplastos/ml. Se añadió una muestra de 50  $\mu$ l de un ADN de plásmido no binario, tal como pDAB7151 (tabla 2), a 200  $\mu$ l de solución de protoplastos contenida en un tubo de centrifugación estéril desechable de 15 ml y se mezcló. Se añadió un volumen igual de solución de PEG-3350 al 40 % (Sigma Chemical Co) a la solución de protoplastos y se incubó a temperatura ambiente durante 20 minutos. Posteriormente, se añadieron 0,8 ml de medio W5 (CaCl<sub>2</sub>·H<sub>2</sub>O 125 mM, NaCl 154 mM, KCl 5 mM y glucosa 5 mM) y se incubó durante 10 minutos adicionales seguido de centrifugación a 180  $\times$  g durante 3 minutos. Se retiró la solución de PEG con una pipeta y se resuspendieron los protoplastos en 1 ml de solución WI. Entonces se incubaron los tubos en la oscuridad durante aproximadamente 18 horas seguido de centrifugación a 180  $\times$  g durante 3 minutos. Se retiró el sobrenadante y se transfirieron 100  $\mu$ l de suspensión de protoplastos a un tubo Eppendorf de 2 ml. Se recogieron los protoplastos ( $10^5$ /ml) a los 0, 1, 2, 3 días después del tratamiento de ADN y se almacenaron a -80 °C hasta su análisis. Se mezcló una muestra de 10  $\mu$ l de la suspensión de protoplastos con colorante de diacetato de fluoresceína y se contaron los protoplastos viables en un hemocitómetro.

En otro ejemplo, el sistema de suministro transitorio usó transformación mediada por *Agrobacterium* de los segmentos de hipocotilo. Los segmentos se cultivaron en papel de filtro estéril en medio de inducción de callos MSK1D1 y se les administró un pretratamiento de 3 días tal como se ha descrito anteriormente para el protocolo de whiskers. El día antes del tratamiento con *Agrobacterium*, se inoculó un cultivo bacteriano (1 bucle) de un plásmido binario, tal como pDAB7147 (tabla 2) en un matraz que contenía 35 ml de medio YEP que contenía los antibióticos adecuados. Se dejó que creciese el cultivo bacteriano durante toda la noche durante  $\sim 16$  h en la oscuridad a 28 °C con agitación constante a 200 rpm. Al día siguiente, se preparó la solución de *Agrobacterium* a una concentración final de Klett 50 en medio M líquido. Se transfirieron los segmentos de hipocotilo del papel de filtro a placas de Petri de 100  $\times$  25 mm que contenían 40 ml de suspensión de *Agrobacterium* y se inocularon durante 30 min a temperatura ambiente con agitación en remolino periódica cada 10 minutos. Al final del periodo de tratamiento, se retiró la solución de *Agrobacterium* y se transfirieron de nuevo los segmentos de hipocotilo a las placas originales que contenían medio MSK1D1 con papel de filtro. Se cultivaron conjuntamente los segmentos durante 3 días en Percival o en la sala de cultivo con intensidad de luz reducida cubriendo las placas con film de aluminio.

Después de 3 días de cultivo conjunto, se transfirieron los segmentos a un medio de inducción de callos MSK1D1TC (MS, 1 mg/l de cinetina, 1 mg/l de 2,4-D, 0,5 mg/l de MES, 5 mg/l de AgNO<sub>3</sub>, 300 mg/l de Timentina, 200 mg/l de Carbenicilina, sacarosa al 3 %, Phytagar al 0,7 %). Se extrajeron aproximadamente 100 mg de tejidos de hipocotilo a los 0, 2, 3, 4 y 7 días después del tratamiento con *Agrobacterium* y se almacenaron a -80 °C hasta su análisis.

En un ejemplo de suministro de ZFN usando un sistema de expresión transgénica estable, se esterilizaron superficialmente semillas de *Brassica napus* var. Nex710, se germinaron durante 5 días, se prepararon segmentos de hipocotilo de  $\leq 1$  mm y se pretrataron durante 3 días tal como se ha descrito para el tratamiento de whiskers. Después de 3 días, se trataron los segmentos de hipocotilo con cualquiera de las cepas binarias de *Agrobacterium* (tabla 2) y se cultivaron conjuntamente durante 3 días tal como se ha descrito para el sistema de expresión transitoria mediada por *Agrobacterium* como se ha descrito anteriormente.

Después de 3 días de cultivo conjunto, se transfirieron 300 segmentos de hipocotilo a medio de inducción de callos MSK1D1H1 (MS, 1 mg/l de cinetina, 1 mg/l de 2,4-D, 0,5 mg/l de MES, 5 mg/l de AgNO<sub>3</sub>, 300 mg/l de Timentina, 200 mg/l de Carbenicilina, 1 mg/l de Herbiace, sacarosa al 3 %, Phytagar al 0,7 %) con un bajo nivel de selección de herbicida durante 7 días. Los segmentos de hipocotilo se transfirieron entonces a medio MSK1D1H3 que contenía niveles mayores de selección (MS, 1 mg/l de cinetina, 1 mg/l de 2,4-D, 0,5 mg/l de MES, 5 mg/l de AgNO<sub>3</sub>, 300 mg/l de Timentina, 200 mg/l de Carbenicilina, 3 mg/l de Herbiace, sacarosa al 3 %, Phytagar al 0,7 %) durante 2 semanas y después se transfirieron a medio MSK1D1H5 (MS, 1 mg/l de cinetina, 1 mg/l de 2,4-D, 0,5 mg/l de MES, 5 mg/l de AgNO<sub>3</sub>, 300 mg/l de Timentina, 200 mg/l de Carbenicilina, 5 mg/l de Herbiace, sacarosa al 3 %, Phytagar al 0,7 %) durante otras 2 semanas. Se obtuvieron un total de 203 y 227 líneas de callos para cada una de las construcciones

binarias (tabla 2), proporcionando una frecuencia de callos del 67,6 % y del 75,5 % respectivamente. Se sometió a cincuenta líneas de callos aleatorias a análisis de ADN después de 5-7 semanas tras el tratamiento con *Agrobacterium*.

C: Análisis de parálogo de EPSPS para la escisión bicatenaria mediada por ZFN dirigida

5 Se entiende en este ejemplo que la funcionalidad de una ZFN incluye la capacidad de una ZFN para expresarse en células de una especie de cultivo y para esa ZFN mediar una rotura bicatenaria (ds) en el genoma endógeno de ese cultivo mediante el reconocimiento de, la unión a y la escisión de su diana deseada. También se entiende que, en este ejemplo, la diana de la ZFN es un gen en un locus endógeno y su conformación en el genoma del cultivo. Para evaluar si las ZFN modificadas tienen funcionalidad contra el gen diana predicho en un contexto genómico, se emplearon ensayos basados en la secuencia de ADN. Se predice que las roturas ds inducidas por la ZFN inducen mecanismos de reparación, tales como unión de extremos no homólogos (NHEJ) (revisado por Cahill et al. (2006) Front Biosci. 1:1958-1976). Un resultado de la NHEJ es que una proporción de las hebras de ADN rotas se repararán de una manera imperfecta, dando como resultado pequeñas eliminaciones, inserciones o sustituciones en el sitio de escisión. Un experto en la materia puede detectar estos cambios en la secuencia de ADN mediante una diversidad de métodos.

Para la identificación de las NHEJ en los parálogos de EPSPS, se desarrollaron ensayos específicos de genes con estrategias basadas en la PCR. Las suficientes diferencias de secuencia en cuatro de los cinco parálogos de EPSPS, A, B, C y D, permitieron el desarrollo de ensayos específicos de parálogo. Las secuencias de los parálogos D y E no pudieron diferenciarse lo suficiente en el locus usado como diana, lo que dio como resultado el desarrollo de solo un ensayo que representase a ambos parálogos. Se llevó a cabo amplificación PCR usando cebadores oligonucleotídicos específicos para el gen diana y que flanqueaban el sitio de escisión predicho en la ZFN. Los cebadores PCR específicos de parálogo fueron los siguientes:

Parálogo A:

Cebador orientación directa: 5'-TCCCAGCTTCTTTAGATTCTAAGG-3' (SEQ ID NO: 17)

25 Cebador orientación inversa: 5'-CTGCAACTTTTCACATAGCAA-3' (SEQ ID NO: 18)

Parálogo B:

Cebador orientación directa: 5'-CAAGAGTGATATCGAGTTGTACCTTGGGAATGCT-3' (SEQ ID NO: 19)

Cebador PCR orientación inversa: 5'-AGGCCATCATATCGAGCAAACGCAGT-3' (SEQ ID NO: 20)

Parálogo C:

30 Cebador orientación directa: 5'-GGGTAAACAACCGTGCTGTA-3' (SEQ ID NO: 21)

Cebador orientación inversa: 5'-AAAGACTGCTGCAAACAAGATC-3' (SEQ ID NO: 22)

Parálogo D/E:

Cebador orientación directa: 5'-GGTTGTTGAAGGATGCGGT-3' (SEQ ID NO: 23)

Cebador orientación inversa: 5'-GCAAACAATTCATAGAGTAAATGTG-3' (SEQ ID NO: 24)

35 Todos los cebadores PCR de orientación directa e inversa se usaron en combinaciones para un parálogo dado para amplificar bien ADN genómico purificado o ADN plasmídico de control positivo que contenía cada uno de los parálogos en las siguientes condiciones: volumen de reacción de 25 µl que contenía 2,5 µl de molde de ADN (10 ng/µl) o control positivo de ADN plasmídico (1 ng/µl), 0,625 µl de cada cebador (a 10 µM cada uno), 2,5 µl de tampón PCR ACCUPRIME II 10x y 0,15 µl (0,75 unidades) de ADN polimerasa Taq ACCUPRIME (Invitrogen, Carlsbad, CA) en el tampón enzimático del fabricante. La amplificación se llevó a cabo en el dispositivo ICYCLER IQ (Bio-Rad, Hercules, CA) usando los siguientes parámetros: 94 °C durante 2 minutos, 35 ciclos de (94 °C durante 30 segundos, hibridación (véase el gradiente a continuación) durante 30 segundos, 68 °C durante 1 minuto), 68 °C durante 10 minutos, 4 °C retención indefinida.

45 Se ejecutó un gradiente para determinar las condiciones de reacción óptimas. La temperatura del gradiente fue de entre 65,0 °C y 50,0 °C. Los parálogos A, B, C y D/E mostraron la mejor amplificación a temperaturas de hibridación de 62,1 °C, 65,0 °C, 65,0 °C y 59,3 °C, respectivamente (figuras 3A-3D) y estas temperaturas se usaron en estudios posteriores. Los productos de la PCR para los 4 parálogos se clonaron en el vector TOPO PCR 2.1 (Invitrogen, Carlsbad, CA) y se confirmaron mediante secuenciación para amplificación específica de parálogo.

50 Se compararon los tres métodos transitorios y un método estable de suministro de ZFN a células de hipocotilo, como se ha descrito anteriormente, para identificar el método más eficaz para evaluar la eficacia de la ZFN (determinada por la presencia del mayor número de NHEJ). Se observó que las ZFN específicas del parálogo D eran eficaces en

las células 293 de riñón usadas en este estudio. Se predijo que estas proteínas ZFN se unirían a 2 secuencias cortas específicas del gen EPSPS del parálogo D para crear una nucleasa heterodimérica que escinde el ADN bicatenario (figura 4). Estos genes de ZFN estaban presentes en cuatro construcciones; dos construcciones binarias, pDAB7147 y pDAB7150, que eran específicas para transformación mediada por *Agrobacterium* y las otras dos construcciones restantes, pDAB7151 y pDAB7154, para transformación transitoria, (tabla 2). El tejido calloso transformado de manera estable se categorizó adicionalmente en muestras "verdes" y "marrones", con la posibilidad de mayor expresión de ZFN en el tejido "marrón" y de este modo, la probabilidad de mayores frecuencias de NHEJ. Esto último puede haber dado lugar a toxicidad celular que hace que los tejidos viren a "marrón." Se recogieron todas las muestras (véase la sección 4B), incluyendo los controles no tratados, se congelaron y se liofilizaron a excepción de las muestras de protoplastos, que se usaron directamente para el aislamiento de ADN. Se aisló el ADN genómico mediante el método Qiagen como se ha descrito anteriormente. Se digirieron 3 µg de todo el ADN genómico con 5 unidades de endonucleasa de restricción *Mae III* (Roche Applied Science, Indianapolis, IN) durante toda la noche según las recomendaciones del fabricante.

Entonces se purificó el ADN mediante precipitación de etanol añadiendo 0,1 volúmenes de acetato de sodio 3 M, pH 5,2 y 2 volúmenes de etanol al 100 % seguido de centrifugación en una microcentrifugadora durante 5 minutos a 10.000 g. Entonces se lavó el ADN con etanol al 70 % y el sedimento se secó en un evaporador SPEEDVAC (Savant) y se resuspendió en agua. Entonces se sometió al ADN a una segunda digestión de *Mae III* durante toda la noche y se precipitó con etanol como se ha indicado anteriormente. El sitio de la enzima de restricción está localizado entre las dos ubicaciones de unión de la ZFN monomérica de un par (figura 4) y próximo a donde dimerizan los dominios de *FokI* e induce roturas bicatenarias en el ADN genómico. De este modo, la digestión de enzima de restricción enriquece respecto de fragmentos que han sufrido NHEJ dando como resultado la pérdida de sitios de reconocimiento de enzimas de restricción.

Entonces se llevó a cabo la amplificación PCR usando cebadores oligonucleotídicos específicos para el parálogo D y flanqueando al sitio de escisión previsto de la ZFN. Se usaron en combinación el cebador PCR de orientación directa (5'-GGTTGTTGAAGGATGCGGT-3') (SEQ ID NO: 23) y el cebador PCR de orientación inversa (5'-GCAAACAATTCATAGAGTAAATGTG-3') (SEQ ID NO: 24) específicos para el parálogo D de EPSPS usado como diana para amplificar el ADN genómico en las siguientes condiciones: volumen de reacción de 55 µl que contenía 10 µl de molde de ADN digerido con *Mae III* (26,4 ng), 1,25 µl de cada cebador (a 10 µM cada uno), 5 µl de tampón PCR ACCUPRIME II 10x, 5 µl de PVP-40 al 10 % y 1 µl (5 unidades) de ADN polimerasa Taq ACCUPRIME (Invitrogen, Carlsbad, CA). Los productos de amplificación del tamaño esperado fueron el resultado de ciclos de amplificación que consistían en 94 °C durante 2 minutos, 40 ciclos de (94 °C durante 30 segundos, 59,3 °C durante 30 segundos, 68 °C durante 1 minuto), 68 °C durante 10 minutos y 4 °C retención indefinida. Los fragmentos amplificados se clonaron directamente en el vector pCR2.1 usando el kit de clonación TA de Invitrogen (Carlsbad, CA).

Se secuenciaron aproximadamente 90 fragmentos clonados individuales por instante por tratamiento usando los sitios de cebado M13 directo y M13 inverso presentes en el vector PCR 2.1. Se incluyeron controles no tratados para un tratamiento dado. De este modo se secuenciaron aproximadamente 3000 clones.

El análisis de todos los resultados de secuenciación entre dos tratamientos de ZFN diferentes reveló 13 clones (confirmados mediante cebadores de secuenciación tanto directos como inversos) que contenían una eliminación pequeña precisamente en el sitio de escisión predicho de la ZFN presente en pDAB7151, lo que indica que el mecanismo de NHEJ había mediado una reparación imperfecta de la secuencia de ADN en ese sitio (figura 5). Estos clones particulares se obtuvieron de las muestras de ADN de protoplasto 3 días después de la transformación mediante ZFN. Estos resultados demostraron la capacidad de las ZFN modificadas para inducir roturas bicatenarias dirigidas de un modo específico en un locus génico diana dentro de una especie de cultivo. No se observaron NHEJ en cualquier otro método de tratamiento con ZFN con este tipo de secuenciación.

#### D. Análisis de secuenciación masivamente paralelo

En otro ejemplo, se aplicó una combinación de métodos PCR y de pirosecuenciación masivamente paralela interrogar al parálogo D en las muestras obtenidas anteriormente. Se usó el mismo conjunto de cebadores directos e inversos específicos del parálogo D (SEQ ID NO: 23 y SEQ ID NO: 24) para amplificar el ADN de todas las muestras que representaban 3 métodos transitorios y un método estable de transformación por ZFN en células de hipocotilo de colza. Las condiciones de amplificación fueron como las descritas anteriormente.

Entonces se purificó este producto de amplificación primario usando el kit de purificación MINELUTE PCR (Qiagen, Valencia, CA) y se eluyó el ADN en 10 µl. Se diseñó un segundo conjunto de cebadores anidados para amplificar un fragmento de aproximadamente 100 pb adecuado para secuenciación masivamente paralela. Se sintetizaron seis variantes del cebador PCR de orientación directa (5'-XXX AGTTGTACCTTGGAATG-3') (SEQ ID NO: 25) en el que XXX = GGC, CGC, GGC, CGG, CCG o GCG y seis variantes del cebador PCR de orientación inversa (5'-XXX ATCAATTTCTTGACAATAACA-3') (SEQ ID NO: 26) en el que XXX = GGC, CGC, GGC, CGG, CCG o GCG y se purificaron mediante HPLC (IDT, Coralville, IA). Los marcadores de 3 pb en el extremo 5' de cada cebador sirvieron como clave de identificación e indicaron de qué muestra de células procedía el amplicón. Se usaron pares de cebadores con etiquetas de identificación emparejadas (claves) en combinación para amplificar el amplicón de PCR

primario purificado procedente de las muestras descritas anteriormente en las siguientes condiciones: volumen de reacción de 50 µl que contenía 10 µl de amplicón PCR purificado diluido a 1:10, 1,25 µl de cada cebador (10 µM cada uno), 5 µl de tampón PCR ACCUPRIME I 10x y 0,3 µl (1,5 unidades) de ADN polimerasa de alta fidelidad Taq ACCUPRIME (Invitrogen, Carlsbad, CA) en el tampón enzimático del fabricante. Los productos de amplificación del tamaño esperado fueron el resultado de los ciclos de amplificación que consistían en 94 °C durante 2 minutos, 30 ciclos de (94 °C durante 30 segundos, 62 °C durante 30 segundos, de 68 °C durante 30 segundos), 68 °C durante 5 minutos y 4 °C de retención indefinida y se purificaron usando el kit de purificación MINELUTE PCR (Qiagen, Valencia, CA) según las recomendaciones del fabricante.

Se llevaron a cabo las reacciones de pirosecuenciación masivamente paralelas (también conocida como secuenciación 454) directamente en los productos de la PCR según se describe en (Margulies et al. (2005) Nature 437:376-380). El análisis de los resultados de la secuenciación 454 se llevó a cabo identificando lecturas de secuencia que contenían eliminaciones del tamaño y en la posición esperada dentro de la molécula de ADN. Los resultados de estos análisis indicaron la presencia de múltiples eliminaciones pequeñas de 9-12 pb en el sitio de escisión previsto para estas ZFN, tal como se muestra en la figura 6. Se observaron cuarenta y seis de las cuarenta y ocho eliminaciones en las lecturas de secuencia obtenidas a partir de los callos verdes transformados de manera estable con la construcción de ZFN, pDAB7147 (tabla 3). Se obtuvieron dos eliminaciones adicionales con el mismo par de ZFN, una del ADN de protoplasto tratado de manera transitoria (pDAB7151) y otra del tejido de hipocotilo tratado de manera transitoria con *Agrobacterium* (pDAB7147). Estas eliminaciones se ubicaron de manera precisa en el sitio diana de la ZFN e indicaron que se generaron las roturas ds inducidas por la ZFN, que posteriormente se repararon mediante el mecanismo de NHEJ.

Ya que los parálogos D y E eran indistinguibles mediante el ensayo PCR usado en este ejemplo, es posible que se escindiese uno cualquiera o ambos de los parálogos por la ZFN. Estos resultados demuestran además la capacidad de las ZFN modificadas para inducir roturas bicatenarias dirigidas de una manera específica en un locus génico endógeno en una especie de cultivo. Además, se demostró que el método de transformación por ZFN estable era el método más eficaz para explorar NHEJ en las presentes condiciones experimentales. Las ZFN presentes en las construcciones pDAB7150 y pDAB7154 no mostraron eliminaciones entre las múltiples muestras tratadas con los diferentes métodos de transformación (véase la tabla 4).

**Tabla 4.** Los resultados de secuenciación masivamente paralela que muestran NHEJ en las secuencias diana del parálogo D de EPSPS obtenidas mediante la transformación transitoria y estable de segmentos de hipocotilo de *B. napus* con las ZFN pDAB7147 y pDAB7151. La muestra de control estaba compuesta de tejidos no tratados con las ZFN.

N.M.	Muestra	n.º de construcción de ZFN	n.º de muestra - cebador	Secuencias analizadas	n.º de NHEJ
1	Control	pDAB7147 + pDAB7151	1-Directo	18.215	0
2	Control	pDAB7147 + pDAB7151	1-Inverso	18.922	0
3	Protoplasto	pDAB7151	2-Directo	45.896	0
4	Protoplasto	pDAB7151	2-Inverso	48.606	1
5	Whiskers	pDAB7151	3-Directo	19.601	0
6	Whiskers	pDAB7151	3-Inverso	19.628	0
7	Agro. transitoria	pDAB7147	4-Directo	31.281	1
8	Agro. transitoria	pDAB7147	4-Inverso	31.595	0
9	Agro estable -G	pDAB7147	5-Directo	13.795	29
10	Agro estable -G	pDAB7147	5-Inverso	13.221	17
11	Agro estable - B	pDAB7147	6-Directo	8.167	0
12	Agro estable - B	pDAB7147	6-Inverso	7.549	0
	<b>Total</b>			<b>276.476</b>	<b>48</b>

En un esfuerzo para analizar las roturas bicatenarias inducidas por ZFN en otros parálogos de EPSPS, se llevó a cabo una combinación de PCR y de pirosecuenciación masivamente paralela para interrogar al ADN de los demás parálogos de EPSPS respecto de roturas bicatenarias inducidas por ZFN. Se empleó el mismo ADN genómico

digerido con *MaeIII* de los mismos callos "verdes" estables transformados con la cepa de *Agrobacterium* que contenía pDAB7147 y pDAB7150, como se describe en las secciones 4D-E. Entonces se llevó a cabo la amplificación PCR con cebadores oligonucleotídicos específicos para los parálogos de EPSPS A, B, C y D que se anclaron al ADN genómico que flanqueaba al sitio de escisión predicho de la ZFN. Se usaron en combinación un cebador PCR de orientación directa para el parálogo A (5'-TCCCAGCTTCTTTAGATTCTAAGG-3') (SEQ ID NO: 17) y un cebador de PCR de orientación inversa (5'-CTGCAACTTTTCACATAGCAA -3') (SEQ ID NO: 18), un cebador PCR de orientación directa para el parálogo B (5'-CAAGAGTGATATCGAGTTGTACCTTGGGAATGCT-3') (SEQ ID NO: 19) y un cebador PCR de orientación inversa (5'-AGGCCATCATATCGAGCAAACGCAGT-3') (SEQ ID NO: 20), un cebador PCR de orientación directa para el parálogo C (5'-GGGTAAACAACCGTGCTGTA-3') (SEQ ID NO: 21) y un cebador PCR de orientación inversa (5'-AAAGACTGCTGCAAACAAGATC -3') (SEQ ID NO: 22) y el mismo conjunto de cebadores directos e inversos para el parálogo D (SEQ ID NO: 23 y SEQ ID NO: 24), como se ha descrito en las secciones 4 D-E, para amplificar el ADN genómico para cada uno de los parálogos por separado en las siguientes condiciones: volumen de reacción de 50 µl que contiene 200 ng de molde de ADN digerido con *Mae III* (10 µl), 1,25 µl de cada cebador (a 10 µM cada uno), 5 µl de tampón PCR Accuprime II 10x, 5 µl de PVP-40 al 10 % y 0,3 µl (1,5 unidades) de ADN polimerasa de alta fidelidad Taq ACCUPRIME (Invitrogen, Carlsbad, CA) en el tampón enzimático del fabricante. Los productos de amplificación se generaron a partir de ciclos de amplificación que consistían en 94 °C durante 2 minutos, 25 ciclos de (94 °C durante 30 segundos, hibridación durante 30 segundos, 68 °C durante 1 minuto), 68 °C durante 5 minutos, 4 °C retención indefinida. Las temperaturas de hibridación fueron como sigue para los siguientes parálogos A = 62,1 °C, B = 65 °C, C = 65 °C y D = 59,3 °C.

Entonces se purificó este producto de amplificación primario usando el kit de purificación PCR MinElute de Qiagen (Qiagen, Valencia, CA) y se eluyó en 10 µl de tampón EB. Se sintetizaron tres variantes de un cebador de PCR de orientación directa de parálogo A (5'-XXX ATCGAGTTGTACCTTGGGAATG -3') (SEQ ID NO: 27) en el que XXX = GGC, CGG o GCC y tres variantes de un cebador PCR de orientación inversa de parálogo A (5'-XXX AATAAGTCCTTAACCTTACCTT-3') (SEQ ID NO: 28) en el que XXX = GGC, CGG o GCC y se purificaron mediante HPLC (IDT, Coralville, IA). Se sintetizaron tres variantes de cebador PCR de orientación directa de parálogo B (5'-XXX AGAGTGATATCGAGTTGTACCTTG -3') (SEQ ID NO: 29) en el que XXX = CGG, CGC o GCC y tres variantes de un cebador PCR de orientación inversa de parálogo B (5'-XXX AACTCCTTAACCTTACCTT -3') (SEQ ID NO: 30) en el que XXX = CGG, CGC o GCC y se purificaron mediante HPLC (IDT, Coralville, IA). Se sintetizaron tres variantes de cebador PCR de orientación directa de parálogo C (5'-XXX AGAGTGATATTGAGTTGTACCTTG-3') (SEQ ID NO: 31) en el que XXX = CGG, GGC o GCC y tres variantes de un cebador PCR de orientación inversa de parálogo C (5'-XXX AAAGCTCCTTAACCTTTACCT -3') (SEQ ID NO: 32) en el que XXX = CGG, GGC o GCC y se purificaron mediante HPLC (IDT, Coralville, IA). Los cebadores para la amplificación PCR secundaria (SEQ ID NO: 25 and SEQ ID NO: 26) para el parálogo D se describen en la sección 4D. Las etiquetas de 3 pb en el extremo 5' de cada cebador sirvieron como clave de identificación e indicaron de qué muestra de *B. napus* provenía el amplicón. Se usaron pares de cebador con etiquetas identificadoras emparejadas (claves) en combinación para amplificar el amplicón de PCR primario purificado procedente de muestras descritas anteriormente en las siguientes condiciones: volumen de reacción de 50 µl que contenía 10 µl de amplicón PCR purificado diluido a 1:10, 1,25 µl de cada cebador (10 µM cada uno), 5 µl de tampón PCR ACCUPRIME I 10x y 0,3 µl (1,5 unidades) de ADN polimerasa de alta fidelidad Taq ACCUPRIME (Invitrogen, Carlsbad, CA) en el tampón enzimático del fabricante. Los productos de amplificación del tamaño esperado fueron el resultado de los ciclos de amplificación que consistían en 94 °C durante 2 minutos, 33 ciclos de (94 °C durante 30 segundos, 62 °C durante 30 segundos, 68 °C durante 30 segundos), 68 °C durante 5 minutos y 4 °C retención indefinida y se purificaron usando el kit de purificación MINELUTE PCR de Qiagen (Valencia, CA) según las recomendaciones del fabricante.

Las reacciones de pirosecuenciación masivamente paralela se llevaron a cabo directamente en los productos de la PCR descritos en la sección 4D. El análisis de los resultados de la secuenciación se llevó a cabo identificando lecturas de secuencia que contenían eliminaciones del tamaño y la posición esperada en la molécula de ADN.

Los resultados de estos análisis indicaron la presencia de múltiples eliminaciones pequeñas en los sitios de escisión de ZFN esperados en los parálogos C y D (figura 7). De nuevo, la ZFN pDAB7147 fue eficaz para escindir a los dos parálogos, C y D (y E). Estas eliminaciones de 5 - 32 pb se localizaron de manera precisa en el sitio diana de ZFN y demostraron la escisión de 2 o más parálogos de EPSPS por la ZFN pDAB7147. Estos resultados demuestran además la capacidad de las ZFN modificadas para inducir roturas bicatenarias dirigidas de una manera específica en un locus génico endógeno en una especie de cultivo.

También se observó una NHEJ en cada uno de los parálogos A y B de las muestras 5, 6 y 9 tratadas con pDAB7147 y pDAB7150 (tabla 5). Estas NHEJ se observaron en las ubicaciones esperadas. Sin embargo, ya que algunas de las muestras de control también contenían una NHEJ (muestras 2, 7 y 13 en la tabla 5), no se consideraron eficaces las ZFN para escindir estos parálogos.

**Tabla 5.** Los resultados de la secuenciación masivamente paralela que muestran las moléculas de ADN que han sufrido roturas bicatenarias mediadas por ZFN seguidas de reparación de NHEJ en las secuencias diana de los cuatro parálogos de EPSPS en *B. napus*. Las muestras de control representan callos transgénicos no tratados con ZFN.

Número muestra	de	Parálogo diana	Construcción de ZFN	Cebador	Secuencias alineadas	NHEJ totales
1		A	Control	Directo	18.451	0
2		A	Control	Inverso	19.005	1
3		A	pDAB7150	Directo	14.711	0
4		A	pDAB7150	Inverso	15.616	0
5		A	pDAB7147	Directo	21.478	1
6		A	pDAB7147	Inverso	22.459	1
7		B	Control	Directo	12.682	1
8		B	Control	Inverso	18.033	0
9		B	pDAB7150	Directo	11.361	1
10		B	pDAB7150	Inverso	13.017	0
11		B	pDAB7147	Directo	16.565	0
12		B	pDAB7147	Inverso	17.499	0
13		C	Control	Directo	20.516	1
14		C	Control	Inverso	14.790	0
15		C	pDAB7150	Directo	9.073	0
16		C	pDAB7150	Inverso	9.096	1
17		C	pDAB7147	Directo	12.696	3
18		C	pDAB7147	Inverso	14.719	19
23		D	pDAB7147	Directo	7.028	5
24		D	pDAB7147	Inverso	6.832	9

5

#### Ejemplo 5: Una segunda ZFN escinde los dos parálogos restantes de los cinco de *B. napus*

A continuación, se intentó la escisión bicatenaria mediada por ZFN, centrada en la inducción de roturas bicatenarias en los dos parálogos restantes A y B. Se usaron dos nuevas ZFN modificadas que se dirigían a diferentes secuencias, localizadas a aproximadamente 350 pb 5' respecto de la primera ubicación de unión de ZFN (figura 4). Estas construcciones de ZFN particulares, pDAB7185 y pDAB7186 (tabla 3), se usaron en la transformación estable de segmentos de hipocotilo de *B. napus*, tal como se describe en el ejemplo 4, sección B. Los callos transformados de manera estable con las ZFN se congelaron, se liofilizaron y se extrajo el ADN tal como se ha descrito anteriormente en el ejemplo 4, sección C. Entonces se digirió el ADN con *BsoB1* (New England Biolabs, Ipswich, MA) o *Lwe I* (Fermentas, Inc., Hanover, MD) para enriquecer respecto de los fragmentos que habían sufrido NHEJ (figura 4). Las digestiones se llevaron a cabo según las instrucciones del fabricante durante toda la noche y se purificaron mediante precipitación de etanol, tal como se ha descrito anteriormente. Entonces se llevó a cabo la amplificación PCR usando cebadores oligonucleotídicos específicos para los genes diana que flanquean al sitio de escisión predicho en la ZFN. Se usaron en combinación un cebador PCR de orientación directa para el parálogo A (5'-CAGCGTGGAGCTTATCAGA-3') (SEQ ID NO: 33) y un cebador de PCR de orientación inversa (5'-AAACGCAACTAAGCAAAC-3') (SEQ ID NO: 35), un cebador PCR de orientación directa para el parálogo B (5'-GAAGAGTAACAACGGCTCTGTG -3') (SEQ ID NO: 34) y un cebador PCR de orientación inversa (5'-GAAAGAAAGAAGCAAACCGAC-3') (SEQ ID NO: 90), específicos para los parálogos del gen EPSPS para amplificar el ADN genómico purificado en las siguientes condiciones.

20

Para el parólogo A, un volumen de reacción de 50 µl que contenía 420-700 ng de molde de ADN genómico digerido con *BsoBI* (10 µl), 1,25 µl de cada cebador (a 10 µM cada uno), 5 µl de tampón de ACCUPRIME PCR II 10x, 5 µl de PVP-40 al 10 % y 0,3 µl (1,5 unidades) de ADN polimerasa de alta fidelidad Taq ACCUPRIME (Invitrogen, Carlsbad, CA) en el tampón enzimático del fabricante. Los productos de amplificación se generaron a partir de ciclos de amplificación que consistían en 94 °C durante 2 minutos, 28 ciclos de (94 °C durante 30 segundos, 60 °C durante 30 segundos, 68 °C durante 1 minuto), 68 °C durante 5 minutos, 4 °C de retención indefinida.

Para el parólogo B, un volumen de reacción de 50 µl que contenía 420 - 700 ng de molde de ADN digerido con *BsoBI* (10 µl), 1,25 µl de cada cebador (a 10 µM cada uno), 5 µl de tampón Accuprime PCR II 10x, 5 µl de PVP-40 al 10 % y 0,3 µl (1,5 unidades) de ADN polimerasa de alta fidelidad Taq ACCUPRIME (Invitrogen, Carlsbad, CA) en el tampón enzimático del fabricante. Los productos de amplificación se generaron a partir de ciclos de amplificación que consistían en 94 °C durante 2 minutos, 28 ciclos de (94 °C durante 30 segundos, 58 °C durante 30 segundos, 68 °C durante 1 minuto), 68 °C durante 5 minutos, 4 °C de retención indefinida. Entonces se aisló este producto primario de amplificación usando el kit de purificación MINELUTE PCR (Qiagen, Valencia, CA) eluyendo en 10 µl de tampón EB. Se sintetizaron tres variantes de cebador PCR de orientación directa de parólogo A (5'-XXX TCTGTTTCCACGGCGGAG -3') (SEQ ID NO: 36) en el que XXX = CCG, GCG o CGC y tres variantes de un cebador PCR de orientación inversa de parólogo A (5'-XXX AAGCGGCAAGAAGAAGAATC -3') (SEQ ID NO: 37) en el que XXX = CCG, GCG o CGC y se purificaron mediante HPLC (IDT, Coralville, IA). Se sintetizaron tres variantes de un cebador de PCR de orientación directa de parólogo B (5'-XXX TCTGTTTCCACGGCTGAG -3') (SEQ ID NO: 38) en el que XXX = GGC, GCC o CGG y tres variantes de un cebador PCR de orientación inversa de parólogo B (5'-XXX ATTGGACAGAGATTTGGGTC -3') (SEQ ID NO: 39) en el que XXX = GGC, GCC o CGG y se purificaron mediante HPLC (IDT, Coralville, IA). Las etiquetas de 3 pb en el extremo 5' de cada cebador sirven como clave identificadora e indican de qué muestra proceden los fragmentos amplificados. Se usaron pares de cebador con etiquetas identificadoras emparejadas (claves) en combinación para amplificar el amplicón de PCR primario purificado procedente de muestras descritas anteriormente en las siguientes condiciones: volumen de reacción de 50 µl que contenía 10 µl de amplicón PCR purificado diluido a 1:10, 1,25 µl de cada cebador (10 µM cada uno), 5 µl de tampón PCR ACCUPRIME I10x y 0,3 µl (1,5 unidades) de ADN polimerasa de alta fidelidad Taq ACCUPRIME (Invitrogen, Carlsbad, CA) en el tampón enzimático del fabricante.

Los productos de amplificación del tamaño esperado fueron el resultado de ciclos de amplificación que consistían en 94 °C durante 2 minutos, 25 ciclos de (94 °C durante 30 segundos, hibridación durante 30 segundos, 68 °C durante 30 segundos), 68 °C durante 5 minutos, 4 °C de retención indefinida y se purificaron usando el kit de purificación MINELUTE PCR de Qiagen (Valencia, CA) según las recomendaciones del fabricante. Las temperaturas de hibridación para la reacción PCR secundaria listada fueron las siguientes: parólogo A a 66 °C y parólogo B a 64 °C.

Se llevaron a cabo reacciones de pirosecuenciación masivamente paralela directamente en los productos de la PCR. El análisis de los resultados de la secuenciación se llevó a cabo identificando lecturas de secuencia que contenían eliminaciones del tamaño y la posición esperada en la molécula de ADN.

Los resultados de estos análisis indicaron la presencia de múltiples eliminaciones pequeñas en el sitio de escisión esperado (figura 8, tabla 6). De nuevo, estas eliminaciones se localizaron de manera precisa en el sitio diana de ZFN e indicaron que se generaron roturas bicatenarias inducidas por la ZFN en el genoma y posteriormente se repararon mediante NHEJ. La ZFN clonada en pDAB7185 fue en general más eficaz para causar roturas bicatenarias que la ZFN de pDAB7186.

Estos resultados demostraron adicionalmente la capacidad de estas ZFN modificadas para inducir roturas bicatenarias dirigidas de una manera específica en un locus génico endógeno en una especie de cultivo. Los resultados también demostraron la capacidad de la misma ZFN para escindir dos parálogos de EPSPS, A y B.

**Tabla 6.** Los resultados de la secuenciación masivamente paralela que muestran moléculas de ADN que han sufrido roturas bicatenarias mediadas por ZFN seguido de reparación por NHEJ en las secuencias diana de los parálogos de EPSPS A y B en *B. napus*. Las muestras de control representan callos transgénicos no tratados con ZFN.

N.S.	Parólogo diana	Construcción de ZFN	Cebador	Secuencias alineadas	NHEJ totales
1	A	Control	Directo	12903	0
2	A	Control	Inverso	12027	0
3	A	pDAB7185	Directo	10432	9
4	A	pDAB7185	Inverso	9883	3
5	A	pDAB7186	Directo	20.496	2
6	A	pDAB7186	Inverso	18697	3

7	B	Control	Directo	20655	0
8	B	Control	Inverso	22733	0
9	B	pDAB7185	Directo	15663	32
10	B	pDAB7185	Inverso	15864	26
11	B	pDAB7186	Directo	21333	4
12	B	pDAB7186	Inverso	24003	12

**Ejemplo 6: Dos ZFN pueden escindir a todos los parálogos de EPSPS**

5 En otro ejemplo, se transformaron de manera conjunta segmentos de hipocotilo de *B. napus* con dos ZFN y se crearon eventos transgénicos estables que contenían ambas ZFN, que mostraron NHEJ en todos los parálogos de EPSPS. Las construcciones específicas de ZFN usadas para la transformación conjunta en este experimento fueron pDAB7147 y pDAB7185 (tabla 3). Se generaron eventos de callos transgénicos estables, se aisló el ADN y se analizó tal como se describe en los ejemplos 4 y 5. Se identificaron NHEJ en todos los cuatro parálogos de EPSPS A, B, C y D, similares a aquellos descritos en las secciones 4 y 5. De nuevo, las NHEJ en los parálogos D y E no pudieron diferenciarse debido a la similitud de secuencia. Todas las NHEJ se ubicaron en las secuencias diana predichas de los diversos parálogos.

10 Estos resultados validan dos puntos: 1. Las ZFN pueden diseñarse contra diferentes secuencias de una familia de múltiples genes para que escinde de manera específica 1-2 genes/parálogos. 2. Pueden usarse múltiples ZFN, en caso necesario, para escindir todos los parálogos del gen simultáneamente.

15 Puede encontrarse información adicional relacionada con la escisión dirigida, la recombinación dirigida y la integración dirigida en las Publicaciones de Solicitud de Patente de Estados Unidos US-2003-0232410; US-2005-0026157; US-2005-0064474; US-2005-0208489 y US-2007-0134796.

**REIVINDICACIONES**

1. Una proteína de dedo de cinc (ZFP) que se une a una región genómica diana de EPSPS de interés, comprendiendo la ZFP uno o más dominios de unión de dedo de cinc, en donde la ZFP comprende las regiones de hélice de reconocimiento mostradas en una sola fila de la tabla A.
- 5 2. La ZFP de la reivindicación 1, en donde la ZFP es una proteína de fusión que comprende uno o más dominios funcionales seleccionados del grupo que consiste en un represor transcripcional, una endonucleasa, una metil transferasa, una histona desacetilasa, un activador transcripcional y una histona acetiltransferasa.
- 10 3. La ZFP de la reivindicación 2, en donde la ZFP es una nucleasa de dedo de cinc (ZFN) que escinde la región genómica diana de EPSPS de interés, comprendiendo dicha ZFN uno o más dominios de unión de dedo de cinc modificados y un dominio de escisión de nucleasa y además en donde la ZFN escinde opcionalmente una o más secuencias génicas parálogas u ortólogas de EPSPS que comprenden una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o cualquier secuencia que tenga una identidad de secuencia de al menos 80-100 % con las mismas.
- 15 4. Un polinucleótido que codifica una o más proteínas de dedo de cinc (ZFP) de acuerdo con las reivindicaciones 1 a 3.
5. Una célula vegetal hospedadora que comprende uno o más polinucleótidos de acuerdo con la reivindicación 4, en donde la célula se transfecta de manera estable o transitoria con los uno o más polinucleótidos.
6. Un método para escindir uno o más genes de EPSPS en una célula vegetal, comprendiendo el método:  
introducir uno o más polinucleótidos según la reivindicación 4 en las células vegetales; y  
expresar las una o más ZFN en la célula, en donde las ZFN escinden uno o más genes de EPSPS.
- 20 7. Un vector donante que comprende primeras y segundas secuencias de ADN; en donde la primera secuencia es homóloga de una tercera secuencia y la segunda secuencia es homóloga de una cuarta secuencia; y en donde las secuencias tercera y cuarta son secuencias de ADN cromosómico de EPSPS, además en donde los extremos próximos de las secuencias tercera y cuarta son contiguos o están separados por al menos 1 par de nucleótidos y opcionalmente en donde el vector comprende además una quinta secuencia, en donde la quinta secuencia:
- 25 (a) está interpuesta entre las secuencias primera y segunda; y  
(b) es una secuencia de ácido nucleico exógena;  
y en donde las secuencias homólogas tienen una identidad de secuencia de 80-100 % entre sí,  
y en donde las secuencias de ADN cromosómico de EPSPS pertenecen a un gen de EPSPS que comprende una secuencia nucleotídica seleccionada del grupo que consiste en las SEQ ID NO: 10-14 o una secuencia que tiene  
30 una identidad de secuencia de al menos 80-100 % con las mismas.
8. El vector de la reivindicación 7, en donde la quinta secuencia comprende secuencias que codifican un marcador de selección seleccionado del grupo que consiste en proteína verde fluorescente (GFP),  $\beta$ -glucuronidasa (GUS), fosfinotricin N-acetil transferasa (PAT, BAR), neomicina fosfotransferasa, higromicina fosfotransferasa,  $\beta$ -lactamasa, catecol dioxigenasa,  $\alpha$ -amilasa, tirosinasa,  $\beta$ -galactosidasa, luciferasa, aecuorina, EPSP sintasa, nitrilasa, acetolactato sintasa (ALS), dihidrofolato reductasa (DHFR), dalapon deshalogenasa y antranilato sintasa o en donde la quinta secuencia comprende una secuencia que codifica una proteína distinta de un marcador de selección; una o más secuencias reguladoras transcripcionales; una o más secuencias de direccionamiento a proteínas que potencian o reducen el transporte de proteínas; un ARN pequeño interferente dirigido a un gen que se va a reprimir; y un micro ARN dirigido a un gen que se va a reprimir.
- 35 9. El vector de cualquiera de la reivindicación 7 o la reivindicación 8, en donde la quinta secuencia comprende una secuencia que codifica una secuencia cromosómica de EPSPS mutante que aumenta la tolerancia de una planta al herbicida glifosato.
10. Un método para introducir una secuencia de ácido nucleico exógena en el genoma de una célula vegetal, comprendiendo el método la etapa de:
- 45 expresar una o más nucleasas de dedo de cinc (ZFN) a partir de un polinucleótido según la reivindicación 4 en la célula, en presencia de un vector donante según cualquiera de las reivindicaciones 7-9;  
en donde la ZFN escinde un gen EPSPS en ADN cromosómico a 3 pares de kilobases de cualquiera de las secuencias tercera o cuarta;  
de tal forma que la escisión del ADN cromosómico en la etapa (b) estimula la incorporación del vector donante en el  
50 genoma mediante recombinación de homólogos.

11. Un método para expresar el producto de una molécula de secuencia de ácido nucleico exógena en una célula vegetal, comprendiendo el método las etapas de:  
introducir la molécula de secuencia exógena según el método de la reivindicación 10 de tal forma que se expresa el producto de la molécula de secuencia exógena en la célula vegetal.
- 5 12. Una célula vegetal transgénica o una planta transgénica obtenida según el método de la reivindicación 10 u 11.
13. Un método para estimular la recombinación de homólogos intramolecular en el genoma de una célula vegetal, comprendiendo el método la etapa de:  
introducir una ZFN según la reivindicación 3 o un polinucleótido según la reivindicación 4 en la célula vegetal en presencia de un segmento de ADN que comprende un gen EPSPS y una primera secuencia que es homóloga de una segunda secuencia en la célula vegetal, de tal forma que la ZFN escinde el segmento de ADN y estimula la recombinación de homólogos intramolecular, en donde la recombinación de homólogos da como resultado una adición, eliminación o reemplazo de secuencias dentro del segmento de ADN endógeno.
- 10 14. El método de la reivindicación 13, en donde las secuencias eliminadas se seleccionan del grupo que consiste en secuencias que codifican la totalidad o parte de un gen EPSPS y secuencias que codifican la totalidad o parte de un marcador de selección.
- 15 15. El método de la reivindicación 13 o 14, en donde el ADN eliminado se reemplaza por una secuencia exógena, comprendiendo el método, además:  
introducir un polinucleótido en la célula, en donde el polinucleótido comprende:  
(a) las secuencias cuarta y quinta, en donde la cuarta secuencia es homóloga de secuencias no eliminadas en la proximidad de la primera secuencia y la quinta secuencia es homóloga de secuencias no eliminadas en la proximidad de la segunda secuencia; y  
(b) la secuencia exógena.
- 20 16. El método de la reivindicación 15, en donde la secuencia exógena introducida mediante recombinación de homólogos comprende una secuencia cromosómica de EPSPS mutante que aumenta la tolerancia de una planta al herbicida glifosato.
- 25 17. Un método para eliminar una secuencia génica de EPSPS del genoma de una célula vegetal, comprendiendo el método:  
expresar las nucleasas de dedo de cinc (ZFN) primera y segunda según la reivindicación 3 o un polinucleótido según la reivindicación 4 en la célula, en donde la primera ZFN escinde en un primer sitio de escisión en un gen EPSPS y la segunda ZFN escinde en un segundo sitio de escisión en el gen EPSPS, en donde la secuencia génica de EPSPS se localiza entre el primer sitio de escisión y el segundo sitio de escisión, en donde la escisión de los sitios de escisión primero y segundo da como resultado la eliminación de la secuencia génica de EPSPS, en donde la secuencia génica de EPSPS se elimina mediante unión de extremos no homólogos de los sitios de escisión primero y segundo.
- 30 18. El método de la reivindicación 17, en donde la célula vegetal es una célula vegetal transgénica que comprende una secuencia de EPSPS exógena.
- 35 19. Un método para modular la expresión de un gen de EPSPS, comprendiendo el método:  
expresar una ZFP en la célula, siendo la ZFP según la reivindicación 1 o 2, en donde la ZFP se une a un sitio diana en el gen EPSPS, modulando de este modo la expresión del gen EPSPS.
- 40 20. El método de la reivindicación 19, en donde la ZFP aumenta o disminuye la transcripción del gen EPSPS o en donde la ZFP aumenta o disminuye la tolerancia de una planta al herbicida glifosato.

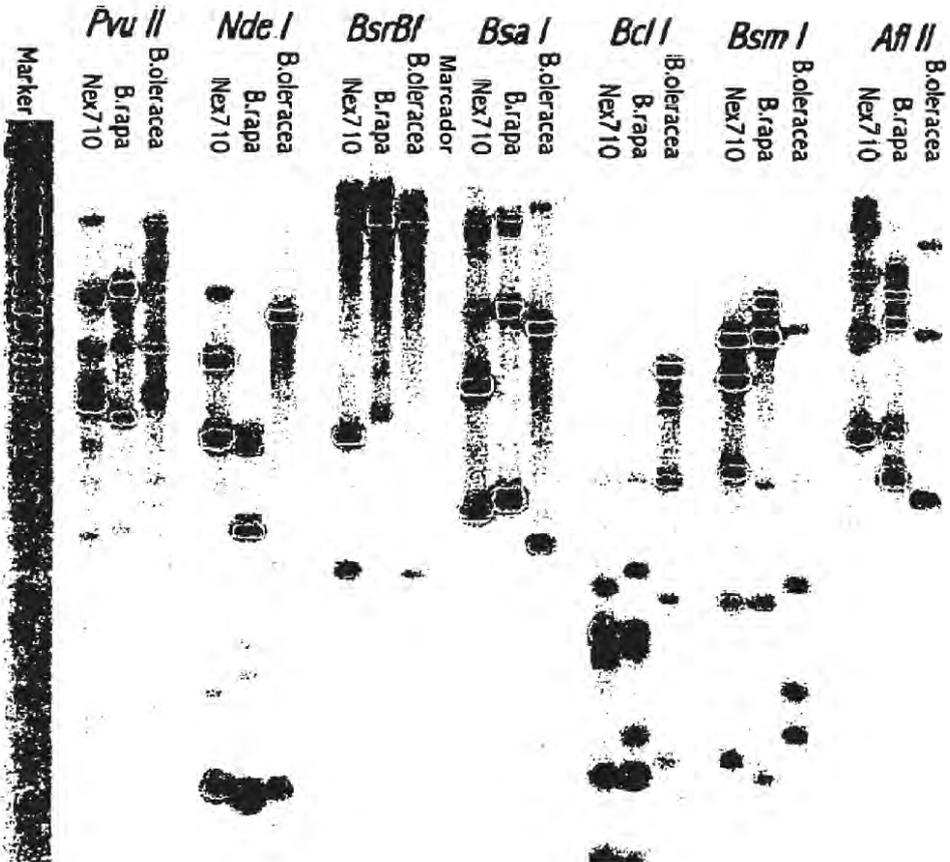


Fig. 1

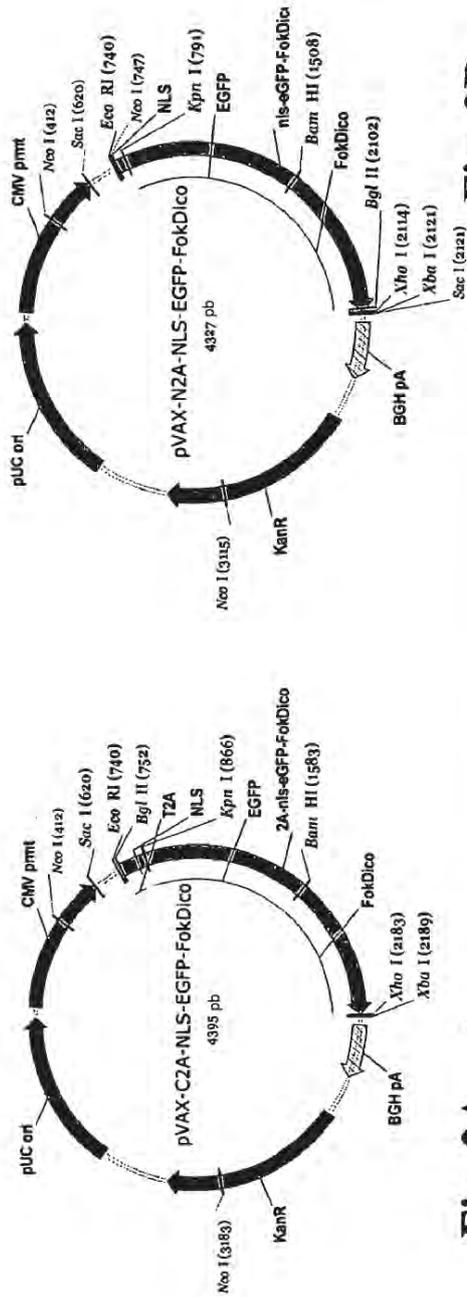


Fig. 2A

Fig. 2B

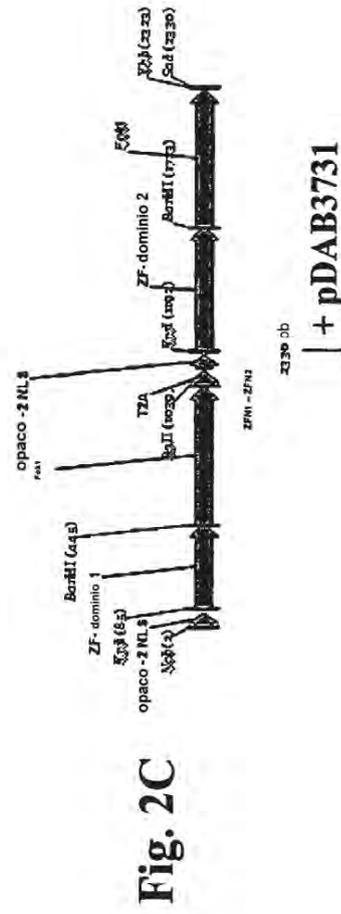
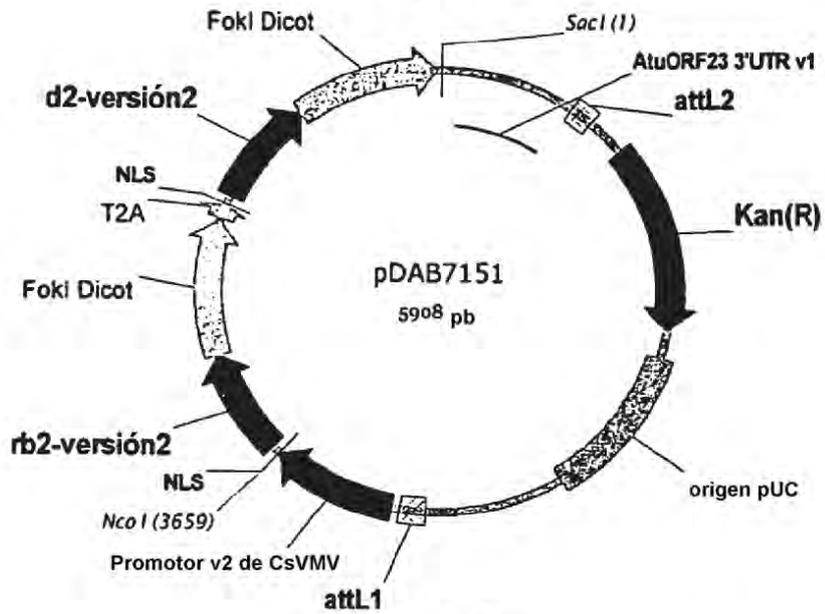
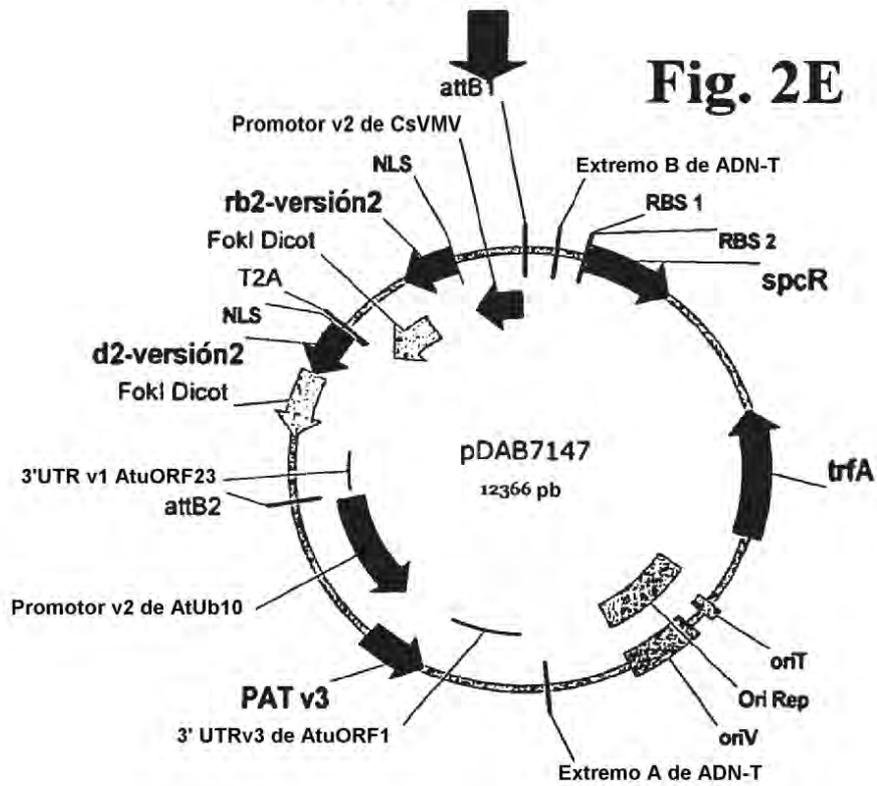


Fig. 2C

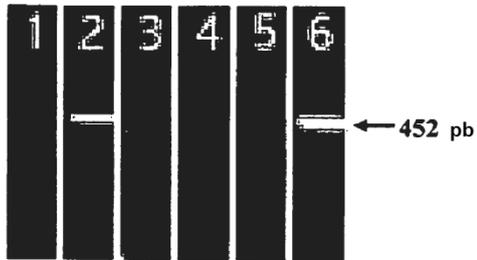
**Fig. 2D**



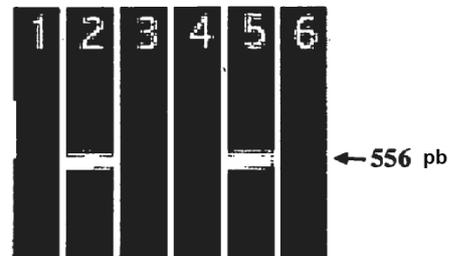
**Fig. 2E**



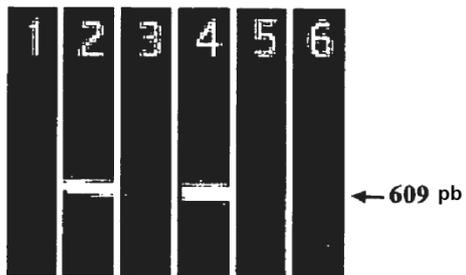
**Fig. 3A**



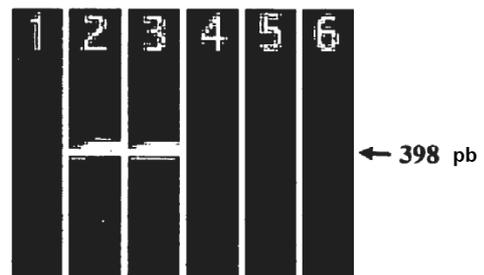
**Fig. 3B**



**Fig. 3C**



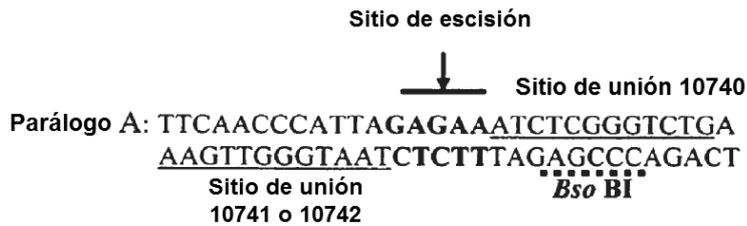
**Fig. 3D**





Comparación de secuencia de los sitios de unión de la ZFN 10654/10657 entre los parálogos de EPSPS. Las bases en letras minúsculas se saltaron en el diseño de la ZFN.

Parálogo A: GTCCACTTACAGCTGCAGTTACTIGCTGCTGGTGGCA  
 Parálogo B: GTCCACTTACCCGCTGCAGTTACTIGCTgCAGGTGGCA  
 Parálogo C: GTCCACTCACCCGCTGCAGTTACTIGCTgCAGGTGGCA  
 Parálogo D: GTCCACTCACCCGCTGCAGTTACAGCTgCAGGTGGCA  
 Parálogo E: GTCCACTCACCCGCTGCAGTTACAGCTgCAGGTGGCA



Comparación de secuencia de los sitios de unión de la ZFN 10740/10741 entre los parálogos de EPSPS.

Parálogo A: TTCAACCATTAGAGAAATCTCGGGTCTG  
 Parálogo B: TTCAACCATTAGAGAAATCTCGGGTCTG  
 Parálogo C: TTCAACCATTAGAGAAATCTCGGGTCTG  
 Parálogo D: TTCAACCAATCAGAGAAATCTCGGGTCTC  
 Parálogo E: TTCAACCAATCAGAGAAATCTCGGGTCTC

**Fig. 4**

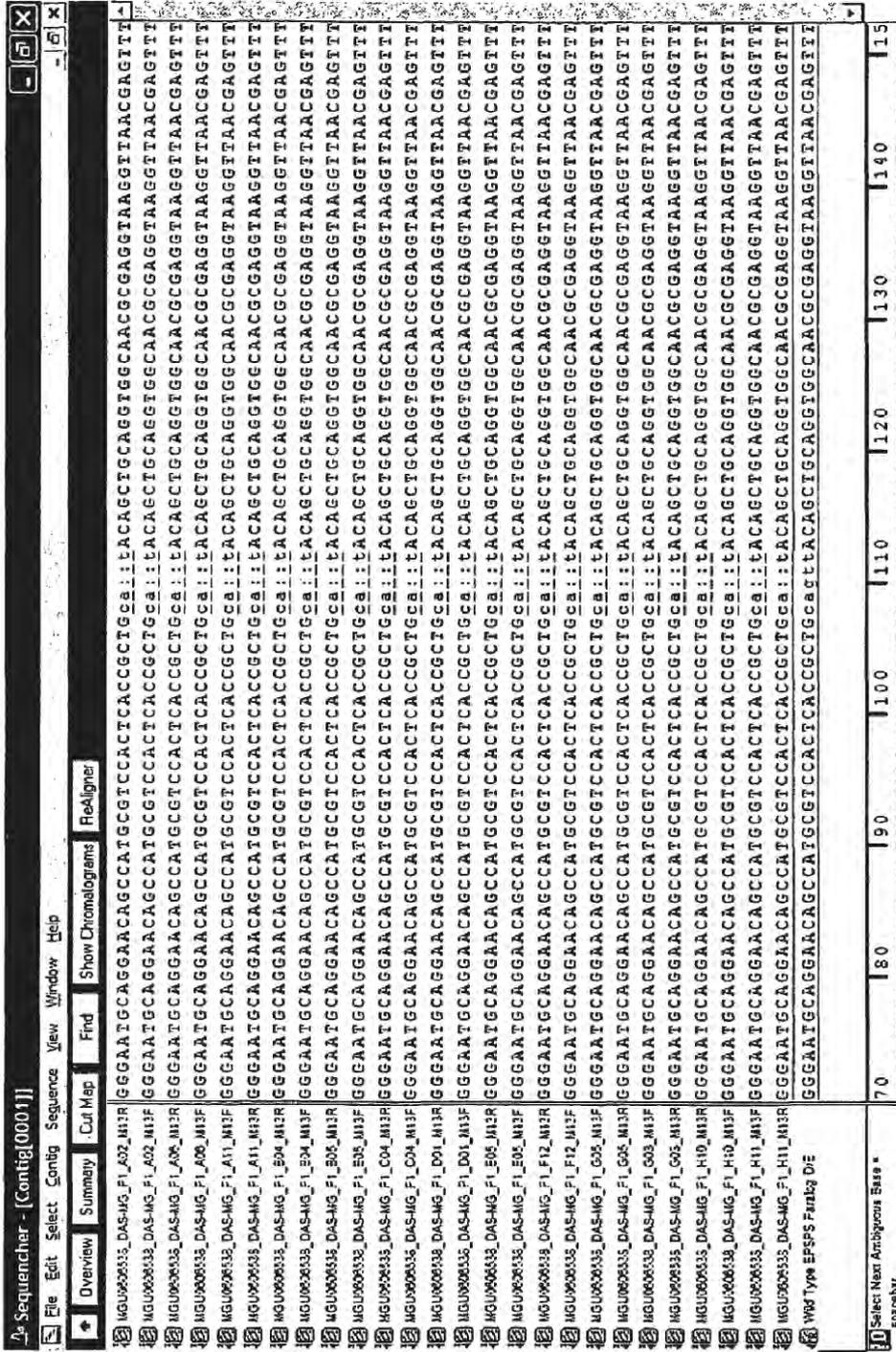


Fig. 5

AACAGCCATGCGTCCACTCACCCTGCAGTTACAGCTGCAGTGGCAACCGGAGGTAAGGTTAACGAGTTTTTTGTTA	Tipo silvestre
AACAGCCATGCGTCCACTCACCCTGCAGTGGCAACCGGAGGTAAGGTTAACGAGTTTTTTGTTA	muestra 2
AACAGCCATGCGTCCACTCACCCTGCAGTGGCAACCGGAGGTAAGGTTAACGAGTTTTTTGTTA	muestra 4
AACAGCCATGCGTCCACTCACCCTGCAGTGGCAACCGGAGGTAAGGTTAACGAGTTTTTTGTTA	muestra 5 (35)
AACAGCCATGCGTCCACTCACCCTGCAGTGGCAACCGGAGGTAAGGTTAACGAGTTTTTTGTTA	muestra 5
AACAGCCATGCGTCCACTCACCCTGCAGTGGCAACCGGAGGTAAGGTTAACGAGTTTTTTGTTA	muestra 5
AACAGCCATGCGTCCACTCACCCTGCAGTGGCAACCGGAGGTAAGGTTAACGAGTTTTTTGTTA	muestra 5 (8)
AACAGCCATGCGTCCACTCACCCTGCAGTGGCAACCGGAGGTAAGGTTAACGAGTTTTTTGTTA	muestra 5

Fig. 6

Parálogo C - Secuencia inversa de tipo silvestre  
 AAAGCTCCTTAACCTTTACCTTGGCTTGCACCTGCAGCAGTAACGGCAGCGGTGAGTGGACGCATGGCTGTTCCTGCATTCCCAAGGTA

Parálogo C - Muestra 16 NHEJ  
 AAAGCTCCTTAACCTTTACCTTGGC::::::::::::::::::::::::::TGGACGCATGGCTGTTCCTGCATTCCCAAGGTA

Parálogo C - Secuencia directa de tipo silvestre  
 AGAGTGATATTGAGTTGTACCTTGGGAATGCAGGAACAGCCATGCCGTCCACTACCGCTGCCGTTACTGTGCAGGTGGCAACGGCAAGGT

Parálogo C - Muestra 17 NHEJ  
 AGAGTGATATTGAGTTGTACCTT:GGAATGCAGGAACAGCCATGGCTCCACTACCGCTG:::::::::ACTGTGCAGGTGGCAACGGCAAGGT  
 AGAGTGATATTGAGTTGTACCTTGGGAATGCAGGAACAGCCATGCCGTCCACTACCGCTG:::::::::ACTGTGCAGGTGGCAACGGCAAGGT  
 AGAGTGATATTGAGTTGTACCTTGGGAATGCAGGAACAGCCATGCCGTCCACTACCGCTG:::::::::ACTGTGCAGGTGGCAACGGCAAGGT

Parálogo D - Secuencia de tipo silvestre  
 AGTTGTACCTTGGGAATGCAGGAACAGCCATGCCGTCCACTACCGCTGCAGTTACAGCTGCAGGTGGCAACGGCAGGTAAGGTTAACCGAG

Parálogo D - Secuencia directa de muestra  
 AGTTGTACCTTGGGAATGCAGGAACAGCCATGC:::::::::::::::::::::::::::AGGTGGCAACGGCAGGTAAGGTTAACCGAG  
 AGTTGTACCTTGGGAATGCAGGAACAGCCATGC:::::::::::::::::::::::::::AGGTGGCAACGGCAGGTAAGGTTAACCGAG  
 AGTTGTACCTTGGGAATGCAGGAACAGCCATGC:::::::::::::::::::::::::::AGGTGGTAACGGCAGGTAAGGTTAACCGAG  
 AGTTGTACCTTGGGAATGCAGGAACAGCCATGCCGTCCACTACCGCTG:::::::::::::CTGCAGGTGGCAACGGCAGGTAAGGTTAACCGAG  
 AGTTGTACCTTGGGAATGCAGGAACAGCCATGCCGTCCACTACCGCTG:::::::::::::CTGCAGGTGGCAACGGCAGGTAAGGTTAACCGAG

Fig. 7A

**Parágrafo D - Secuencia inversa de tipo silvestre**

ATCAATTTCTTTGACAAATAACAAAAAACTCGTTAACCTTACCTCGCGTTGCCACCTGCAGCTGTAACTGCAGCGGTGAGTGGACGCATGGC

**Parágrafo D - Muestra 124HEJ**

ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC  
ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC  
ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC  
ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC  
ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC  
ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC  
ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC  
ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC  
ATCAATTTCTTTGACAAATAACAAAAAACTACGTTAACCTTACCTCGCGTTGCCACCTGCAG:::::CAGCGGTGAGTGGACGCATGGC

**Fig. 7B**





CCGCTGAGAAATCTTCGGAGATTGT:.....:CTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGT:.....:CTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCA:.....:CTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCA:.....:GGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCA:.....:GGGTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCC:.....:AATCTCGGCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCC:.....:AATCTCGGCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCATT:.....:GTCCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCATT:.....:CGGCTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCATT:.....:CGGCTCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCATT:.....:TCTGATCAAGCTACCCGGACCCCAAATC  
 CCGCTGAGAAATCTTCGGAGATTGTGCTTCAGCCCATT:..AGAAATCTCGGCTGATCAAGCTACCCGGACCCCAAATC

**Fig. 8C**

(SEQ ID NO:10)

```

1   GCAGCAGCGT GGAGCTTATC AGATATCTTC GCGGGGGTTG AAGAAGAGCG
51  CGATGGTGCT AAACCGTTCT GTAACTCGTC CGGTTAAGGT TATGGCCTCT
101 GTTCCACGG CGGAGAAAGC TTCGGAGATT GTGCTTCAAC CCATTAGAGA
151 AATCTCGGGT CTGATCAAGC TACCCGGATC CAAATCTCTG TCCAACCGGA
201 TTCTTCTTCT TGCCGCTTTA TCCGAGGTTT GCTTCTTTCT TTGTTTGCTT
251 AGTGTGCGT TTTTAACGGC GTGAGGATGA AGAAAGGTTT TGACTTTGTT
301 GTGGTTTTAT AGGGAACACT TGTAGTTGAC AACTTGTTGA ACAGTGATGA
351 CATTAACACT ATGCTTGATG CGTTGAACAA GTTGGGGCTT AATGTGGAAC
401 GTGACAGTGA GAACAACCGT GCGGTTGTTG AAGGATGTGG CGGGATATTC
451 CCAGCTTCTT TAGATTCTAA GGGTGATATC GAGTTGTACC TTGGGAATGC
501 AGGAACAGCC ATGCGTCCAC TTACAGCTGC AGTTACTGCT GCTGGTGGCA
551 ACGCAAGGTA AGGTTAAGGA CTTATTCTGT TAGTTAGTTT TGATTATTTT
601 AAGAATCGGT CTTGTACTGA CTGTTTTTAG TTGGGTTTGT TTACCAGTTA
651 TGTGCTTGAT GGGGTGCCTA GAATGAGGGA AAGACCTATA GGAGATTTGG
701 TTGTTGGTCT TAAGCAGCTT GGTGCTGATG TTGAATGTAC TCTTGGTACT
751 AACTGTCCCTC CTGTTCTGT CAATGCTAAT GGTGGCCTGC CCGGTGGAAA
801 GGTGAGTTTG TAATTTTCAGC ATTTGCTATG TGAAAAGTTG CAGCAATCTT
851 TGTTTCATCAC ACTGCGTTAG CTTGACATGA TTTTAGCTTT TGTATGGTTT
901 CTTGATTGAC ACATTAGACA TGTTTTTGCA TTTTTCAGGT GAAGCTTTCT
951 GGATCAATCA GTAGTCAATA CTTGACTGCA CTGCTCATGG CAGCTCCCTT
1001 AGCTCTTGGA GACGTTGAGA TTGAGATCAT TGATAAATTG ATTTCTGTTC
1051 CATATGTTGA AATGACATTG AAGTTGATGG AACGTTTTGG TGTTAGTGCC
1101 GAGCATAGTG ACAGTTGGGA TCGTTTCTTT GTCAAGGGCG GTCAGAAATA
1151 CAAGTAAGAG TTGTTTCTAA AATCACTGAA CTTATAATTA GATTGACAGA
1201 AGAGTGACTA ACCAAATGGT AAAATTTGAT TCAGGTCGCC TGGTAATGCT
1251 TACGTAGAAG GTGATGCTTC TAGTGCTAGT TATTTCTTGG CTGGTGCTGC
1301 CATTACTGGT GAAACCGTTA CTGTTGAAGG TTGTGGAACA ACCAGCCTGC
1351 AGGTAACACT AAGTTTATAA TAAATTTGC TTAGTTCAAT TTTTTTTTGT
1401 CTTTCTAAGG CTTGGCTAGT TGTGTCACTT GTGTGTAACA TATGAAGAAT
1451 CTAAGTTTAG TTTTTTTTGG TGATGAATCT CAAAGGGAGA TGTGAAGTTC
1501 GCTGAGGTTT TTGAGAAAAT GGGATGTAAA GTGTCATGGA CAGAGAACAG
1551 TGTGACTGTG ACTGGACCAT C

```

**Fig. 9**

(SEQ ID NO:11)

```

1   ATGGCGCAAG CTAGCAGAAT CTGCCAGAAC CCATGTGTGA TCTCCAATCT
51  CTCCAAATCA AACCAACGCA AATCGCCCTT GTCTGTCTCG ATGAAGACGC
101 ACCAGATATC TTCGTGGGGG TTGAAGAAGA GTAACAACGG CTCTGTGATT
151 CGTCCGGTTC GGGTAACGGC GTCTGTTTCC ACGGCTGAGA AATCTTCGGA
201 GATTGTGCTT CAGCCCATTA GAGAAATCTC GGGTCTGATC AAGCTACCCG
251 GACCCAAATC TCTGTCCAAT CGAATCCTTC TTCTAGCCGC TCTATCCGAG
301 GTCGGTTTGC TTCTTTCTTT CTTTGTTAGC TTAGTGTGTC GTTTTTAACG
351 GCGTGAGATT GAAGAAAGGT TCACACTTTG TTGTGGTGT ATAGGGAACC
401 ACTGTAGTTG ACAACTTGTT GAACAGTGAT GACATCAATT ACATGCTTGA
451 TGC GTTGAAG AAATTGGGGC TTAATGTGGA ACGTGACAGT GAGAATAACC
501 GTGCGGTTGT TGAAGGATGT GGCGGGATAT TCCCAGCTTC TTTAGATTCC
551 AAGAGTGATA TCGAGTTGTA CCTTGGGAAT GCTGGAACAG CCATGCGTCC
601 ACTTACCGCT GCAGTTACTG CTGCAGGTGG CAACGCAAGG TAAGGTTAAG
651 GAGTGTGATT TTGTTAGTTA GTTTTGIGTT ATGTCAAGAA CCGATCTTGT
701 CCTCATGCTT TTAGTTCGGT TTATTTTCCA GTTATATTCT TGGTGGGGTG
751 CCTAGAATGA GGGAAAGGCC TATTGGAGAT TTGGTTGTTG GTCTTAAGCA
801 GCTTGGTGCT GATGTTGAAT GTACTCTTGG AACTAACTGC CCTCCTGTTC
851 GCGTCAATGC TAATGGTGGC CTTCCCGGTG GAAAGGTGAG TTTGTAAATCT
901 CAGCATCTAC TATGTGGAAA GTTGACAGGAA TTTTGTTC TAACATGCGC
951 TTTGCTCGAT ATGATGGCCT TTGTATGGTT TCTTGATTGA CATATTAGAT
1001 ATGATTTGCA TTTTTCAGGT GAAGCTATCT GGTTNAATCA GTAGTCAATA
1051 CTTGACTGCT CTGCTCATGG CAGCTCCTTT AGCTCTTGGA GACGTTGAGA
1101 TTGAGATCGT TGATAAACTG ATCTCTGTTC CGTATGTTGA AATGACATTG
1151 AAGTTGATGG AACGTTTTGG TGTTAGTGCC GAGCATAGTG ACAGTTGGGA
1201 TCGTTTCTTT GTCAAGGGCG GTCAGAAATA CAAGTAAGCG TTGTTTCTGA
1251 AATCACTGAA CTTATAGTTA GATTGACAGA AGAGTACTA ACCAAATGGT
1301 AAAATTTGAT TCAGGTCGCC TGGTAATGCT TACGTAGAAG GTGATGCTTC
1351 TAGTGCTAGT TATTTCTTGG CTGGTGCCGC CATTACTGGT GAGACTGTTA
1401 CTGTTGAAGG TTGTGGAACA ACCAGCCTGC AGGTAACACT AAGTTTATAA
1451 TGAAATTTGC TTAGTTCAAT TTGTTTTTTT GTCTTTCTAA GGCTTTGGCT
1501 AGTTATGTGT AACATATGTT AGAATCTAAG CTCATTTTTG TTGTTGTGAT
1551 GAATCTCAA GGGAGATGTG AAGTTCGCTG AGGTTCTTGA GAAAATGGGA
1601 GTAAAGTGT CATGGACAGA GAACAGTGTG ACTGTGACTG GACCATCTAG
1651 AGATGCTTTT GGAATGAGAC ACTTGCGCGC TGTTGATGTC AACATGAACA
1701 AAATGCCTGA TGTAGCCATG ACTCTTGCCG TTGTTGCTCT CTTTGCAGAT
1751 GGTCCAACCA CCATTAGAGA TGGTAAGTAC TCCCTCTAAC CATCTAATTG
1801 AGGTTTTTAA GATTCATAGT CACTTAGTTC TCCTCTCATC CAATCGTTTT
1851 ATCATATATA GTGGCTAGCT GGAGAGTAAA GGAGACAGAA AGGATGATTG
1901 CCATTTGCAC AGAGCTTAGG AAGGTA AAC AATTTTCTTT CTGTCCCGCT
1951 CTCACTCTCT TGGTTTTATG TGCTCAGTCT AGGTTAAGTT CTGCATAACT
2001 TTTGCGTGCA GCTTGGAGCT ACAGTGGAAG AGGGTTCAGA TTATTGTGTG
2051 ATA ACTCCAC CAGCAAAGCT GAAACCGGCG GAGATTGACA CATATGATGA
2101 TCATAGAATG GCAATGGCAT TCTCCCTTGC AGCTTGTGCT GATGTTCCAG
2151 TAACCATCAA AGATCCTGGT TGCACCAGGA AA ACTTTCCC T GACTACTTC
2201 CAGGTCCTTG AAAGTATCAC AAAGCACTAA ACAGACCTTA AAGCCCATTT
2251 GTCTTTTCTT TTTGATCCAA TTGGGATCAG TTTCTCTGT TATCACTGTA
2301 AGATTACGAA AAACAAAGAG TATTAAGATT GTTTGCTTGT ACCTTAAACT
2351 GTTTGATGCA ATCGTTGAAT CAGTTTTGGG CCAAGGGC

```

**Fig. 10**

(SEQ ID NO:12)

```

1   ATGGCGCAAG CTAGCAGAAT CTGCCATGGC GTGCAGAACC CATGTGTTAT
51  CATCTCCAAT CTCTCCAAAT CAAACCAAAA CAAATCACCT TTCTCCGTCT
101 CGCTGAAGAC GCAGCAGTCT CGAGCTTCTT CGTGGGGACT AAAGAAGAGT
151 GGAACGATGC TAAACGGTTC TGTAATTCGC CCGGTTAAGG TAACAGCTTC
201 CGTTTCCACG GCCGAGAAAG CTTCAGAGAT TGTGCTTCAA CCAATTAGAG
251 AAATCTCGGG TCTCATTAAG CTACCCGGAT CCAAATCTCT CTCCAATCGG
301 ATCCTCCTTC TTGCTGCTCT ATCTGAGGTA CATATACTTG ATTAGTGTTA
351 GGCCTTTGCT GTGAGATTTT GGGAACTATA GACAATTTAG TAAGAATTTA
401 TATATTATTT TTAAAAAATT AAAAGCCTAT ATATATATAT ATTTAAAATT
451 TTCAAAAAAT TATGGAGGTT TGAGACTGAA GAAAGTTTTT TTTTAATTAT
501 TATTATAGGG AACTACTGTA GTGGACAACT TGTGTAACAG TGATGACATC
551 AACTACATGC TTGATGCGTT GAAGAAGCTG GGGCTTAACG TGGAACGTGA
601 CAGGGTAAAC AACCGTGCTG TAGTTGAAGG ATGTGGTGGA ATATTCCCAG
651 CTCCTTAGA TTCCAAGAGT GATATTGAGT TGTACCTTGG GAATGCAGGA
701 ACAGCCATGC GTCCACTCAC CGCTGCCGTT ACTGCTGCAG GTGGCAACGC
751 AAGGTAAAGG TTAAGGAGCT TTTTGTTATT GTCAAGAAAT TGATTTTGTG
801 TTTGATGCTT TTAGTTTGGT TTGTTTTCTA GTTATGTGCT TGATGGGGTG
851 CCTAGAATGA GGGAGAGACC TATAGGAGAT TTGGTTGTTG GTCTTAAGCA
901 GCTTGGTGCT GATGTTGAAT GTACTCTCGG CACTAACTGT CCTCCTGPTC
951 GTGTCAATGC TAATGGTGGC CTTCCCGGTG GAAAGGTGAT CTTGTTTGCA
1001 GCAGTCTTTG TTCATCACAG CCTTTGCTTC ACATTATTAC ATCTTTTAGT
1051 TTGTTGTTGT GACTTGATGG ATCTTAAAAA AAGGAATTGG GAACTGGTGT
1101 GAAAGTGATT AGCAATCTTT CTCGATTCCT TGCAGGGCCG TGGGCATTAC
1151 TAAGTGAAC ATTAGCCTAT TAACCCCAA ATATTTTGAA AAAAATTTAG
1201 TATATGGCCC CAAAATAGTT TTTAAGAAAT TAGAAAACT TTTAATAAAT
1251 CGTCTACGGT CCCCATTTTA GAGCCGACCC TGCTTGATG GTTCTTGAG
1301 TGAGATATTT TACATGTTTT GCATTTTCAG GTGAAGCTTT CTGGATCAAT
1351 CAGTAGTCAA TACTTGACTG CTTTGCTCAT GGCAGCTCCT TTAGCTCTTG
1401 GAGACGTGGA GATTGAGATC ATTGATAAAC TGATTTCTGT TCCATATGTT
1451 GAAATGACAT TGAAGTTGAT GGAACGTTTT GGTGTTAGTG CCGAGCATAG
1501 TGATAGCTGG GATCGTTTCT TTGTCAAGGG CGGTCAGAAG TACAAGTAAG

```

**Fig. 11A**

```

1551 AATTCTTTAA ATTAAAGAAT TAGATTGAAG AAAATGACTG ATTAACCAAA
1601 TGGCAAAACT GATTCAGGTC GCCTGGTAAT GCTTATGTAG AAGGTGATGC
1651 TTCTAGTGCT AGCTACTTCT TGGCTGGTGC TGCTATTACC GGTGAAACCG
1701 TCACTGTTGA AGGTTGTGGA ACAACTAGCC TCCAGGTAGT TTCTCCACTC
1751 TGAATCATCA AATATTATAC TCCCTCCGTT TTGTATTAAG TGTCATTTTA
1801 GCTTTTAAAT TTTGTCTCAT TAAAAGTGTC ATTTTACATT TTCAATGTAT
1851 ATATTAAATA AATTTTCCAG TTTTACTAA TTCATTATAT TAAATAATAT
1901 AAAACAGAAA ATTTAACAAT TATCGTAATT CGTGTGCAA GTTGATTAGT
1951 TCAAAGTTGT GTGTAACATG TTTTGAAGAA TCTAAGCTCA TTCTCTTTTT
2001 ATTTTTTTTG TGATGAATCC CAAAGGGAGA TGTGAAATC GCAGAGGTAC
2051 TTGAGAAAAAT GGGATGTAAA GTGTCATGGA CAGAGAACAG TGTGACTGTG
2101 ACTGGACCAT CTAGAGATGC TTTTGGAAATG AGACACTTGC GTGCTGTTGA
2151 TGTCAACATG AATAAAATGC CCGATGTAGC CATGACTCTT GCCGTTGTTG
2201 CTCTCTTTGC CGATGGTCCA ACCACCATCA GAGATGGTAA AGCAAACCC
2251 TCTCTTTGAA TCAGCGTCTT TAAAAGATT CATGGTTGCT TTAActCTAT
2301 TTGGTCAATG TAGTGGCTAG CTGGAGAGTT AAGGAGACAG AAAGGATGAT
2351 AGCCATCTGC ACAGAGCTTC GAAAGGTAAG TTTCTTTTC TCTCATGCTC
2401 TCATTCTAAG TTAATCGTTG CATAACTTTT TGGGGTTTTT TTTTTGCGTT
2451 CAGCTTGGAG CTACAGTGGA AGAAGGTTCA GATTATTGTG TGATAACTCC
2501 ACCAGCGAAG GTGAAACCGG CGGAGATTGA TACGTATGAT GATCATAGAA
2551 TGGCGATGGC GTTCTCGCTT GCAGCATGTG CTGATGTTCC AGTCACCATC
2601 AAGGATCCTG GCTGCACCAG AAAGACTTTC CCTGACTACT TTCAAGTCCT
2651 TGAAAGTATC ACAAAGCACT AAAAAGATCA TTTCTTTGA ATCCAAATGT
2701 GAGAATGTGT TTCTTCTCT CTCTGTTGCC ACTGTAAACAT TTATTAGAAG
2751 AACAAAGTGT GTGTGTTTAA GAGTGTGTTT GCTTGTAATG AACTGAGTGA
2801 GATGCAATCG TTGAATCAGT TTTGGGCCAA GGGC

```

**Fig. 11B**

(SEQ ID NO:13)

```

1  ATGGCGCAAG CTAGCAGAAT CTGCCATGGC GTGCAGAACC CATGTGTTAT
51  CATCTCCAAT CTCTCCAAAT CCAACCAAAA CAAATCACCT CTCTCCGTCT
101 CTTGAAGAC GCATCAGCCT CGAGCTTCTT CGTGGGGATT GAAGAAGAGT
151 GGAACGACGC TAAACGGTTC TGTAATTCGC CCGGTTAAGG TAACAGCTTC
201 TGTTTCCACG TCCGAGAAAG CTTCAGAGAT TGTGCTTCAA CCAATCAGAG
251 AAATCTCGGG TCTCATTAAG CTACCCGGAT CCAAATCTCT CTCCAATCGG
301 ATCCTCCTTC TTGCCGCTCT ATCTGAGGTA CATATACTTG CTTAGTGTTA
351 GGCCTTTGCT GTGAGATTTT GGGAACTATA GACAATTTAG TAAGAATTTA
401 TATATAATTT TTTTAAAAAA AATCAGAAGC CTATATATAT TTAAATTTTT
451 CCAAATTTT TGGAGGTTAT AGGCTTGTGT TACACCATTC TAGTCTGCAT
501 CTTTCGGTTT GAGACTGAAG AATTTTATTT TTTAAAAAAT TATTATAGGG
551 AACTACTGTA GTGGACAAC TGTGAACAG TGATGACATC AACTACATGC
601 TTGATGCGTT GAAGAAGCTG GGGCTTAACG TGGAACGTGA CAGTGTA AAC
651 AACCGTGCGG TTGTTGAAGG ATGCGGTGGA ATATTCCCAG CTTCCCTAGA
701 TTCCAAGAGT GATATTGAGT TGTACCTTGG GAATGCAGGA ACAGCCATGC
751 GTCCACTCAC CGCTGCAGTT ACAGCTGCAG GTGGCAACGC GAGGTAAGGT
801 TAACGAGTTT TTTGTTATTG TCAAGAAATT GATCTTGTGT TTGATGCTTT
851 TAGTTTGGTT TGTTTTCTAG TTATGTA CTGATGGGGTGC CTAGAATGAG
901 GGAAAGACCT ATAGGAGATT TGGTTGTTGG TCTTAAGCAG CTTGGTGCTG
951 ATGTTGAGTG TACTCTTGGC ACTAACTGTC CTCCTGTTCG TGCAATGCT
1001 AATGGTGGCC TTCCCAGTGG AAAGGTGATC TTCACATTTA CTCTATGAAT
1051 TGTTTGCAGC AGTCTTTGTT CATCACAGCC TTTGCTTCAC ATTATTTTCA
1101 CTTTTAGTTT GTTGTATAT TACTTGATGG ATCTTTAAAA AGGAATTGGG
1151 TCTGGTGTGA AAGTGATTAG CAATCTTTCT CGATTCCCTG CAGGGCCGTG
1201 GGCATTACTA AGTGAAACAT TAGCCTATTA ACCCCCAAAA TTTTGA AAA
1251 AAATTTAGTA TATGGCCCA AAATAGTTTT TAAAAAATTA GAAAACTTT
1301 TAATAAATCG TCTACAGTCC CAAAATCTT AGAGCCGGCC CTGCTTGTAT
1351 GGTTTCTCGA TTGATATATT AGACTATGTT TTGAATTTTC AGGTGAAGCT
1401 TTCTGGATCG ATCAGTAGTC AGTACTTGAC TGCCCTCCTC ATGGCAGCTC
1451 CTTTAGCTCT TGGAGACGTG GAGATTGAGA TCATTGATAA ACTGATATCT
1501 GTTCCATATG TTGAAATGAC ATTGAAGTTG ATGGAACGTC TTGGTGTTAG
1551 TGCCGAGCAT AGTGATAGCT GGGATCGTTT CTTTGTC AAG GGCGGTCAGA
1601 AGTACAAGTA AGAATCTTT AAATTAAGA ATTAGATTGA AGAAAATGAC
1651 TGATTAACCA AATGGCAAAA CTGATTCAGG TCGCCTGGTA ATGCTTATGT
1701 AGAAGGTGAT GCTTCTAGTG CTAGCTACTT CTTGGCTGGT GCTGCTATTA
1751 CCGGTGAAAC CGTCACTGTT GAAGGTTGTG GAACAAC TAG CCTCCAGGTA

```

**Fig. 12A**

```

1801  GTTTCTCCAC TCTGAATCAT CAAATATTAT ACTCCCTCCG TTTTGTATTA
1851  AGTGTCAATTT TAGCTTTTAA ATTTTGTCTC ATTAAAAGTG TCATTTTACA
1901  TTTTCAATGT ATATATTAAA TAAATTTTCC AGTTTTTACT AATTCATTAT
1951  ATTAAATAAT ATAAAACAGA AAATTTAACA ATTATCGTAA TTCGTGTGCA
2001  AAGTTGATTA GTTCAAAGTT GTGTGTAACA TGTTTTGAAG AATCTAAGCT
2051  CATTCTCTTC TTATTTTTTT TGTGATGAAT CCCAAAGGGA GATGTGAAAT
2101  TCGCAGAGGT ACTTGAGAAA ATGGGATGTA AAGTGTCTATG GACAGAGAAC
2151  AGTGTGACTG TGAATGAGAC ATCTAGAGAT GCTTTTGGAA TGAGACACTT
2201  GCGTGCTGTT GATGTCAACA TGAATAAAAT GCCCGATGTA GCCATGACTC
2251  TTGCCGTTGT TGCTCTCTTT GCCGATGGTC CAACCACCAT CAGAGATGGT
2301  AAAGCAAAAC CCTCTCTTTG AATCAGCGTC TTTTAAAAGA TTCATGGTTG
2351  CTTTAACTCT ATTTGGTCAA TGTAGTGGCT AGCTGGAGAG TTAAGGAGAC
2401  AGAAAGGATG ATAGCCATCT GCACAGAGCT TCGAAAGGTA AGTTTCCTTT
2451  TCTCTCATGC TCTCATTCTA AGTTAATCGT TGCATAACTT TTTGGGGTTT
2501  TTTTTTGCGT TCAGCTTGGA GCTACAGTGG AAGAAGGTTT AGATTATTGT
2551  GTGATAACTC CACCAGCGAA GGTGAAACCG GCGGAGATTG ATACGTATGA
2601  TGATCATAGA ATGGCGATGG CGTTCTCGCT TGCAGCATGT GCTGATGTTC
2651  CAGTCACCAT CAAGGATCCT GGCTGCACCA GAAAGACTTT CCCTGACTAC
2701  TTTCAAGTCC TTGAAAGTAT CACAAAGCAC TAAAAAGATC ATTTTCCTTTG
2751  AATCCAAATG TGAGAATGTG TTTCTTCTC TCTCTGTTGC CACTGTAACA
2801  TTTATTAGAA GAACAAAGTG TGTGTGTTTA AGAGTGTGTT TGCTTGTAAT
2851  GAACTGAGTG AGATGCAATC GTTGAATCAG TTTTGGGCC

```

**Fig. 12B**

(SEQ ID NO:14)

```

1  ATGGCGCAAG CTAGCAGAAT CTGCCATGGC GTGCAGAACC CATGTGTTAT
51  CATCTCCAAT CTCTCCAAAT CCAACCAAAA CAAATCACCT TTCTCCGTCT
101 CCTTGAAGAC GCATCAGCCT CGAGCTTCTT CGTGGGGATT GAAGAAGAGT
151 GGAACGATGC TAAACGGTTC TGTAATTCGC CCGGTAAAGG TAACAGCTTC
201 TGTTTCCACG TCCGAGAAAG CTTCAGAGAT TGTGCTTCAA CCAATCAGAG
251 AAATCTCGGG TCTCATTAAG CTACCCGGAT CCAAATCTCT CTCCAATCGG
301 ATCCTCCTTC TTGCCGCTCT ATCTGAGGTA CATATACTTG CTTAGTGTTA
351 GGCCTTTGCT GTGAGATTTT GGGAECTATA GACAATTTAG TAAGAATTTA
401 TATATAATTT TTTTAAAAAA AATCAGAAGC CTATATATAT TTAATTTTTT
451 CCAAAATTTT TGGAGGTTAT AGGCTTATGT TACACCATTC TAGTCTGCAT
501 CTTTCGGTTT GAGACTGAAG AATTTTATTT TTTAAAAAAT TATTATAGGG
551 AACTACTGTA GTGGACAAC TGTGTAACAG TGATGACATC AACTACATGC
601 TTGATGCGTT GAAGAAGCTG GGGCTTAACG TGGAACGTGA CAGTGTA AAC
651 AACCGTGCGG TTGTTGAAGG ATGCGGTGGA ATATTCCAG CTTCCCTAGA
701 TTCCAAGAGT GATATTGAGT TGTACCTTGG GAATGCAGGA ACAGCCATGC
751 GTCCACTCAC CGCTGCAGTT ACAGCTGCAG GTGGCAACGC GAGGTAAGGT
801 TAACGAGTTT TTTGTTATTG TCAAGAAATT GATCTTGTGT TTGATGCTTT
851 TAGTTTGGTT TGTTTTCTAG TGATGTACTT GATGGGGTGC CTAGAATGAG
901 GGAAAGACCT ATAGGAGATT TGGTTGTTGG TCTTAAGCAG CTTGGTGCTG
951 ATGTTGAGTG TACTCTTGGC ACTAECTGTC CTCCTGTTCC TGTC AATGCT
1001 AATGGTGGCC TTCCCGGTGG GAAGGTGATC TTCACATTTA CTCTATGAAT
1051 TGTTTGCAGC AGTCTTTGTT CATCACAGCC TTTGCTTAC ATTATTT CAT
1101 CTTT TAGTTT GTTGTATAT TACTTGATGG ATCTTTAAAA AGGAATTGGG
1151 TCTGGTGTGA AAGTGATTAG CAATCTTCT CGATTCCTTG CAGGGCCGTG
1201 GGCATTAATA AGTGAACAT TAGCCTATTA ACCCCAAAA TTTTGA AAA
1251 AAATTTAGTA TATGGCCCA AAATAGTTT TAAAAAATA GAAAAACTTT
1301 TAATAAATCG TCTACAGTCC CAAAATCTT AGAGCCGGCC CTGCTTG TAT
1351 GGTTTCTCGA TTGATATATT AGACTATGTT TTGAATTTTC AGGTGAAGCT
1401 TCCTGGATCG ATCAGTAGTC AGTACTTGAC TGCCCTCCTC ATGGCAGCTC
1451 CTTTAGCTCT TGGAGACGTG GAGATTGAGA TCATTGATAA ACTGATATCT
1501 GTTCCATATG TTGAAATGAC ATTGAAGTTG ATGGAGCGTT TTGGTGTTAG
1551 TGCCGAGCAT AGTGATAGCT GGGATCGTTT CTTTGTCAAG GGCGGTCAGA
1601 AATACAAGTA ATGAGTTCTT TTAAGTTGAG AGTTAGATTG AAGAATGAAT

```

**Fig. 13A**

```

1651 GACTGATTAA CCAAATGGCA AACTGATTC AGGTCGCCTG GTAATGCTTA
1701 TGTAGAAGGT GATGCTTCTA GTGCTAGCTA CTTCTTGGCT GGTGCTGCCA
1751 TTAGTGGTGA AACTGTTACT GTCGAAGGTT GTGGAACAAC TAGCCTCCAG
1801 GTAGTTTATC CACTCTGAAT CATCAAATAT TATACTCCCT CCGTTTTATG
1851 TTAAGTGTCA TTAGCTTTTA AATTTTGTTT CATTAAAAGT GTCATTTTAC
1901 ATTTTCAATG CATATATTAA ATAAATTTTC CAGTTTTTAC TAATTCATTA
1951 ATTAGCAAAA TCAAACAAA ATTATATTAA ATAATGTAAA ATTTCGTAATT
2001 TGTGTGCAAA TACCTTAAAC CTTATGAAAC GGAAACCTTA TGAAACAGAG
2051 GGAGTACTAA TTTTATAATA AAATTTGATT AGTTCAAAGT TGTGTATAAC
2101 ATGTTCTGTA AGAATCTAAG CTCATTCTCT TTTTATTTTT TGTGATGAAT
2151 CCCAAAGGGA GATGTGAAAT TCGCAGAGGT TCTTGAGAAA ATGGGATGTA
2201 AGGTGTCATG GACAGAGAAC AGTGTGACTG TGACTGGACC ATCAAGAGAT
2251 GCTTTTGGAA TGAGGCACTT GCGTGCTGTT GATGTCAACA TGAACAAAAT
2301 GCCTGATGTA GCCATGACTC TAGCCGTTGT TGCTCTCTTT GCCGATGGTC
2351 CAACCACCAT CAGAGATGGT AAAGCAAAAC CCTCTCTTTG AATCAGCGTG
2401 TTTTAAAAGA TTCATGGTTG CTTAAACTCT ATTTGGTCAA TGTAGTGGCT
2451 AGCTGGAGAG TTAAGGAGAC AGAGAGGATG ATTGCCATTT GCACAGAGCT
2501 TAGAAAGGTA AGTTTCCTTT TCTCTCATGC TCTCTCATTG GAAGTTAATC
2551 GTTGCCATAAC TTTTTCGGT TTTTTFTTTT GCGTTCAGCT TGGAGCTACA
2601 GTGGAAGAAG GTTCAGATTA TTGTGTGATA ACTCCACCAG CAAAGGTGAA
2651 ACCGGCGGAG ATTGATACGT ATGATGATCA TAGAATGGCG ATGGCGTTCT
2701 CGCTTGCAGC TTGTGCTGAT GTTCCAGTCA CCATCAAGGA TCCTGGCTGC
2751 ACCAGGAAGA CTTTCCCTGA CTACTTCCAA GTCCTTGAAA GTATCACAAA
2801 GCATTAAAAG ACCCTTTCCT CTGATCCAAA TGTGAGAATC TGTGCTTTC
2851 TCTTTGTTGC CACCGTAACA TTTATTAGAA GAACAAAGTG TGTGTGTAA
2901 GAGTGTGTTT GCTTGTAAATG AACTGAGTGA GATGCAATCG TTGAATCAGT
2951 TTTGGGCC

```

**Fig. 13B**

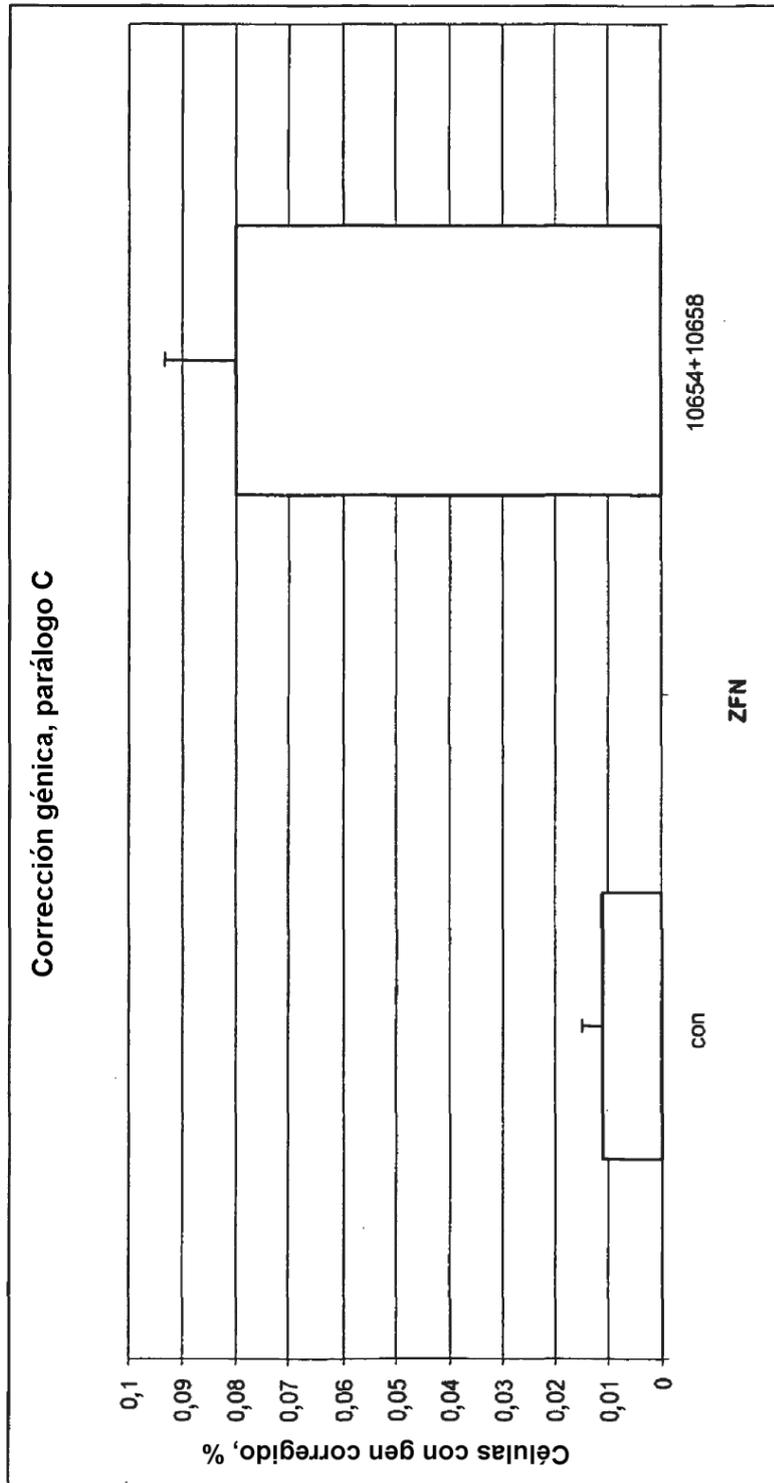


Fig. 14A

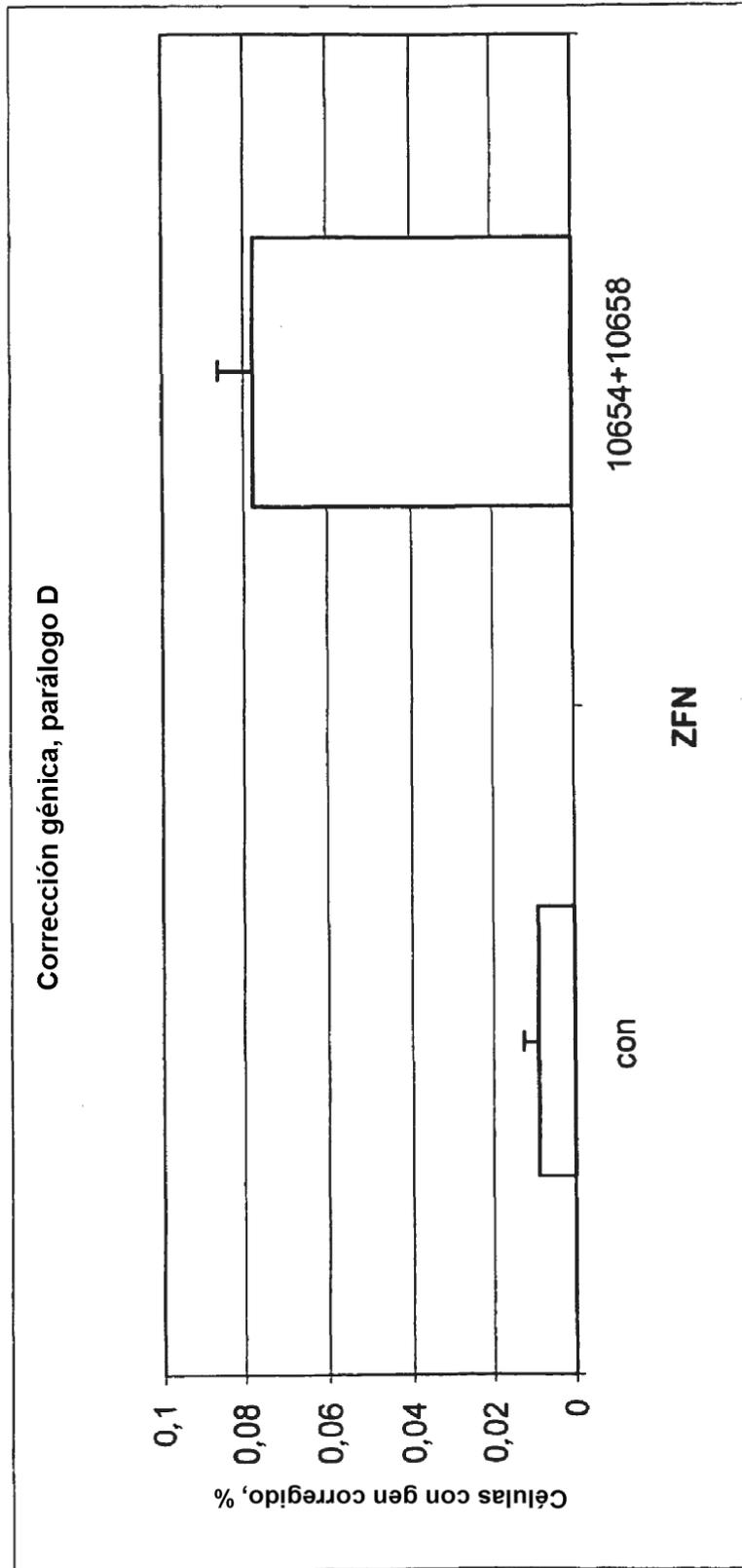


Fig. 14B

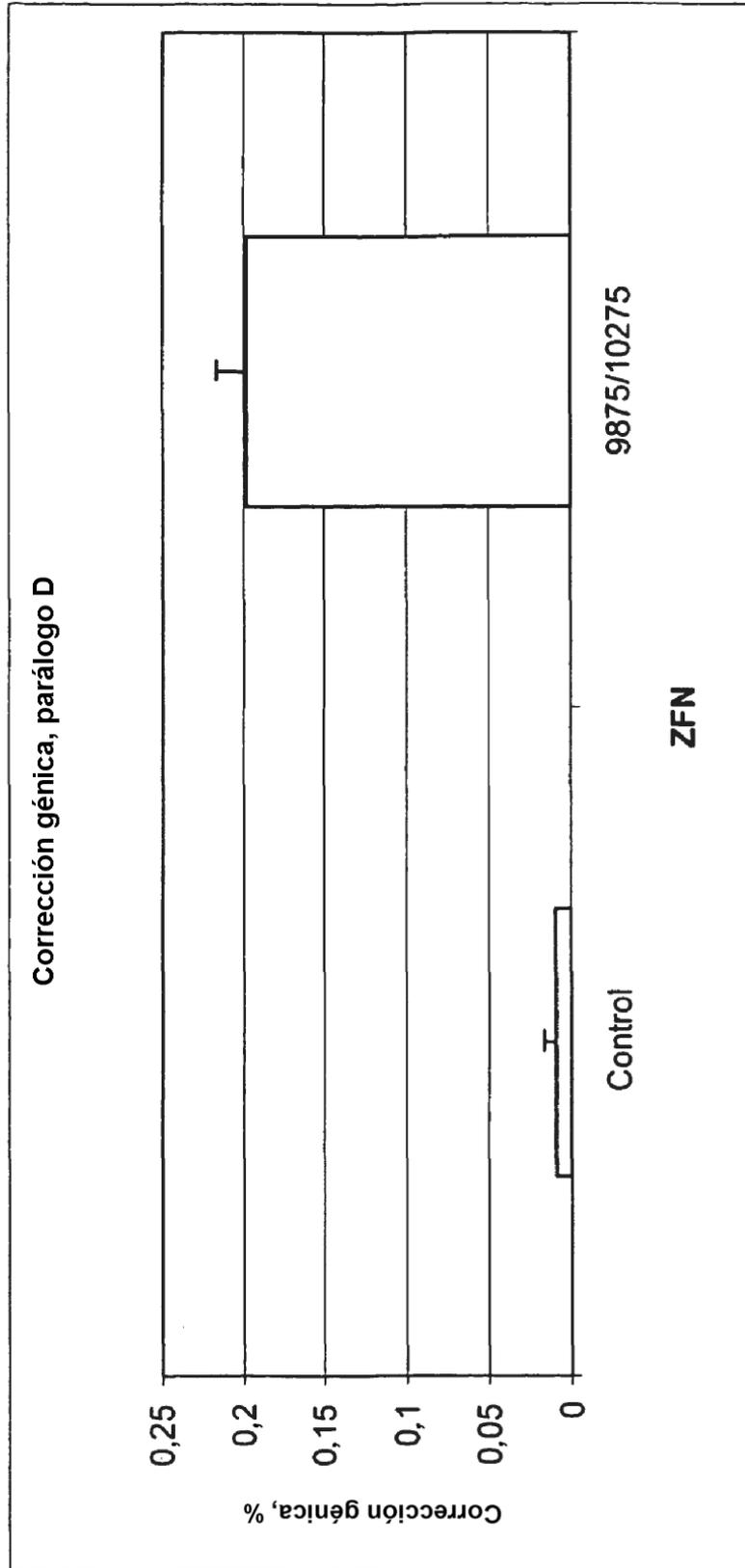
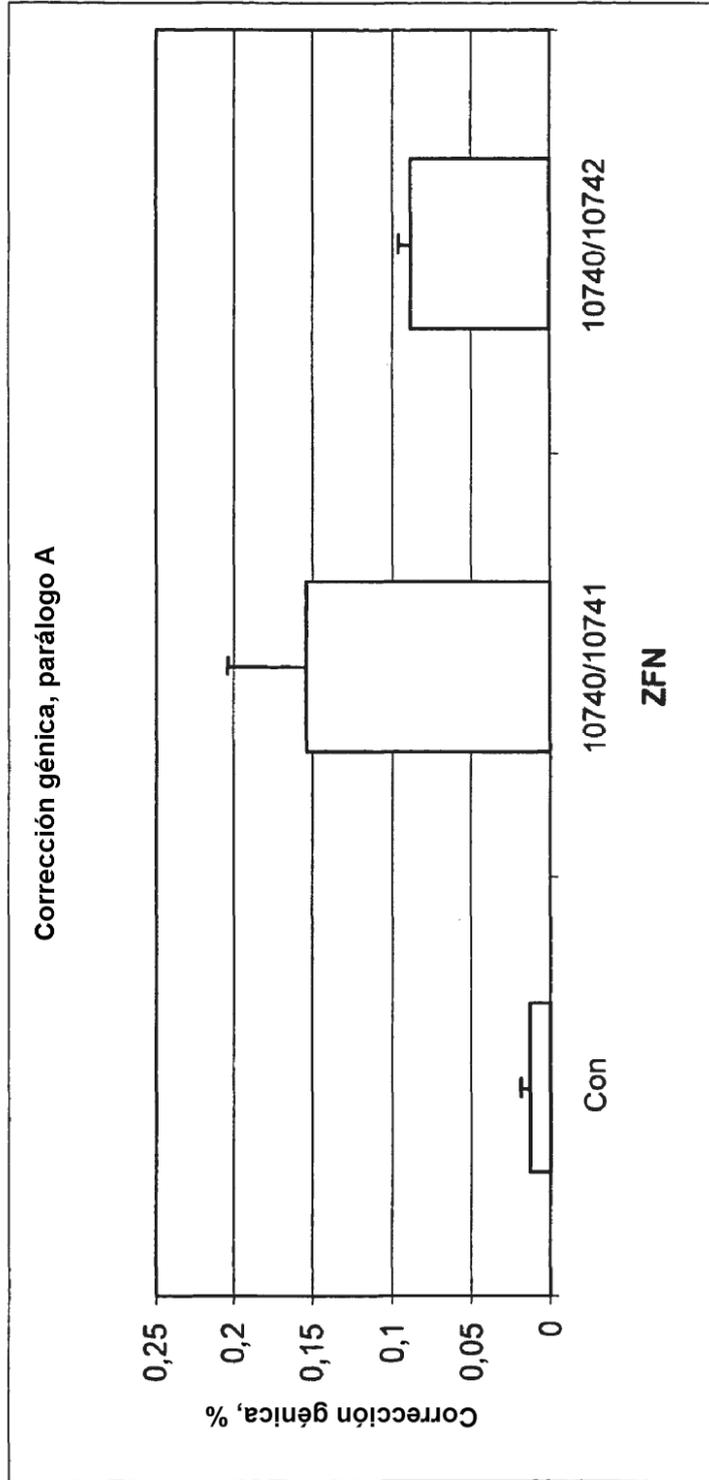


Fig. 14C



#Nota: Estas proteínas no se ensayaron con la línea celular indicadora de parágrafo B de EPSPS

**Fig. 14D**

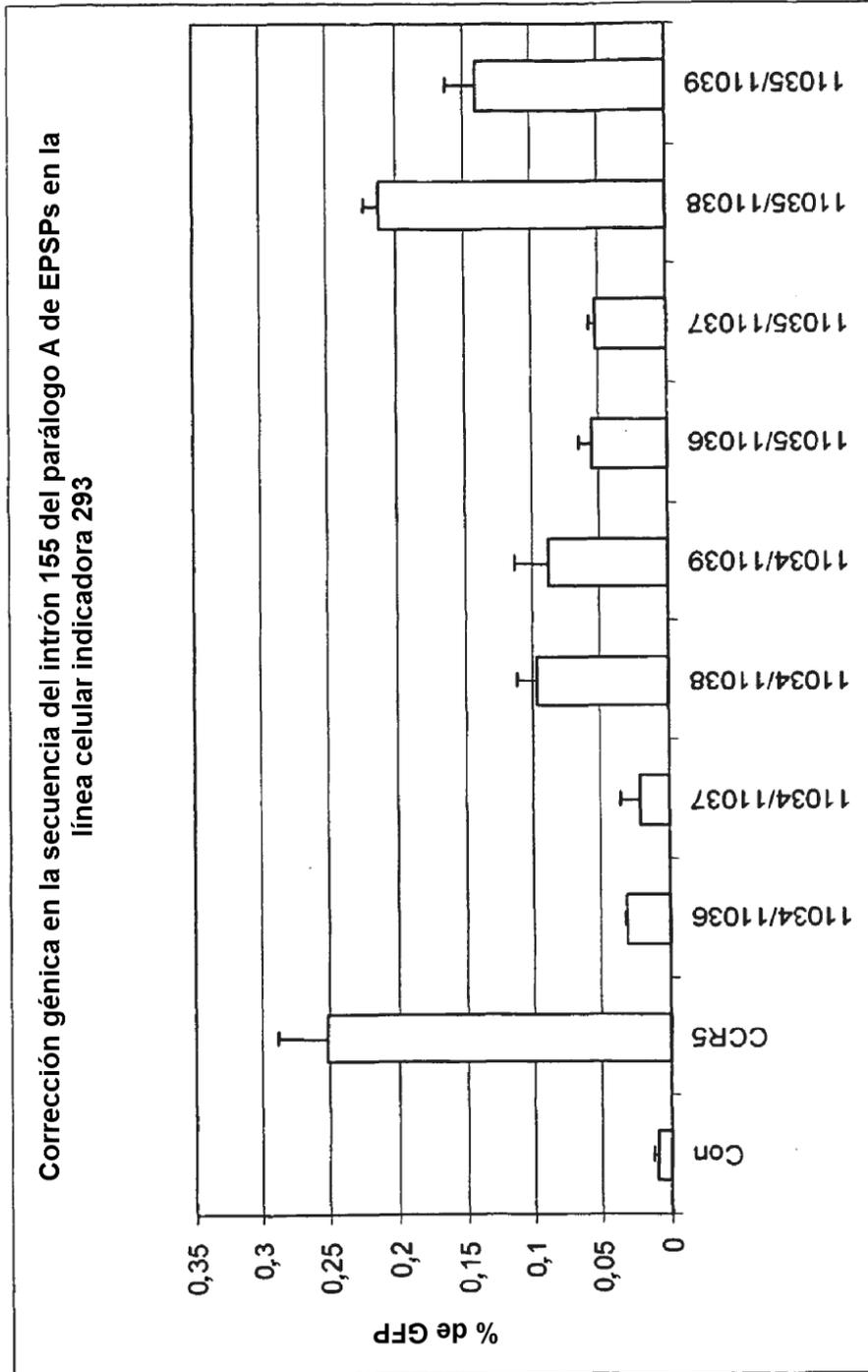


Fig. 14E