

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 368 962**

51 Int. Cl.:
C12N 15/82 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **08075865 .9**
96 Fecha de presentación: **17.11.2004**
97 Número de publicación de la solicitud: **2025756**
97 Fecha de publicación de la solicitud: **18.02.2009**

54 Título: **INSERCIÓN MEJORADA DIRECCIONADA DE DNA EN LAS PLANTAS.**

30 Prioridad:
18.11.2003 EP 03078700

45 Fecha de publicación de la mención BOPI:
24.11.2011

45 Fecha de la publicación del folleto de la patente:
24.11.2011

73 Titular/es:
**BAYER BIOSCIENCE N.V.
TECHNOLOGIEPARK 38
9052 GENT, BE**

72 Inventor/es:
**D'Halluin, Kathleen;
Vanderstraeten, Chantal y
Ruiter, Rene**

74 Agente: **Lehmann Novo, Isabel**

ES 2 368 962 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

Inserción mejorada direccionada de DNA en las plantas

Campo de la invención

5 La presente invención se refiere al campo de la biología molecular de las plantas, más específicamente al campo de la ingeniería genómica de las plantas. Se proporcionan métodos para la introducción dirigida de un fragmento de DNA extraño en un sitio de inserción preseleccionado en el genoma de una planta. Las plantas que contienen el DNA extraño insertado en un sitio particular se pueden obtener ahora con una mayor frecuencia y con mayor exactitud que lo que es posible con los métodos de inserción direccionada de DNA disponibles actualmente. Además, en una gran proporción de las plantas resultantes, el DNA extraño se ha insertado solamente en el sitio de inserción preseleccionado, sin que el DNA extraño se haya insertado aleatoriamente en otras localizaciones en el genoma de la planta. Los métodos de la invención son así una mejora, tanto cuantitativa como cualitativamente, con respecto a los métodos de la técnica anterior. Se proporcionan también genes quiméricos, plásmidos, vectores y otros medios para ser usados en los métodos de la invención.

Antecedentes de la invención

15 La primera generación de plantas transgénicas a principios de los años 80 del siglo pasado por tecnología de transformación mediada por *Agrobacterium* ha estimulado el desarrollo de otros métodos para introducir un DNA extraño de interés o un transgén en el genoma de una planta, tales como la incorporación de DNA mediada por PEG en protoplastos, bombardeo con microproyectiles, transformación mediada por triquitas de silicio, etc.

20 Sin embargo, todos los métodos de transformación de plantas tienen en común que los transgenes incorporados en el genoma de la planta se integran de una manera aleatoria y en número de copias impredecible. Con frecuencia, los transgenes pueden integrarse en forma de repeticiones, ya sea del transgén completo o de partes del mismo. Un patrón de integración compleja de este tipo puede influir en el nivel de expresión de los transgenes, v.g. por destrucción del DNA transcrito por mecanismos de silenciamiento de genes posteriores a la transcripción o por inducción de metilación del DNA introducido, regulando con ello en sentido decreciente la actividad de transcripción en el transgén. Asimismo, el sitio de integración per se puede influir en el nivel de expresión del transgén. La combinación de estos factores da como resultado una gran variación en el nivel de expresión de los transgenes o DNA extraño de interés entre células de plantas y líneas de plantas transgénicas diferentes. Además, la integración del DNA extraño de interés puede tener un efecto disruptivo en la región del genoma en la que ocurre la integración, y puede influir o perturbar la función normal de dicha región diana, conduciendo con ello a efectos secundarios a menudo indeseables.

30 Por esta razón, siempre que se investiga el efecto de introducción de un DNA extraño particular en una planta, se requiere que se generen y se analicen un gran número de líneas de plantas transgénicas a fin de obtener resultados significativos. Análogamente, en la generación de plantas de cosecha transgénicas, en las que se introduce un DNA de interés particular en las plantas para proporcionar a la planta transgénica un fenotipo deseado conocido, se crea una gran población de líneas de plantas transgénicas creadas independientemente o los denominados sucesos, para permitir la selección de aquellas líneas de plantas que exhiben una expresión óptima de los transgenes, y con efectos secundarios mínimos o inexistentes, en el fenotipo global de la planta transgénica. Particularmente en este campo, sería ventajoso si este proceso de tanteos pudiera reemplazarse por un método más directo, teniendo en cuenta los requisitos reguladores engorrosos y los costes elevados asociados con las pruebas en campo repetidas necesarias para la eliminación de los sucesos transgénicos no deseados. Adicionalmente, estará claro que la posibilidad de inserción de DNA direccionada podría ser también beneficiosa en el proceso de la denominada superposición de transgenes.

40 La necesidad de controlar la integración del transgén en plantas ha sido pronto reconocida, y se han desarrollado varios métodos en un esfuerzo para satisfacer esta necesidad (para una revisión véase Kumar y Fladung, 2001, Trends in Plant Science, 6, p. 155-159). Estos métodos están basados fundamentalmente en integración de transgenes basada en recombinación homóloga, una estrategia que ha sido aplicada con éxito en procariontes y eucariotes inferiores (véase v.g. EP 0317509 o la publicación correspondiente por Paszkowski *et al.*, 1988, EMBO J., 7, p. 4021-4026). Sin embargo, para las plantas, el mecanismo predominante para la integración de transgenes está basado en la recombinación ilegítima que implica poca homología entre las cadenas de DNA recombinantes. Un desafío importante en esta área es por consiguiente la detección de los sucesos raros de recombinación homóloga, que se ven enmascarados por la integración mucho más eficiente del DNA extraño introducido por recombinación ilegítima.

Una forma de resolver este problema es por selección contra los sucesos de integración que se han producido por recombinación ilegítima, tal como se ilustra en WO 94/17176.

55 Otra forma de resolver el problema por activación del locus diana y/o del DNA de reparación o donante por la inducción de roturas de DNA bicatenario por medio de endonucleasas de corte raro, tales como I-SceI. Se ha

demostrado que esta técnica aumenta la frecuencia de recombinación homóloga al menos en dos órdenes de magnitud utilizando Agrobacterias para suministrar el DNA de reparación a las células de las plantas (Puchta et al., 1996, Proc. Natl. Acad. Sci. U.S.A., 93, p. 5055-5060; Chilton y Que, Plant Physiol., 2003).

5 WO 96/14408 describe un DNA aislado que codifica la enzima I-SceI. Esta secuencia de DNA puede incorporarse en vectores de clonación y expresión, líneas de células transformadas y animales transgénicos. Los vectores son útiles en mapeado génico e inserción de genes dirigida al sitio.

10 WO 00/46386 describe métodos de modificación, reparación, atenuación e inactivación de un gen u otro DNA cromosómico en una célula por rotura de las dobles cadenas con I-SceI. Se describen también métodos de tratamiento o profilaxis de una enfermedad genética en un individuo que se encuentra en necesidad de ello. Se describen adicionalmente endonucleasas de restricción quiméricas.

Sin embargo, persiste todavía la necesidad de mejorar la frecuencia de la inserción direccionada de un DNA extraño en el genoma de una célula eucariota, particularmente en el genoma de una célula vegetal. Estos y otros problemas se resuelven como se describe más adelante en esta memoria en las diferentes realizaciones detalladas de la invención, así como en las reivindicaciones.

15 Sumario de la invención

En una realización, la invención proporciona un método para la introducción de un DNA extraño de interés, que puede estar flanqueado por una región de DNA que tiene al menos 80% de identidad de secuencia con una región de DNA que flanquea un sitio preseleccionado, en un sitio preseleccionado, tal como un sitio I-SceI de un genoma de una célula vegetal, tal como una célula de maíz, que comprende las etapas de

20 (a) inducir una rotura de DNA bicatenario en el sitio preseleccionado en el genoma de la célula mediante introducción en dicha célula vegetal de un gen expresable en plantas que codifica una enzima que induce una rotura de DNA bicatenario de corte raro que reconoce dicho sitio preseleccionado, por ejemplo introduciendo un gen que codifica I-SceI;

(b) introducir el DNA extraño de interés en la célula vegetal;

25 caracterizado porque el DNA extraño se suministra mediante transferencia directa de DNA, que se puede lograr mediante bombardeo de microproyectiles revestidos con el DNA extraño de interés. El gen que codifica I-SceI puede comprender una secuencia nucleotídica que codifica la secuencia de aminoácidos de SEQ ID No. 1, en el que dicha secuencia nucleotídica tiene un contenido de GC de alrededor de 50% a alrededor de 60%, con la condición de que

30 i) la secuencia nucleotídica no comprende una secuencia nucleotídica seleccionada del grupo que consiste en GATAAT, TATAAA, AATATA, AATATT, GATAAA, AATGAA, AATAAG, AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAATA, AATTAA, AATACA y CATAAA;

ii) el nucleótido no comprende una secuencia nucleotídica seleccionada del grupo que consiste en CCAAT, ATTGG, GCAAT y ATTGC;

35 iii) la secuencia nucleotídica no comprende una secuencia seleccionada del grupo que consiste en ATTTA, AAGGT, AGGTA, GGTA o GCAGG;

iv) la secuencia nucleotídica no comprende un tramo GC que consiste en 7 nucleótidos consecutivos seleccionados del grupo de G o C;

40 v) la secuencia nucleotídica no comprende un tramo AT que consiste en 5 nucleótidos consecutivos seleccionados del grupo de A o T; y

vi) la secuencia nucleotídica no comprende los codones TTA, CTA, ATA, GTA, TCG, CCG, ACG y GCG. Un ejemplo de tal gen que codifica I-SceI comprende la secuencia nucleotídica de SEQ ID 4.

45 La célula vegetal se puede incubar, antes de la etapa a), en un compuesto fenólico de origen vegetal, que se puede seleccionar del grupo de acetosiringona (3,5-dimetoxi-4-hidroxiacetofenona), α -hidroxi-acetosiringona, ácido sinapínico (ácido 3,5-dimetoxi-4-hidroxicinámico), ácido sirínico (ácido 4-hidroxi-3,5-dimetoxibenzoico), ácido ferúlico (ácido 4-hidroxi-3-metoxicinámico), catecol (1,2-dihidroxibenceno), ácido p-hidroxibenzoico (ácido 4-hidroxibenzoico), ácido β -resorcílico (ácido 2,4-dihidroxibenzoico), ácido protocatéquico (ácido 3,4-dihidroxibenzoico), ácido pirogálico (ácido 2,3,4-trihidroxibenzoico), ácido gálico (ácido 3,4,5-trihidroxibenzoico) y vanillina (3-metoxi-4-hidroxibenzaldehído).

50 La enzima que induce una rotura de DNA bicatenario de corte raro puede comprender una señal de localización

nuclear.

Breve descripción de las figuras

La Tabla 1 representa las posibles elecciones de trinucleótidos (codones) para una región codificante de I-SceI sintética (véase también la secuencia nucleotídica en SEQ ID No. 2).

5 La Tabla 2 representa las posibles elecciones preferidas de trinucleótidos para una región codificante de I-SceI sintética (véase también la secuencia nucleotídica en SEQ ID No. 3).

10 Figura 1: Representación esquemática del locus diana (A) y el DNA de reparación (B) utilizado en el ensayo para la inserción de DNA direccionada mediada por recombinación homóloga. Se representa también el locus diana después de recombinación (C). Sitio DSB: sitio de rotura del DNA bicatenario; 3'g7: señal de poliadenilación y de terminación de la transcripción del gen 7 de *A. tumefaciens*; neo: neomicina fosfotransferasa expresable en plantas; 35S: promotor del transcrito 35S de CaMV; 5' bar: región de DNA que codifica la porción amino terminal de la fosfotricina acetiltransferasa; 3'nos: señal de poliadenilación y de terminación de la transcripción del gen de nopalina sintetasa de *A. tumefaciens*; Pnos: promotor del gen de nopalina sintetasa de *A. tumefaciens*; 3'ocs: señal de poliadenilación y de terminación de la transcripción en 3' del gen de octopina sintasa de *A. tumefaciens*.

15 Descripción detallada

Se ha encontrado que:

20 a) La introducción en las células de la planta del DNA extraño a insertar por transferencia directa de DNA, particularmente bombardeo con microproyectiles, aumentaba inesperadamente la frecuencia de sucesos de inserción direccionada. La totalidad de los sucesos de inserción obtenidos eran sucesos de inserción de DNA direccionada, que ocurrían en el sitio de la rotura de DNA bicatenario inducida. Además, la totalidad de estos sucesos de inserción direccionada parecían ser sucesos de recombinación exacta entre la homología de secuencia proporcionada que flanqueaba la rotura del DNA bicatenario. Sólo alrededor de la mitad de estos sucesos tuvieron una inserción adicional del DNA extraño en un sitio diferente del sitio de la rotura del DNA bicatenario inducida.

25 b) La inducción de la rotura del DNA bicatenario por expresión transitoria de una endonucleasa inductora de rotura bicatenaria de corte raro, tal como I-SceI, codificada por un gen quimérico que comprende una región codificante sintética para una endonucleasa de corte raro tal como I-SceI, diseñada de acuerdo con un conjunto preseleccionado de reglas, aumentaba sorprendentemente la calidad de los sucesos de inserción del DNA direccionada resultantes (es decir la frecuencia de sucesos de inserción de DNA perfectamente direccionada). Adicionalmente, la endonucleasa se había equipado con una señal de localización nuclear.

30 c) La preincubación de las células diana en un compuesto fenólico vegetal, tal como acetosiringona, aumentaba adicionalmente la frecuencia de inserción direccionada en roturas de DNA bicatenario inducidas en el genoma de una célula vegetal.

35 Cualquiera de los hallazgos anteriores, aislados o en combinación, mejora la frecuencia con la cual pueden obtenerse sucesos de inserción direccionada basados en recombinación homóloga, así como la calidad de los sucesos recuperados.

De este modo, en un aspecto, la invención se refiere a un método para introducir un DNA extraño de interés en un sitio preseleccionado de un genoma de una célula vegetal, que comprende las etapas de

40 (a) inducir una rotura de DNA bicatenario en el sitio preseleccionado en el genoma de la célula mediante introducción en dicha célula vegetal de un gen expresable en plantas que codifica una enzima que induce una rotura de DNA bicatenario de corte raro que reconoce dicho sitio preseleccionado;

(b) introducir el DNA extraño de interés en la célula vegetal;

caracterizado porque el DNA extraño se suministra mediante transferencia directa de DNA.

45 Como se utiliza en esta memoria, "transferencia directa de DNA" es cualquier método de introducción de DNA en células vegetales que no implica el uso de *Agrobacterium spp.* natural que es capaz de introducir DNA en células vegetales. Esto incluye métodos bien conocidos en la técnica, tales como la introducción de DNA por electroporación en protoplastos, introducción de DNA por electroporación en células vegetales intactas o tejidos o células vegetales parcialmente degradados, introducción en protoplastos de DNA por la acción de agentes tales como PEG y análogos, y particularmente bombardeo con microproyectiles recubiertos con DNA. La introducción de DNA por transferencia directa en células vegetales difiere de la introducción de DNA mediada por *Agrobacterium* al menos en que el DNA bicatenario entra en la célula vegetal, en que el DNA que entra no está recubierto con ninguna proteína, y en que la cantidad de DNA que entra en la célula vegetal puede ser considerablemente mayor. Adicionalmente, el

DNA introducido por métodos de transferencia directa, tales como el gen quimérico introducido que codifica una endonucleasa inductora de una rotura del DNA bicatenario, puede ser más susceptible a la transcripción, dando como resultado una mejor sincronización de la inducción de la rotura del DNA bicatenario. Aunque no se pretende limitar la invención a un modo de acción particular, se cree que la inserción eficiente basada en recombinación homóloga del DNA de reparación o DNA extraño en el genoma de una célula vegetal puede ser debida a una combinación de cualquiera de estos parámetros.

Convenientemente, la rotura del DNA bicatenario puede inducirse en el sitio preseleccionado por expresión transitoria después de la introducción de un gen expresable en plantas que codifica una enzima de corte raro inductora de la rotura bicatenaria. Como se expone en otro lugar en este documento, se puede utilizar I-SceI para dicho propósito a fin de introducir un DNA extraño en un sitio de reconocimiento de I-SceI. Sin embargo, resultará inmediatamente evidente para la persona experta en la técnica que también pueden utilizarse otras enzimas inductoras de rotura bicatenaria para insertar el DNA extraño en sus sitios de reconocimiento respectivos. Una lista de enzimas de corte raro inductoras de DSB (rotura bicatenaria) y sus sitios de reconocimiento respectivos se proporciona en la Tabla I del documento WO 03/004659 (páginas 17 a 20) (incorporado aquí como referencia). Adicionalmente, están disponibles métodos para diseñar endonucleasas de corte raro fabricadas por encargo que reconocen básicamente cualquier secuencia nucleotídica diana de elección. Tales métodos se han descrito p.ej. en WO 03/080809, WO 94/18313 o WO 95/09933, y en Isalan et al., 2001, Nature Biotechnology 19, 656- 660; Liu et al. 1997, Proc. Natl. Acad. Sci. USA 94, 5525-5530.)

Así pues, como se utiliza en esta memoria, "un sitio preseleccionado" indica una secuencia nucleotídica particular en el genoma nuclear de la planta en cuya localización se desea insertar el DNA extraño. Una persona experta en la técnica sería perfectamente capaz de seleccionar una enzima inductora de rotura del DNA bicatenario ("DSBI") que reconozca la secuencia nucleotídica diana seleccionada, o de modificar por ingeniería genética una endonucleasa DSBI de este tipo. Alternativamente, un sitio de reconocimiento de una endonucleasa DSBI se puede introducir en el genoma de la planta utilizando cualquier método de transformación convencional o por mejora genética convencional utilizando una línea de plantas que tenga un sitio de reconocimiento de una endonucleasa DSBI en su genoma, y cualquier DNA extraño deseado puede introducirse posteriormente en dicho sitio diana preseleccionado introducido previamente.

La rotura del DNA bicatenario puede inducirse convenientemente por introducción transitoria de un gen quimérico expresable en plantas que comprende una región promotora expresable en plantas enlazada operativamente a una región de DNA que codifique una enzima inductora de rotura bicatenaria. La región de DNA que codifica una enzima inductora de rotura bicatenaria puede ser una región de DNA sintético, tal como, pero sin limitarse a, una región de DNA sintético en la cual los codones se seleccionan de acuerdo con el esquema de diseño que se describe en otro lugar de esta solicitud para las regiones codificantes de I-SceI.

La enzima inductora de rotura bicatenaria puede comprender, pero no comprende necesariamente, una señal de localización nuclear (NLS) [Raikhel, Plant Physiol. 100:1627-1632 (1992) y las referencias allí], tales como la NLS del antígeno T largo de SV40 [Kalderon et al. Cell 39:499-509 (1984)]. La señal de localización nuclear puede localizarse en cualquier lugar de la proteína, pero convenientemente está localizada en el extremo N-terminal de la proteína. La señal de localización nuclear puede reemplazar uno o más de los aminoácidos de la enzima inductora de la rotura bicatenaria.

Como se utiliza en esta memoria, "DNA extraño de interés" indica cualquier fragmento de DNA que se pueda desear introducir en el sitio preseleccionado. Aunque no se requiere estrictamente, el DNA extraño de interés puede estar flanqueado por al menos una región de secuencia nucleotídica que tenga homología con una región de DNA que flanquee el sitio preseleccionado. El DNA extraño de interés puede estar flanqueado en ambos sitios por regiones de DNA que tengan homología con ambas regiones de DNA que flanquean el sitio preseleccionado. Así, la o las moléculas de DNA de reparación introducidas en la célula vegetal pueden comprender un DNA extraño flanqueado por una o dos secuencias flanqueantes que tengan homología con las regiones de DNA situadas respectivamente en dirección 5' o en dirección 3' del sitio preseleccionado. Esto permite un mejor control de la inserción del DNA extraño. De hecho, la integración por recombinación homóloga permitirá la unión precisa del fragmento de DNA extraño al genoma nuclear de la planta hasta el nivel nucleotídico.

Las secuencias nucleotídicas flanqueantes pueden variar en longitud, y deberían tener una longitud de al menos alrededor de 10 nucleótidos. Sin embargo, la región flanqueante puede ser tan larga como sea prácticamente posible (v.g. hasta alrededor de 100-150 kb, tal como cromosomas artificiales bacterianos (BACs) completos). Preferiblemente, la región flanqueante tendrá alrededor de 50 pb a alrededor de 2000 pb. Además, las regiones que flanquean el DNA extraño de interés no precisan ser idénticas a las regiones de DNA que flanquean el sitio preseleccionado, y pueden tener entre alrededor de 80% y alrededor de 100% de identidad de secuencia, con preferencia alrededor de 95% a alrededor de 100% de identidad de secuencia con las regiones de DNA que flanquean el sitio preseleccionado. Cuanto más larga sea la región flanqueante, menos restrictivo será el requisito de homología. Además, se prefiere que la identidad de secuencia sea tan alta como sea prácticamente posible en la

proximidad de la localización de inserción exacta del DNA extraño.

Además, las regiones que flanquean el DNA extraño de interés no precisan tener homología con las regiones que flanquean inmediatamente el sitio preseleccionado, sino que pueden tener homología con una región de DNA del genoma nuclear situada más lejos de dicho sitio preseleccionado. La inserción del DNA extraño dará entonces como resultado una eliminación del DNA diana entre el sitio de inserción preseleccionado y la región de homología del DNA. Dicho de otro modo, el DNA diana localizado entre las regiones de homología se sustituirá por el DNA extraño de interés.

Para el propósito de esta invención, la "identidad de secuencia" de dos secuencias nucleotídicas o de aminoácidos afines, expresada como porcentaje, hace referencia al número de posiciones en las dos secuencias alineadas óptimamente que tienen restos idénticos (x 100) dividido entre el número de posiciones comparadas. Un salto, es decir, una posición en una alineación en la cual está presente un resto en una secuencia pero no en la otra, se considerará como una posición con restos no idénticos. La alineación de las dos secuencias se realiza mediante el algoritmo de Needleman y Wunsch (Needleman y Wunsch, 1970). La alineación de secuencias asistida por ordenador puede realizarse convenientemente utilizando programas de software estándar tales como GAP, que forma parte del paquete Wisconsin, versión 10.1 (Genetics Computer Group, Madison, Wisconsin, EE.UU.) utilizando la matriz de registro por defecto con una penalización por creación de salto de 50 y una penalización por extensión de salto de 3.

También se describe un fragmento de DNA modificado que codifica I-SceI y el uso del mismo para introducir eficientemente un DNA extraño de interés en un sitio preseleccionado de un genoma de una célula vegetal, con lo que el fragmento de DNA modificado que codifica I-SceI tiene una secuencia nucleotídica que se ha diseñado para cumplir los criterios siguientes:

- a) la secuencia nucleotídica codifica una endonucleasa I-SceI funcional, tal como una endonucleasa I-SceI que tiene la secuencia de aminoácidos como se proporciona en SEQ ID No. 1;
- b) la secuencia nucleotídica tiene un contenido de GC de alrededor de 50% a alrededor de 60%;
- c) la secuencia nucleotídica no comprende una secuencia nucleotídica seleccionada del grupo que consiste en GATAAT, TATAAA, AATATA, AATATT, GATAAA, AATGAA, AATAAG, AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA y CATAAA;
- d) el nucleótido no comprende una secuencia nucleotídica seleccionada del grupo que consiste en CCAAT, ATTGG, GCAAT y ATTGC;
- e) la secuencia nucleotídica no comprende una secuencia seleccionada del grupo que consiste en ATTTA, AAGGT, AGGTA, GGTA o GCAGG;
- f) la secuencia nucleotídica no comprende un tramo de GC que consiste en 7 nucleótidos consecutivos seleccionados del grupo de G o C;
- g) la secuencia nucleotídica no comprende un tramo de GC que consiste en 5 nucleótidos consecutivos seleccionados del grupo de A o T; y
- h) la secuencia nucleotídica no comprende codones que codifiquen Leu, Ile, Val, Ser, Pro, Thr, Ala que comprenden dobletes TA o GC en posiciones 2 y 3 (es decir, la secuencia nucleotídica no comprende los codones TTA, CTA, ATA, GTA, TCG, CCG, ACG y GCG).

I-SceI es una endonucleasa específica del sitio, responsable de la movilidad de los intrones en las mitocondrias en *Saccharomyces cerevisiae*. La enzima es codificada por el intrón opcional Sc LSU.1 del gen de rRNA 21S, e inicia una rotura del DNA bicatenario en el sitio de inserción del intrón, generando un corte escalonado de 4 pb con salientes 3'OH. El sitio de reconocimiento de la endonucleasa I-SceI se extiende a lo largo de una secuencia no simétrica de 18 pares de bases (Colleaux et al. 1988 Proc. Natl. Acad. Sci. USA 85:6022-6026). La secuencia de aminoácidos para I-SceI y un equivalente de código universal del gen mitocondrial de I-SceI han sido proporcionados, v.g., por el documento WO 96/14408.

El documento WO 96/14408 describe que las variantes siguientes de la proteína I-SceI son todavía funcionales:

- las posiciones 1 a 10 pueden eliminarse
- posición 36: se tolera Gly (G)
- posición 40: se toleran Met (M) o Val (V)

- posición 41: se toleran Ser (S) o Asn (N)
 - posición 43: se tolera Ala (A)
 - posición 46: se toleran Val (V) o N (Asn)
 - posición 91: se tolera Ala (A)
- 5
- posiciones 123 y 156: se tolera Leu (L)
 - posición 223 : se toleran Ala (A) y Ser (S)

y también pueden diseñarse y utilizarse de acuerdo con la presente invención secuencias sintéticas de nucleótidos que codifican tales enzimas variantes de I-SceI.

10 Una secuencia nucleotídica que codifica la secuencia de aminoácidos de I-SceI, en la cual los cuatro aminoácidos localizados en posición amino terminal se han reemplazado por una señal de localización nuclear (SEQ ID No. 1) consiste así en 244 trinucleótidos que pueden representarse como R1 a R244. Para cada una de estas posiciones son posibles entre 1 y 6 posibles elecciones de trinucleótidos que codifican el mismo aminoácido. La Tabla 1 presenta las elecciones posibles para los trinucleótidos que codifican la secuencia de aminoácidos de SEQ ID 1, y
 15 proporciona los requisitos estructurales (sean condicionales o absolutos) que permiten evitar la inclusión en la secuencia de DNA sintética de las "secuencias nucleotídicas prohibidas" arriba mencionadas. Se proporciona también la secuencia nucleotídica de los trinucleótidos contiguos en el código UIPAC.

Como se utilizan en esta memoria, los símbolos del código UIPAC tienen su significado habitual, es decir: N= A o C o G o T; R= A o G; Y= C o T; B= C o G o T (no A); V= A o C o G (no T); D= A o G o T (no C); H= A o C o T (no G); K= G o T; M= A o C; S= G o C; W=A o T.

20 Un fragmento de DNA sintético aislado adecuado es aquel que comprende una secuencia nucleotídica como la representada en SEQ ID No. 2, en la que los codones se seleccionan entre las opciones proporcionadas, de tal manera para obtener una secuencia nucleotídica con un contenido global de GC de alrededor de 50% a alrededor de 60%, con preferencia alrededor de 54% a 55%, con la condición de que la secuencia nucleotídica desde la posición 28 a la posición 30 no es AAG; si la secuencia nucleotídica en la posición 34 a la posición 36 es AAT, entonces la secuencia nucleotídica desde la posición 37 a la posición 39 no es ATT o ATA; si la secuencia nucleotídica desde la posición 34 a la posición 36 es AAC, entonces la secuencia nucleotídica desde la posición 37 a la posición 39 no es ATT simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 40 a la posición 42 sea AAA; si la secuencia nucleotídica desde la posición 34 a la posición 36 es AAC, entonces la secuencia nucleotídica desde la posición 37 a la posición 39 no es ATA; si la secuencia nucleotídica desde la posición 37 a la posición 39 es ATT o ATA, entonces la secuencia nucleotídica desde la posición 40 a la 42 no es AAA; la secuencia nucleotídica desde la posición 49 a la posición 51 no es CAA; la secuencia nucleotídica desde la posición 52 a la posición 54 no es GTA; los codones de la secuencia nucleotídica desde la posición 58 a la posición 63 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 67 a la posición 69 es CCC, entonces la secuencia nucleotídica desde la posición 70 a la posición 72 no es AAT; si la secuencia nucleotídica desde la posición 76 a la posición 78 es AAA, entonces la secuencia nucleotídica desde la posición 79 a la posición 81 no es TTG simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 82 a la 84 sea CTN; si la secuencia nucleotídica desde la posición 79 a la posición 81 es TTA o CTA, entonces la secuencia nucleotídica desde la posición 82 a la posición 84 no es TTA; la secuencia nucleotídica desde la posición 88 a la posición 90 no es GAA; si la secuencia nucleotídica desde la posición 91 a la posición 93 es TAT, entonces la secuencia nucleotídica desde la posición 94 a la posición 96 no es AAA; si la secuencia nucleotídica desde la posición 97 a la posición 99 es TCC o TCG o AGC, entonces la secuencia nucleotídica desde la posición 100 a la 102 no es CCA simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 103 a 105 sea TTR; si la secuencia nucleotídica desde la posición 100 a la 102 es CAA, entonces la secuencia nucleotídica desde la posición 103 a la 105 no es TTA; si la secuencia nucleotídica desde la posición 109 a la posición 111 es GAA, entonces la secuencia nucleotídica desde la posición 112 a la 114 no es TTA; si la secuencia nucleotídica desde la posición 115 a 117 es AAT, entonces la secuencia nucleotídica desde la posición 118 a la posición 120 no es ATT o ATA; si la secuencia nucleotídica desde la posición 121 a 123 es GAG, entonces la secuencia nucleotídica desde la posición 124 a la posición 126; la secuencia nucleotídica desde la posición 133 a 135 no es GCA; la secuencia nucleotídica desde la posición 139 a la posición 141 no es ATT; si la secuencia nucleotídica desde la posición 142 a la posición 144 es GGA, entonces la secuencia nucleotídica desde la posición 145 a la posición 147 no es TTA; si la secuencia nucleotídica desde la posición 145 a la posición 147 es TTA, entonces la secuencia nucleotídica desde la posición 148 a la posición 150 no es ATA simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 151 a 153 sea TTR; si la secuencia nucleotídica desde la posición 145 a la posición 147 es CTA, entonces la secuencia nucleotídica desde la posición 148 a la posición 150 no es ATA simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 151 a 153 sea TTR; si la secuencia nucleotídica desde la posición 148 a la posición 150 es ATA, entonces la secuencia nucleotídica desde la posición 151 a la posición 153 no es CTA o TTG;

si la secuencia nucleotídica desde la posición 160 a la posición 162 es GCA, entonces la secuencia nucleotídica desde la posición 163 a la posición 165 no es TAC; si la secuencia nucleotídica desde la posición 163 a la posición 165 es TAT, entonces la secuencia nucleotídica desde la posición 166 a la posición 168 no es ATA simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 169 a la posición 171 sea AGR; los codones de la secuencia nucleotídica desde la posición 172 a la posición 177 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda GCAGG; los codones de la secuencia nucleotídica desde la posición 178 a la posición 186 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda AGGTA; si la secuencia nucleotídica desde la posición 193 a la posición 195 es TAT, entonces la secuencia nucleotídica desde la posición 196 a la posición 198 no es TGC; la secuencia nucleotídica desde la posición 202 a la posición 204 no es CAA; la secuencia nucleotídica desde la posición 217 a la posición 219 no es AAT; si la secuencia nucleotídica desde la posición 220 a la posición 222 es AAA, entonces la secuencia nucleotídica desde la posición 223 a la posición 225 no es GCA; si la secuencia nucleotídica desde la posición 223 a la posición 225 es GCA, entonces la secuencia nucleotídica desde la posición 226 a la posición 228 no es TAC; si la secuencia nucleotídica desde la posición 253 a la posición 255 es GAC, entonces la secuencia nucleotídica desde la posición 256 a la posición 258 no es CAA; si la secuencia nucleotídica desde la posición 277 a la posición 279 es CAT, entonces la secuencia nucleotídica desde la posición 280 a la posición 282 no es AAA; los codones de la secuencia nucleotídica desde la posición 298 a la posición 303 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 304 a la posición 306 es GGC, entonces la secuencia nucleotídica desde la posición 307 a la posición 309 no es AAT; los codones de la secuencia nucleotídica desde la posición 307 a la posición 312 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; los codones de la secuencia nucleotídica desde la posición 334 a la posición 342 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 340 a la posición 342 es AAG, entonces la secuencia nucleotídica desde la posición 343 a 345 no es CAT; si la secuencia nucleotídica desde la posición 346 a la posición 348 es CAA, entonces la secuencia nucleotídica desde la posición 349 a la posición 351 no es GCA; los codones de la secuencia nucleotídica desde la posición 349 a la posición 357 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; la secuencia nucleotídica desde la posición 355 a la posición 357 no es AAT; si la secuencia nucleotídica desde la posición 358 a la posición 360 es AAA, entonces la secuencia nucleotídica desde la posición 361 a 363 no es TTG; si la secuencia nucleotídica desde la posición 364 a la posición 366 es GCC, entonces la secuencia nucleotídica desde la posición 367 a la posición 369 no es AAT; los codones de la secuencia nucleotídica desde la posición 367 a la posición 378 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 382 a la posición 384 es AAT, entonces la secuencia nucleotídica desde la posición 385 a la posición 387 no es AAT; la secuencia nucleotídica desde la posición 385 a la posición 387 no es AAT; si la secuencia nucleotídica desde la posición 400 a 402 es CCC, entonces la secuencia nucleotídica desde la posición 403 a 405 no es AAT; si la secuencia nucleotídica desde la posición 403 a 405 es AAT, entonces la secuencia nucleotídica desde la posición 406 a 408 no es AAT; los codones de la secuencia nucleotídica desde la posición 406 a la posición 411 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; los codones de la secuencia nucleotídica desde la posición 421 a la posición 426 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; la secuencia nucleotídica desde la posición 430 a la posición 432 no es CCA; si la secuencia nucleotídica desde la posición 436 a la posición 438 es TCA, entonces la secuencia nucleotídica desde la posición 439 a la posición 441 no es TTG; la secuencia nucleotídica desde la posición 445 a la posición 447 no es TAT; la secuencia nucleotídica desde la posición 481 a 483 no es AAT; si la secuencia nucleotídica desde la posición 484 a la posición 486 es AAA, entonces la secuencia nucleotídica desde la posición 487 a la posición 489 no es AAT simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 490 a la posición 492 sea AGY; si la secuencia nucleotídica desde la posición 490 a la posición 492 es TCA, entonces la secuencia nucleotídica desde la posición 493 a la posición 495 no es ACC simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 496 a 498 sea AAY; si la secuencia nucleotídica desde la posición 493 a la posición 495 es ACC, entonces la secuencia nucleotídica desde la posición 496 a 498 no es AAT; la secuencia nucleotídica desde la posición 496 a la posición 498 no es AAT; si la secuencia nucleotídica desde la posición 499 a la posición 501 es AAA, entonces la secuencia nucleotídica desde la posición 502 a la posición 504 no es TCA o AGC; si la secuencia nucleotídica desde la posición 508 a posición 510 es GTA, entonces la secuencia nucleotídica desde la posición 511 a 513 no es TTA; si la secuencia nucleotídica desde la posición 514 a la posición 516 es AAT, entonces la secuencia nucleotídica desde la posición 517 a la posición 519 no es ACA; si la secuencia nucleotídica desde la posición 517 a la posición 519 es ACC o ACG, entonces la secuencia nucleotídica desde la posición 520 a la posición 522 no es CAA simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 523 a la posición 525 sea TCN; los codones de la secuencia nucleotídica desde la posición 523 a la posición 531 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 544 a la posición 546 es GAA, entonces la secuencia nucleotídica desde la posición 547 a la posición 549 no es TAT, simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 550 a la posición 552 sea TTR; los codones de la secuencia nucleotídica desde la posición 547 a la posición 552 se

seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 559 a la posición 561 es GGA, entonces la secuencia nucleotídica desde la posición 562 a la posición 564 no es TTG simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 565 a 567 sea CGN; si la secuencia nucleotídica desde la posición 565 a la posición 567 es CGC, entonces la secuencia nucleotídica desde la posición 568 a la posición 570 no es AAT; la secuencia nucleotídica desde la posición 568 a la posición 570 no es AAT; si la secuencia nucleotídica desde la posición 574 a la posición 576 es TTC, entonces la secuencia nucleotídica desde la posición 577 a la posición 579 no es CAA simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 580 a la posición 582 sea TTR; si la secuencia nucleotídica desde la posición 577 a la posición 579 es CAA, entonces la secuencia nucleotídica desde la posición 580 a la posición 582 no es TTA; si la secuencia nucleotídica desde la posición 583 a la posición 585 es AAT la secuencia nucleotídica desde la posición 586 a 588 no es TGC; la secuencia nucleotídica desde la posición 595 a la posición 597 no es AAA; si la secuencia nucleotídica desde la posición 598 a la posición 600 es ATT, entonces la secuencia nucleotídica desde la posición 601 a la posición 603 no es AAT; la secuencia nucleotídica desde la posición 598 a la posición 600 no es ATA; la secuencia nucleotídica desde la posición 601 a la posición 603 no es AAT; si la secuencia nucleotídica desde la posición 604 a la posición 606 es AAA, entonces la secuencia nucleotídica desde la posición 607 a la posición 609 no es AAT; la secuencia nucleotídica desde la posición 607 a la posición 609 no es AAT; la secuencia nucleotídica desde la posición 613 a la posición 615 no es CCA; si la secuencia nucleotídica desde la posición 613 a la posición 615 es CCG, entonces la secuencia nucleotídica desde la posición 616 a la posición 618 no es ATA; si la secuencia nucleotídica desde la posición 616 al nucleótido de la posición 618 es ATA, entonces la secuencia nucleotídica desde la posición 619 a 621 no es ATA; si la secuencia nucleotídica desde la posición 619 a la posición 621 es ATA, entonces la secuencia nucleotídica desde la posición 622 a la posición 624 no es TAC; la secuencia nucleotídica desde la posición 619 a la posición 621 no es ATT; los codones de la secuencia nucleotídica desde la posición 640 a la posición 645 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 643 a la posición 645 es TTA, entonces la secuencia nucleotídica desde la posición 646 a la posición 648 no es ATA; si la secuencia nucleotídica desde la posición 643 a la posición 645 es CTA, entonces la secuencia nucleotídica desde la posición 646 a la posición 648 no es ATA; los codones de la secuencia nucleotídica desde la posición 655 a la posición 660 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 658 a 660 es TTA o CTA, entonces la secuencia nucleotídica desde la posición 661 a la posición 663 no es ATT o ATC; la secuencia nucleotídica desde la posición 661 a la posición 663 no es ATA; si la secuencia nucleotídica desde la posición 661 a la posición 663 es ATT, entonces la secuencia nucleotídica desde la posición 664 a la posición 666 no es AAA; los codones de la secuencia nucleotídica desde la posición 670 a la posición 675 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda ATTTA; si la secuencia nucleotídica desde la posición 691 a la posición 693 es TAT, entonces la secuencia nucleotídica desde la posición 694 a la posición 696 no es AAA; si la secuencia nucleotídica desde la posición 694 a la posición 696 es AAA, entonces la secuencia nucleotídica desde la posición 697 a la posición 699 no es TTG; si la secuencia nucleotídica desde la posición 700 a la posición 702 es CCC, entonces la secuencia nucleotídica desde la posición 703 a la posición 705 no es AAT; si la secuencia nucleotídica desde la posición 703 a la posición 705 es AAT, entonces la secuencia nucleotídica desde la posición 706 a la posición 708 no es ACA o ACT; si la secuencia nucleotídica desde la posición 706 a la posición 708 es ACA, entonces la secuencia nucleotídica desde la posición 709 a 711 no es ATA simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 712 a la posición 714 sea AGY; la secuencia nucleotídica no comprende los codones TTA, CTA, ATA, GTA, TCG, CCG, ACG y GCG; dicha secuencia nucleotídica no comprende un tramo GC que consiste en 7 nucleótidos consecutivos seleccionados del grupo de G o C; y la secuencia nucleotídica no comprende un tramo AT que consiste en 5 nucleótidos consecutivos seleccionados del grupo de A o T.

También son adecuadas las secuencias nucleotídicas sintéticas como se presentan en la Tabla 2, y que corresponden a un fragmento de DNA sintético aislado como se proporciona que comprende una secuencia nucleotídica como se presenta en SEQ ID No 3, en la que los codones se seleccionan entre las opciones proporcionadas de tal manera para obtener una secuencia nucleotídica con un contenido global de GC de alrededor de 50% a alrededor de 60%, con preferencia alrededor de 54%-55%, con la condición de que si la secuencia nucleotídica desde la posición 121 a la posición 123 es GAG, entonces la secuencia nucleotídica desde la posición 124 a 126 no sea CAA; si la secuencia nucleotídica desde la posición 253 a la posición 255 es GAC, entonces la secuencia nucleotídica desde la posición 256 a 258 no es CAA; si la secuencia nucleotídica desde la posición 277 a la posición 279 es CAT, entonces la secuencia nucleotídica desde la posición 280 a 282 no es AAA; si la secuencia nucleotídica desde la posición 340 a la posición 342 es AAG, entonces la secuencia nucleotídica desde la posición 343 a la posición 345 no es CAT; si la secuencia nucleotídica desde la posición 490 a la posición 492 es TCA, entonces la secuencia nucleotídica desde la posición 493 a la posición 495 no es ACC; si la secuencia nucleotídica desde la posición 499 a la posición 501 es AAA, entonces la secuencia nucleotídica desde la posición 502 a 504 no es TCA o AGC; si la secuencia nucleotídica desde la posición 517 a la posición 519 es ACC, entonces la secuencia nucleotídica desde la posición 520 a la posición 522 no es CAA simultáneamente con el hecho de que la secuencia nucleotídica desde la posición 523 a 525 sea TCN; si la secuencia nucleotídica desde la posición 661 a la posición 663 es ATT, entonces la secuencia nucleotídica desde la posición 664 a la posición 666 no es AAA; los

5 codones de la secuencia nucleotídica desde la posición 7 a la posición 15 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de siete nucleótidos contiguos del grupo de G o C; los codones de la secuencia nucleotídica desde la posición 61 a la posición 69 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de siete nucleótidos contiguos del grupo de G o C; los codones de la secuencia nucleotídica desde la posición 130 a la posición 138 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de siete nucleótidos contiguos del grupo de G o C; los codones de la secuencia nucleotídica desde la posición 268 a la posición 279 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de siete nucleótidos contiguos del grupo de G o C; los codones de la secuencia nucleotídica desde la posición 322 a la posición 333 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de siete nucleótidos contiguos del grupo de G o C; los codones de la secuencia nucleotídica desde la posición 460 a la posición 468 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de siete nucleótidos contiguos del grupo de G o C; los codones de la secuencia nucleotídica desde la posición 13 a la posición 27 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 37 a la posición 48 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 184 a la posición 192 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 214 a la posición 219 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 277 a la posición 285 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; y los codones de la secuencia nucleotídica desde la posición 388 a la posición 396 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 466 a la posición 474 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 484 a la posición 489 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 571 a la posición 576 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 598 a la posición 603 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 604 a la posición 609 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 613 a la posición 621 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 646 a la posición 651 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; los codones de la secuencia nucleotídica desde la posición 661 a la posición 666 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T; y los codones de la secuencia nucleotídica desde la posición 706 a la posición 714 se seleccionan de acuerdo con las opciones proporcionadas de tal manera que la secuencia nucleotídica resultante no comprenda un tramo de cinco nucleótidos contiguos del grupo de A o T.

55 La secuencia nucleotídica de SEQ ID No 4 es un ejemplo de una secuencia nucleotídica sintética de este tipo que codifica una endonucleasa I-SceI que ya no contiene ninguna de las secuencias nucleotídicas o codones a evitar. Sin embargo, estará claro que una persona experta en la técnica puede obtener fácilmente una secuencia similar codificante de I-SceI reemplazando uno o más (entre dos y veinte) de los nucleótidos a seleccionar por cualquiera de las alternativas proporcionadas en la secuencia nucleotídica de SEQ ID 3 (excluyendo cualquiera de las combinaciones prohibidas descritas en el párrafo anterior), y utilizarla para obtener un efecto similar.

Para la expresión en célula vegetal, los fragmentos de DNA sintéticos que codifican I-SceI se pueden enlazar operativamente a un promotor expresable en plantas, a fin de obtener un gen quimérico expresable en plantas.

En todavía otro aspecto, la invención se refiere a un método para introducir un DNA extraño de interés en un sitio preseleccionado de un genoma de una célula vegetal, que comprende las etapas de

(a) inducir una rotura bicatenaria en el sitio preseleccionado en el genoma de la célula mediante introducción en dicha célula vegetal de un gen expresable en plantas que codifica una enzima que induce una rotura de DNA bicatenario de corte raro que reconoce dicho sitio preseleccionado;

5 (b) introducir el DNA extraño de interés en la célula vegetal mediante transferencia directa de DNA; caracterizado porque, antes de la etapa (a), las células vegetales se incuban en un compuesto fenólico de origen vegetal.

“Compuestos fenólicos de origen vegetal” o “fenólicos de origen vegetal” adecuados para la invención son aquellas moléculas fenólicas sustituidas que son capaces de inducir una respuesta quimiotáctica positiva, particularmente aquellas que son capaces de inducir expresión incrementada de genes vir en una *Agrobacterium sp.* que contiene un plásmido Ti, particularmente una *Agrobacterium tumefaciens* que contiene un plásmido Ti. Los métodos para medir la respuesta quimiotáctica frente a compuestos fenólicos de origen vegetal han sido descritos por Ashby et al., (1988 J. Bacteriol. 170:4181-4187), y los métodos para medir la inducción de la expresión de genes vir son también muy conocidos (Stachel et al., 1985 Nature 318:624-629; Bolton et al., 1986 Science 232:983-985). Los compuestos fenólicos de origen vegetal preferidos son los encontrados en exudados de lesiones de células vegetales. Uno de los compuestos fenólicos de origen vegetal mejor conocidos es la acetosiringona, que está presente en cierto número de células lesionadas e intactas de diversas plantas, si bien en diferentes concentraciones. Sin embargo, la acetosiringona (3,5-dimetoxi-4-hidroxiacetofenona) no es el único compuesto fenólico de origen vegetal que puede inducir la expresión de genes vir. Otros ejemplos son α -hidroxi-acetosiringona, ácido sinapínico (ácido 3,5-dimetoxi-4-hidroxicinámico), ácido siríngico (ácido 4-hidroxi-3,5-dimetoxibenzoico), ácido ferúlico (ácido 4-hidroxi-3-metoxicinámico), catecol (1,2-dihidroxibenceno), ácido p-hidroxibenzoico (ácido 4-hidroxibenzoico), ácido β -resorcílico (ácido 2,4-dihidroxibenzoico), ácido protocatéquico (ácido 3,4-dihidroxibenzoico), ácido pirogálico (ácido 2,3,4-trihidroxibenzoico), ácido gálico (ácido 3,4,5-trihidroxibenzoico) y vanillina (3-metoxi-4-hidroxibenzaldehído). Como se utilizan en esta memoria, las moléculas mencionadas se denominan como compuestos fenólicos de origen vegetal. Los compuestos fenólicos de origen vegetal pueden añadirse al medio de cultivo vegetal, ya sea aisladamente o en combinación con otros compuestos fenólicos de origen vegetal. Aunque no se pretende limitar la invención a un modo de acción particular, se cree que el efecto estimulante aparente de estos compuestos fenólicos de origen vegetal sobre la división celular (y de este modo también sobre la replicación genómica) pueden mejorar la inserción direccionada del DNA extraño.

Las células vegetales se incuban preferiblemente en el compuesto fenólico de origen vegetal durante alrededor de una semana, aunque se espera que la incubación durante alrededor de uno o dos días en o sobre un compuesto fenólico de origen vegetal será suficiente. Las células vegetales deberían incubarse durante un tiempo suficiente para estimular la división celular. De acuerdo con Guivarc'h et al., (1993), Protoplasma 174:10-18), dicho efecto puede obtenerse fácilmente por incubación de las células vegetales durante un tiempo tan reducido como 10 minutos.

Los métodos mejorados arriba mencionados para inserción de DNA direccionada basada en la recombinación homóloga pueden aplicarse también para mejorar la calidad de las células de plantas transgénicas y plantas obtenidas por métodos de transferencia directa de DNA, particularmente por bombardeo con microproyectiles. Es bien conocido en la técnica que la introducción de DNA por bombardeo con microproyectiles conduce frecuentemente a patrones de integración complejos del DNA introducido (integración de copias múltiples del DNA extraño de interés, sea completa o parcial, generación de estructuras repetidas). No obstante, algunos genotipos o variedades de plantas pueden ser más aptos para transformación utilizando bombardeo con microproyectiles que para transformación utilizando v.g. *Agrobacterium tumefaciens*. De este modo, sería ventajoso si pudiera mejorarse la calidad de las células vegetales o plantas transgénicas obtenidas por bombardeo con microproyectiles, es decir, si pudiera influirse en el patrón de integración del DNA extraño para hacerlo más simple.

El descubrimiento arriba mencionado de que la introducción de DNA extraño por bombardeo con microproyectiles en presencia de una rotura inducida del DNA bicatenario en el genoma celular, con lo que el DNA extraño tiene homología con las secuencias que flanquean la rotura del DNA bicatenario, conduce frecuentemente (alrededor de 50% de los sucesos obtenidos) a patrones de integración simples (inserción de una sola copia de una manera predecible y sin inserción de fragmentos adicionales del DNA extraño) proporciona la base para un método de simplificación de la complejidad de inserción del DNA extraño en el genoma nuclear de las células vegetales.

De este modo, la invención también se refiere a un método de producción de una planta transgénica por bombardeo con microproyectiles, comprende las etapas de

(a) inducir una rotura de DNA bicatenario en un sitio preseleccionado en el genoma de una célula vegetal, de acuerdo con los métodos descritos en cualquier otro lugar en este documento o disponibles en la técnica; y

55 (b) introducir el DNA extraño de interés en la célula vegetal por bombardeo con microproyectiles, en el que

dicho DNA extraño de interés está flanqueado por dos regiones de DNA que tienen al menos 80% de identidad de secuencia con las regiones de DNA que flanquean el sitio preseleccionado en el genoma de la planta.

5 Una porción significativa de la población de plantas transgénicas así obtenida tendrá un patrón de integración simple del DNA extraño en el genoma de las células vegetales; más particularmente, una porción significativa de las plantas transgénicas tendrán solamente una inserción de una copia del DNA extraño, comprendida exactamente entre las dos regiones de DNA que flanquean el sitio preseleccionado en el genoma de la planta. Esta porción es mayor que la población de plantas transgénicas con patrones de integración simples, cuando las plantas se obtienen por bombardeo simple con microproyectiles sin inducir una rotura del DNA bicatenario, y sin proporcionar el DNA
10 extraño con homología a las regiones genómicas que flanquean el sitio preseleccionado.

La célula vegetal diana puede comprender en su genoma un gen marcador, flanqueado por dos sitios de reconocimiento para una endonucleasa de corte raro inductora de rotura del DNA bicatenario, una en cada lado. Este DNA marcador puede introducirse en el genoma de la célula vegetal de interés utilizando cualquier método de transformación, o puede haber sido introducido en el genoma de una célula vegetal de otra línea o variedad de plantas (tal como una línea o variedad de una planta fácilmente susceptible de transformación) e introducirse en la célula de la planta de interés por técnicas clásicas de mejora genética. Preferiblemente, la población de plantas o células vegetales transgénicas que comprenden un gen marcador flanqueado por dos sitios de reconocimiento para una endonucleasa de corte raro inductora de una rotura bicatenaria ha sido analizada respecto al patrón de expresión del gen marcador (tal como expresión elevada, expresión regulada temporal o espacialmente) y se han
15 identificado las líneas de plantas con el patrón de expresión deseado. La producción de una planta transgénica por bombardeo con microproyectiles, que comprende las etapas de

(a) inducir una rotura del DNA bicatenario en un sitio preseleccionado en el genoma de una célula vegetal, de acuerdo con los métodos descritos en cualquier otro lugar de este documento o disponibles en la técnica; e

(b) introducir el DNA extraño de interés en la célula vegetal por bombardeo con microproyectiles, en el que
25 dicho DNA extraño de interés está flanqueado por dos regiones de DNA que tienen al menos 80% de identidad de secuencia con las regiones de DNA que flanquean el sitio preseleccionado en el genoma de la planta;

conducirá a células vegetales y plantas transgénicas en las cuales el gen marcador ha sido reemplazado por el DNA extraño de interés.

30 El gen marcador puede ser cualquier gen marcador seleccionable o un gen marcador expresable en plantas y susceptible de cribado, que es preferiblemente un gen marcador quimérico convencional. El gen marcador quimérico puede comprender un DNA marcador que se encuentra bajo el control de, y enlazado operativamente en su extremo 5', a un promotor, preferiblemente un promotor constitutivo expresable en plantas, tal como un promotor 35S de CaMV, o un promotor ligero inducible tal como el promotor del gen que codifica la subunidad pequeña de Rubisco; y enlazado operativamente en su extremo 3' a señales adecuadas de poliadenilación y terminación de la transcripción
35 en plantas. El DNA marcador codifica preferiblemente un RNA, proteína o polipéptido que, cuando se expresa en las células de una planta, permite que tales células se separen fácilmente de aquellas células en las cuales no se expresa el DNA marcador. La elección del DNA marcador no es crítica, y puede seleccionarse cualquier DNA marcador adecuado de una manera bien conocida. Por ejemplo, un DNA marcador puede codificar una proteína que proporciona un color distinguible a la célula vegetal transformada, tal como el gen A1 (Meyer et al., (1987), Nature 330:677), puede codificar una proteína fluorescente [Chalfie et al., Science 263; 802-805 (1994); Crameri et al., Nature Biotechnology 14:315-319 (1996)], puede codificar una proteína que proporciona a la célula vegetal transformada resistencia a los herbicidas, tal como el gen bar, codificante de PAT, que proporciona resistencia a la fosfinotricina (documento EP 0242246), o puede codificar una proteína que proporciona a las células transformadas
40 resistencia a los antibióticos, tal como el gen aac(6'), que codifica GAT, que proporciona resistencia a la gentamicina (documento WO 94/01560). Un gen marcador seleccionable de este tipo codifica generalmente una proteína que confiere a la célula resistencia a un antibiótico u otro compuesto químico que es normalmente tóxico para las células. En las plantas, el gen marcador seleccionable puede codificar también así una proteína que confiere resistencia a un herbicida, tal como un herbicida que comprende un inhibidor de la glutamina sintetasa (v.g. fosfinotricina) como ingrediente activo. Un ejemplo de tales genes son genes que codifican fosfinotricin acetil transferasa, tales como los genes sfr o sfrv (EP 242236; EP 242246; De Block et al., 1987 EMBO J. 6:2513-2518).
45

El DNA de reparación introducido puede comprender adicionalmente un gen marcador que permite discriminar mejor entre la integración por recombinación homóloga en el sitio preseleccionado y la integración en cualquier otra parte del genoma. Genes marcadores de este tipo están disponibles en la técnica, e incluyen genes marcadores por cuyo medio la ausencia del gen marcador puede seleccionarse positivamente en condiciones selectivas (v.g. *codA*, citosina desaminasa de *E. coli* que confiere sensibilidad a 5-fluoro-citosina, Perera *et al.* 1993 Plant Mol. Biol. 23, 793; Stougaard (1993) Plant J.: 755). El DNA de reparación tiene que comprender el gen marcador de una manera tal que la integración del DNA de reparación en el genoma nuclear de manera aleatoria da como resultado la
55

presencia del gen marcador, mientras que la integración del DNA de reparación por recombinación homóloga da como resultado la ausencia del gen marcador.

5 Quedará inmediatamente claro que pueden obtenerse también los mismos resultados utilizando un solo sitio preseleccionado en el que inducir la rotura bicatenaria, que está localizado en o cerca de un gen marcador. Las regiones de homología flanqueantes se seleccionan luego preferiblemente de tal manera que o bien desactivan el gen marcador, o eliminan el gen marcador y se sustituye por el DNA extraño a insertar.

10 Se apreciará que los medios y métodos de la invención son particularmente útiles para maíz, pero pueden ser utilizados también en otras plantas con efectos similares, particularmente en plantas de cereales que incluyen trigo, avena, cebada, centeno, arroz, césped de hipódromo, sorgo, mijo o caña de azúcar. Los métodos de la invención pueden aplicarse también a cualquier planta, incluyendo, pero sin limitarse a, algodón, tabaco, cáñola, colza oleaginosa, soja, hortalizas, patatas, Lemna spp., Nicotiana spp., Arabidopsis, alfalfa, cebada, habichuelas, maíz, algodón, lino, guisante, colza, arroz, centeno, cártamo, sorgo, soja, girasol, tabaco, trigo, espárragos, remolacha, brócoli, col, zanahoria, coliflor, apio, pepino, berenjena, lechuga, cebolla, colza oleaginosa, pimiento, patata, calabaza, rábano, espinaca, calabaza de cuello curvo, tomate, calabacín, almendra, manzana, albaricoque, plátano, mora, arándano común, cacao, cereza, coco, arándano agrio, dátil, uva, pomelo, guayaba, kiwi, limón, lima, mango, melón, nectarina, naranja, papaya, fruto de la pasión, melocotón, cacahuete, pera, piña, pistacho, ciruela, frambuesa, fresa, mandarina, nogal y sandía.

20 De acuerdo con los métodos de la invención pueden obtenerse células vegetales y plantas que comprenden moléculas de DNA extraño insertadas en sitios preseleccionados. También pueden obtenerse gametos, semillas, embriones, sean cigóticos o somáticos, progenie o híbridos de plantas que comprenden los sucesos de inserción de DNA direccionada, que se producen por métodos de mejora genética tradicionales.

Las plantas obtenidas por los métodos descritos en esta memoria pueden cruzarse adicionalmente por técnicas tradicionales de mejora genética con otras plantas para obtener plantas de progenie que comprenden los sucesos de inserción de DNA direccionada obtenidos de acuerdo con la presente invención.

25 Los siguientes ejemplos no limitantes describen el diseño de un gen quimérico modificado codificante de I-SceI, y el uso del mismo para insertar DNA extraño en un sitio preseleccionado del genoma vegetal.

30 A no ser que se indique de otro modo en los ejemplos, todas las técnicas de DNA recombinante se llevan a cabo de acuerdo con protocolos estándar como se describen en Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, Segunda Edición, Cold Spring Harbor Laboratory Press, NY y en los Volúmenes 1 y 2 de Ausubel et al. (1994) *Current Protocols in Molecular Biology*, Current Protocols, USA. Materiales y métodos estándar para trabajo molecular con plantas se describen en *Plant Molecular Biology Labfax* (1993) por R.D.D. Croy, publicado conjuntamente por BIOS Scientific Publications Ltd (Reino Unido) y Blackwell Scientific Publications, Reino Unido. Otras referencias para técnicas estándar de biología molecular incluyen Sambrook y Russell (2001) *Molecular Cloning: A Laboratory Manual*, Tercera Edición, Cold Spring Harbor Laboratory Press, NY, Volúmenes I y II de Brown (1998) *Molecular Biology LabFax*, Segunda Edición, Academic Press (UK). Materiales y métodos estándar para las reacciones en cadena de la polimerasa pueden encontrarse en Dieffenbach y Dveksler (1995) *PCR Primer: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, y en McPherson et al. (2000) *PCR - Basics: From Background to Bench*, Primera Edición, Springer Verlag, Alemania.

A lo largo de la descripción y los ejemplos, se hace referencia a las siguientes secuencias:

40 SEQ ID No 1: secuencia de aminoácidos de una I-SceI quimérica que comprende una señal de localización nuclear enlazada a una proteína I-SceI que carece de los 4 aminoácidos amino-terminales.

SEQ ID No 2: secuencia nucleotídica de la región codificante de I-SceI (código UIPAC).

SEQ ID No 3: secuencia nucleotídica de la región codificante de I-SceI sintética (código UIPAC).

SEQ ID No 4: secuencia nucleotídica de la región codificante de I-SceI sintética.

45 SEQ ID No 5: secuencia nucleotídica del T-DNA de pTTAM78 (locus diana).

SEQ ID No 6: secuencia nucleotídica del T-DNA de pTTA82 (DNA de reparación).

SEQ ID No 7: secuencia nucleotídica de pCV78.

Tabla 1. (correspondiente a SEC ID 2)

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R1	M	ATG	ATG	
R2	A	GCA GCC GCG GCT	GCN	
R3	K	AAA AAG	AAR	
R4	P	CCA CCC CCG CCT	CCN	
R5	P	CCA CCC CCG CCT	CCN	
R6	K	AAA AAG	AAR	
R7	K	AAA AAG	AAR	
R8	K	AAA AAG	AAR	
R9	R	AGA AGG CGA CGC CGG CGT	AGR o CGN	
R10	K	AAA AAG	AAR	NO AAG
R11	V	GTA GTC GTG GTT	GTN	
R12	N	AAC AAT	AAY	SI R12 AAT NO (R13 ATT O R13 ATA). SI R12 AAC NO (R13 ATT Y R14 AAA) SI R12 AAC NO R13 ATA
R13	I	ATA ATC ATT	ATH	SI R13 ATT NO R14 AAA SI R13 ATA NO R14 AAA
R14	K	AAA AAG	AAR	
R15	K	AAA AAG	AAR	
R16	N	AAC AAT	AAY	
R17	Q	CAA CAG	CAR	NO CAA
R18	V	GTA GTC GTG GTT	GTN	NO GTA
R19	M	ATG	ATG	
R20	N	AAC AAT	AAY	EVITAR ATTTA
R21	L	TTA TTG CTA CTC CTG CTT	TTR o CTN	
R22	G	GGA GGC GGG GGT	GGN	
R23	P	CCA CCC CCG CCT	CCN	SI R23 CCC NO R24 AAT
R24	N	AAC AAT	AAY	
R25	S	AGC AGT TCA TCC TCG TCT	AGY o TCN	
R26	K	AAA AAG	AAR	SI R26 AAA NO (R27 TTG Y R28 CTN)
R27	L	TTA TTG CTA CTC CTG CTT	TTR o CTN	SI R27 (TTA O CTA) NO R28 TTA

ES 2 368 962 T3

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R28	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R29	K	AAA AAG	AAR	
R30	E	GAA GAG	GAR	NO GAA
R31	Y	TAC TAT	TAY	SI R31 TAT NO R32 AAA
R32	K	AAA AAG	AAR	
R33	S	AGC AGT TCA TCC TCG TCT	AGY TCN	o SI R33 (TCC O TCG O AGC) NO (R34 CAA Y R35 TTR)
R34	Q	CAA CAG	CAR	SI R34 CAA NO R35 TTA
R35	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R36	I	ATA ATC ATT	ATH	
R37	E	GAA GAG	GAR	SI R37 GAA NO R38 TTA
R38	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R39	N	AAC AAT	AAV	SI R39 AAT NO R40 (ATT O ATA)
R40	I	ATA ATC ATT	ATH	
R41	E	GAA GAG	GAR	SI R41 GAG NO R42 CAA
R42	Q	CAA CAG	CAR	
R43	F	TTC TTT	TTY	
R44	E	GAA GAG	GAR	
R45	A	GCA GCC GCG GCT	GCN	NO GCA
R46	G	GGA GGC GGG GGT	GGN	
R47	I	ATA ATC ATT	ATH	NO ATT
R48	G	GGA GGC GGG GGT	GGN	SI R48 GGA NO R49 TTA
R49	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o SI R49 TTA NO (R50 ATA Y R51 TTR) SI R49 CTA NO (R50 ATA Y R51 TTR)
R50	I	ATA ATC ATT	ATH	SI R50 ATA NO R51 (CTA O TTG)
R51	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R52	G	GGA GGC GGG GGT	GGN	
R53	D	GAC GAT	GAY	
R54	A	GCA GCC GCG GCT	GCN	SI R54 GCA NO R55 TAC
R55	Y	TAC TAT	TAY	SI R55 TAT NO (R56 ATA Y R57 AGR)
R56	I	ATA ATC ATT	ATH	

ES 2 368 962 T3

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R57	R	AGA AGG CGA CGC CGG CGT	AGR CGN	
R58	S	AGC AGT TCA TCC TCG TCT	AGY TCN	EVITAR GCAGG
R59	R	AGA AGG CGA CGC CGG CGT	AGR CGN	
R60	D	GAC GAT	GAY	
R61	E	GAA GAG	GAR	EVITAR AAGGT
R62	G	GGA GGC GGG GGT	GGN	
R63	K	AAA AAG	AAR	
R64	T	ACA ACC ACG ACT	ACN	
R65	Y	TAC TAT	TAY	SI R65 TAT NO R66 TGC
R66	C	TGC TGT	TGY	
R67	M	ATG	ATG	
R68	Q	CAA CAG	CAR	NO CAA
R69	F	TTC TTT	TTY	
R70	E	GAA GAG	GAR	
R71	W	TGG	TGG	
R72	K	AAA AAG	AAR	
R73	N	AAC AAT	AAY	NO AAT
R74	K	AAA AAG	AAR	SI R74 AAA NO R75 GCA
R75	A	GCA GCC GCG GCT	GCN	SI R75 GCA NO R76 TAC
R76	Y	TAC TAT	TAY	
R77	M	ATG	ATG	
R78	D	GAC GAT	GAY	
R79	H	CAC CAT	CAY	
R80	V	GTA GTC GTG GTT	GTN	
R81	C	TGC TGT	TGY	
R82	L	TTA TTG CTA CTC CTG CTT	TTR CTN	
R83	L	TTA TTG CTA CTC CTG CTT	TTR CTN	
R84	Y	TAC TAT	TAY	
R85	D	GAC GAT	GAY	SI R85 GAC NO R86 CAA
R86	Q	CAA CAG	CAR	

ES 2 368 962 T3

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R87	W	TGG	TGG	
R88	V	GTA GTC GTG GTT	GTN	
R89	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R90	S	AGC AGT TCA TCC TCG TCT	AGY TCN	o
R91	P	CCA CCC CCG CCT	CCN	
R92	P	CCA CCC CCG CCT	CCN	
R93	H	CAC CAT	CAY	SI R93 CAT NO R94 AAA
R94	K	AAA AAG	AAR	
R95	K	AAA AAG	AAR	
R96	E	GAA GAG	GAR	
R97	R	AGA AGG CGA CGC CGG CGT	AGR CGN	o
R98	V	GTA GTC GTG GTT	GTN	
R99	N	AAC AAT	AAY	
R100	H	CAC CAT	CAY	EVITAR ATTTA
R101	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R102	G	GGA GGC GGG GGT	GGN	SI R102 GGC NO R103 AAT
R103	N	AAC AAT	AAY	EVITAR ATTTA
R104	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R105	V	GTA GTC GTG GTT	GTN	
R106	I	ATA ATC ATT	ATH	
R107	T	ACA ACC ACG ACT	ACN	
R108	W	TGG	TGG	
R109	G	GGA GGC GGG GGT	GGN	
R110	A	GCA GCC GCG GCT	GCN	
R111	Q	CAA CAG	CAR	
R112	T	ACA ACC ACG ACT	ACN	EVITAR ATTTA
R113	F	TTC TTT	TTY	
R114	K	AAA AAG	AAR	SI R114 AAG NO R115 CAT
R115	H	CAC CAT	CAY	
R116	Q	CAA CAG	CAR	SI R116 CAA NO R117 GCA

ES 2 368 962 T3

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R117	A	GCA GCC GCG GCT	GCN	EVITAR ATTTA
R118	F	TTC TTT	TTY	
R119	N	AAC AAT	AAY	NO AAT
R120	K	AAA AAG	AAR	SI R120 AAA NO R121 TTG
R121	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R122	A	GCA GCC GCG GCT	GCN	SI R122 GCC NO R123 AAT
R123	N	AAC AAT	AAY	EVITAR ATTTA
R124	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R125	F	TTC TTT	TTY	
R126	I	ATA ATC ATT	ATH	
R127	V	GTA GTC GTG GTT	GTN	
R128	N	AAC AAT	AAY	SI R128 AAT NO R129 AAT
R129	N	AAC AAT	AAY	NO AAT
R130	K	AAA AAG	AAR	
R131	K	AAA AAG	AAR	
R132	T	ACA ACC ACG ACT	ACN	
R133	I	ATA ATC ATT	ATH	
R134	P	CCA CCC CCG CCT	CCN	SI R134 CCC NO R135 AAT
R135	N	AAC AAT	AAY	SI R135 AAT NO R136 AAT
R136	N	AAC AAT	AAY	EVITAR ATTTA
R137	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R138	V	GTA GTC GTG GTT	GTN	
R139	E	GAA GAG	GAR	
R140	N	AAC AAT	AAY	
R141	Y	TAC TAT	TAY	EVITAR ATTTA
R142	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R143	T	ACA ACC ACG ACT	ACN	
R144	P	CCA CCC CCG CCT	CCN	NO CCA
R145	M	ATG	ATG	
R146	S	AGC AGT TCA TCC TCG TCT	AGY TCN	o SI R146 TCA NO R147 TTG

ES 2 368 962 T3

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R147	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R148	A	GCA GCC GCG GCT	GCN	
R149	Y	TAC TAT	TAY	NO TAT
R150	W	TGG	TGG	
R151	F	TTC TTT	TTY	
R152	M	ATG	ATG	
R153	D	GAC GAT	GAY	
R154	D	GAC GAT	GAY	
R155	G	GGA GGC GGG GGT	GGN	
R156	G	GGA GGC GGG GGT	GGN	
R157	K	AAA AAG	AAR	
R158	W	TGG	TGG	
R159	D	GAC GAT	GAY	
R160	Y	TAC TAT	TAY	
R161	N	AAC AAT	AAY	NO AAT
R162	K	AAA AAG	AAR	SI R162 AAA NO (R163 AAT Y R164 AGY)
R163	N	AAC AAT	AAY	
R164	S	AGC AGT TCA TCC TCG TCT	AGY TCN	o SI R164 TCA NO (R165 ACC Y R166 AAY)
R165	T	ACA ACC ACG ACT	ACN	SI R165 ACC NO R166 AAT
R166	N	AAC AAT	AAY	NO AAT
R167	K	AAA AAG	AAR	SI R167 AAA R168 NO TCA O R168 NO AGC
R168	S	AGC AGT TCA TCC TCG TCT	AGY TCN	o
R169	I	ATA ATC ATT	ATH	
R170	V	GTA GTC GTG GTT	GTN	SI R170 GTA NO R171 TTA
R171	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R172	N	AAC AAT	AAY	SI R172 AAT NO R173 ACA
R173	T	ACA ACC ACG ACT	ACN	SI R173 (ACC O ACG) NO (R174 CAA Y R175 TCN)
R174	Q	CAA CAG	CAR	

ES 2 368 962 T3

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R175	S	AGC AGT TCA TCC TCG TCT	AGY TCN	o EVITAR ATTTA
R176	F	TTC TTT	TTY	
R177	T	ACA ACC ACG ACT	ACN	
R178	F	TTC TTT	TTY	
R179	E	GAA GAG	GAR	
R180	E	GAA GAG	GAR	
R181	V	GTA GTC GTG GTT	GTN	
R182	E	GAA GAG	GAR	SI R182 GAA NO (R183 TAT Y R184 TTR)
R183	Y	TAC TAT	TAY	EVITAR ATTTA
R184	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R185	V	GTA GTC GTG GTT	GTN	
R186	K	AAA AAG	AAR	
R187	G	GGA GGC GGG GGT	GGN	SI R187 GGA NO (R188 TTG Y R189 CGN)
R188	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R189	R	AGA AGG CGA CGC CGG CGT	AGR CGN	o SI R189 CGC NO R190 AAT
R190	N	AAC AAT	AAY	NO AAT
R191	K	AAA AAG	AAR	
R192	F	TTC TTT	TTY	SI R192 TTC NO (R193 CAA Y R194 TTR)
R193	Q	CAA CAG	CAR	SI R193 CAA NO R194 TTA
R194	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R195	N	AAC AAT	AAY	SI R195 AAT NO R196 TGC
R196	C	TGC TGT	TGY	
R197	Y	TAC TAT	TAY	
R198	V	GTA GTC GTG GTT	GTN	
R199	K	AAA AAG	AAR	NO AAA
R200	I	ATA ATC ATT	ATH	SI R200 ATT NO R201 AAT NO ATA
R201	N	AAC AAT	AAY	NO AAT

ES 2 368 962 T3

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R202	K	AAA AAG	AAR	SI R202 AAA NO R203 AAT
R203	N	AAC AAT	AAY	NO AAT
R204	K	AAA AAG	AAR	
R205	P	CCA CCC CCG CCT	CCN	NO CCA SI R205 CCG NO R206 ATA
R206	I	ATA ATC ATT	ATH	SI R206 ATA NO R207 ATA
R207	I	ATA ATC ATT	ATH	SI R207 ATA NO R208 TAC NO ATT
R208	Y	TAC TAT	TAY	
R209	I	ATA ATC ATT	ATH	
R210	D	GAC GAT	GAY	
R211	S	AGC AGT TCA TCC TCG TCT	AGY TCN	o
R212	M	ATG	ATG	
R213	S	AGC AGT TCA TCC TCG TCT	AGY TCN	o
R214	Y	TAC TAT	TAY	EVITAR ATTTA
R215	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o SI R215 (TTA O CTA) NO R216 ATA
R216	I	ATA ATC ATT	ATH	
R217	F	TTC TTT	TTY	
R218	Y	TAC TAT	TAY	
R219	N	AAC AAT	AAY	EVITAR ATTTA
R220	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o SI R220 (TTA O CTA) NO R221 ATT SI R220 (TTA O CTA) NO R221 ATC
R221	I	ATA ATC ATT	ATH	SI R221 ATT NO R222 AAA NO ATA
R222	K	AAA AAG	AAR	
R223	P	CCA CCC CCG CCT	CCN	
R224	Y	TAC TAT	TAY	EVITAR ATTTA
R225	L	TTA TTG CTA CTC CTG CTT	TTR CTN	o
R226	I	ATA ATC ATT	ATH	
R227	P	CCA CCC CCG CCT	CCN	
R228	Q	CAA CAG	CAR	

Trinucleótido	AA	Trinucleótidos posibles	Código UIPAC	CONDICIÓN
R229	M	ATG	ATG	
R230	M	ATG	ATG	
R231	Y	TAC TAT	TAY	SI R231TAT NO R232 AAA
R232	K	AAA AAG	AAR	SI R232 AAA NO R233 TTG
R233	L	TTA TTG CTA CTC CTG CTT	TTR CTN	
R234	P	CCA CCC CCG CCT	CCN	SI 234 CCC NO R235 AAT
R235	N	AAC AAT	AAY	SI R235 AAT NO R236 ACA SI R235 AAT NO R236 ACT
R236	T	ACA ACC ACG ACT	ACN	SI R236 ACA NO (R237 ATA Y R238 AGY)
R237	I	ATA ATC ATT	ATH	
R238	S	AGC AGT TCA TCC TCG TCT	AGY TCN	
R239	S	AGC AGT TCA TCC TCG TCT	AGY TCN	
R240	E	GAA GAG	GAR	
R241	T	ACA ACC ACG ACT	ACN	
R242	F	TTC TTT	TTY	
R243	L	TTA TTG CTA CTC CTG CTT	TTR CTN	
R244	K	AAA AAG	AAR	

Tabla 2. (correspondiente a SEC ID N° 3)

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-Scel ejemplificada (SEC ID N° 4)
R1	M	ATG	ATG		ATG
R2	A	GCC GCT	GCY		GCC
R3	K	AAA AAG	AAR		AAG
R4	P	CCA CCC CCT	CCH		CCT
R5	P	CCA CCC CCT	CCH		CCC
R6	K	AAA AAG	AAR		AAG
R7	K	AAA AAG	AAR		AAG
R8	K	AAA AAG	AAR		AAG

ES 2 368 962 T3

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-Scel ejemplificada (SEC ID N° 4)
R9	R	AGA CGC CGG	AGA CGS	o	CGC
R10	K	AAA	AAA		AAA
R11	V	GTC GTG	GTS		GTG
R12	N	AAC	AAC		AAC
R13	I	ATC ATT	ATY		ATC
R14	K	AAA AAG	AAR		AAG
R15	K	AAA AAG	AAR		AAG
R16	N	AAC	AAC		AAC
R17	Q	CAG	CAG		CAG
R18	V	GTC GTG	GTS		GTG
R19	M	ATG	ATG		ATG
R20	N	AAC	AAC		AAC
R21	L	CTC CTG	CTS		CTG
R22	G	GGC GGA	GGM		GGA
R23	P	CCA CCC CCT	CCH		CCT
R24	N	AAC	AAC		AAC
R25	S	AGC TCA TCC	AGC TCM	o	AGC
R26	K	AAA AAG	AAR		AAG
R27	L	CTC CTG	CTS		CTC
R28	L	CTC CTG	CTS		CTG
R29	K	AAA AAG	AAR		AAG
R30	E	GAG	GAG		GAG
R31	Y	TAC	TAC		TAC
R32	K	AAA AAG	AAR		AAG
R33	S	AGC TCA TCC	AGC TCM	o	AGC
R34	Q	CAA CAG	CAR		CAG
R35	L	CTC CTG	CTS		CTG
R36	I	ATC ATT	ATY		ATC
R37	E	GAA GAG	GAR		GAA
R38	L	CTC CTG	CTS		CTG
R39	N	AAC	AAC		AAC

ES 2 368 962 T3

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-SceI ejemplificada (SEC ID Nº 4)
R40	I	ATC ATT	ATY		ATC
R41	E	GAA GAG	GAR	SI R41 GAG NO R42 CAA	GAG
R42	Q	CAA CAG	CAR		CAG
R43	F	TTC	TTC		TTC
R44	E	GAA GAG	GAR		GAA
R45	A	GCC GCT	GCY		GCT
R46	G	GGC GGA	GGM		GGC
R47	I	ATC	ATC		ATC
R48	G	GGC GGA	GGM		GGC
R49	L	CTC CTG	CTS		CTG
R50	I	ATC ATT	ATY		ATC
R51	L	CTC CTG	CTS		CTG
R52	G	GGC GGA	GGM		GGC
R53	D	GAC GAT	GAY		GAT
R54	A	GCC GCT	GCY		GCC
R55	Y	TAC	TAC		TAC
R56	I	ATC ATT	ATY		ATC
R57	R	AGA CGC CGG	AGA CGS	o	AGA
R58	S	AGC TCA TCC	AGC TCM	o	TCC
R59	R	AGA CGC CGG	AGA CGS	o	CGG
R60	D	GAC GAT	GAY		GAC
R61	E	GAA GAG	GAR		GAA
R62	G	GGC GGA	GGM		GGC
R63	K	AAA AAG	AAR		AAG
R64	T	ACC ACT	ACY		ACC
R65	Y	TAC	TAC		TAC
R66	C	TGC TGT	TGY		TGC
R67	M	ATG	ATG		ATG
R68	Q	CAG	CAG		CAG
R69	F	TTC	TTC		TTC
R70	E	GAA GAG	GAR		GAG

ES 2 368 962 T3

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-SceI ejemplificada (SEC ID N° 4)
R71	W	TGG	TGG		TGG
R72	K	AAA AAG	AAR		AAG
R73	N	AAC	AAC		AAC
R74	K	AAA AAG	AAR		AAG
R75	A	GCC GCT	GCY		GCC
R76	Y	TAC	TAC		TAC
R77	M	ATG	ATG		ATG
R78	D	GAC GAT	GAY		GAC
R79	H	CAC CAT	CAY		CAC
R80	V	GTC GTG	GTS		GTG
R81	C	TGC TGT	TGY		TGT
R82	L	CTC CTG	CTS		CTG
R83	L	CTC CTG	CTS		CTG
R84	Y	TAC	TAC		TAC
R85	D	GAC GAT	GAY	SI R85 GAC NO R86 CAA	GAC
R86	Q	CAA CAG	CAR		CAG
R87	W	TGG	TGG		TGG
R88	V	GTC GTG	GTS		GTC
R89	L	CTC CTG	CTS		CTG
R90	S	AGC TCA TCC	AGC TCM	o	AGC
R91	P	CCA CCC CCT	CCH		CCT
R92	P	CCA CCC CCT	CCH		CCT
R93	H	CAC CAT	CAY	SI R93 CAT NO R94 AAA	CAC
R94	K	AAA AAG	AAR		AAG
R95	K	AAA AAG	AAR		AAG
R96	E	GAA GAG	GAR		GAG
R97	R	AGA CGC CGG	AGA CGS	o	CGC
R98	V	GTC GTG	GTS		GTG
R99	N	AAC	AAC		AAC
R100	H	CAC CAT	CAY		CAT
R101	L	CTC CTG	CTS		CTG

ES 2 368 962 T3

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-SceI ejemplificada (SEC ID N° 4)
R102	G	GGC GGA	GGM		GGC
R103	N	AAC	AAC		AAC
R104	L	CTC CTG	CTS		CTC
R105	V	GTC GTG	GTS		GTG
R106	I	ATC ATT	ATY		ATC
R107	T	ACC ACT	ACY		ACC
R108	W	TGG	TGG		TGG
R109	G	GGC GGA	GGM		GGA
R110	A	GCC GCT	GCY		GCC
R111	Q	CAA CAG	CAR		CAG
R112	T	ACC ACT	ACY		ACC
R113	F	TTC	TTC		TTC
R114	K	AAA AAG	AAR	SI R114 AAG NO R115 CAT	AAG
R115	H	CAC CAT	CAY		CAC
R116	Q	CAA CAG	CAR		CAG
R117	A	GCC GCT	GCY		GCC
R118	F	TTC	TTC		TTC
R119	N	AAC	AAC		AAC
R120	K	AAA AAG	AAR		AAG
R121	L	CTC CTG	CTS		CTG
R122	A	GCC GCT	GCS		GCC
R123	N	AAC	AAC		AAC
R124	L	CTC CTG	CTS		CTG
R125	F	TTC	TTC		TTC
R126	I	ATC ATT	ATY		ATC
R127	V	GTC GTG	CTS		GTG
R128	N	AAC	AAC		AAC
R129	N	AAC	AAC		AAC
R130	K	AAA AAG	AAR		AAG
R131	K	AAA AAG	AAR		AAG
R132	T	ACC ACT	ACY		ACC
R133	I	ATC ATT	ATY		ATC

ES 2 368 962 T3

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-SceI ejemplificada (SEC ID N° 4)
R134	P	CCA CCC CCT	CCH		CCC
R135	N	AAC	AAC		AAC
R136	N	AAC	AAC		AAC
R137	L	CTC CTG	CTS		CTC
R138	V	GTC GTG	GTS		GTG
R139	E	GAA GAG	GAR		GAG
R140	N	AAC	AAC		AAC
R141	Y	TAC	TAC		TAC
R142	L	CTC CTG	CTS		CTC
R143	T	ACC ACT	ACY		ACT
R144	P	CCC CCT	CCY		CCC
R145	M	ATG	ATG		ATG
R146	S	AGC TCA TCC	AGC TCM	o	AGC
R147	L	CTC CTG	CTS		CTG
R148	A	GCC GCT	GCY		GCC
R149	Y	TAC	TAC		TAC
R150	W	TGG	TGG		TGG
R151	F	TTC	TTC		TTC
R152	M	ATG	ATG		ATG
R153	D	GAC GAT	GAY		GAC
R154	D	GAC GAT	GAY		GAC
R155	G	GGC GGA	GGM		GGA
R156	G	GGC GGA	GGM		GGC
R157	K	AAA AAG	AAR		AAG
R158	W	TGG	TGG		TGG
R159	D	GAC GAT	GAY		GAC
R160	Y	TAC	TAC		TAC
R161	N	AAC	AAC		AAC
R162	K	AAA AAG	AAR		AAG
R163	N	AAC	AAC		AAC
R164	S	AGC TCA TCC	AGC TCM	o SI R164 TCA NO R165 ACC	AGC

ES 2 368 962 T3

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-SceI ejemplificada (SEC ID N° 4)
R165	T	ACC ACT	ACY		ACC
R166	N	AAC	AAC		AAC
R167	K	AAA AAG	AAR	SI R167 AAA R168 NO TCA o R168 NO AGC	AAG
R168	S	AGC TCA TCC	AGC TCM o		TCA
R169	I	ATC ATT	ATY		ATT
R170	V	GTC GTG	GTS		GTG
R171	L	CTC CTG	CTS		CTG
R172	N	AAC	AAC		AAC
R173	T	ACC ACT	ACY	SI R173 ACC NO (R174 CAA Y R175 TCN)	ACC
R174	Q	CAA CAG	CAR		CAA
R175	S	AGC TCA TCC	AGC TCM o		AGC
R176	F	TTC	TTC		TTC
R177	T	ACC ACT	ACY		ACC
R178	F	TTC	TTC		TTC
R179	E	GAA GAG	GAR		GAA
R180	E	GAA GAG	GAR		GAA
R181	V	GTC GTG	GTS		GTG
R182	E	GAA GAG	GAR		GAG
R183	Y	TAC	TAC		TAC
R184	L	CTC CTG	CTS		CTC
R185	V	GTC GTG	GTS		GTC
R186	K	AAA AAG	AAR		AAG
R187	G	GGC GGA	GGM		GGC
R188	L	CTC CTG	CTS		CTG
R189	R	AGA CGC CGG	AGA CGS o		CGC
R190	N	AAC	AAC		AAC
R191	K	AAA AAG	AAR		AAG
R192	F	TTC	TTC		TTC
R193	Q	CAA CAG	CAR		CAG

ES 2 368 962 T3

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-Scel ejemplificada (SEC ID N° 4)
R194	L	CTC CTG	CTS		CTG
R195	N	AAC	AAC		AAC
R196	C	TGC TGT	TGY		TGC
R197	Y	TAC	TAC		TAC
R198	V	GTC GTG	GTS		GTG
R199	K	AAG	AAG		AAG
R200	I	ATC ATT	ATY		ATC
R201	N	AAC	AAC		AAC
R202	K	AAA AAG	AAR		AAG
R203	N	AAC	AAC		AAC
R204	K	AAA AAG	AAR		AAG
R205	P	CCC CCT	CCY		CCT
R206	I	ATC ATT	ATY		ATC
R207	I	ATC	ATC		ATC
R208	Y	TAC	TAC		TAC
R209	I	ATC ATT	ATY		ATC
R210	D	GAC GAT	GAY		GAC
R211	S	AGC TCA TCC	AGC TCM	o	AGC
R212	M	ATG	ATG		ATG
R213	S	AGC TCA TCC	AGC TCM	o	AGC
R214	Y	TAC	TAC		TAC
R215	L	CTC CTG	CTS		CTG
R216	I	ATC ATT	ATY		ATC
R217	F	TTC	TTC		TTC
R218	Y	TAC	TAC		TAC
R219	N	AAC	AAC		AAC
R220	L	CTC CTG	CTS		CTG
R221	I	ATC ATT	ATY	SI R221 ATT NO R222 AAA	ATC
R222	K	AAA AAG	AAR		AAG
R223	P	CCA CCC CCT	CCH		CCA
R224	Y	TAC	TAC		TAC

Trinucleótido	AA	Opciones	UIPAC	CONDICIÓN	I-SceI ejemplificada (SEC ID N° 4)
R225	L	CTC CTG	CTS		CTG
R226	I	ATC ATT	ATY		ATC
R227	P	CCA CCC CCT	CCH		CCT
R228	Q	CAA CAG	CAR		CAG
R229	M	ATG	ATG		ATG
R230	M	ATG	ATG		ATG
R231	Y	TAC	TAC		TAC
R232	K	AAA AAG	AAR		AAG
R233	L	CTC CTG	CTS		CTG
R234	P	CCA CCC CCT	CCH		CCC
R235	N	AAC	AAC		AAC
R236	T	ACC ACT	ACY		ACC
R237	I	ATC ATT	ATY		ATC
R238	S	AGC TCA TCC	AGC TCM	o	AGC
R239	S	AGC TCA TCC	AGC TCM	o	AGC
R240	E	GAA GAG	GAR		GAG
R241	T	ACC ACT	ACY		ACC
R242	F	TTC	TTC		TTC
R243	L	CTC CTG	CTS		CTG
R244	K	AAA AAG	AAR		AAG

Ejemplos

Ejemplo 1: Diseño, síntesis y análisis de un gen quimérico expresable en plantas que codifica I-SceI.

5 La región codificante de I-SceI en la cual se han reemplazado los 4 aminoácidos amino terminales por una señal de localización nuclear se optimizó utilizando el proceso siguiente:

1. Cambiar los codones al uso de codones más preferido para maíz sin alterar la secuencia de aminoácidos de la proteína I-SceI, utilizando el equipo Synergy Geneoptimizer™;
2. Ajustar la secuencia para crear o eliminar sitios de restricción específicos para cambiar la región codificante de I-SceI sintética con el gen de I-SceI de código universal;
- 10 3. Eliminar todos los tramos GC más largos que 6 pb y los tramos AT más largos que 4 pb, para evitar la formación de estructuras de RNA secundarias que pueden efectuar el corte y empalme de pre-mRNA;
4. Evitar dobletes CG y TA en las posiciones de los codones 2 y 3;
- 15 5. Evitar otros elementos reguladores, tales como posibles señales prematuras de poliadenilación (GATAAT, TATAAA, AATATA, AATATT, GATAAA, AATGAA, AATAAG, AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATAAAA, AATTAA, AATACA y CATAAA),

sitios de corte y empalme de intrones crípticos (AAGGTAAGT y TGCAGG), pentámeros ATTTA y secuencias de la caja CCAAT (CCAAT, ATTGG, CGAAT y ATTGC);

6. Comprobar de nuevo si la región codificante adaptada cumple la totalidad de los criterios arriba mencionados.

5 Un posible ejemplo de una secuencia nucleotídica de este tipo se representa en SEQ ID No 4. Un fragmento de DNA sintético que tenía la secuencia nucleotídica de SEQ ID No 4 se sintetizó y se enlazó operativamente a un promotor CaMV35S y una señal de terminación y poliadenilación en 3' de CaMV35S (produciendo el plásmido pCV78; SEQ ID No 7).

10 La región codificante de I-SceI sintética se clonó también en un vector de expresión bacteriano (como una proteína de fusión que permitía el enriquecimiento de la proteína en perlas de amilosa). Se comprobó la capacidad de la proteína I-SceI semi-purificada para escindir in vitro un plásmido que contuviera un sitio de reconocimiento de I-SceI.

Ejemplo 2. Aislamiento de líneas de células de maíz que contienen un gen bar sin promotor precedido por un sitio I-SceI.

15 Con objeto de desarrollar un ensayo para la recombinación mediada por homología inducida por la rotura de DNA bicatenario, se aislaron suspensiones de células de maíz que contenían un gen bar sin promotor precedido por un sitio de reconocimiento de I-SceI integrado en el genoma nuclear en una sola copia. Con la inducción de la rotura del DNA bicatenario por suministro de un gen quimérico expresable en plantas que codifica la endonucleasa I-SceI, y el suministro simultáneo de DNA de reparación que comprende un promotor 35S de CaMV enlazado operativamente al extremo 5' del gen bar, se puede insertar el promotor 35S por inserción del DNA direccionada mediada por homología, dando como resultado un gen bar funcional que confiere resistencia a fosfotricina (PPT). El ensayo se representa esquemáticamente en la Figura 1.

20 Se construyó el locus diana enlazando operativamente mediante técnicas convencionales de clonación las regiones de DNA siguientes:

- a) una señal de terminación y poliadenilación en el extremo 3' del gen de nopalina-sintetasa
- b) una región de DNA codificante de bar sin promotor
- 25 c) una región de DNA que comprendía un sitio de reconocimiento de I-SceI
- d) una señal de terminación y poliadenilación en el extremo 3' del gen 7 de *A. tumefaciens* (3'g7)
- e) un gen de resistencia a neomicina expresable en plantas que comprende un promotor de nopalina-sintetasa, un gen de neomicina-fosfotransferasa, y una señal 3' ocs.

30 Esta región de DNA se insertó en un vector de T-DNA entre los límites del T-DNA. El vector de T-DNA se designó pTTAM78 (para la secuencia nucleotídica del T-DNA véase SEQ ID No 5).

35 El vector de T-DNA se utilizó directamente para transformar protoplastos de maíz de acuerdo con los métodos descritos en EP 0469273, utilizando una suspensión de células de maíz derivada de He89. El vector T-DNA se introdujo también en *Agrobacterium tumefaciens* C58C1Rif(pEHA101) y la cepa de *Agrobacterium* resultante se utilizó para transformar una línea de células derivada de He89. Se identificaron varias líneas diana que contenían una sola copia del constructo del locus diana pTTAM78, tales como T24 (obtenida por transformación de protoplastos) y las líneas 14-1 y 1-20 (obtenidas por transformación mediada por *Agrobacterium*).

40 Se establecieron suspensiones de células a partir de estas líneas diana en un medio de suspensión de células N6M, y se hicieron crecer en presencia de luz en un agitador (120 rpm) a 25°C. Las suspensiones se subcultivaron cada semana.

Ejemplo 3: Inserción direccionada basada en homología

El DNA de reparación pTTA82 es un vector de T-DNA que contiene entre los límites del T-DNA las siguientes regiones de DNA enlazadas operativamente:

- a) una región de DNA que codifica solamente la parte amino terminal del gen bar
- 45 b) una región de DNA que comprende un sitio de reconocimiento parcial de I-SceI (13 nucleótidos localizados en el extremo -5' del sitio de reconocimiento)
- c) una región promotora 35S de CaMV
- d) una región de DNA que comprende un sitio de reconocimiento parcial de I-SceI (9 nucleótidos localizados en el extremo 3' del sitio de reconocimiento)

- e) una señal de terminación y de poliadenilación en el extremo 3' del gen 7 de *A. tumefaciens* (3'g7)
- f) un gen quimérico de resistencia a neomicina expresable en plantas
- g) un gen que codifica una endonucleasa I-SceI defectuosa, bajo control de un promotor 35S de CaMV

La secuencia nucleotídica del T-DNA de pTTA82 se representa en SEQ ID No 6.

- 5 Este DNA de reparación se suministró juntamente con pCV78 (véase el Ejemplo 1) por bombardeo con partículas en células derivadas de suspensión que se extendieron en placas sobre papel de filtro como una capa delgada. El papel de filtro se extendió sobre sustrato Mahq1 VII.

10 El DNA se bombardeó a las células utilizando un dispositivo PDS-1000/He Biolistics. La preparación de los microvehículos y el recubrimiento del DNA sobre los microvehículos se realizó esencialmente como ha sido descrito por Sanford *et al.* 1992. Los parámetros del bombardeo de partículas fueron: distancia a la diana 9 cm; presión de bombardeo de 1350 psi (94,8 kg/cm²), distancia de salto de ¼" (6,35 mm) y distancia de vuelo del macrovehículo 11 cm. Inmediatamente después del bombardeo, el tejido se transfirió al sustrato no selectivo Mh1VII. Como control para el suministro con éxito del DNA por bombardeo con partículas, las tres líneas diana se bombardearon también con microvehículos recubiertos con DNA plasmídico que comprendía un gen bar quimérico bajo el control de un
15 promotor CaMV35S (pRVA52).

Cuatro días después del bombardeo, los filtros se transfirieron a un sustrato Mh1 VII complementado con 25 mg/l de PPT o sobre sustrato Ahx1.5VIIino1000 complementado con 50 mg/l de PPT.

Catorce días después, los filtros se transfirieron sobre medio Mh1 VII nuevo con 10 mg/l de PPT para las líneas diana T24 y 14-1, y sustrato Mh1 VII con 25 mg/l de PPT para la línea diana 1-20.

- 20 Dos semanas después, se registraron los sucesos potenciales de inserción direccionada basándose en su resistencia a PPT. Estos sucesos resistentes a PPT fueron también positivos en el test Liberty Link de Hojas/Semillas de Maíz (Strategic Diagnostics Inc.).

Número de callos resistentes a PPT 38 días después del bombardeo:

Línea diana	pRVA52		pTTA82+pCV78	
	Número total de sucesos PPT ^R	Número medio de sucesos PPT ^R /cápsula Petri	Número total de sucesos PPT ^R	Número medio de sucesos PPT ^R /cápsula Petri
1-20	75	25	115	7,6
14-1	37	12,3	38	2,2
24	40	13,3	2	0,13

- 25 Los sucesos resistentes a PPT se subcultivaron ulteriormente en sustrato Mh1 VII que contenía 10 mg/l de PTT, y se utilizó material de callo para el análisis molecular. Se analizaron 20 candidatos independientes TSI mediante análisis Southern utilizando el promotor 35S y la señal de poliadenilación y de terminación en el extremo 3' del gen de nopalina-sintasa como sonda. Basándose en el tamaño del fragmento esperado, todos los sucesos parecían ser
30 sucesos de inserción de secuencia direccionada perfectos. Además, el análisis ulterior de aproximadamente la mitad de los sucesos de inserción de secuencia direccionada no mostró integración adicional no direccionada del DNA de reparación o el DNA codificante de I-SceI.

El análisis de secuencias de DNA amplificado a partir de ocho de los sucesos de inserción direccionada demostró que estos sucesos fueron de hecho sucesos TSI perfectos basados en recombinación homóloga.

- 35 Basándose en estos datos, la relación de inserción de DNA basada en recombinación homóloga frente a la recombinación "normal" ilegítima varía desde aproximadamente 30% para 1-20 hasta aproximadamente 17% para 14-1 y hasta aproximadamente 1% para 24.

40 Cuando se utilizan vectores similares a los descritos en Puchta *et al.* 1996 (más arriba) suministrados por electroporación a protoplastos de tabaco en presencia de roturas de DNA bicatenario inducidas por I-SceI, la relación de inserción de DNA basada en recombinación homóloga frente a inserción normal fue aproximadamente 15%. Sin embargo, solamente uno de 33 sucesos caracterizados fue un suceso de inserción de secuencia direccionada mediada por homología, con lo cual la recombinación homóloga fue perfecta a ambos lados de la

rotura bicatenaria.

5 Utilizando los vectores del Ejemplo 2, pero con un “constructo de I-SceI de código universal” que comprende una señal de localización nuclear, la relación de inserción de DNA basada en HR frente a inserción normal varió entre 0,032% y 16% para líneas diana diferentes, utilizando tanto electroporación como suministro de DNA mediado por *Agrobacterium*. La frecuencia relativa de sucesos de inserción direccionada perfectos difirió entre las diferentes líneas diana, y varió desde 8 hasta 70% para el suministro de DNA mediado por electroporación, y entre 73 y 90% para suministro de DNA mediado por *Agrobacterium*.

Ejemplo 4. La preincubación con acetosiringona mejora la frecuencia de recuperación de los sucesos de inserción direccionada.

10 Una semana antes del bombardeo como se describe en el Ejemplo 3, se diluyeron suspensiones de células en medio N6M o en medio LSIDhy1.5 complementado con acetosiringona 200 µM. En caso contrario, se utilizó el método como se describe en el Ejemplo 3. Como puede verse por los resultados resumidos en la tabla siguiente, la preincubación de las células a transformar con acetosiringona tuvo un efecto beneficioso sobre la recuperación de sucesos de inserción resistente a PPT direccionada.

Línea diana	Preincubación con acetosiringona		Sin preincubación	
	Número total de sucesos PPT ^R	Número medio de sucesos PPT ^R /cápsula Petri	Número total de sucesos PPT ^R	Número medio de sucesos PPT ^R /cápsula Petri
1-20	89	7,6	26	3,7
14-1	32	3,6	6	0,75
24	0	0	2	0,3

15 **Ejemplo 5: Inserción de la secuencia direccionada mediada por DSB en maíz por suministro de DNA de reparación mediado por *Agrobacterium***

20 Para realizar la inserción de secuencia direccionada mediada por DSB en el maíz, con lo cual el DNA de reparación es suministrado por transformación mediada por *Agrobacterium*, se construyeron vectores de T-DNA similares a pTTA82 (véase Ejemplo 3), en los cuales la I-SceI defectuosa se reemplazó por el gen sintético codificante de I-SceI del Ejemplo 1. El vector de T-DNA contenía adicionalmente una copia de *Agrobacterium tumefaciens virG* y *virC* (pTCV83) o *virG*, *virC* y *virB* (pTCV87) fuera de los límites del T-DNA. Estos vectores de T-DNA se insertaron en LBA4404, que contenía el plásmido Ti adyuvante pAL4404, produciendo las cepas de *Agrobacterium* A4995 y A4996, respectivamente.

25 Cultivos en suspensión de las líneas de células diana del Ejemplo 2, así como otras líneas de células diana obtenidas de manera similar a la descrita en el Ejemplo 2, se cultivaron conjuntamente con las cepas de *Agrobacterium*, y se extendieron después en varias placas. El número de extensiones en placa se determinó por la densidad de la suspensión de células. Como control para la eficiencia de transformación, se cultivaron conjuntamente la suspensión de células en un experimento paralelo con una cepa de *Agrobacterium* LBA4404 que contenía el plásmido Ti adyuvante pAL4404 y un vector de T-DNA con un gen quimérico de resistencia a fosfotricina (gen *bar*) bajo control de un vector 35S de CaMV. El vector de T-DNA contenía adicionalmente una copia de los genes de *Agrobacterium tumefaciens virG*, *virC* y *virD*, fuera del límite de T-DNA. Los resultados de cuatro experimentos independientes diferentes se resumen en las tablas siguientes:

30

Experimento I con *Agrobacterium*:

Línea diana	Control		A4495	
	No. de extensiones en placa	No. de transformantes	No. de extensiones en placa ⁽¹⁾	No. de sucesos de TSI
T24	26	10	32	0
T26	36	44	36	1
14-1	20	18	28	0
T1 F155	26	7	24	0

Experimento II con *Agrobacterium*:

Línea diana	Control		A4495	
	No. de extensiones en placa	No. de transformantes	No. de extensiones en placa ⁽¹⁾	No. de sucesos de TSI
1-20	18	~200	27	11
T79	24	~480	24	6
T66	26	73	31	0
T5	28	35	18	0
T1 F154	22	65	16	1

Experimento III con *Agrobacterium*:

Línea diana	Control		A4495	
	No. de extensiones en placa	No. de transformantes	No. de extensiones en placa ⁽¹⁾	No. de sucesos de TSI
T24	5	~2250	30	1
T26	44	~220	32	1
14-1	20	~1020	13	1
T1 F155	3	~1870	32	0

Experimento IV con Agrobacterium:

Línea diana	Control		A4495	
	No. de extensiones en placa	No. de transformantes	No. de extensiones en placa ⁽¹⁾	No. de sucesos de TSI
T1 F154			28	1
T5	12	~600	28	1
T66			28	0
T79			24	0
1-20	18	~400	40	9

5 Por consiguiente, está claro que, si bien el suministro de DNA de reparación mediado por Agrobacterium es claramente factible, la frecuencia de sucesos de Inserción de Secuencia Direccionada (TSI) es menor en comparación con el suministro de DNA de reparación mediado por bombardeo con partículas. El análisis Southern realizado sobre 23 sucesos de TSI supuestos demostraron que 20 sucesos de TSI son perfectos, basándose en el tamaño del fragmento. Sin embargo, en contraste con sucesos obtenidos por bombardeo con microproyectiles como en el Ejemplo 3, únicamente 6 sucesos de 20 no contenían insertos adicionales del DNA de reparación, 9 sucesos contenían 1 a 3 insertos adicionales del DNA de reparación, y 5 sucesos contenían muchos insertos adicionales del DNA de reparación.

10 El suministro mediado por bombardeo con partículas de DNA de reparación da también como resultado una mejor calidad de los sucesos de TSI mediados por DSB, comparado con el suministro del DNA de reparación por Agrobacterium. Esto contrasta con el suministro mediado por bombardeo con partículas de "DNA transformante normal", que se caracteriza por la menor calidad de los transformantes (patrón de integración complejo) en comparación con la transformación mediada por Agrobacterium.

15 Esto indica que la calidad de los transformantes obtenidos por bombardeo con partículas u otros métodos de suministro directo de DNA puede mejorarse por inserción de secuencias mediada por DSB. Este resultado se confirma también por el experimento siguiente: con la inserción de secuencia direccionada de un promotor 35S mediada por DSB, en ausencia de secuencias flanqueantes con homología para el locus diana en el DNA de reparación, se observó que, con el suministro de DNA de reparación mediado por electroporación, solamente una
20 minoría de los sucesos de TSI contenían inserciones adicionales no direccionadas de promotor 35S (2 sucesos de TSI de un total de 16 sucesos de TSI analizados muestran una o unas inserciones adicionales aleatorias del promotor 35S). En contraste, la inserción aleatoria del promotor 35S era considerablemente mayor en los sucesos de TSI obtenidos por suministro del promotor 35S mediado por Agrobacterium (17 de 22 sucesos de TSI analizados mostraban una o unas inserciones adicionales aleatorias del promotor 35S).

25 **Ejemplo 6: Composición de los medios**

Mahq1VII: Medio N6 (Chu *et al.* 1975) complementado con 100 mg/l de hidrolizado de caseína, L-prolina 6 mM, 0,5 g/l de ácido 2-(N-morfolino)etanosulfónico (MES), manitol 0,2 M, sorbitol 0,2 M, 2% de sacarosa, 1 mg/l de ácido 2,4-diclorofenoxi-acético (2,4-D), ajustado a pH 5,8, solidificado con 2,5 g/l de Gelrite®.

30 **Mhi1VII:** Medio N6 (Chu *et al.* 1975) complementado con 0,5 g/l de ácido 2-(N-morfolino)etanosulfónico (MES), manitol 0,2 M, 2% de sacarosa, 1 mg/l de ácido 2,4-diclorofenoxi-acético (2,4-D), ajustado a pH 5,8 y solidificado con 2,5 g/l de Gelrite®.

Mh1VII: Igual que el sustrato Mhi1VII, pero sin manitol 0,2 M.

35 **Ahx1.5VIIino1000:** Sales MS, complementadas con 1000 mg/l de mio-inositol, 0,1 mg/l de hidrocloreuro de tiamina, 0,5 mg/l de ácido nicotínico, 0,5 mg/l de hidrocloreuro de piridoxina, 0,5 g/l de MES, 30 g/l de sacarosa, 10 g/l de glucosa, 1,5 mg/l de 2,4-D, ajustado a pH 5,8 y solidificado con 2,5 g/l de Gelrite®.

LSIDhy1.5: Sales MS complementadas con 0,5 mg/l de ácido nicotínico, 0,5 mg/l de hidrocloreuro de piridoxina, 1 mg/l de hidrocloreuro de tiamina, 100 mg/l de mio-inositol, L-prolina 6 mM, 0,5 g/l de MES, 20 g/l de sacarosa, 10 g/l de glucosa, 1,5 mg/l de 2,4-D, ajustado a pH 5,2.

N6M: Macro-elementos: 2830 mg/l de KNO₃; 433 mg/l de (NH₄)₂SO₄; 166 mg/l de CaCl₂·2H₂O; 250 mg/l de

ES 2 368 962 T3

MgSO₄.7H₂O; 400 mg/l de KH₂PO₄; 37,3 mg/l de Na₂EDTA; 27,3 mg/l de FeSO₄.7H₂O, micro-elementos MS, 500 mg/l de Bactotripton, 0,5 g/l de MES, 1 mg/l de hidrocloreuro de tiamina, 0,05 mg/l de ácido nicotínico, 0,5 mg/l de hidrocloreuro de piridoxina, 2 mg/l de glicina, 100 mg/l de mio-inositol, 3% de sacarosa, 0,5 mg/l de 2,4-D, ajustado a pH 5,8.

5

LISTADO DE SECUENCIAS

<110> Bayer BioScience N.V.

D'Halluin, Kathleen

10 Vanderstraeten, Chantal

Ruiter, Rene

<120> Inserción de DNA direccionada mejorada en plantas

15 <130> BCS 03-2007 EPw2

<150> EP 03078700.6

<151> 2003-11-18

20 <160> 7

<170> PatentIn version 3.0

<210> 1

25 <211> 244

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 1

30 Met Ala Lys Pro Pro Lys Lys Lys Arg Lys Val Asn Ile Lys Lys Asn
1 5 10 15

Gln Val Met Asn Leu Gly Pro Asn Ser Lys Leu Leu Lys Glu Tyr Lys
20 25 30

35 Ser Gln Leu Ile Glu Leu Asn Ile Glu Gln Phe Glu Ala Gly Ile Gly
35 40 45

40 Leu Ile Leu Gly Asp Ala Tyr Ile Arg Ser Arg Asp Glu Gly Lys Thr
50 55 60

Tyr Cys Met Gln Phe Glu Trp Lys Asn Lys Ala Tyr Met Asp His Val

ES 2 368 962 T3

65		70		75		80										
	Cys	Leu	Leu	Tyr	Asp	Gln	Trp	Val	Leu	Ser	Pro	Pro	His	Lys	Lys	Glu
				85						90					95	
5	Arg	Val	Asn	His	Leu	Gly	Asn	Leu	Val	Ile	Thr	Trp	Gly	Ala	Gln	Thr
			100						105					110		
	Phe	Lys	His	Gln	Ala	Phe	Asn	Lys	Leu	Ala	Asn	Leu	Phe	Ile	Val	Asn
10			115					120					125			
	Asn	Lys	Lys	Thr	Ile	Pro	Asn	Asn	Leu	Val	Glu	Asn	Tyr	Leu	Thr	Pro
		130					135					140				
15	Met	Ser	Leu	Ala	Tyr	Trp	Phe	Met	Asp	Asp	Gly	Gly	Lys	Trp	Asp	Tyr
	145					150					155				160	
	Asn	Lys	Asn	Ser	Thr	Asn	Lys	Ser	Ile	Val	Leu	Asn	Thr	Gln	Ser	Phe
20				165						170					175	
	Thr	Phe	Glu	Glu	Val	Glu	Tyr	Leu	Val	Lys	Gly	Leu	Arg	Asn	Lys	Phe
			180						185					190		
	Gln	Leu	Asn	Cys	Tyr	Val	Lys	Ile	Asn	Lys	Asn	Lys	Pro	Ile	Ile	Tyr
25			195					200					205			
	Ile	Asp	Ser	Met	Ser	Tyr	Leu	Ile	Phe	Tyr	Asn	Leu	Ile	Lys	Pro	Tyr
		210					215					220				
30	Leu	Ile	Pro	Gln	Met	Met	Tyr	Lys	Leu	Pro	Asn	Thr	Ile	Ser	Ser	Glu
	225					230					235					240
	Thr	Phe	Leu	Lys												

35

<210> 2

<211> 732

<212> DNA

40 <213> Artificial

<220>

<223> Secuencia sintética de DNA codificante de I-SceI (código UIPAC)

45 <220>

<221> característica diversa

<222> (6)..(6)

<223> N= A, G, C o T

<220>

<221> variación

<222> (25)..(27)

5 <223> AGR

<220>

<221> variación

10 <222> (61)..(63)

<223> TTR

<220>

15 <221> variación

<222> (73)..(75)

<223> AGY

20 <220>

<221> variación

<222> (79)..(81)

<223> TTR

25

<220>

<221> variación

<222> (82)..(84)

<223> TTR

30

<220>

<221> variación

<222> (97)..(99)

35 <223> AGY

<220>

<221> variación

5 <222> (103)..(105)

<223> TTR

<220>

10 <221> variación

<222> (112)..(114)

<223> TTR

15 <220>

<221> variación

<222> (145)..(147)

<223> TTR

20

<220>

<221> variación

<222> (151)..(153)

<223> TTR

25

<220>

<221> variación

<222> (169)..(171)

30 <223> AGR

<220>

<221> variación

35 <222> (172)..(174)

<223> AGY

<220>

5 <221> variación

<222> (175)..(177)

<223> AGR

10 <220>

<221> variación

<222> (244)..(246)

<223> TTR

15

<220>

<221> variación

<222> (247)..(249)

<223> TTR

20

<220>

<221> variación

<222> (265)..(267)

25 <223> TTR

<220>

<221> variación

30 <222> (268)..(270)

<223> AGY

<220>

35 <221> variación

<222> (289)..(291)

<223> AGR

5 <220>

<221> variación

<222> (301)..(303)

<223> TTR

10

<220>

<221> variación

<222> (310)..(312)

<223> TTR

15

<220>

<221> variación

<222> (361)..(363)

20 <223> TTR

<220>

<221> variación

25 <222> (370)..(372)

<223> TTR

<220>

30 <221> variación

<222> (409)..(411)

<223> TTR

35 <220>

<221> variación
<222> (424)..(426)
<223> TTR

5

<220>
<221> variación
<222> (436)..(438)
<223> AGY

10

<220>
<221> variación
<222> (439)..(441)

15

<223> TTR

<220>
<221> variación
<222> (490)..(492)
<223> AGY

20

<220>
<221> variación
<222> (502)..(504)
<223> AGY

25

30

<220>
<221> variación
<222> (511)..(513)
<223> TTR

35

<220>

<221> variación

<222> (523)..(525)

<223> AGY

5

<220>

<221> variación

<222> (550)..(552)

10 <223> TTR

<220>

<221> variación

15 <222> (562)..(564)

<223> TTR

<220>

20 <221> variación

<222> (565)..(567)

<223> AGR

25 <220>

<221> variación

<222> (580)..(582)

<223> TTR

30

<220>

<221> variación

<222> (631)..(633)

<223> AGY

35

<220>

<221> variación

<222> (637)..(639)

5 <223> AGY

<220>

<221> variación

10 <222> (643)..(645)

<223> TTR

<220>

15 <221> variación

<222> (658)..(660)

<223> TTR

20 <220>

<221> variación

<222> (673)..(675)

<223> TTR

25

<220>

<221> variación

<222> (697)..(699)

<223> TTR

30

<220>

<221> variación

<222> (712)..(714)

35 <223> AGY

<220>

<221> variación

5 <222> (715)..(717)

<223> AGY

<220>

10 <221> variación

<222> (727)..(729)

<223> TTR

15 <220>

<221> característica diversa

<222> (12)..(12)

<223> N = A, G, C o T

20

<220>

<221> característica diversa

<222> (15)..(15)

<223> N = A, G, C o T

25

<220>

<221> característica diversa

<222> (27)..(27)

30 <223> N = A, G, C o T

<220>

<221> característica diversa

35 <222> (33)..(33)

<223> N = A, G, C o T

<220>

5 <221> característica diversa

<222> (54)..(54)

<223> N = A, G, C o T

10 <220>

<221> característica diversa

<222> (63)..(63)

<223> N = A, G, C o T

15

<220>

<221> característica diversa

<222> (66)..(66)

<223> N = A, G, C o T

20

<220>

<221> característica diversa

<222> (69)..(69)

25 <223> N = A, G, C o T

<220>

<221> característica diversa

30 <222> (75)..(75)

<223> N = A, G, C o T

<220>

35 <221> característica diversa

<222> (81)..(81)

<223> N = A, G, C o T

5 <220>

<221> característica diversa

<222> (84)..(84)

<223> N = A, G, C o T

10

<220>

<221> característica diversa

<222> (99)..(99)

<223> N = A, G, C o T

15

<220>

<221> característica diversa

<222> (105)..(105)

20 <223> N = A, G, C o T

<220>

<221> característica diversa

25 <222> (114)..(114)

<223> N = A, G, C o T

<220>

30 <221> característica diversa

<222> (135)..(135)

<223> N = A, G, C o T

35 <220>

<221> característica diversa

<222> (138)..(138)

<223> N = A, G, C o T

5

<220>

<221> característica diversa

<222> (144)..(144)

<223> N = A, G, C o T

10

<220>

<221> característica diversa

<222> (147)..(147)

15 <223> N = A, G, C o T

<220>

<221> característica diversa

20 <222> (153)..(153)

<223> N = A, G, C o T

<220>

25 <221> característica diversa

<222> (156)..(156)

<223> N = A, G, C o T

30 <220>

<221> característica diversa

<222> (162)..(162)

<223> N = A, G, C o T

35

<220>

<221> característica diversa

<222> (171)..(171)

<223> N = A, G, C o T

5

<220>

<221> característica diversa

<222> (174)..(174)

10 <223> N = A, G, C o T

<220>

<221> característica diversa

15 <222> (177)..(177)

<223> N = A, G, C o T

<220>

20 <221> característica diversa

<222> (186)..(186)

<223> N = A, G, C o T

25 <220>

<221> característica diversa

<222> (192)..(192)

<223> N = A, G, C o T

30

<220>

<221> característica diversa

<222> (225)..(225)

<223> N = A, G, C o T

35

<220>

<221> característica diversa

<222> (240)..(240)

5 <223> N = A, G, C o T

<220>

<221> característica diversa

10 <222> (246)..(246)

<223> N = A, G, C o T

<220>

15 <221> característica diversa

<222> (249)..(249)

<223> N = A, G, C o T

20 <220>

<221> característica diversa

<222> (264)..(264)

<223> N = A, G, C o T

25

<220>

<221> característica diversa

<222> (267)..(267)

<223> N = A, G, C o T

30

<220>

<221> característica diversa

<222> (270)..(270)

35 <223> N = A, G, C o T

<220>

<221> característica diversa

5 <222> (273)..(273)

<223> N = A, G, C o T

<220>

10 <221> característica diversa

<222> (276)..(276)

<223> N = A, G, C o T

15 <220>

<221> característica diversa

<222> (291)..(291)

<223> N = A, G, C o T

20

<220>

<221> característica diversa

<222> (294)..(294)

<223> N = A, G, C o T

25

<220>

<221> característica diversa

<222> (303)..(303)

30 <223> N = A, G, C o T

<220>

<221> característica diversa

35 <222> (306)..(306)

<223> N = A, G, C o T

<220>

5 <221> característica diversa

<222> (312)..(312)

<223> N = A, G, C o T

10 <220>

<221> característica diversa

<222> (315)..(315)

<223> N = A, G, C o T

15

<220>

<221> característica diversa

<222> (321)..(321)

<223> N = A, G, C o T

20

<220>

<221> característica diversa

<222> (327)..(327)

25 <223> N = A, G, C o T

<220>

<221> característica diversa

30 <222> (330)..(330)

<223> N = A, G, C o T

<220>

35 <221> característica diversa

<222> (336)..(336)

<223> N = A, G, C o T

5 <220>

<221> característica diversa

<222> (351)..(351)

<223> N = A, G, C o T

10

<220>

<221> característica diversa

<222> (363)..(363)

<223> N = A, G, C o T

15

<220>

<221> característica diversa

<222> (366)..(366)

20 <223> N = A, G, C o T

<220>

<221> característica diversa

25 <222> (372)..(372)

<223> N = A, G, C o T

<220>

30 <221> característica diversa

<222> (381)..(381)

<223> N = A, G, C o T

35 <220>

<221> característica diversa

<222> (396)..(396)

<223> N = A, G, C o T

5

<220>

<221> característica diversa

<222> (402)..(402)

<223> N = A, G, C o T

10

<220>

<221> característica diversa

<222> (411)..(411)

15 <223> N = A, G, C o T

<220>

<221> característica diversa

20 <222> (414)..(414)

<223> N = A, G, C o T

<220>

25 <221> característica diversa

<222> (426)..(426)

<223> N = A, G, C o T

30 <220>

<221> característica diversa

<222> (429)..(429)

<223> N = A, G, C o T

35

<220>

<221> característica diversa

<222> (432)..(432)

<223> N = A, G, C o T

5

<220>

<221> característica diversa

<222> (438)..(438)

10 <223> N = A, G, C o T

<220>

<221> característica diversa

15 <222> (441)..(441)

<223> N = A, G, C o T

<220>

20 <221> característica diversa

<222> (444)..(444)

<223> N = A, G, C o T

25 <220>

<221> característica diversa

<222> (465)..(465)

<223> N = A, G, C o T

30

<220>

<221> característica diversa

<222> (468)..(468)

<223> N = A, G, C o T

35

<220>

<221> característica diversa

<222> (492)..(492)

5 <223> N = A, G, C o T

<220>

<221> característica diversa

10 <222> (495)..(495)

<223> N = A, G, C o T

<220>

15 <221> característica diversa

<222> (504)..(504)

<223> N = A, G, C o T

20 <220>

<221> característica diversa

<222> (510)..(510)

<223> N = A, G, C o T

25

<220>

<221> característica diversa

<222> (513)..(513)

<223> N = A, G, C o T

30

<220>

<221> característica diversa

<222> (519)..(519)

35 <223> N = A, G, C o T

<220>

<221> característica diversa

5 <222> (525)..(525)

<223> N = A, G, C o T

<220>

10 <221> característica diversa

<222> (531)..(531)

<223> N = A, G, C o T

15 <220>

<221> característica diversa

<222> (543)..(543)

<223> N = A, G, C o T

20

<220>

<221> característica diversa

<222> (552)..(552)

<223> N = A, G, C o T

25

<220>

<221> característica diversa

<222> (552)..(552)

30 <223> N = A, G, C o T

<220>

<221> característica diversa

35 <222> (555)..(555)

<223> N = A, G, C o T

<220>

5 <221> característica diversa

<222> (561)..(561)

<223> N = A, G, C o T

10 <220>

<221> característica diversa

<222> (564)..(564)

<223> N = A, G, C o T

15

<220>

<221> característica diversa

<222> (567)..(567)

<223> N = A, G, C o T

20

<220>

<221> característica diversa

<222> (582)..(582)

25 <223> N = A, G, C o T

<220>

<221> característica diversa

30 <222> (594)..(594)

<223> N = A, G, C o T

<220>

35 <221> característica diversa

<222> (615)..(615)

<223> N = A, G, C o T

5 <220>

<221> característica diversa

<222> (633)..(633)

<223> N = A, G, C o T

10

<220>

<221> característica diversa

<222> (639)..(639)

<223> N = A, G, C o T

15

<220>

<221> característica diversa

<222> (645)..(645)

20 <223> N = A, G, C o T

<220>

<221> característica diversa

25 <222> (660)..(660)

<223> N = A, G, C o T

<220>

30 <221> característica diversa

<222> (669)..(669)

<223> N = A, G, C o T

35 <220>

<221> característica diversa

<222> (675)..(675)

<223> N = A, G, C o T

5

<220>

<221> característica diversa

<222> (681)..(681)

<223> N = A, G, C o T

10

<220>

<221> característica diversa

<222> (699)..(699)

15 <223> N = A, G, C o T

<220>

<221> característica diversa

20 <222> (702)..(702)

<223> N = A, G, C o T

<220>

25 <221> característica diversa

<222> (708)..(708)

<223> N = A, G, C o T

30 <220>

<221> característica diversa

<222> (714)..(714)

<223> N = A, G, C o T

35

ES 2 368 962 T3

<220>

<221> característica diversa

<222> (717)..(717)

<223> N = A, G, C o T

5

<220>

<221> característica diversa

<222> (723)..(723)

10 <223> N = A, G, C o T

<220>

<221> característica diversa

15 <222> (729)..(729)

<223> N = A, G, C o T

<400> 2

20	atggcnaarc cncnaaraa raarcgnaar gtnaayatha araaraayca rgtnatgaay	60
	ctnggncna aytcnaarct nctnaargar tayaartcnc arctnathga rctnaayath	120
	garcarttyg argcnggnat hggncnath ctnggngayg cntayathcg ntcncngay	180
25	garggnaara cntaytgyat gcarttygar tggaaraaya argcntayat ggaycaygtn	240
	tgycnctnt aygaycartg ggtncntcn ccncncaya araargarcg ngtnaaycay	300
30	ctnggnaayc tngtnathac ntgggngcn caracnttya arcaycargc nttyaayaar	360
	ctngcnaayc tnttyathgt naayaayaar aaracnathc cnaayaayct ngtngaraay	420
	tayctnacnc cnatgtcnct ngcntaytgg ttyatggayg ayggnggnaa rtgggaytay	480
35	aayaaraayt cnacnaayaa rtcnathgtn ctnaayacnc artcnttyac nttygargar	540
	gtngartayc tngtnaargg nctncgnaay aarttycarc tnaaytgyta ygtnaarath	600
40	aayaaraaya arccnathat htayathgay tcnatgtcnt ayctnathtt ytayaayctn	660
	athaarcnt ayctnathcc ncaratgatg tayaarctnc cnaayacnat htcntcngar	720
45	acnttyctna ar	732

<210> 3

<211> 732

<212> DNA

5 <213> Artificial

<220>

<223> Secuencia sintética de DNA preferida codificante de I-SceI (código UIPAC)

10 <220>

<221> variación

<222> (25)..(27)

<223> AGA

15

<220>

<221> variación

<222> (73)..(75)

<223> AGC

20

<220>

<221> variación

<222> (97)..(99)

25 <223> AGC

<220>

<221> variación

30 <222> (169)..(171)

<223> AGA

<220>

35 <221> variación

<222> (172)..(174)

<223> AGC

5 <220>

<221> variación

<222> (175)..(177)

<223> AGA

10

<220>

<221> variación

<222> (268)..(270)

<223> AGC

15

<220>

<221> variación

<222> (289)..(291)

20 <223> AGA

<220>

<221> variación

25 <222> (436)..(438)

<223> AGC

<220>

30 <221> variación

<222> (490)..(492)

<223> AGC

35 <220>

<221> variación
<222> (502)..(504)
<223> AGC

5

<220>
<221> variación
<222> (523)..(525)
<223> AGC

10

<220>
<221> variación
<222> (565)..(567)

15 <223> AGA

<220>
<221> variación
<222> (631)..(633)
<223> AGC

20

<220>
<221> variación
<222> (637)..(639)
<223> AGC

25

30

<220>
<221> variación
<222> (712)..(714)
<223> AGC

35

ES 2 368 962 T3

<220>

<221> variación

<222> (715)..(717)

<223> AGC

5

<400> 3

	atggcyaarc chcchaaraa raarcgsaaa gtsaacatya araaraacca ggtsatgaac	60
10	ctsggmcccha actcmaarct sctsaargag tacaartcmc arctsatyga rctsaacaty	120
	garcarttcg argcyggmat cggmctsaty ctsggmgayg cytacatycg stcmcgsgay	180
15	garggmaara cytactgyat gcagttcgar tggaraaaca argcytacat ggaycaygts	240
	tgycststst acgaycartg ggtsctstcm cchcchcaya araargarcg sgtsaaccay	300
	ctsggmaacc tsgtsatyac ytgggmgcy caracyttca arcaycargc yttcaacaar	360
20	ctsgcsaacc tsttcatyct saacaacaar aaracyatyc chaacaacct sgtsgaraac	420
	tacctsacyc cyatgtcmct sgcytactgg ttcattggayg ayggmgmaa rtgggaytac	480
25	aacaaraact cmacyaaca rctmatygts ctsaacacyc artcmttcac yttcgargar	540
	gtsgartacc tsgtsaargg mctscgsaac aarttccarc tsaactgyta cgtsaagaty	600
	aacaaraaca arccyatyat ctacatygay tcmatgtomt acctsatytt ctacaaccts	660
30	atyaarccht acctsatycc hcaratgatg tacaarctsc chaacacyat ytcmtcmgar	720
	acyttcctsa ar	732

35 <210> 4

<211> 732

<212> DNA

<213> Artificial

40 <220>

<223> Secuencia sintética de DNA preferida codificante de I-SceI (código UIPAC)

<400> 4

45	atggccaagc ctccaagaa gaagcgcaaa gtgaacatca agaagaacca ggtgatgaac	60
	ctgggacctt acagcaagct cctgaaggag tacaagagcc agctgatcga actgaacatc	120

ES 2 368 962 T3

	gagcagttcg aagctggcat cggcctgatc ctgggogatg cctacatcag atcccgggac	180
	gaaggcaaga cctactgcat gcagttcgag tggaagaaca aggcctacat ggaccacgtg	240
5	tgtctgctgt acgaccagtg ggtcctgagc cctcctcaca agaaggagcg cgtgaaccat	300
	ctgggcaacc tcgtgatcac ctggggagcc cagaccttca agcaccaggc cttcaacaag	360
10	ctggccaacc tgttcatcgt gaacaacaag aagaccatcc ccaacaacct cgtggagAAC	420
	tacctcactc ccatgagcct ggcctactgg ttcattggacg acggaggcaa gtgggactac	480
	aacaagaaca gcaccaaca gtcaattgtg ctgaacaccc aaagcttcac cttcgaagaa	540
15	gtggagtacc tcgtcaaggg cctgcgcaac aagttccagc tgaactgcta cgtgaagatc	600
	aacaagaaca agcctatcat ctacatcgac agcatgagct acctgatctt ctacaacctg	660
20	atcaagccat acctgatccc tcagatgatg tacaagctgc ccaacacccat cagcagcgag	720
	accttctga ag	732

<210> 5

25 <211> 3262

<212> DNA

<213> Artificial

<220>

30 <223> T-DNA de pTTAM78 (locus diana)

<220>

<221> característica diversa

<222> (1)..(25)

35 <223> Secuencia del borde derecho del T-DNA

<220>

<221> característica diversa

40 <222> (26)..(72)

<223> Secuencia de polienlazador sintética

<220>

<221> característica diversa

<222> (73)..(333)

<223> 3' nos (complemento)

5

<220>

<221> característica diversa

<222> (334)..(351)

10 <223> Secuencia de polienlazador sintética

<220>

<221> característica diversa

15 <222> (352)..(903)

<223> secuencia de bar (complemento)

<220>

20 <221> característica diversa

<222> (904)..(928)

<223> Secuencia de polienlazador sintética

25 <220>

<221> característica diversa

<222> (929)..(946)

<223> Sitio de reconocimiento de I-Scel

30

<220>

<221> característica diversa

<222> (947)..(967)

<223> Secuencia de polienlazador sintética

35

<220>

<221> característica diversa

<222> (968)..(1171)

5 <223> 3'g7

<220>

<221> característica diversa

10 <222> (1172)..(1290)

<223> Secuencia de polienlazador sintética

<220>

15 <221> característica diversa

<222> (1291)..(1577)

<223> promotor del gen de nopalina-sintetasa

20 <220>

<221> característica diversa

<222> (1578)..(1590)

<223> Secuencia de polienlazador sintética

25

<220>

<221> característica diversa

<222> (1591)..(2394)

<223> nptII

30

<220>

<221> característica diversa

<222> (2395)..(2567)

35 <223> 3' neo

ES 2 368 962 T3

<220>

<221> característica diversa

5 <222> (2568)..(3183)

<223> 3' ocs

<220>

10 <221> característica diversa

<222> (3184)..(3234)

<223> Secuencia de polienlazador sintética

15 <220>

<221> característica diversa

<222> (3235)..(3262)

<223> Secuencia del borde izquierdo del T-DNA

20

<400> 5

aattacaacg gtatatatcc tgccagtact cggccgtcga cctgcaggca attggtacct	60
agaggatctt cccgatctag taacatagat gacaccgcgc gcgataatth atcctagtht	120
25 gcgcgctata thttgtthttc tatcgcgtat thaatgtata attgcgggac tctaatacata	180
aaaaccatc tcataaataa cgtcatgcat tacatgthaa thattacatg cthaacgthaa	240
30 thcaacagaa attatatgat aatcatcgcga agaccggcaa caggattcaa ththaaagaaa	300
ctthattgcc aatgtthtga acgatctgct tcggatccta gacgcgtgag atcagatctc	360
ggtgacgggc aggaccggac ggggcggtac cggcaggctg aagthccagct gccagaaacc	420
35 cacgtcatgc cagthtcccgth gctthgaagcc ggccgcccgc agcatgcccgc ggggggcata	480
thccgagcgc tcgtgcatgc gcacgctcgg gthcgtthggc agcccgatga cagcgcaccac	540
40 gctctthgaag cctgtgctc thcagggacth cagcaggtgg gthgtagagcg thgagcccag	600
thccgthccgc thggtggcggg gggagacgth cacggtcgcac thcggccgthc agthcgtaggc	660
45 gthcgcgtgcc ththcaggggc cgcgctaggc gatgcgcggcg acctcgcctg thcacctcggc	720

ES 2 368 962 T3

gacgagccag ggatagcgct cccgcagacg gacgaggctg tccgtccact cctgcgggtc 780
 ctgcggctcg gtacggaagt tgaccgtgct tgtctcgatg tagtggttga cgatggtgca 840
 5 gaccgccggc atgtccgct cgggtggcacg gcggatgtcg gccgggcgtc gttctgggtc 900
 catggttata gagagagaga tagatttaat taccctgtta tccctaggcc gctgtacagg 960
 gcccgggatc ttgaaagaaa tatagtttaa atatttattg ataaaataac aagtcaggta 1020
 10 ttatagtcca agcaaaaaca taaatttatt gatgcaagtt taaattcaga aatatttcaa 1080
 taactgatta tatcagctgg tacattgccg tagatgaaag actgagtgcg atattatgtg 1140
 15 taatacataa attgatgata tagctagctt aggcgcgcca tagatcccgt caattctcac 1200
 tcattaggca cccagggctt tacactttat gcttccggct cgtataatgt gtggaattgt 1260
 gagcggataa caatttcaca caggaaacag gatcatgagc ggagaattaa gggagtcacg 1320
 20 ttatgacccc cgccgatgac gcggggacaag ccgttttacg tttggaactg acagaaccgc 1380
 aacgattgaa ggagccactc agccgcgggt ttctggagtt taatgagcta agcacatacg 1440
 25 tcagaaacca ttattgcbg ttcaaaagtc gcctaaggct actatcagct agcaaatatt 1500
 tcttgtcaaa aatgctccac tgacgttoca taaattcccc tcggtatcca attagagtct 1560
 catattcact ctcaatcaaa gatccggccc atgatcatgt ggattgaaca agatggattg 1620
 30 cacgcagggt ctccggccgc ttgggtggag aggcatttcg gctatgactg ggcacaacag 1680
 acaatcggct gctctgatgc cgccgtgttc cggctgtcag cgcagggcg cccggttctt 1740
 35 tttgtcaaga ccgacctgtc cggtgccctg aatgaactgc aggacgaggc agcgcggcta 1800
 tcgtggctgg ccacgacggg cgttccttgc gcagctgtgc tcgacgttgt cactgaagcg 1860
 ggaagggact ggctgctatt gggcgaagtg ccggggcagg atctcctgtc atctcacctt 1920
 40 gctcctgccg agaaagtatc catcatggct gatgcaatgc ggcggctgca tacgcttgat 1980
 ccggctacct gccattcga ccaccaagcg aaacatcgca tcgagcgagc acgtactcgg 2040
 45 atggaagccg gtcttgtcga tcaggatgat ctggacgaag agcatcaggg gctcgcgcca 2100
 gccgaactgt tcgccaggct caaggcgcgc atgcccagc gcgaggatct cgtcgtgacc 2160
 catggcgatg cctgcttgcc gaatatcatg gtggaaaatg gccgcttttc tggattcatc 2220
 50 gactgtggcc ggctgggtgt ggcggaccgc tatcaggaca tagcgttggc taccctgat 2280
 attgctgaag agcttggcgg cgaatgggct gaccgcttcc tcgtgcttta cggtatcgcc 2340
 55 gctcccgatt cgcagcgcac cgccttctat cgccttcttg acgagttctt ctgagcggga 2400
 ctctgggggt cgaaatgacc gaccaagcga cgcccaacct gccatcacga gatttcgatt 2460

ES 2 368 962 T3

```

ccaccgccgc cttctatgaa aggttgggct tcggaatcgt tttccgggac gccggctgga 2520
tgatcctcca gcgcggggat ctcatgctgg agttcttcgc ccaccccctg ctttaatgag 2580
5  atatgcgaga cgcctatgat cgcgatgat ttgctttcaa ttctgttgtg cacgttgtaa 2640
aaaacctgag catgtgtagc tcagatcctt accgccgggt tcggttcatt ctaatgaata 2700
10  tatcaccctg tactatcgta tttttatgaa taatattctc cgttcaattt actgattgta 2760
ccctactact tatatgtaca atattaaaat gaaaacaata tattgtgctg aataggttta 2820
tagcgacatc tatgatagag cgccacaata acaaacaatt gcgttttatt attacaaatc 2880
15  caattttaaa aaaagcggca gaaccgggtc aacctaaaag actgattaca taaatcttat 2940
tcaaatttca aaaggcccca ggggctagta tctacgacac accgagcggc gaactaataa 3000
20  cgttcaactga agggaactcc ggttccccgc cggcgcgcac gggtgagatt ccttgaagtt 3060
gagtattggc cgtccgctct accgaaagtt acgggcacca ttcaaccggg tccagcacgg 3120
cggccgggta accgacttgc tgccccgaga attatgcagc atttttttgg tgtatgtggg 3180
25  ccctgtacag cggccgcggt aacgcgtata ctctagagcg atcgccatgg agccatttac 3240
aattgaatat atcctgccgc cg 3262

```

```

30  <210> 6
    <211> 5345
    <212> DNA
    <213> Artificial

```

```

35  <220>
    <223> T-DNA de pTTA82 (DNA de reparación)

```

```

    <220>
    <221> característica diversa

```

```

40  <222> (1)..(25)
    <223> Secuencia del borde derecho del T-DNA

```

```

    <220>

```

```

45  <221> característica diversa
    <222> (26)..(62)

```

<223> Secuencia de polienlazador sintética

<220>

5 <221> característica diversa

<222> (63)..(578)

<223> bar 3' suprimido (complemento)

10 <220>

<221> característica diversa

<222> (579)..(603)

<223> Secuencia de polienlazador sintética

15

<220>

<221> característica diversa

<222> (604)..(616)

<223> sitio de I-Scel parcial

20

<220>

<221> característica diversa

<222> (617)..(1429)

25 <223> P35S3 (complemento)

<220>

<221> característica diversa

30 <222> (1430)..(1438)

<223> sitio de I-Scel parcial

<220>

35 <221> característica diversa

<222> (1460)..(1663)

<223> gen 7 de 3'

5 <220>

<221> característica diversa

<222> (1664)..(1782)

<223> Secuencia de polienlazador sintética

10

<220>

<221> característica diversa

<222> (1783)..(2069)

<223> promotor del gen de nopalina-sintetasa

15

<220>

<221> característica diversa

<222> (2070)..(2082)

20 <223> Secuencia de polienlazador sintética

<220>

<221> característica diversa

25 <222> (2083)..(2886)

<223> nptII

<220>

30 <221> característica diversa

<222> (2887)..(3059)

<223> 3' neo

35 <220>

<221> característica diversa

<222> (3060)..(3675)

<223> 3' ocs

5

<220>

<221> característica diversa

<222> (3676)..(3731)

<223> Secuencia de polienlazador sintética

10

<220>

<221> característica diversa

<222> (3732)..(4246)

15 <223> P35S2

<220>

<221> característica diversa

20 <222> (4247)..(4289)

<223> Ats1BL

<220>

25 <221> característica diversa

<222> (4290)..(4322)

<223> NLS

30 <220>

<221> característica diversa

<222> (4323)..(5023)

<223> I-Scel defectuosa

35

ES 2 368 962 T3

<220>

<221> característica diversa

<222> (5024)..(5260)

<223> 3' 35S

5

<220>

<221> característica diversa

<222> (5261)..(5317)

10 <223> Secuencia de polienlazador sintética

<220>

<221> característica diversa

15 <222> (5318)..(5345)

<223> Secuencia del borde izquierdo del T-DNA

<400> 6

20	aattacaacg gtatatatcc tgccagtact cggccgctcga cctgcaggca attggtacga	60
	tcctagacgc gtgagatcag atcctgccag aaaccacgt catgccagtt cccgtgcttg	120
	aagccggccg cccgcagcat gccgcggggg gcatatccga gcgcctcgtg catgcgcacg	180
25	ctcgggtcgt tgggcagccc gatgacagcg accacgctct tgaagccctg tgcctccagg	240
	gacttcagca ggtgggtgta gagcgtggag cccagtcccg tccgctggtg gcggggggag	300
30	acgtacacgg tcgactcggc cgtccagtcg taggcgttgc gtgccttcca ggggcccgcg	360
	taggcgatgc cggcgacctc gccgtccacc tcggcgacga gccagggata gcgctcccgc	420
	agacggacga ggtcgtccgt ccaactcctgc ggttctcgcg gctcggtagc gaagttgacc	480
35	gtgcttgtct cgatgtagtg gttgacgatg gtgcagaccg ccggcatgtc cgctcggtg	540
	gcacggcgga tgcggccgg gcgctcgttct ggggtccatgg ttatagagag agagatagat	600
40	ttaattaccc tgttattaga gagagactgg tgatttcagc gtgtcctctc caaatgaaat	660
	gaacttcctt atatagagga agggctcttgc gaaggatagt gggattgtgc gtcacccctt	720
	acgtcagtgg agatgtcaca tcaatccact tgctttgaag acgtggttgg aacgtcttct	780
45	ttttccacga tgctcctcgt ggggtgggggt ccatctttgg gaccactgtc ggcagaggca	840

ES 2 368 962 T3

tcttgaatga tagcctttcc tttatcgcaa tgatggcatt tgtaggagcc accttccttt 900
 5 tctactgtcc tttcgatgaa gtgacagata gctgggcaat ggaatccgag gaggtttccc 960
 gaaattatcc tttgttgaaa agtctcaata gccctttggt cttctgagac tgtatctttg 1020
 acatthtttg agtagaccag agtgtcgtgc tccaccatgt tgacgaagat tttcttcttg 1080
 10 tcattgagtc gtaaaagact ctgtatgaac tgttcgccag tcttcacggc gagttctggt 1140
 agatcctcga tttgaatctt agactccatg catggcctta gattcagtag gaactacctt 1200
 tttagagact ccaatctcta ttacttgcoct tggtttatga agcaagcctt gaatcgtcca 1260
 15 tactggaata gtacttctga tcttgagaaa tatgtctttc tctgtgttct tgatgcaatt 1320
 agtcctgaat cttttgactg catctttaac cttcttggga aggtatthga tctcctggag 1380
 20 attgttactc gggtagatcg tcttgatgag acctgctgcg taggaacgct tatccctagg 1440
 ccgctgtaca gggcccggga tcttgaaaga aatatagttt aaatatttat tgataaaata 1500
 acaagtcagg tattatagtc caagcaaaaa cataaattta ttgatgcaag tttaaattca 1560
 25 gaaatatttc aataactgat tatatcagct ggtacattgc cgtagatgaa agactgagtg 1620
 cgatattatg tgtaatacat aaattgatga tatagctagc ttaggcgcgc catagatccc 1680
 30 gtcaattctc actcattagg cccccaggc tttacacttt atgcttccgg ctcgtataat 1740
 gtgtggaatt gtgagcggat aacaatttca cacaggaaac aggatcatga gcggagaatt 1800
 aagggagtca cgttatgacc cccgccgatg acgcgggaca agccgthtta cgthttggaac 1860
 35 tgacagaacc gcaacgattg aaggagccac tcagccgcggt gthttctggag thtaatgagc 1920
 taagcacata cgtcagaaac cattattgcy cgttcaaaaag tgcctaaagg tcaactatcag 1980
 40 ctagcaaaata thtcttgtaaaa aaatgctcc actgacgttc cataaattcc cctcggatc 2040
 caattagagt ctcatattca ctctcaatca aagatccggc ccatgatcat gtggattgaa 2100
 caagatggat tgcacgcagg thctccggcc gcttgggtgg agaggctatt cggctatgac 2160
 45 tgggcacaac agacaatcgg ctgctctgat gccgcctgt tccggctgtc agcgcagggg 2220
 cggccggthc thtttgtaaaa gaccgacctg tccggtgccc tgaatgaaact gcaggacgag 2280
 50 gcagcgcggc tatcgtggct ggccacgagc ggcgttccct gcgcagctgt gctcagctt 2340
 gtcactgaag cgggaagggg ctggctgcta ttgggcgaag tgccggggca ggatctcctg 2400
 tcatctcacc ttgctcctgc cgagaaaagta tccatcatgg ctgatgcaat gcggcggctg 2460
 55 catacgttg atccggctac ctgcccattc gaccaccaag cgaaacatcg catcagagcga 2520
 gcacgtactc ggatggaagc cggctctgtc gatcaggatg atctggacga agagcatcag 2580

ES 2 368 962 T3

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55

gggctcgcgc cagccgaact gttcgccagg ctcaaggcgc gcatgcccga cggcgaggat 2640
 ctcgtcgtga cccatggcga tgcoctgcttg ccgaatatca tgggtggaaaa tggccgcttt 2700
 tctggattca tcgactgtgg ccggctgggt gtggcggacc gctatcagga catagcgttg 2760
 gctacccgtg atattgctga agagcttggc ggcgaatggg ctgaccgctt cctcgtgctt 2820
 tacggtatcg ccgctcccga ttcgcagcgc atcgcocttct atcgcocttct tgacgagttc 2880
 ttctgagcgg gactctgggg ttcgaaatga ccgaccaagc gacgcccac ctgccatcac 2940
 gagatttcga ttccaccgcc gccttctatg aaaggttggg cttcggaatc gttttccggg 3000
 acgccggctg gatgatcctc cagcgcgggg atctcatgct ggagttcttc gccaccccc 3060
 tgctttaatg agatatgcga gacgcctatg atcgcatgat atttgctttc aattctgttg 3120
 tgcacgttgt aaaaaacctg agcatgtgta gctcagatcc ttaccgccgg tttcggttca 3180
 ttctaataaa tatatcacc gttactatcg ttttttatg aataatattc tccgttcaat 3240
 ttactgattg taccctacta cttatatgta caatattaa atgaaaaca tatattgtgc 3300
 tgaatagggt tatagcgaca tctatgatag agcgcacaa taacaaaca ttgcgtttta 3360
 ttattacaaa tccaatttta aaaaagcgg cagaaccggt caaacctaaa agactgatta 3420
 cataaatctt attcaaattt caaaaggccc caggggctag tatctacgac acaccgagcg 3480
 gcgaactaat aacgttact gaaggaact ccggttccc gccggcgcgc atgggtgaga 3540
 ttccttgaag ttgagtattg gccgtccgct ctaccgaaag ttacgggcac cattcaacc 3600
 ggtccagcac ggcggccggg taaccgactt gctgccccga gaattatgca gcattttttt 3660
 ggtgtatgtg ggccctgtac agcggccgcyg ttaacgcgta tactctagta tgcaccatac 3720
 atggagtcaa aaattcagat cgaggatcta acagaactcg ccgtgaagac tggcgaacag 3780
 ttcatacaga gtcttttacg actcaatgac aagaagaaaa tcttcgtcaa catggtggag 3840
 cacgacactc tcgtctactc caagaatata aaagatacag tctcagaaga ccaaagggct 3900
 attgagactt ttcaacaaag ggtaatatcg ggaaacctcc tcggattcca ttgccagct 3960
 atctgtcact tcatcaaaag gacagtagaa aaggaaggtyg gcacctaca atgccatcat 4020
 tgcgataaag gaaaggctat cgttcaagat gcctctgccg acagtgggtcc caaagatgga 4080
 cccccacca cgaggagcat cgtggaaaaa gaagacgctc caaccacgtc ttcaaagcaa 4140
 gtggattgat gtgatatact cactgacgta agggatgacg cacaatccca ctatccttcg 4200
 caagaccctt cctctatata aggaagttca tttcatttgg agaggactcg agaattaagc 4260
 aaaagaagaa gaagaagaag tccaaaacca tggctaaacc cccaagaag aagcgaagg 4320

ES 2 368 962 T3

	ttaacatcaa aaaaaaccag gtaatgaacc tgggtccgaa ctctaaactg ctgaaagaat	4380
5	acaaatccca gctgatcgaa ctgaacatcg aacagttcga agcaggatc ggtctgatcc	4440
	tgggtgatgc ttacatccgt tctcgtgatg aaggtaaaac ctactgtatg cagttcgagt	4500
	ggaaaaacaa agcatacatg gaccacgtat gtctgctgta cgatcagtgg gtactgtccc	4560
10	cgccgcacaa aaaagaacgt gttaaccacc tgggtaacct ggtaatcacc tggggcgccc	4620
	agactttcaa acaccaagct ttcaacaaac tgggtaacct gttcatcggt aacaacaaaa	4680
15	aaaccatccc gaacaacctg gttgaaaact acctgacccc gatgtctctg gcatactggt	4740
	tcatggatga tggtggtaaa tgggattaca aaaaaactc taccaacaaa gtattgtact	4800
	gaacaccag tctttcactt tcgaagaagt agaatacctg gttaagggtc tgcgtaacaa	4860
20	attccaactg aactgttacg taaaaatcaa caaaaacaaa ccgatcatct acatcgattc	4920
	tatgtcttac ctgatcttct acaacctgat caaacctgac ctgatcccgc agatgatgta	4980
25	caaactgccg aacactatct cctccgaaac tttcctgaaa tagggctagc aagcttggac	5040
	acgctgaaat caccagtctc tctctacaaa tctatctctc tctatcttct ccataataat	5100
	gtgtgagtag ttcccagata agggaattag ggttcctata gggtttcgct catgtgttga	5160
30	gcatataaga aacccttagt atgtatctgt atttgtaaaa tacttctatc aataaaaattt	5220
	ctaattccta aaacccaaat ccagtactaa aatccagatc atgcatggta cagcggccgc	5280
35	gttaacgcgt atactctaga gcgatcgcca tggagccatt tacaattgaa tatatcctgc	5340
	cgccg	5345

<210> 7

40 <211> 4066

<212> DNA

<213> Artificial

<220>

45 <223> pCV78

<220>

<221> característica diversa

<222> (234)..(763)

<223> promotor P35S2

<220>

5 <221> característica diversa

<222> (764)..(805)

<223> Ats1b'

10 <220>

<221> característica diversa

<222> (808)..(839)

<223> señal de localización nuclear

15

<220>

<221> característica diversa

<222> (840)..(1541)

<223> I-SceI sintética

20

<220>

<221> característica diversa

<222> (1544)..(1792)

25 <223> 3' 35S

<220>

<221> característica diversa

30 <222> (3006)..(3886)

<223> Resistencia a ampicilina (complemento)

<400> 7

35 t c g c g c g t t t c g g t g a t g a c g g t g a a a a c c t c t g a c a c a t g c a g c t c c c g g a g a c g g t c a

60

ES 2 368 962 T3

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
 ttggcgggtg tcggggctgg cttactatg cggcatcaga gcagattgta ctgagagtgc 180
 5 accatacctg caggcaattg gtacctacgt atgcatggcg cgccatatgc accatacatg 240
 gagtcaaaaa ttcagatcga ggatctaaca gaactcgccg tgaagactgg cgaacagttc 300
 10 atacagagtc ttttacgact caatgacaag aagaaaatct tcgtcaacat ggtggagcac 360
 gacactctcg tctactccaa gaatatcaaa gatacagtct cagaagacca aagggctatt 420
 gagacttttc aacaaagggg aatatcggga aacctcctcg gattccattg cccagctatc 480
 15 tgtcacttca tcaaaaggac agtagaaaag gaaggtggca cctacaaatg ccatcattgc 540
 gataaaggaa aggctatcgt tcaagatgcc tctgccgaca gtggtcccaa agatggacct 600
 ccaccacga ggagcatcgt ggaaaaagaa gacgttccaa ccacgtcttc aaagcaagtg 660
 20 gattgatgtg atatctccac tgacgtaagg gatgacgcac aatcccacta tccttcgcaa 720
 gacccttct ctatataagg aagttcattt catttgagaga ggactcgaga attaagcaaa 780
 25 agaagaagaa gaagaagtcc aaaaccatgg ccaagcctcc caagaagaag cgcaaagtga 840
 acatcaagaa gaaccaggtg atgaacctgg gacctaacag caagctcctg aaggagtaca 900
 agagccagct gatcgaactg aacatcgagc agttogaagc tggcatcggc ctgatcctgg 960
 30 gcgatgccta catcagatcc cgggacgaag gcaagacctc ctgcatgcag ttcgagtgga 1020
 agaacaaggc ctacatggac cacgtgtgtc tgctgtacga ccagtgggtc ctgagccctc 1080
 35 ctcaacaagaa ggagcgcgtg aacctctgg gcaacctcgt gatcacctgg ggagcccaga 1140
 ccttcaagca ccaggccttc aacaagctgg ccaacctggt catcgtgaac aacaagaaga 1200
 ccattcccaa caacctcgtg gagaactacc tcaactccat gagcctggcc tactggttca 1260
 40 tggacgacgg aggcaagtgg gactacaaca agaacagcac caacaagtca attgtgctga 1320
 acacccaaag cttcaccttc gaagaagtgg agtacctcgt caagggcctg cgcaacaagt 1380
 45 tccagctgaa ctgctacgtg aagatcaaca agaacaagcc tatcatctac atcgacagca 1440
 tgagctacct gatcttctac aacctgatca agccatacct gatccctcag atgatgtaca 1500
 agctgcccaa caccatcagc agcgagacct tcctgaagtg aggctagcaa gcttggacac 1560
 50 gctgaaatca ccagtctctc tctacaaatc tatctctctc tattttctcc ataataatgt 1620
 gtgagtagtt cccagataag ggaattaggg ttocatatagg gtttcgctca tgtgttgagc 1680
 55 atataagaaa cccttagtat gtatttgtat ttgtaaaata cttctatcaa taaaatttct 1740
 aattcctaaa accaaaatcc agtactaaaa tccagatcat gcatggtaca gcggccgcgt 1800

ES 2 368 962 T3

	taacgcgtat	actctagagc	gatcgcgaagc	ttggcgtaat	catggtcata	gctgtttctc	1860
	gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	1920
5	aaagcctggg	gtgcctaattg	agtgagctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	1980
	gctttccagt	cgggaaacct	gtcgtgccag	ctgcattaat	gaatcggcca	acgcgcgggg	2040
10	agaggcgggt	tgcgtattgg	gcgctcttcc	gcttctctgc	tactgactc	gctgcgctcg	2100
	gtcgttcggc	tgcggcgagc	ggtatcagct	cactcaaagg	cggtaatacg	gttatccaca	2160
	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaaag	gccagcaaaa	ggccaggaac	2220
15	cgtaaaaagg	ccgcgttgct	ggcgtttttc	cataggctcc	gccccctga	cgagcatcac	2280
	aaaaatcgac	gctcaagtca	gaggtggcga	aaccgcagac	gactataaag	ataccaggcg	2340
20	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	ccctgccgct	taccggatac	2400
	ctgtccgcct	ttctcccttc	gggaagcgtg	gcgctttctc	aaagctcacg	ctgtaggtat	2460
	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	cccgttcag	2520
25	cccgaccgct	gcgccttatac	cggtaacctat	cgtcttgagt	ccaaccgggt	aagacacgac	2580
	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	tgtaggcgggt	2640
30	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaagaac	agtatttggt	2700
	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	2760
	aaacaaacca	ccgctggtag	cggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	2820
35	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	ggtctgacgc	tcagtggaac	2880
	gaaaactcac	gttaagggat	tttggatcatg	agattatcaa	aaaggatctt	cacctagatc	2940
40	cttttaaatt	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	aacttggtct	3000
	gacagttacc	aatgcttaat	cagtgaggca	cctatctcag	cgatctgtct	atcttggtca	3060
	tccatagttg	cctgactccc	cgctcgtgtag	ataactacga	tacgggaggg	cttaccatct	3120
45	ggccccagtg	ctgcaatgat	accgcgagac	ccacgctcac	cggtcccaga	tttatcagca	3180
	ataaaccagc	cagccggaag	ggccgagcgc	agaagtggtc	ctgcaacttt	atccgcctcc	3240
50	atccagtcta	ttaattgttg	ccgggaagct	agagtaagta	gttcgccagt	taatagtttg	3300
	cgcaacgttg	ttgccattgc	tacaggcatc	gtgggtgtcac	gctcgtcggt	tggtatggct	3360
	tcattcagct	ccggttccca	acgatcaagg	cgagttacat	gatcccccat	gttgtgcaaa	3420
55	aaagcggtta	gctccttcgg	tcctccgatc	gttgtcagaa	gtaagttggc	cgcagtgtta	3480
	tcactcatgg	ttatggcagc	actgcataat	tctcttactg	tcatgccatc	cgtaagatgc	3540

ES 2 368 962 T3

	ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg	3600
	agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa	3660
5	gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg	3720
	agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc	3780
10	accagcgttt ctgggtgagc aaaaacagga aggcaaatg ccgcaaaaaa ggaataagg	3840
	gcgacacgga aatggtgaat actcatactc ttcctttttc aatattattg aagcatttat	3900
	cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata	3960
15	ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc	4020
	atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtc	4066

REIVINDICACIONES

1. Un método para introducir un DNA extraño de interés en un sitio preseleccionado de un genoma nuclear de una célula vegetal, que comprende las etapas de
 - 5 (a) inducir una rotura de DNA bicatenario en un sitio preseleccionado en dicho genoma nuclear mediante introducción en dicha célula vegetal de un gen expresable en plantas que codifica una enzima que induce una rotura de DNA bicatenario de corte raro que reconoce dicho sitio preseleccionado;
 - (b) introducir el DNA extraño de interés en la célula vegetal mediante transferencia directa de DNA.
2. El método según la reivindicación 1, en el que dicha enzima que induce una rotura de DNA bicatenario comprende una señal de localización nuclear.
- 10 3. El método según las reivindicaciones 1 y 2, en el que dicha enzima que induce una rotura de DNA bicatenario es la endonucleasa I-SceI.
4. El método según las reivindicaciones 1-3, en el que dicha transferencia directa de DNA se logra por bombardeo de microproyectiles recubiertos con el DNA extraño de interés.
- 15 5. El método según las reivindicaciones 1-3, en el que dicha transferencia directa de DNA se logra mediante introducción de DNA por electroporación en protoplastos.
6. El método según las reivindicaciones 1-3, en el que dicha transferencia directa de DNA se logra mediante introducción de DNA por electroporación en células vegetales intactas o tejidos o células vegetales parcialmente degradados.
- 20 7. El método de una cualquiera de las reivindicaciones 1 a 6, en el que dicho DNA extraño de interés está flanqueado por una región de DNA que tiene al menos 80% de identidad de secuencia con una región de DNA que flanquea el sitio preseleccionado.
8. El método de una cualquiera de las reivindicaciones 1 a 7, por el cual la célula vegetal es una célula de maíz.
9. El método de la reivindicación 8, en el que la célula de maíz está comprendida en una suspensión celular.
- 25 10. El método de una cualquiera de las reivindicaciones 1 a 9, por el cual dicha célula vegetal se incuba en un compuesto fenólico vegetal antes de la etapa (a).
11. El método de la reivindicación 10, en el que dicho compuesto fenólico vegetal es acetosiringona.

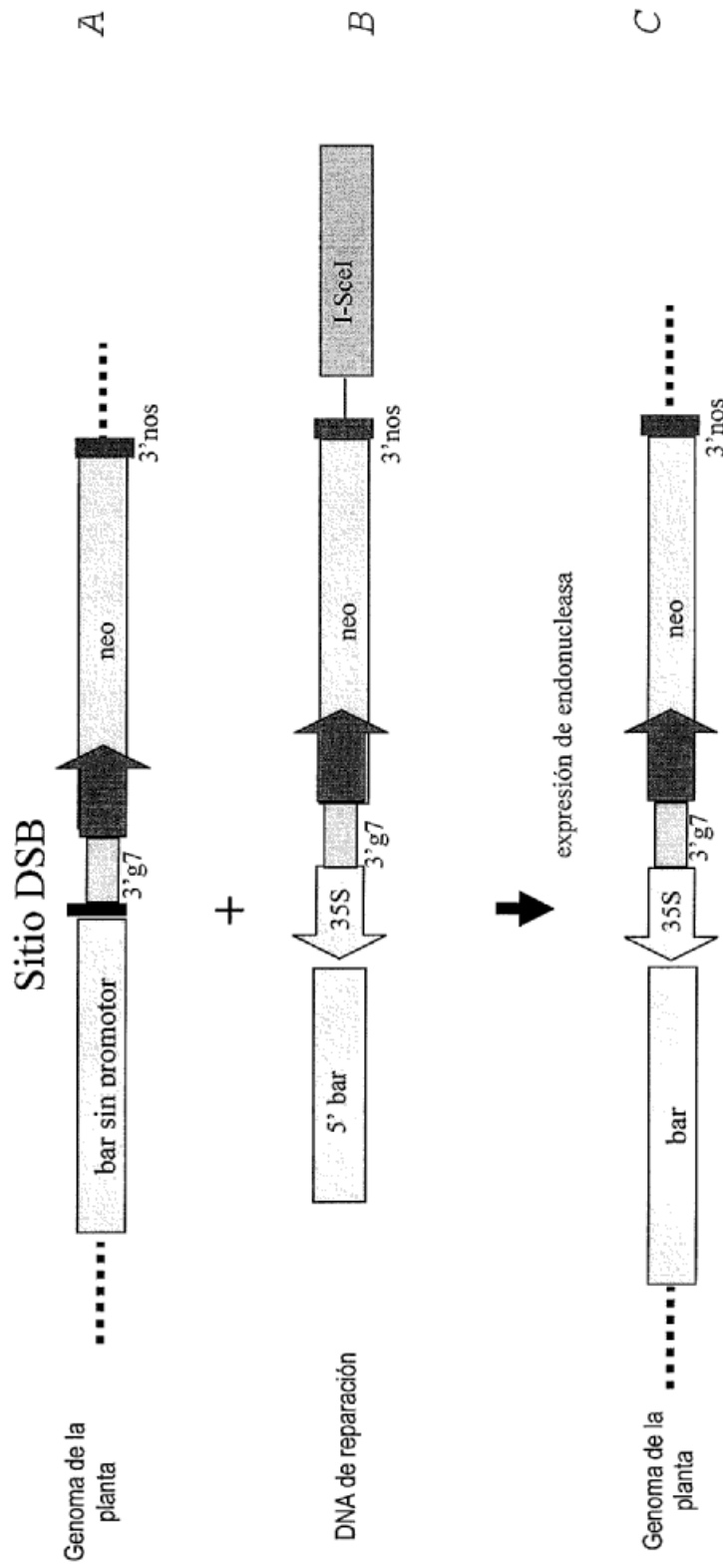


Figura 1