

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 625 941**

51 Int. Cl.:

**C12N 9/22** (2006.01)

**C12N 5/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.07.2009 PCT/US2009/050566**

87 Fecha y número de publicación internacional: **21.01.2010 WO10009147**

96 Fecha de presentación y número de la solicitud europea: **14.07.2009 E 09798667 (3)**

97 Fecha y número de publicación de la concesión europea: **15.03.2017 EP 2313498**

54 Título: **Secuencias de reconocimiento para meganucleasas derivadas de i-crei y sus usos**

30 Prioridad:

**14.07.2008 US 80453 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**21.07.2017**

73 Titular/es:

**PRECISION BIOSCIENCES, INC. (100.0%)  
302 East Pettigrew Street, Dibrell Building, Suite  
A-100  
Durham, NC 27701, US**

72 Inventor/es:

**JANTZ, DEREK y  
SMITH, JAMES, JEFFERSON**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

ES 2 625 941 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

Secuencias de reconocimiento para meganucleasas derivadas de I-Crel y sus usos

5 **Referencia a las solicitudes relacionadas**

Esta solicitud reivindica el beneficio de prioridad a la Solicitud de Patente Provisional de Estados Unidos Nº 61/080.453, presentada el 14 de julio de 2008.

10 **Campo de la invención**

La invención se refiere al campo de la biología molecular y de la tecnología de ácido nucleico recombinante. En particular, la invención se refiere a secuencias de ADN que pueden ser reconocidas y escindidas por una endonucleasa de asentamiento (*homing*) derivada de I-Crel, racionalmente diseñada, de origen no natural y métodos de uso de la misma. La invención también se refiere a métodos de producción de ácidos nucleicos recombinantes, células y organismos que usan tales meganucleasas que escinden tales sitios de ADN. Además, la invención se refiere a métodos para dirigir una actividad meganucleasa derivada de I-Crel, diseñada a medida.

20 **Antecedentes de la invención**

La ingeniería genómica requiere la capacidad de insertar, someter a delección, sustituir y si no manipular secuencias genéticas específicas dentro de un genoma, y tiene numerosas aplicaciones terapéuticas y biotecnológicas. El desarrollo de medios eficaces para la modificación genómica queda como un objetivo principal en la terapia génica, agrotecnología, y biología sintética (Porteus y col. (2005), *Nat. Biotechnol.* 23:967-73; Tzfira y col. (2005), *Trends Biotechnol.* 23:567-9; McDaniel y col. (2005), *Curr. Opin. Biotechnol.* 16:476-83). Un método común para insertar o modificar una secuencia de ADN implica introducir una secuencia de ADN transgénico flanqueada por secuencias homólogas a la diana genómica y seleccionar o someter a cribado para un evento de recombinación homóloga con éxito. La recombinación con el ADN transgénico se da rara vez, pero se puede estimular por una rotura de la doble cadena en el ADN genómico en el sitio diana. Se han empleado numerosos métodos para crear roturas de la doble cadena de ADN, incluyendo irradiación y tratamientos químicos. Aunque estos métodos estimulan eficazmente la recombinación, las roturas de la doble cadena se dispersan al azar en el genoma, lo cual puede ser altamente mutagénico y tóxico. En la actualidad, la incapacidad de modificaciones genéticas diana a sitios únicos dentro de un ambiente cromosómico es un impedimento muy importante para la ingeniería genómica con éxito.

Un enfoque para conseguir este objetivo es estimular la recombinación homóloga en una rotura de la doble cadena en un locus diana usando una nucleasa con especificidad para una secuencia que es suficientemente grande para estar presente en solamente un sitio único dentro del genoma (véase, por ejemplo, Porteus y col. (2005), *Nat. Biotechnol.* 23:967-73). La eficacia de esta estrategia se ha demostrado en diversos organismos usando fusiones quiméricas entre un dominio de unión a ADN de dedo de zinc modificado por ingeniería y el dominio nucleasa no específico de la enzima de restricción FokI (Porteus (2006), *Mol. Ther.* 13:438-46; Wright y col. (2005), *Plant J.* 44:693-705; Urnov y col. (2005), *Nature* 435:646-51). Aunque estas nucleasas dedo de zinc artificiales estimulan la recombinación específica a sitio, retienen actividad de escisión no específica residual resultante de la poca regulación del dominio nucleasa y frecuentemente escinden en sitios no previstos (Smith y col. (2000), *Nucleic Acids Res.* 28:3.361-9). Tal escisión no prevista puede causar mutaciones y toxicidad en el organismo tratado (Porteus y col. (2005), *Nat. Biotechnol.* 23:967-73).

Un grupo de nucleasas de origen natural que reconocen sitios de escisión de 15 a 40 pares de bases encontrados comúnmente en los genomas de plantas y hongos pueden proporcionar una alternativa de modificación por ingeniería de genoma menos tóxica. Tales "meganucleasas" o "endonucleasas de asentamiento" frecuentemente están asociadas con elementos de ADN parásito, tales como intrones e inteínas de auto corte y empalme (*self-splicing*) de grupo 1. Promueven de manera natural la recombinación homóloga o la inserción génica en localizaciones específicas en el genoma hospedador produciendo una rotura de la doble cadena en el cromosoma, que recluta la maquinaria de reparación de ADN celular (Stoddard (2006), *Q. Rev. Biophys.* 38:49-95). Las meganucleasas se agrupan normalmente en cuatro familias: la familia LAGLIDADG, la familia GIY-YIG, la familia His-Cys box y la familia HNH. Estas familias se caracterizan por motivos estructurales, que afectan la actividad catalítica y la secuencia de reconocimiento. Por ejemplo, los miembros de la familia LAGLIDADG se caracterizan por tener o bien una o dos copias del motivo de LAGLIDADG conservado (véase, Chevalier y col. (2001), *Nucleic Acids Res.* 29(18):3.757-3.774). Las meganucleasas LAGLIDADG con una única copia del motivo de LAGLIDADG forman homodímeros, mientras que los miembros con dos copias del motivo de LAGLIDADG se encuentran como monómeros.

Se han usado meganucleasas naturales, principalmente de la familia LAGLIDADG, para promover eficazmente la modificación genómica específica a sitio en plantas, levadura, *Drosophila*, células de mamíferos y ratones, pero este enfoque se ha limitado a la modificación de o bien genes homólogos que conservan la secuencia de reconocimiento de meganucleasa (Monnat y col. (1999), *Biochem. Biophys. Res. Commun.* 255:88-93) o bien genomas previamente modificados por ingeniería en los cuales se ha introducido una secuencia de reconocimiento (Rouet y col. (1994),

*Mol. Cell. Biol.* 14: 8.096-106; Chilton y col. (2003), *Plant Physiol.* 133:956-65; Puchta y col. (1996), *Proc. Natl Acad. Sci. USA* 93:5.055-60; Rong y col. (2002), *Genes Dev.* 16:1.568-81; Gouble y col. (2006), *J. Gene Med.* 8(5):616-622).

5 La ejecución sistemática de la modificación génica estimulada por nucleasa requiere el uso de enzimas modificadas por ingeniería con especificidades personalizadas a roturas de ADN diana a sitios existentes en un genoma y, por lo tanto, ha habido un gran interés en la adaptación de meganucleasas para promover modificaciones génicas a sitios médicamente o biotecnológicamente relevantes (Porteus y col. (2005), *Nat. Biotechnol.* 23:967-73; Sussman y col. (2004), *J. Mol. Biol.* 342:31-41; Epinat y col. (2003), *Nucleic Acids Res.* 31:2.952-62).

10 I-Crel (SEQ ID NO: 1) es un miembro de la familia LAGLIDADG que reconoce y escinde una secuencia de reconocimiento de 22 pares de bases en el cromosoma del cloroplasto, y que representa una diana atractiva para el rediseño de meganucleasa. Las técnicas de selección genética se han usado para modificar la preferencia de sitio de reconocimiento de I-Cre tipo natural (Sussman y col. (2004), *J. Mol. Biol.* 342:31-41; Chames y col. (2005), *Nucleic Acids Res.* 33:e178; Seligman y col. (2002), *Nucleic Acids Res.* 30:3.870-9, Arnould y col. (2006), *J. Mol. Biol.* 355:443-58). Más recientemente, se describió un método de diseño de manera racional de mono-

15 meganucleasas LAGLIDADG que es capaz de rediseñar completamente I-Crel y otras tales meganucleasas a sitios de ADN muy divergentes dianas, incluyendo sitios en genomas de mamífero, levadura, planta, bacterianos y víricos (documento WO 2007/047859).

20 Las secuencias de ADN reconocidas por I-Crel tienen 22 pares de bases de longitud. Un ejemplo de un sitio de reconocimiento de I-Crel de origen natural se proporciona en SEQ ID NO: 2 y SEQ ID NO: 3, pero la enzima se unirá a diversas secuencias relacionadas con afinidad variante. La enzima se une a ADN como un homodímero en el cual cada monómero hace contactos directos con un "medio-sitio" de nueve pares de bases y los dos medio-sitios se separan por cuatro pares de bases que no están directamente contactadas por la enzima (Figura 1a). Como todas las meganucleasas de la familia LAGLIDADG, I-Cre produce una rotura de la doble cadena escalonada en el centro de sus secuencias de reconocimiento que da como resultado la producción de un saliente 3' de cuatro pares de bases (Figura 1a). La presente invención concierne a los cuatro pares de bases centrales en las secuencias de reconocimiento de I-Crel (es decir, los cuatro pares de bases que llegan a ser el saliente 3' después de la escisión por I-Crel, o "secuencia central", Figura 1b). En el caso de la secuencia de reconocimiento de I-Crel nativa en el gen de ARNr 23S de *Chlamydomonas reinhardtii*, esta secuencia de cuatro pares de bases es 5'-GTGA-3'. En el interés de producir meganucleasas modificadas por ingeniería genética que reconocen secuencias de ADN que se desvían de las secuencias de reconocimiento de I-Crel de tipo natural, es deseable saber el grado al cual la secuencia central de cuatro pares de bases puede desviarse de las secuencias de tipo natural. Un número de estudios publicados que conciernen a I-Crel o sus derivados evaluaron la enzima, o bien tipo natural o modificada por ingeniería genética, usando sustratos de ADN que emplearon o bien la secuencia central 5'-GTGA-3' nativa o la secuencia palindrómica 5'-GTAC-3'. Recientemente, Arnould y col. (Arnould y col. (2007), *J. Mol. Biol.* 371:49-65) informaron que un conjunto de meganucleasas modificadas por ingeniería genética derivadas de sustratos de ADN escindido por I-Crel con eficacias variantes dependiendo de si las secuencias de sustrato estaban centradas

35 alrededor de 5'-GTAC-3', 5'-TTGA-3', 5'-GAAA-3', o 5'-ACAC-3' (eficacia de escisión: GTAC > ACAC >> TTGA ≈ GAAA).

### Compendio de la invención

45 La presente invención se basa, en parte, en la identificación y caracterización de un subconjunto de secuencias de reconocimiento de ADN que pueden actuar como sustratos eficaces para la escisión por las meganucleasas derivadas de I-Crel, racionalmente diseñadas (más adelante, "meganucleasas derivadas de I-Crel"). La invención está definida por las reivindicaciones adjuntas.

50 Se describen métodos de identificación de conjuntos de secuencias de ADN de 22 pares de bases que se pueden escindir por meganucleasas derivadas de I-Crel y que tienen, en su centro, una de un conjunto limitado de secuencias centrales de ADN de cuatro pares de bases que contribuyen a escisión más eficaz por las meganucleasas derivadas de I-Crel. La invención también proporciona métodos que usan tales secuencias de ADN para producir ácidos nucleicos recombinantes, células y organismos utilizando las secuencias de reconocimiento como sustratos para meganucleasas derivadas de I-Crel, y productos que incorporan tales secuencias de ADN.

55

Por tanto, en un aspecto, la invención proporciona un método *ex vivo* para escindir un ADN de doble cadena que comprende: (a) identificar en el ADN al menos un sitio de reconocimiento para una meganucleasa derivada de I-Crel racionalmente diseñada con especificidad alterada en relación a I-Crel, en el que el sitio de reconocimiento no se escinde por una I-Crel de origen natural, en el que el sitio de reconocimiento se identifica basándose en la presencia de una secuencia central de cuatro pares de bases seleccionada entre el grupo que consiste en TTGT, TTAT, TTTC, TTCC, TTAG, TTAC, y TTGC; (b) proporcionar la meganucleasa racionalmente diseñada; y (c) poner en contacto el ADN con la meganucleasa racionalmente diseñada; por lo cual la meganucleasa racionalmente diseñada escinde el ADN:

60

65

- En otro aspecto, la invención proporciona un método para escindir un ADN de doble cadena que comprende: (a) incorporar una secuencia central de cuatro pares de bases en un sitio de reconocimiento para una meganucleasa derivada de I Crel racionalmente diseñada con especificidad alterada en relación a I-Crel, en el que el sitio de reconocimiento no se escinde por una I-Crel de origen natural, en el que la secuencia central de cuatro pares de bases se selecciona entre el grupo que consiste en TTGT, TTAT, TTTC, TTCC, TTAG, TTAC, y TTGC; y (b) introducir dicho sitio de reconocimiento en dicho ADN; y (c) proporcionar la meganucleasa racionalmente diseñada; y (d) poner en contacto el ADN con la meganucleasa racionalmente diseñada; por lo cual la meganucleasa racionalmente diseñada escinde el ADN.
- 5
- 10 En algunas realizaciones, la escisión de ADN es *in vitro*.
- En algunas realizaciones, el ADN se selecciona entre el grupo que consiste en un producto de PCR; un cromosoma artificial; ADN genómico aislado de bacterias, hongos, plantas o células animales; y ADN vírico.
- 15 En algunas realizaciones, el ADN está presente en una célula seleccionada entre el grupo que consiste en una célula bacteriana, fúngica, vegetal y animal.
- En algunas realizaciones, el ADN está presente en un ácido nucleico seleccionado entre el grupo que consiste en un plásmido, un profago y un cromosoma.
- 20 En ciertas realizaciones, el método comprende además diseñar racionalmente la meganucleasa derivada de I Crel para reconocer el sitio de reconocimiento.
- En algunas realizaciones, el método comprende además producir la meganucleasa derivada de I-Crel racionalmente diseñada.
- 25 Se describe una célula transformada con un ácido nucleico que comprende, en orden: a) una primera secuencia de ADN de 9 pares de bases que se puede unir por un monómero de meganucleasa derivada de I Crel o por un primer dominio de una meganucleasa derivada de I Crel de cadena sencilla; b) una secuencia de ADN de cuatro pares de bases seleccionada entre el grupo que consiste en GTGT, GTAT, TTAG, GTAG, TTAC, TCTC, TCAC, GTCC, GTAC, TCGC, AAGC, GAGC, GCGC, GTGC, TAGC, TTGC, ATGC, ACAC, ATAC, CTAA, CTAC, GTAA, GAGA, GTGA, GGAC, GTAC, GCGA, GCTT, GCTC, GCGC, GCAC, GCTA, GCAA y GCAT; y c) una segunda secuencia de ADN de 9 pares de bases que se puede unir por un monómero de meganucleasa derivada de I Crel o por un segundo dominio de la meganucleasa derivada de I Crel de cadena sencilla, en el que la segunda secuencia de ADN de 9 pares de bases está en dirección inversa en relación a la primera.
- 30
- 35 También se describe una célula que contiene una secuencia de ácidos nucleicos exógena integrada en su genoma, que comprende, en orden: a) una primera secuencia de ADN de 9 pares de bases exógena que se puede unir por un monómero de meganucleasa derivada de I Crel o por un primer dominio de una meganucleasa derivada de I Crel de cadena sencilla; b) una secuencia de ADN de cuatro pares de bases exógena seleccionada entre el grupo que consiste en GTGT, GTAT, TTAG, GTAG, TTAC, TCTC, TCAC, GTCC, GTAC, TCGC, AAGC, GAGC, GCGC, GTGC, TAGC, TTGC, ATGC, ACAC, ATAC, CTAA, CTAC, GTAA, GAGA, GTGA, GGAC, GTAC, GCGA, GCTT, GCTC, GCGC, GCAC, GCTA, GCAA y GCAT; y a) una segunda secuencia de ADN de 9 pares de bases exógena que se puede unir por un monómero de meganucleasa derivada de I Crel o por un segundo dominio de la meganucleasa derivada de I Crel de cadena sencilla, en la que la segunda secuencia de ADN de 9 pares de bases está en dirección inversa en relación a la primera.
- 40
- 45 El ácido nucleico puede ser un plásmido, un cromosoma artificial o un ácido nucleico vírico.
- 50 La célula puede ser una célula animal no humana, una célula vegetal, una célula bacteriana o una célula fúngica.
- La secuencia de ADN de cuatro pares de bases puede ser TTGT, TTAT, TCTT, TCGT, TCAT, GTTT, GTCT, GGAT, GAGT, GAAT, ATGT, TTTC, TTCC, TGAC, TAAC, GTTC, ATAT, TCGA, TTAA, GGGC, ACGC, CCGC, CTGC, ACAA, ATAA, AAGA, ACGA, ATGA, AAAC, AGAC, ATCC, ACTC, ATTC, ACAT, GAAA, GGAA, GTCA, GTTA, GAAC, ATAT, TCGA, TTAA, GCCC, GCGT, GCGG o GCAG.
- 55
- La secuencia de ADN de cuatro pares de bases puede ser GTGT, GTAT, TTAG, GTAG, TTAC, TCTC, TCAC, GTCC, GTAC, TCGC, AAGC, GAGC, GCGC, GTGC, TAGC, TTGC, ATGC, ACAC, ATAC, CTAA, CTAC, GTAA, GAGA, GTGA, GGAC, GTAC, GCGA, GCTT, GCTC, GCGC, GCAC, GCTA, GCAA o GCAT.
- 60
- Se describe un método para dirigir una actividad meganucleasa derivada de I-Crel, diseñada a medida, que comprende: (a) recibir una secuencia de ADN en la cual se ha de introducir una rotura de la doble cadena por una meganucleasa derivada de I Crel racionalmente diseñada; (b) identificar en la secuencia de ADN al menos un sitio de reconocimiento para una meganucleasa derivada de I Crel racionalmente diseñada con especificidad alterada en relación a I-Crel, en el que el sitio de reconocimiento no se escinde por una I-Crel de origen natural, en el que el sitio de reconocimiento tiene una secuencia central de cuatro pares de bases seleccionada entre el grupo que consiste
- 65

en TTGT, TTAT, TCTT, TCGT, TCAT, GTTT, GTCT, GGAT, GAGT, GAAT, ATGT, TTTC, TTCC, TGAC, TAAC, GTTC, ATAT, TCGA, TTAA, GGGC, ACGC, CCGC, CTGC, ACAA, ATAA, AAGA, ACGA, ATGA, AAAC, AGAC, ATCC, ACTC, ATTC, ACAT, GAAA, GGAA, GTCA, GTTA, GAAC, ATAT, TCGA, TTAA, GCCC, GCGT, GCGG y GCAG; y (c) proporcionar la meganucleasa racionalmente diseñada.

5 Además, el método puede comprender diseñar racionalmente la meganucleasa derivada de I-Crel para reconocer el sitio de reconocimiento.

Además, el método puede comprender producir la meganucleasa racionalmente diseñada.

10 La meganucleasa racionalmente diseñada se puede suministrar al mismo grupo del cual se ha recibido la secuencia de ADN.

15 Estos y otros aspectos y realizaciones de la invención serán aparentes a los expertos de la técnica a partir de la siguiente descripción detallada de la invención, figuras y reivindicaciones adjuntas.

### Breve descripción de las figuras

20 Figura 1. (A) Ilustración esquemática de las interacciones entre el homodímero de I-Crel de origen natural y una secuencia de reconocimiento de doble cadena, basado en datos cristalográficos. Esta representación esquemática representa una secuencia de reconocimiento (SEQ ID NO: 2 y SEQ ID NO: 3), mostrada como desenrollada solamente con fines de ilustración, unida por el homodímero, mostrado como dos óvalos. Las bases de cada medio-sitio de ADN se numeran de -1 a -9, y los residuos de aminoácidos de I-Crel que forman la superficie de reconocimiento se indican por denominaciones de aminoácido de una letra y números que indican la posición de residuo. Los cuatro pares de bases que comprenden la secuencia central se numeran de +1 a +4. Líneas negras sólidas: enlaces de hidrógeno a bases de ADN. (B) Secuencia de reconocimiento de I-Crel de tipo natural (SEQ ID NO: 2 y SEQ ID NO: 3) que muestra las localizaciones de los medio-sitios invertidos y la secuencia central.

30 Figura 2. (A) Diagrama esquemático de los sustratos plásmidos evaluados para determinar la preferencia de la secuencia central. Se produjeron un conjunto de plásmidos pUC-19 que portaban secuencias de reconocimiento potenciales para la meganucleasa modificada por ingeniería genética DJ 1. Estas secuencias de reconocimiento potenciales comprendían un par de medio-sitios DJ1 invertidos separados por diversas secuencias centrales diferentes de cuatro pares de bases (numeradas +1 a +4), como se describe más adelante. (B) Ejemplo de datos de electroforesis en gel que muestra escisión por meganucleasa DJ1 de sustratos plásmidos descritos en (A). La flecha "no cortado" indica el sustrato plásmido linearizado con XmnI. Las flechas "cortado" indican sustratos plásmido linearizado con XmnI que también se han escindido con éxito por DJ1.

35 Figura 3. (A) Diagrama esquemático de un ADN-T que estaba integrado establemente en el genoma de *Arabidopsis thaliana* como se describe en el Ejemplo 1. En esta construcción de ADN-T, un gen optimizado con codones que codifica la meganucleasa BRP2 modificada por ingeniería genética (BRP2) (SEQ ID NO: 8) está bajo el control de un promotor Hsp70 (HSP) y un terminador NOS (TERM). Un par de secuencias de reconocimiento de BRP2 potenciales (Sitio1, Sitio2) están alojadas adyacentes al terminador separado por 7 pares de bases que contienen un sitio de enzima de restricción PstI (PstI). Un marcador de resistencia a kanamicina (Kan) también está alojado en el ADN-T para permitir la selección para transformantes estables. (B) El producto esperado después de escisión por meganucleasa BRP2 del Sitio 1 y el Sitio 2 que muestra la pérdida del fragmento de 7 pares de bases intermedio y el sitio de restricción de PstI. Las flechas muestran la localización de los cebadores de PCR usados para investigar la escisión del ADN-T. (C) Secuencias de las secuencias de reconocimiento de BRP2 alojadas en o bien la construcción GTAC (GTAC) o la construcción TAGA (TAGA) con las secuencias centrales subrayadas. (D) Ejemplo de los datos de la electroforesis de una planta transformada con la construcción GTAC. El ADN genómico se aisló de las hojas de plantones de *Arabidopsis* establemente transformados con cada construcción de ADN-T GTAC antes y después de 2 horas del "choque térmico" para inducir la expresión de BRP2. A continuación, las muestras de ADN se añadieron a reacciones de PCR usando los cebadores mostrados en (B). Las reacciones de PCR se digirieron con PstI y se visualizaron por electroforesis en gel. C: pista control que carece de PstI. 44, 45 y 46: muestras de PCR de tres plantas representativas que muestran digestión casi completa por PstI en muestras tomadas antes del choque térmico (-pistas) y muy poca digestión por PstI en muestras tomadas después del choque térmico (+pistas). Estos resultados indican que la meganucleasa BRP2 era capaz de escindir la secuencia de reconocimiento de BRP2 que incorporaba una secuencia central GTAC *in vivo*.

50 Figura 4. (A) Diagrama esquemático de un ADN-T que estaba establemente integrado en el genoma de *Arabidopsis thaliana* como se describe en el Ejemplo 2. En esta construcción de ADN-T, un gen optimizado con codones que codifica la meganucleasa BRP12-SC (BRP12-SC) (SEQ ID NO: 15) está bajo el control de un promotor Hsp70 (HSP) y un terminador NOS (TERM). Un par de secuencias de reconocimiento de BRP12-SC potenciales (Sitio 1, Sitio 2) están alojadas adyacentes al terminador separado por 7 pares de bases que contienen un sitio de enzima de restricción PstI (PstI). Un marcador de resistencia a kanamicina (Kan) también está alojado en el ADN-T para permitir la selección para transformantes estables. (B) El producto esperado después de la escisión por meganucleasa BRP12-SC del Sitio 1 y el Sitio 2 que muestra pérdida del fragmento de 7 pares de bases intermedio y el sitio de restricción PstI. Las flechas muestran la localización de los

cebadores de PCR usados para investigar la escisión del ADN-T. (C) Secuencias de las secuencias de reconocimiento de BRP12-SC alojadas en o bien la construcción GTAC (GTAC) o la construcción de TAGA (TAGA) con las secuencias centrales subrayadas.

5 Figura 5. Representación gráfica de los efectos de la concentración de meganucleasa y la secuencia central sobre la escisión de meganucleasa *in vitro*. La meganucleasa BRP2 (SEQ ID NO: 8, véase el Ejemplo 1) se añadió a la concentración indicada a una reacción de digestión que contenía 0,25 picomoles de un sustrato plásmido que portaba o bien una secuencia de reconocimiento de BRP2 con la secuencia central GTAC o una  
10 secuencia de reconocimiento de BRP2 con la secuencia central TAGA. Las reacciones eran 25 microlitros en tampón SA (Tris-HCL 25 mM, pH 8,0, NaCl 100 mM, MgCl<sub>2</sub> 5 mM, EDTA 5 mM). Las reacciones se incubaron a 37 °C durante 2 horas y, a continuación, se visualizaron por electroforesis en gel y el porcentaje de sustrato plásmido escindido por la meganucleasa se trazó como función de la concentración de la meganucleasa.

## Descripción detallada de la invención

### 15 1.1 Introducción

La presente invención se basa, en parte, en la identificación y caracterización de secuencias de ADN particulares que se escinden más eficazmente por las meganucleasas derivadas de I-Crel racionalmente diseñadas. Específicamente, la invención se basa en el descubrimiento de que ciertas secuencias de ADN de cuatro pares de bases, cuando se incorporan como los cuatro pares de bases centrales de una secuencia de reconocimiento de meganucleasa derivada de I-Crel, racionalmente diseñada, puede impactar significativamente en la escisión por la correspondiente meganucleasa, aunque la meganucleasa no parezca que, basándose en el análisis de estructuras de cristal, contacte los cuatro pares de bases centrales. Ya que hay cuatro bases de ADN (A, C, G, y T), hay 4<sup>4</sup> o 256 posibles secuencias de ADN que sean de cuatro pares de bases de longitud. Todas estas posibles secuencias se examinaron para determinar los subconjuntos de secuencias que se escinden más eficazmente por meganucleasas derivadas de I-Crel. Los resultados de este análisis permiten predicción más exacta de si un sitio de ADN de doble cadena particular de 22 pares de bases de longitud se puede o no escindir más eficazmente por la meganucleasa derivada de I-Crel.

### 30 1.2 Referencias y Definiciones

La patente y la bibliografía científica referida en el presente documento establecen el conocimiento que está disponible para los expertos en la técnica. Las patentes americanas expedidas, solicitudes permitidas, solicitudes extranjeras publicadas, y referencias, incluyendo secuencias de la base de datos GenBank, que se citan en el presente documento se incorporan por la presente por referencia al mismo alcance que si cada uno se indicara específicamente e individualmente a incorporar por referencia.

Tal como se usa en el presente documento, el término "meganucleasa derivada de I-Crel" se refiere a una meganucleasa racionalmente diseñada (es decir, modificada por ingeniería genética) que se deriva de I-Crel. El término meganucleasa modificada por ingeniería genética, como se usa en el presente documento, se refiere a una variante recombinante de una endonucleasa de asentamiento I-Crel que se ha modificado por una o más inserciones, deleciones o sustituciones de aminoácidos que afectan una o más de las propiedades de especificidad de unión a ADN, actividad de escisión de ADN, afinidad de unión a ADN y/o dimerización. Algunas meganucleasas modificadas por ingeniería genética son conocidas en la técnica (véase, por ejemplo, Porteus y col. (2005), *Nat. Biotechnol.* 23:967-73; Sussman y col. (2004), *J. Mol. Biol.* 342:31-41; Epinat y col. (2003), *Nucleic Acids Res.* 31:2.952-62) y un método general para diseñar racionalmente tales variantes se describe en el documento WO 2007/047859. Métodos adicionales para modificar por ingeniería genética tales variantes se describen en los documentos WO 04/067736, WO 07/060495, WO 06/097853, WO 07/049095, WO 08/102198, WO 08/010093, WO 08/010009, WO 07/093918, WO 07/093836, WO 08/102274, WO 08/059317, WO 09/013622, WO 09/019614, WO 09/019528, WO 08/152523, WO 04/067753, WO 03/078619, WO 06/097784, WO 07/034262, WO 07/049156, WO 07/057781, WO 08/093152, WO 08/102199, WO 08/102274, WO 08/149176, WO 09/013559, WO 09/013622, y WO 09/019528.

Una meganucleasa puede unirse a ADN de doble cadena como homodímero, como es el caso para I-Crel de tipo natural, o puede unirse a ADN como heterodímero. Una meganucleasa también puede ser un "heterodímero de cadena sencilla" en el que un par de dominios de unión a ADN derivados de I-Crel se juntan dentro de un polipéptido sencillo usando un conector peptídico. El término "endonucleasa de asentamiento" es sinónimo del término "meganucleasa".

60 Tal como se usa en el presente documento, el término "racionalmente diseñado" significa que no es de origen natural y/o que está modificado por ingeniería genética. Las meganucleasas racionalmente diseñadas de la invención difieren de las meganucleasas de tipo natural o de origen natural en su secuencia de aminoácidos o estructura primaria, y también pueden diferir en su estructura secundaria, terciaria o cuaternaria. Además, las meganucleasas racionalmente diseñadas de la invención también difieren de las meganucleasas de tipo natural o de  
65 origen natural en la especificidad y/o actividad de la secuencia de reconocimiento.

Tal como se usa en el presente documento, con respecto a una proteína, el término “recombinante” significa que tiene una secuencia de aminoácidos alterada como resultado de la aplicación de técnicas de ingeniería genética a ácidos nucleicos que codifican la proteína, y células u organismos que expresan la proteína. Con respecto a un ácido nucleico, el término “recombinante” significa que tiene una secuencia de ácidos nucleicos alterada como resultado de la aplicación de técnicas de ingeniería genética. Las técnicas de ingeniería genética incluyen, pero no se limitan a, técnicas de PCR y de clonación de ADN; transfección, transformación y otras tecnologías de transferencia de gen; recombinación homóloga; mutagénesis dirigida a sitio; y fusión de gen. De acuerdo con esta definición, una proteína que tiene una secuencia de aminoácidos idéntica a una proteína de origen natural, pero producida por clonación y expresión en un hospedador heterólogo, no se considera recombinante.

Tal como se usa en el presente documento, el término “genéticamente modificado” se refiere a una célula u organismo en el cual, o en un predecesor del cual, una secuencia de ADN genómico se ha modificado deliberadamente por tecnología recombinante. Tal como se usa en el presente documento, el término “genéticamente modificado” abarca el término “transgénico”.

Tal como se usa en el presente documento, el término “tipo natural” se refiere a cualquier forma de origen natural de una meganucleasa. El término “tipo natural” no pretende significar la variante alélica más común de la enzima en la naturaleza si no, más bien, cualquier variante alélica de origen natural. Las meganucleasas de tipo natural se distinguen de las meganucleasas recombinantes o que no son de origen natural.

Tal como se usa en el presente documento, el término “medio-sitio de la secuencia de reconocimiento” o simplemente “medio-sitio” significa una secuencia de ADN de 9 pares de bases que se reconoce por un monómero de meganucleasa, en el caso de una meganucleasa dimérica, o por un dominio de una meganucleasa de cadena sencilla.

Tal como se usa en el presente documento, el término “secuencia de reconocimiento” se refiere a un par de medio-sitios que se unen y se escinden por una meganucleasa. Una secuencia de reconocimiento comprende un par de medio-sitios de 9 pares de bases invertidos separados por cuatro pares de bases. Por lo tanto, la secuencia de reconocimiento es de 22 pares de bases de longitud. Los pares de bases de cada medio-sitio se denominan -9 a -1, siendo la posición -9 la más distal del sitio de escisión y siendo la posición -1 adyacente a la secuencia central de 4 pares de bases, cuyos pares de bases se denominan +1 a +4. La cadena de cada medio-sitio que se orienta 5' a 3' en la dirección desde -9 a -1 (es decir, hacia el sitio de escisión), se denomina la cadena “sentido”, y la cadena opuesta se denomina la “cadena antisentido”, aunque ninguna cadena puede codificar proteína. Por tanto, la cadena “sentido” de un medio-sitio es la cadena antisentido (opuesta) del otro medio-sitio. Véase, por ejemplo, la Figura 1a.

Tal como se usa en el presente documento, el término “secuencia central” se refiere a los cuatro pares de bases que separan medio-sitios en la secuencia de reconocimiento de meganucleasa. Estas bases se numeran +1 a +4 en la Figura 1a. La secuencia central comprende las cuatro bases que llegan a ser los salientes de cadena sencilla 3' después de escisión por meganucleasa. “Secuencia central” puede referirse a la secuencia de la cadena sentido o la cadena antisentido (opuesta).

Tal como se usa en el presente documento, el término “especificidad” se refiere a la capacidad de una meganucleasa de reconocer y escindir moléculas de ADN de doble cadena solamente en un subconjunto particular de todas las posibles secuencias de reconocimiento. El conjunto de secuencias de reconocimiento compartirá ciertas posiciones o motivos de secuencia conservados, pero pueden degenerarse en una o más posiciones. Una meganucleasa más específica es capaz de unir y escindir un subconjunto más pequeño de las posibles secuencias de reconocimiento, mientras que una meganucleasa menos específica es capaz de unir y escindir un subconjunto mayor de las posibles secuencias de reconocimiento.

Tal como se usa en el presente documento, el término “palindrómico” se refiere a una secuencia de reconocimiento que consiste en repeticiones invertidas de medio-sitios idénticos. Sin embargo, en este caso, la secuencia palindrómica no necesita ser palindrómica con respecto a la secuencia central, que no está contactada por la enzima. En el caso de meganucleasas diméricas, las secuencias de ADN palindrómicas se reconocen por homodímeros en los cuales los dos monómeros hacen contactos con medio-sitios idénticos.

Tal como se usa en el presente documento, el término “pseudopalindrómico” se refiere a una secuencia de reconocimiento que consiste en repeticiones invertidas de medio-sitios no idénticos o imperfectamente palindrómicos. En este caso, la secuencia pseudopalindrómica no necesita ser palindrómica con respecto a la secuencia central, y también puede desviarse de una secuencia perfectamente palindrómica entre los dos medio-sitios. Las secuencias de ADN pseudopalindrómicas son típicas de los sitios de ADN naturales reconocidos por meganucleasas homodiméricas de tipo natural en las cuales dos monómeros enzimáticos idénticos hacen contactos con diferentes medio-sitios.

Tal como se usa en el presente documento, el término “no palindrómico” se refiere a una secuencia de reconocimiento compuesta de dos medio-sitios no relacionados de una meganucleasa. En este caso, la secuencia no palindrómica no necesita ser palindrómica ni respecto a la secuencia central ni a los dos medio-sitios de

monómero. Las secuencias de ADN no palindrómicas se reconocen por o bien meganucleasas heterodiméricas o meganucleasas de cadena sencilla que comprenden un par de dominios que reconocen medio-sitios no idénticos.

5 Tal como se usa en el presente documento, el término “actividad” se refiere al índice al cual una meganucleasa de la invención escinde una secuencia de reconocimiento particular. Tal actividad es una reacción enzimática medible, que implica la hidrólisis de enlaces fosfodiéster de ADN de doble cadena. La actividad de una meganucleasa que actúa sobre un sustrato de ADN particular está afectada por la afinidad o la avidez de la meganucleasa para ese sustrato de ADN particular que, por turno, es afectado por interacciones tanto específicas a secuencia como no específicas a secuencia con el ADN.

10 Tal como se usa en el presente documento, el término “recombinación homóloga” se refiere al proceso celular natural en el cual se repara una rotura de ADN de doble cadena usando una secuencia de ADN homóloga como molde de reparación (véase, por ejemplo, Cahill y col. (2006), *Front. Biosci.* 11:1.958-1.976). La secuencia de ADN homóloga puede ser una secuencia cromosómica endógena o un ácido nucleico exógeno que se repartió a la célula. 15 Por tanto, para algunas aplicaciones, una meganucleasa se usa para escindir una secuencia de reconocimiento dentro de una secuencia diana en un genoma y un ácido nucleico exógeno con homología a o similitud de secuencia sustancial con la secuencia diana se reparte dentro de la célula y se usa como un molde para reparar por recombinación homóloga. La secuencia de ADN del ácido nucleico exógeno, que puede diferir significativamente de la secuencia diana, se incorpora de este modo dentro de la secuencia cromosómica. El proceso de la recombinación 20 homóloga se da principalmente en organismos eucariotas. El término “homología” se usa en el presente documento como equivalente a “similitud de secuencia” y no se pretende requerir identidad por ascendencia o relación filogenética.

25 Tal como se usa en el presente documento, el término “unión a extremo no homólogo” se refiere al proceso celular natural en el cual una rotura de ADN de doble cadena se repara por la unión directa de dos segmentos de ADN no homólogos (véase, por ejemplo, Cahill y col. (2006), *Front. Biosci.* 11:1.958-1.976). La reparación de ADN por unión a extremo no homólogo es propensa a error y frecuentemente da como resultado la adición sin molde o delección de secuencias de ADN en el sitio de reparación. Por tanto, para algunas aplicaciones, una meganucleasa se puede usar para producir una rotura de la doble cadena en una secuencia de reconocimiento de meganucleasa dentro de 30 una secuencia diana en un genoma para alterar un gen (por ejemplo, introduciendo inserciones de bases, delecciones de bases o mutaciones por desplazamiento del marco de lectura) por la unión a extremo no homólogo. Para otras aplicaciones, un ácido nucleico exógeno que carece de homología a o similitud sustancial de secuencia con la secuencia diana se puede capturar en el sitio de una rotura del ADN de doble cadena estimulada por meganucleasa mediante la unión a extremo no homólogo (véase, por ejemplo, Salomon y col. (1998), *EMBO J.* 17:6.086-6.095). El proceso de unión a extremo no homólogo se da en tanto eucariotas como procariotas tales como bacterias.

35 Tal como se usa en el presente documento, el término “secuencia de interés” significa cualquier secuencia de ácidos nucleicos, si codifica una proteína, ARN, o elemento regulador (por ejemplo, una secuencia potenciadora, silenciadora o promotora), que se puede insertar en un genoma o usar para reemplazar una secuencia de ADN genómico usando una proteína de meganucleasa. Las secuencias de interés pueden tener secuencias de ADN heterólogas que permiten etiquetar una proteína o ARN que se expresa a partir de la secuencia de interés. Por ejemplo, una proteína se puede etiquetar con etiquetas que incluyen, pero no se limitan a, un epítopo (por ejemplo, c-myc, FLAG) u otro ligando (por ejemplo, poly-His). Además, una secuencia de interés puede codificar una proteína 45 de fusión, según técnicas conocidas en la técnica (véase, por ejemplo, Ausubel y col., *Current Protocols in Molecular Biology*, Wiley 1999). Para algunas aplicaciones, la secuencia de interés está flanqueada por una secuencia de ADN que se reconoce por la meganucleasa por escisión. Por tanto, las secuencias flanqueadoras se escinden permitiendo apropiada inserción de la secuencia de interés dentro de las secuencias de reconocimiento genómico escindidas por una meganucleasa. Para algunas aplicaciones, la secuencia entera de interés es homóloga a o tiene similitud de secuencia sustancial con una secuencia diana en el genoma de modo que la recombinación homóloga reemplaza eficazmente la secuencia diana con la secuencia de interés. Para otras aplicaciones, la secuencia de interés está flanqueada por secuencias de ADN con homología a o similitud de secuencia sustancial con la 50 secuencia diana de modo que la recombinación homóloga inserta la secuencia de interés dentro del genoma en el locus de la secuencia diana. Para algunas aplicaciones, la secuencia de interés es sustancialmente idéntica a la secuencia diana excepto para mutaciones u otras modificaciones en la secuencia de reconocimiento de meganucleasa de modo que la meganucleasa no puede escindir la secuencia diana después de haber sido modificada por la secuencia de interés.

55 Tal como se usa en el presente documento, el término “meganucleasa de cadena sencilla” se refiere a un polipéptido que comprende un par de subunidades de meganucleasa juntas por un conector. Una meganucleasa de cadena sencilla tiene la organización: subunidad N-terminal – conector – subunidad C-terminal. Las dos subunidades de meganucleasa, cada una de las cuales está derivada de I-Crel, generalmente serán no idénticas en secuencia de aminoácidos y reconocerán medio-sitios no idénticos. Por tanto, las meganucleasas de cadena sencilla generalmente escinden secuencias de reconocimiento pseudopalindrómicas o no palindrómicas. Una meganucleasa 60 de cadena sencilla se puede referir como un “heterodímero de cadena sencilla” o “meganucleasa heterodimérica de cadena sencilla” aunque no sea, en realidad, dimérica.

Tal como se usa en el presente documento, a menos que específicamente se indique lo contrario, la palabra “o” se usa en el sentido inclusivo de “y/o” y no en el sentido exclusivo de “o bien/o”.

## 2.1 Secuencias centrales preferidas para meganucleasas derivadas de I-Crel

5 La presente invención se basa, en parte, en la identificación de subconjuntos de las posibles secuencias centrales de cuatro pares de bases que son preferidas por meganucleasas derivadas de I-Crel. Ya que la enzima de tipo natural no hace contactos significantes con las bases en la secuencia central, las mismas preferencias de secuencia central de la nucleasa de asentamiento de I-Crel de tipo natural se aplican a meganucleasas derivadas de I-Crel racionalmente diseñadas que se han rediseñado con respecto a, por ejemplo, preferencia a medio-sitio, afinidad de unión a ADN, y/o capacidad de heterodimerización. Por lo tanto, esta invención proporciona criterios importantes que se pueden considerar en la determinación de si una secuencia de ADN de 22 pares de bases particulares es o no una secuencia de reconocimiento de meganucleasa derivada de I-Crel adecuada.

15 El conjunto preferido de secuencias centrales se determinó usando una meganucleasa modificada por ingeniería genética llamada “DJ1” (SEQ ID NO: 4). La producción de esta meganucleasa se describe en el documento WO 2007/047859. DJ1 es una meganucleasa derivada de I-Crel homodimérica que está diseñada para reconocer una secuencia de reconocimiento de meganucleasa palindrómica (SEQ ID NO: 5, SEQ ID NO: 6) que difiere en 4 posiciones por medio-sitio en relación a I-Crel de tipo natural. Este cambio en la especificidad de medio-sitio se consigue mediante la introducción de 6 sustituciones de aminoácido a I-Crel de tipo natural (K28D, N30R, S32N, Q38E, S40R, y T42R).

25 Para ensayar la actividad de escisión con respecto a diversas secuencias de reconocimiento, se expresó DJ1 en *E. coli* y se purificó como se describe en el Ejemplo 1 del documento WO 2007/047859. A continuación, se añadieron 25 picomoles de proteína meganucleasa purificada a una solución 10 nM de sustrato de ADN plásmido en tampón SA (Tris-HCL 25 mM, pH 8.0, NaCl 100 mM, MgCl<sub>2</sub> 5 mM, EDTA 5mM) en una reacción de 25 microlitros. Se añadió 1 microlitro de enzima de restricción XmnI para linear los sustratos plásmidos. Las reacciones se incubaron a 37 °C durante 4 horas y, a continuación, se visualizaron por electroforesis en gel para determinar el grado al cual cada uno se escindía por la meganucleasa DJ1.

30 Los sustratos plásmidos usados en estos experimentos comprendían un plásmido pUC-19 en el cual se insertó una secuencia de reconocimiento de meganucleasa potencial en el sitio policonector (sitio SmaI). Cada sitio de reconocimiento de meganucleasa potencial comprendía un par de medio-sitios DJ1 invertidos (SEQ ID NO: 7) separados por una secuencia central diferente. Por tanto, evaluando la escisión por DJ1 de los sustratos de ADN múltiples que solamente diferían por la secuencia central, era posible determinar qué secuencias centrales eran las más susceptibles de escisión por meganucleasa (Figura 2).

40 Inicialmente, solamente se evaluó la influencia de las bases N<sub>+2</sub> y N<sub>+3</sub>. La estructura cristalina por rayos X de I-Crel en el complejo con su sitio de ADN natural muestra que el ADN se deforma en estos dos pares de bases centrales (Jurica y col., (1998), *Mol. Cell.* 2:469-76). El modelo realizado por ordenador sugiere que una purina (G o A) en N<sub>+2</sub> es incompatible con una pirimidina (C o T) en N<sub>+3</sub>. Esto es debido a que la deformación introducida por unión a I-Crel causa un choque estérico entre una base de purina en N<sub>+2</sub> y una segunda base de purina emparejada a una pirimidina en N<sub>+3</sub>. Esta incompatibilidad esperada se verificó experimentalmente incubando la proteína DJ1 con sustratos plásmidos que portaban sitios de reconocimiento de meganucleasa con todas las secuencias centrales posibles de la forma A<sub>+1</sub>X<sub>+2</sub>X<sub>+3</sub>T<sub>+4</sub> en la que X es cualquier base. Los resultados se resumen en la Tabla 1. Para las tablas 1 a 5, “Actividad” se refiere a lo siguiente:

- 50 - : no escisión en 4 horas  
 + : 1 % a 25 % de escisión en 4 horas  
 ++ : 26 % a 75 % de escisión en 4 horas  
 +++ : 75 % a 100 % de escisión en 4 horas

**Tabla 1. El efecto de los cambios en N<sub>+2</sub> y N<sub>+3</sub>**

Nº Sec.	N <sub>+1</sub>	N <sub>+2</sub>	N <sub>+3</sub>	N <sub>+4</sub>	Actividad
1	A	A	A	T	+
2	A	A	C	T	-
3	A	A	G	T	+
4	A	A	T	T	-
5	A	C	A	T	+
6	A	C	C	T	+
7	A	C	G	T	+
8	A	C	T	T	+

ES 2 625 941 T3

9	A	G	A	T	+
10	A	G	C	T	-
11	A	G	G	T	+
12	A	G	T	T	-
13	A	T	A	T	++
14	A	T	C	T	+
15	A	T	G	T	++
16	A	T	T	T	+

Coherente con el modelo realizado por ordenador, se encontró que los cuatro sustratos plásmidos con una base de purina en  $N_{+2}$  y una base de pirimidina en  $N_{+3}$  (números de secuencia 2, 4, 10 y 12) no se cortaron eficazmente por DJ1.

5 A continuación, se realizó una evaluación más completa de la preferencia de secuencia central. Hay  $4^4$  o 256 posibles secuencias centrales. De estas, 25 %, o 64, tienen una base de purina en  $N_{+2}$  y pirimidina en  $N_{+3}$  y, por lo tanto, se eliminaron como secuencias centrales basándose en el experimento anteriormente descrito. De las restantes 192, 92 son redundantes debido a que las meganucleasas son simétricas y reconocen bases igualmente sobre tanto la cadena sentido como antisentido. Por ejemplo, la secuencia  $A_{+1}A_{+2}A_{+3}A_{+4}$  en la cadena sentido es reconocida por la meganucleasa como  $T_{+1}T_{+2}T_{+3}T_{+4}$  en la cadena antisentido y, por tanto,  $A_{+1}A_{+2}A_{+3}A_{+4}$  y  $T_{+1}T_{+2}T_{+3}T_{+4}$  son funcionalmente equivalentes. Teniendo en cuenta estas redundancias, así como los conflictos  $N_{+2}/N_{+3}$  anteriormente mencionados, quedaban 100 posibles secuencias centrales. Para determinar cuáles de estas eran preferidas por las meganucleasas, los inventores produjeron 100 sustratos plásmidos que portaban estas 100 secuencias centrales flanqueadas por medio-sitios de reconocimiento invertidos para la meganucleasa DJ1. A continuación, DJ1 se incubó con cada uno de los 100 plásmidos y se evaluó la actividad de escisión como se describió anteriormente. Estos resultados se resumen en la Tabla 2.

**Tabla 2. Secuencias centrales escindibles**

Nº Sec	$N_{+1}$	$N_{+2}$	$N_{+3}$	$N_{+4}$	Actividad
1	T	T	T	T	+
2	T	T	G	T	++
3	T	T	C	T	+
4	T	T	A	T	++
5	T	G	G	T	+
6	T	G	A	T	+
7	T	C	T	T	++
8	T	C	G	T	++
9	T	C	C	T	+
10	T	C	A	T	++
11	T	A	G	T	+
12	T	A	A	T	+
13	G	T	T	T	++
14	G	T	G	T	+++
15	G	T	C	T	++
16	G	T	A	T	+++
17	G	G	G	T	+
18	G	G	A	T	++
19	G	A	G	T	++
20	G	A	A	T	++
21	C	T	T	T	+
22	C	T	G	T	+
23	C	T	C	T	+
24	C	T	A	T	+
25	C	G	G	T	+
26	C	G	A	T	+

ES 2 625 941 T3

27	C	C	T	T	+
28	C	C	G	T	+
29	C	C	C	T	+
30	C	C	A	T	+
31	C	A	G	T	+
32	C	A	A	T	+
33	A	T	T	T	+
34	A	T	G	T	++
35	A	T	C	T	+
36	A	G	G	T	+
37	A	C	T	T	+
38	T	T	T	G	+
39	T	T	G	G	+
40	T	T	C	G	+
41	T	T	A	G	+++
42	T	G	G	G	+
43	T	G	A	G	+
44	T	C	T	G	+
45	T	C	G	G	+
46	T	C	C	G	+
47	T	C	A	G	+
48	T	A	G	G	+
49	T	A	A	G	+
50	G	T	T	G	+
51	G	T	G	G	+
52	G	T	C	G	+
53	G	T	A	G	+++
54	G	G	G	G	+
55	G	G	A	G	+
56	G	A	G	G	+
57	G	A	A	G	+
58	C	T	T	G	+
59	C	T	G	G	+
60	C	T	C	G	+
61	C	G	G	G	+
62	C	C	T	G	+
63	T	T	T	C	++
64	T	T	C	C	++
65	T	T	A	C	+++
66	T	G	A	C	++
67	T	C	T	C	+++
68	T	C	C	C	+
69	T	C	A	C	+++
70	T	A	A	C	++
71	G	T	T	C	++
72	G	T	C	C	+++
73	T	T	T	A	+
74	T	T	G	A	+
75	T	T	C	A	+

76	T	G	G	A	+
77	T	C	T	A	+
78	A	T	A	T	++
79	A	C	G	T	+
80	C	T	A	G	+
81	C	C	G	G	+
82	G	T	A	C	+++
83	T	C	G	A	++
84	T	T	A	A	++
85	T	C	G	C	+++
86	A	A	G	C	+++
87	G	A	G	C	+++
88	G	C	G	C	+++
89	G	G	G	C	++
90	G	T	G	C	+++
91	T	A	G	C	+++
92	T	G	G	C	+
93	T	T	G	C	+++
94	A	C	G	C	++
95	A	G	G	C	+
96	A	T	G	C	+++
97	C	A	G	C	+
98	C	C	G	C	++
99	C	G	G	C	+
100	C	T	G	C	++

Para mayor claridad, cada una de las secuencias centrales enumeradas en la Tabla 2 es equivalente a su secuencia de cadena opuesta debido al hecho de que la meganucleasa I-Crel une su secuencia de reconocimiento como un homodímero simétrico. Por tanto, la secuencia número 100 en la Tabla 2,  $C_{+1}T_{+2}G_{+3}C_{+4}$ , es equivalente a su secuencia de cadena opuesta,  $G_{+1}C_{+2}A_{+3}G_{+4}$ . A partir de estos datos, surge un conjunto general de reglas de preferencia de secuencia central. Estas reglas, que no se suponen que sustituyen la Tabla 1 o la Tabla 2, incluyen:

1. Secuencias centrales con una base de purina en  $N_{+2}$  y una base de pirimidina en  $N_{+3}$  cortadas muy escasamente, como mucho.
2. G se prefiere en  $N_{+1}$ . Esto es equivalente a C en  $N_{+4}$ . Todas las secuencias centrales más preferidas tienen G en  $N_{+1}$  y/o C en  $N_{+4}$ .
3. C se prefiere en  $N_{+2}$ . Esto es equivalente a G en  $N_{+3}$ .
4. Hay una preferencia para las secuencias centrales con una base de pirimidina en  $N_{+2}$  y una base de purina en  $N_{+3}$ .
5. Hay una preferencia para las secuencias con al menos 1 par de bases A-T en la secuencia central.

Por tanto, en general, las secuencias centrales preferidas tienen la forma  $G_{+1}Y_{+2}R_{+3}X_{+4}$  en la que Y es una pirimidina (C o T), R es una purina (A o G), y X es cualquier base (A, C, G, o T).

## 2.2 Aplicaciones *in vitro* usando secuencias centrales preferidas

Las meganucleasas modificadas por ingeniería genética tienen numerosas aplicaciones *in vitro* potenciales incluyendo el mapeo de restricción y la clonación. Estas aplicaciones son conocidas en la técnica y se comentan en el documento WO 2007/047859.

Una ventaja de usar las meganucleasas modificadas por ingeniería genética más que las enzimas de restricción convencionales para aplicaciones tales como clonación es la posibilidad de cortar el ADN para dejar un amplio intervalo de diferentes salientes 3' ("extremos adhesivos") que sean compatibles con, por ejemplo, los salientes 3' producidos por escisión de un vector particular de interés. Por tanto, hay ocasiones cuando es deseable escindir una secuencia de reconocimiento de meganucleasa con una secuencia central subóptima para crear un saliente deseado.

Debido a que las condiciones de escisión de ADN *in vitro* son, en general, menos estrictas que las condiciones *in vivo*, el uso de secuencias centrales subóptimas puede ser aceptable para tales aplicaciones. Por ejemplo, en relación a aplicaciones *in vivo*, digestiones *in vitro* usando meganucleasas modificadas por ingeniería se pueden realizar a una relación superior de meganucleasa y ADN, generalmente hay menos ADN no específico (genómico) que compita por la meganucleasa, y las condiciones de solución se pueden optimizar para favorecer la escisión por meganucleasa (por ejemplo, usando tampón SA como se describió anteriormente). Por tanto, un mayor número de secuencias centrales son adecuadas para aplicaciones *in vitro* que para aplicaciones *in vivo*. Todas las secuencias centrales enumeradas en la Tabla 2 son adecuadas para aplicaciones *in vitro*, pero las secuencias centrales preferidas y las más preferidas para las aplicaciones *in vitro* se enumeran en la Tabla 3 y la Tabla 4, respectivamente, con sus secuencias de cadena opuesta.

**Tabla 3. Secuencias centrales preferidas para aplicaciones *in vitro***

Nº Sec.	$N_1N_2N_3N_4$	Secuencia de cadena opuesta
1	TTGT	ACAA
2	TTAT	ATAA
3	TCTT	AAGA
4	TCGT	ACGA
5	TCAT	ATGA
6	GTTT	AAAC
7	GTCT	AGAC
8	GGAT	ATCC
9	GAGT	ACTC
10	GAAT	ATTC
11	ATGT	ACAT
12	TTTC	GAAA
13	TTCC	GGAA
14	TGAC	GTCA
15	TAAC	GTTA
16	GTTC	GAAC
17	ATAT	ATAT
18	TCGA	TCGA
19	TTAA	TTAA
20	GGGC	GCCC
21	ACGC	GCGT
22	CCGC	GCGG
23	CTGC	GCAG

**Tabla 4. Las secuencias centrales más preferidas para aplicaciones *in vitro***

Nº Sec.	$N_1N_2N_3N_4$	Secuencia de cadena opuesta
1	GTGT	ACAC
2	GTAT	ATAC
3	TTAG	CTAA
4	GTAG	CTAC
5	TTAC	GTAA
6	TCTC	GAGA
7	TCAC	GTGA
8	GTCC	GGAC
9	GTAC	GTAC
10	TCGC	GCGA
11	AAGC	GCTT
12	GAGC	GCTC
13	GCGC	GCGC

14	GTGC	GCAC
15	TAGC	GCTA
16	TTGC	GCAA
17	ATGC	GCAT

Obviamente, no cada secuencia de ADN de 22 pares de bases que tiene una secuencia central preferida o la más preferida es capaz de ser una secuencia de reconocimiento de meganucleasa *in vitro*. La secuencia de los medio-sitios que flanquean la secuencia central también debe ser susceptible de reconocimiento por meganucleasa y escisión. Los métodos para modificar por ingeniería una meganucleasa incluyendo I-Crel, para reconocer un medio-sitio predeterminado son conocidos en la técnica (véase, por ejemplo, el documento WO 2007/047859). Por tanto, una secuencia de reconocimiento de meganucleasa derivada de I-Crel preferida para aplicaciones *in vitro* comprenderá: (1) un primer medio-sitio de 9 pares de bases susceptible de reconocimiento por un monómero de meganucleasa (o un primer dominio de una meganucleasa de cadena sencilla); (2) una secuencia central preferida o la más preferida de la Tabla 2 o Tabla 3; y (3) un segundo medio-sitio de 9 pares de bases susceptible de reconocimiento por un monómero de meganucleasa (o un segundo dominio de una meganucleasa de cadena sencilla) en dirección inversa en relación al primer medio-sitio.

Por tanto, en un aspecto, la invención proporciona métodos para escindir un ADN de doble cadena *in vitro* (a) identificando al menos un sitio de reconocimiento potencial para al menos una meganucleasa derivada de I-Crel dentro del ADN, en los que el sitio de reconocimiento potencial tiene una secuencia central de cuatro pares de bases seleccionada entre el grupo de secuencias centrales de la Tabla 2; (b) identificando una meganucleasa derivada de I-Crel que reconoce ese sitio de reconocimiento en el ADN; y (c) poniendo en contacto la meganucleasa derivada de I-Crel con el ADN; por lo cual la meganucleasa I-Crel escinde el ADN.

En otro aspecto, la invención proporciona métodos para escindir un ADN de doble cadena *in vitro* (a) introduciendo en el ADN un sitio de reconocimiento para una meganucleasa derivada de I-Crel que tiene una secuencia central de cuatro pares de bases seleccionada entre el grupo que consiste en secuencias centrales de la Tabla 2; y (b) poniendo en contacto la meganucleasa derivada de I-Crel con el ADN; por lo cual la meganucleasa derivada de I-Crel escinde el ADN.

En particular, en algunas realizaciones, el ADN se selecciona entre un producto de PCR; un cromosoma artificial; ADN genómico aislado de bacterias, hongos, plantas o células animales; y ADN vírico.

En algunas realizaciones, el ADN está presente en un ácido nucleico seleccionado entre un plásmido, un profago y un cromosoma.

En algunas de las anteriores realizaciones, la secuencia de ADN de cuatro pares de bases se selecciona de la Tabla 3. En otras realizaciones, la secuencia de ADN de cuatro pares de bases se selecciona de la Tabla 4.

En algunas realizaciones, la meganucleasa derivada de I-Crel puede estar específicamente diseñada para su uso con el sitio de reconocimiento elegido en el método.

### 2.3 Aplicaciones *in vivo* usando secuencias centrales preferidas

Aplicaciones tales como la terapia génica, la modificación por ingeniería celular, y la modificación por ingeniería vegetal requieren función de meganucleasa dentro de una célula viva (para mayor claridad, a cualquier aplicación intracelular se referirá como una aplicación "*in vivo*" esté tal célula o no aislada o sea parte de un organismo multicelular). Estas aplicaciones son conocidas en la técnica y están descritas en, por ejemplo, el documento WO 2007/047859. Las aplicaciones *in vivo* significativamente se restringen en relación a las aplicaciones *in vitro* con respecto a la secuencia central. Esto es debido a que las condiciones intracelulares no se pueden manipular de ninguna manera para favorecer la actividad de la meganucleasa y/o debido a que grandes cantidades de ADN genómico compiten por la unión a meganucleasa. Por tanto, solamente se prefieren las secuencias de reconocimiento de meganucleasa con secuencias centrales óptimas para las aplicaciones *in vivo*. Tales secuencias se enumeran en la Tabla 5 con sus secuencias de cadena opuesta.

**Tabla 5. Secuencias centrales preferidas para aplicaciones *in vivo*.**

Nº Sec.	N <sub>1</sub> N <sub>2</sub> N <sub>3</sub> N <sub>4</sub>	Secuencia de cadena opuesta
	GTGT	ACAC
2	GTAT	ATAC
3	TTAG	CTAA
4	GTAG	CTAC
5	TTAC	GTAA

6	TCTC	GAGA
7	TCAC	GTGA
8	GTCC	GGAC
9	GTAC	GTAC
10	TCGC	GCGA
11	AAGC	GCTT
12	GAGC	GCTC
13	GCGC	GCGC
14	GTGC	GCAC
15	TAGC	GCTA
16	TTGC	GCAA
17	ATGC	GCAT

Obviamente, no cada secuencia de ADN de 22 pares de bases que tienen una secuencia central preferida es capaz de ser una secuencia de reconocimiento de meganucleasa *in vivo*. La secuencia de los medio-sitios que flanquean la secuencia central también debe ser susceptible de reconocimiento por meganucleasa y escisión. Los métodos para modificar por ingeniería una meganucleasa, incluyendo I-Crel, para reconocer un medio-sitio predeterminado son conocidos en la técnica (véase, por ejemplo, el documento WO 2007/047859). Por tanto, una secuencia de reconocimiento de meganucleasa *in vivo* preferida comprenderá: (1) un primer medio-sitio de 9 pares de bases susceptible de reconocimiento por un monómero de meganucleasa (o un primer dominio de una meganucleasa de cadena sencilla); (2) una secuencia central preferida de la Tabla 5; y (3) un segundo medio-sitio de 9 pares de bases susceptible de reconocimiento por un monómero de meganucleasa (o un segundo dominio de una meganucleasa de cadena sencilla) en dirección inversa en relación al primer medio-sitio.

Por tanto, en un aspecto, la invención proporciona métodos para escindir un ADN de doble cadena *in vivo* (a) identificando al menos un sitio de reconocimiento potencial para al menos una meganucleasa derivada de I-Crel dentro del ADN, en los que el sitio de reconocimiento potencial tiene una secuencia central de cuatro pares de bases seleccionada entre el grupo de secuencias centrales de la Tabla 2; (b) identificando una meganucleasa derivada de I-Crel que reconoce ese sitio de reconocimiento en el ADN; y (c) poniendo en contacto la meganucleasa derivada de I-Crel con el ADN; por lo cual la meganucleasa I-Crel escinde el ADN.

En otro aspecto, la invención proporciona métodos para escindir un ADN de doble cadena *in vivo* (a) introduciendo en el ADN un sitio de reconocimiento para una meganucleasa derivada de I-Crel que tiene una secuencia central de cuatro pares de bases seleccionada entre el grupo que consiste en secuencias centrales de la Tabla 2; y (b) poniendo en contacto la meganucleasa derivada de I-Crel con el ADN; por lo cual la meganucleasa derivada de I-Crel escinde el ADN.

En algunas realizaciones, el ADN está presente en una célula seleccionada entre una célula bacteriana, fúngica, vegetal y animal.

En algunas realizaciones, el ADN está presente en un ácido nucleico seleccionado entre un plásmido, un profago y un cromosoma.

En algunas de las anteriores realizaciones, la secuencia de ADN de cuatro pares de bases se selecciona de la Tabla 3. En otras realizaciones, la secuencia de ADN de cuatro pares de bases se selecciona de la Tabla 4.

En algunas realizaciones, la meganucleasa derivada de I-Crel está específicamente diseñada para su uso con el sitio de reconocimiento elegido en los métodos de la invención.

En alguna de las anteriores realizaciones, el método incluye la etapa adicional de diseñar racionalmente la meganucleasa derivada de I-Crel para reconocer el sitio de reconocimiento elegido. En algunas realizaciones, el método comprende además producir la meganucleasa derivada de I-Crel.

En otro aspecto, la invención proporciona células transformadas con un ácido nucleico que incluye (a) una primera secuencia de ADN de 9 pares de bases que se puede unir por un monómero de meganucleasa derivada de I-Crel o por un primer dominio de una meganucleasa derivada de I-Crel de cadena sencilla; (b) una secuencia de ADN de cuatro pares de bases seleccionada de la Tabla 2; y (c) una segunda secuencia de ADN de 9 pares de bases que se puede unir por un monómero de meganucleasa derivada de I-Crel o por un segundo dominio de una meganucleasa derivada de I-Crel de cadena sencilla; en el que la segunda secuencia de ADN de 9 pares de bases está en dirección inversa en relación a la primera.

En otro aspecto, la invención proporciona una célula que contiene una secuencia de ácidos nucleicos exógena integrada en su genoma, que incluye, en orden: (a) una primera secuencia de ADN de 9 pares de bases exógena que se puede unir por un monómero de meganucleasa derivada de I-Crel o por un primer dominio de una meganucleasa derivada de I-Crel de cadena sencilla; (b) una secuencia de ADN de cuatro pares de bases exógena seleccionada de la Tabla 2; y (c) una segunda secuencia de ADN de 9 pares de bases exógena que se puede unir por un monómero de meganucleasa derivada de I-Crel o por un segundo dominio de una meganucleasa derivada de I-Crel de cadena sencilla; en la que la segunda secuencia de ADN de 9 pares de bases está en dirección inversa en relación a la primera.

En otro aspecto, la invención proporciona una célula que contiene una secuencia de ácidos nucleicos exógena integrada en su genoma, que incluye, en orden: (a) una primera secuencia de ADN de 9 pares de bases exógena que se puede unir por un monómero de meganucleasa derivada de I-Crel o por un primer dominio de una meganucleasa derivada de I-Crel de cadena sencilla; (b) una secuencia de ADN de dos pares de bases exógena, en la que los dos pares de bases corresponden a las bases  $N_{+1}$  y  $N_{+2}$  de una secuencia de ADN de cuatro pares de bases seleccionada de la Tabla 2; (c) una secuencia de ADN exógena que comprende una secuencia codificante que se expresa en la célula; (d) una secuencia de ADN de dos pares de bases exógena, en la que los dos pares de bases corresponden a las bases  $N_{+3}$  y  $N_{+4}$  de una secuencia de ADN de cuatro pares de bases seleccionada de la Tabla 2; y (e) una segunda secuencia de ADN de 9 pares de bases exógena que se puede unir por el monómero de meganucleasa derivada de I-Crel o por un segundo dominio de la meganucleasa derivada de I-Crel de cadena sencilla; en la que la segunda secuencia de ADN de 9 pares de bases está en dirección inversa en relación a la primera.

En algunas realizaciones, el ácido nucleico es un plásmido. En otras realizaciones, el ácido nucleico es un cromosoma artificial. En otras realizaciones, el ácido nucleico está integrado en el ADN genómico de la célula. En otras realizaciones, el ácido nucleico es un ácido nucleico vírico.

En algunas realizaciones, la célula se selecciona entre el grupo de una célula humana, una célula animal no humana, una célula vegetal, una célula bacteriana, y una célula fúngica.

En alguna de las anteriores realizaciones, la secuencia de ADN de cuatro pares de bases se selecciona de la Tabla 3. En otras realizaciones, la secuencia de ADN de cuatro pares de bases se selecciona de la Tabla 4.

En algunas realizaciones, la meganucleasa I-Crel se diseña específicamente para su uso con el sitio de reconocimiento elegido en los métodos de la invención.

#### 2.4 Métodos para dirigir una actividad meganucleasa derivada de I-Crel, diseñada a medida

Una actividad meganucleasa puede dirigirse basándose en meganucleasas derivadas de I-Crel. Por ejemplo, dicha actividad puede actuar de la siguiente manera. La actividad recibe una secuencia de ADN dentro de la cual hay que introducir una rotura de la doble cadena por una meganucleasa derivada de I-Crel racionalmente diseñada. Esta actividad identifica en la secuencia de ADN al menos un sitio de reconocimiento para una meganucleasa derivada de I-Crel racionalmente diseñada con especificidad alterada en relación a I-Crel, en la que el sitio de reconocimiento no se escinde por una I-Crel de origen natural, en la que el sitio de reconocimiento tiene una secuencia central de cuatro pares de bases seleccionada entre el grupo que consiste en TTGT, TTAT, TCTT, TCGT, TCAT, GTTT, GTCT, GGAT, GAGT, GAAT, ATGT, TTTC, TTCC, TGAC, TAAC, GTTC, ATAT, TCGA, TTAA, GGGC, ACGC, CCGC, CTGC, ACAA, ATAA, AAGA, ACGA, ATGA, AAAC, AGAC, ATCC, ACTC, ATTC, ACAT, GAAA, GGAA, GTCA, GTTA, GAAC, ATAT, TCGA, TTAA, GCCC, GCGT, GCGG y GCAG. A continuación, la actividad proporciona una meganucleasa racionalmente diseñada que escinde el sitio de reconocimiento en el ADN.

Opcionalmente, la actividad diseña racionalmente una meganucleasa derivada de I-Crel que escinde el sitio de reconocimiento en el ADN. Opcionalmente, la actividad produce la meganucleasa derivada de I-Crel racionalmente diseñada.

#### 2.2 Secuencias centrales específicamente excluidas

Previamente se ha mostrado que las secuencias centrales GTAC, ACAC, y GTGA son secuencias centrales eficaces para aplicaciones *in vitro* e *in vivo*. Estas secuencias centrales específicamente se excluyen de algunos aspectos de la presente invención. Además, previamente se ha mostrado que las secuencias centrales TTGA y GAAA son secuencias centrales escasas para aplicaciones *in vivo* (Arnould, y col. (2007). *J. Mol. Biol.* 371:49-65).

#### Ejemplos

Esta invención se ilustra además por los siguientes ejemplos, los cuales no se deberían construir como limitantes. Los expertos en la técnica reconocerán, o serán capaces de determinar, usando no más que experimentación rutinaria, numerosos equivalentes a las sustancias y procedimientos específicos descritos en el presente documento. Tales equivalentes se pretenden que sean abarcados por el alcance de las reivindicaciones que siguen a los

ejemplos de a continuación. Los Ejemplos 1 y 2 se refieren a meganucleasas modificadas por ingeniería que escinden sitios de reconocimiento de meganucleasa optimizados *in vivo* en un sistema vegetal modelo. El Ejemplo 3 se refiere a una meganucleasa modificada por ingeniería que escinde sitios de reconocimiento de meganucleasa optimizada *in vitro*. Ejemplos que no caen dentro del alcance de las reivindicaciones son solamente con fines ilustrativos.

**EJEMPLO 1**

Escisión de un sitio de reconocimiento de meganucleasa optimizada por un homodímero de meganucleasa derivada de I-Crel, racionalmente diseñada *in vivo*

Se produjo una meganucleasa modificada por ingeniería llamada BRP2 (SEQ ID NO: 8) usando el método descrito en el documento WO 2007/047859. Esta meganucleasa está derivada de I-Crel y se modificó por ingeniería para reconocer sitios de ADN que no son reconocidos por la I-Crel de tipo natural (por ejemplo, las secuencias de reconocimiento de BRP2 incluyen SEQ ID NO: 9 y SEQ ID NO: 10, o SEQ ID NO: 11 y SEQ ID NO: 12). Para facilitar la localización nuclear de la meganucleasa modificada por ingeniería, se añadió una señal de localización nuclear SV40 (NLS, SEQ ID NO: 13) al terminal N de la proteína. Se usaron procedimientos de transformación mediada por *Agrobacterium* convencionales para transformar *Arabidopsis thaliana* con un ADN-T que contenía una secuencia codificante de BRP2 optimizada con codones (SEQ ID NO: 14). La expresión de la meganucleasa BRP2 estaba bajo el control de un promotor Hsp70 y un terminador NOS. Un par de secuencias de reconocimiento de BRP2 estaban alojadas en el mismo ADN-T separado por 7 pares de bases que contenían un sitio de enzima de restricción PstI (Figura 3a). El corte por BRP2 del par de las secuencias de reconocimiento de BRP2 en esta construcción se esperó que extirpara la región entre las secuencias de reconocimiento y separara de este modo el sitio de restricción PstI (Figura 3b). Se produjeron dos de tales construcciones de ADN-T que variaron la secuencia central de las secuencias de reconocimiento de meganucleasa que flanqueaban el sitio de enzima de restricción PstI (Figura 3c). En la primera construcción (la "construcción GTAC"), los sitios de reconocimiento de meganucleasa tenían la secuencia central GTAC (una secuencia central *in vivo* preferida, Tabla 5, secuencia 9; SEQ ID NO: 9 y SEQ ID NO: 10). La segunda construcción (la "construcción TAGA") tenía la secuencia central TAGA (una secuencia central no preferida, secuencia de cadena opuesta para Tabla 2, secuencia 77; SEQ ID NO: 11 y SEQ ID NO 12).

Se produjeron plantas de *Arabidopsis* establemente transformadas que llevaban cada construcción por selección para un marcador de resistencia a kanamicina alojado en el ADN-T. A continuación, se aisló el ADN genómico de las plantas transformadas (por perforadora de hoja) antes y después del choque térmico para inducir la expresión de la meganucleasa BRP2. Se añadieron muestras de ADN genómico a reacciones de PCR usando cebadores para amplificar la región del ADN-T que alojaba las secuencias de reconocimiento de meganucleasa. A continuación, se digirieron los productos de PCR con PstI y se visualizaron por electroforesis en gel (Fig. 3d). Los resultados se resumen en la Tabla 6. Cualquier muestra de PCR en la que se encontró que un porcentaje significativo (>25 %) del producto era resistente a PstI se consideró que era indicativa de escisión por meganucleasa *in vivo* en esa planta particular y se registró como "cortada" en la Tabla 6. Se encontró que, antes del choque térmico, la gran mayoría de las muestras de PCR de plantas que llevaban cualquier construcción conservaban el sitio PstI. Sin embargo, después del choque térmico, un gran porcentaje de muestras tomadas de las plantas transformadas con la construcción GTAC, pero no la construcción TAGA, habían perdido el sitio PstI. Los productos de PCR de las plantas transformadas por la construcción GTAC que carecían del sitio PstI se clonaron dentro de un plásmido pUC-19 y se secuenciaron. El 100 % de los clones secuenciados tenían una delección precisa de la región entre los dos sitios cortados por BRP2 (como se muestra en el diagrama en la Fig. 3b). Estos resultados indican que una meganucleasa modificada por ingeniería es capaz de escindir un sitio de reconocimiento de meganucleasa *in vivo* siempre y cuando tenga una secuencia central optimizada.

Tabla 6. Escisión *in vivo* de secuencias de reconocimiento de meganucleasa optimizada por un homodímero de meganucleasa modificada por ingeniería

Construcción	Antes del choque térmico		Después del choque térmico	
	Cortada	No cortada	Cortada	No cortada
GTAC	0	4	3	1
TAGA	0	22	0	22

**EJEMPLO 2**

Escisión de un sitio de reconocimiento de meganucleasa optimizada por un heterodímero de cadena sencilla de meganucleasa derivada de I-Crel racionalmente diseñada *in vivo*

La meganucleasa modificada por ingeniería BRP12-SC (SEQ ID NO: 15) se produjo de acuerdo con el documento WO 2007/047859, excepto que esta meganucleasa es un heterodímero de cadena sencilla. Tal como se comenta en el documento WO 2007/047859, I-Crel de tipo natural se une a y escinde ADN como homodímero. Como consecuencia, la secuencia de reconocimiento natural para I-Crel es pseudopalindrómica. Sin embargo, las

secuencias de reconocimiento de BRP12-SC son no palindrómicas (por ejemplo, SEQ ID NO: 16 y SEQ ID NO: 17 o SEQ ID NO: 18 y SEQ ID NO: 19). Esto requiere el uso de un heterodímero de meganucleasa modificada por ingeniería que comprenda un par de subunidades cada una de las cuales reconozca un medio-sitio dentro de la secuencia de reconocimiento de longitud completa. En el caso de BRP12-SC, los dos monómeros de meganucleasa modificada por ingeniería están unidos físicamente uno a otro usando un conector de aminoácidos para producir un heterodímero de cadena sencilla. El conector comprende aminoácidos 166-204 (SEQ ID NO: 20) de BRP12-SC. La secuencia conectora junta una subunidad de meganucleasa N-terminal terminada en L165 (correspondiente a L55 de I-Crel de tipo natural) con una subunidad de meganucleasa C-terminal que comienza en K204 (correspondiente a K7 de I-Crel de tipo natural). Los beneficios de unir físicamente los dos monómeros de la meganucleasa usando este novedoso conector es doble: Primero, asegura que los monómeros de la meganucleasa solamente pueden asociarse uno con otro (someter a heterodimerización) para cortar la secuencia de reconocimiento de BRP12-SC no palindrómica más que formando también homodímeros que pueden reconocer sitios de ADN palindrómicos o pseudopalindrómicos que difieren de la secuencia de reconocimiento de BRP12-SC. Segundo, la unión física de los monómeros de la meganucleasa obvia la necesidad de expresar dos monómeros simultáneamente en la misma célula para obtener el heterodímero deseado. Esto simplifica significativamente la construcción vector en el sentido de que solamente se requiere un casete de expresión de gen sencillo. Como era el caso con la meganucleasa BRP2 comentado en el Ejemplo 1, la meganucleasa BRP12-SC tiene una señal de localización nuclear SV40 (SEQ ID NO: 13) en su terminal N.

Se usaron procedimientos de transformación mediados por *Agrobacterium* convencionales para transformar *Arabidopsis thaliana* con un ADN-T que contenía una secuencia codificante de BRP-SC optimizada con codones (SEQ ID NO: 21). La expresión de la meganucleasa BRP12-SC era bajo el control de un promotor Hsp70 y un terminador NOS. Un par de secuencias de reconocimiento de BRP12-SC estaban alojadas en el mismo ADN-T separadas por 7 pares de bases que contenía un sitio de enzima de restricción PstI (Figura 4a). El corte de BRP12-SC del par de secuencias de reconocimiento de BRP12-SC en esta construcción se esperó que extirpara la región entre las secuencias de reconocimiento y de este modo separara el sitio de restricción de PstI (Figura 4b). Se produjeron dos de tales construcciones de ADN-T que variaban solamente en las secuencias centrales de las secuencias de reconocimiento de meganucleasa que flanqueaban el sitio de la enzima de restricción PstI (Figura 4c). En la primera construcción (la "construcción GTAC"), los sitios de reconocimiento de meganucleasa tenían la secuencia central GTAC (una secuencia central *in vivo* preferida, Tabla 5, secuencia 9; SEQ ID NO: 16 y SEQ ID NO 17). La segunda construcción (la "construcción TAGA") tenía la secuencia central TAGA (una secuencia central no preferida, secuencia de cadena opuesta para Tabla 2, secuencia 77; SEQ ID NO: 18 y SEQ ID NO 19).

Se produjeron plantas de *Arabidopsis* establemente transformadas que llevaban cada construcción por selección para un marcador de resistencia a kanamicina alojado en el ADN-T. A continuación, se aisló el ADN genómico de las plantas transformadas (por perforadora de hoja) antes y después del choque térmico para inducir la expresión de la meganucleasa BRP12-SC. Se añadieron muestras de ADN genómico a reacciones de PCR usando cebadores para amplificar la región del ADN-T que alojaba las secuencias de reconocimiento de la meganucleasa. A continuación, se digirieron los productos de PCR con PstI y se visualizaron por electroforesis en gel. Los resultados de este análisis se presentan en la Tabla 7. Cualquier muestra de PCR en la que se encontró que un porcentaje significativo (>25 %) del producto era resistente a PstI se consideró que era indicativa de escisión por meganucleasa *in vivo* y se registró como "cortada" en la Tabla 7. Se encontró que, antes del choque térmico, la gran mayoría de las muestras de PCR de plantas que llevaban cualquier construcción conservaban el sitio PstI. Sin embargo, después del choque térmico, un gran porcentaje de muestras tomadas de las plantas transformadas con la construcción GTAC, pero no la construcción TAGA, habían perdido el sitio PstI. Los productos de PCR de las plantas transformadas por la construcción GTAC que carecían de sitio PstI se clonaron dentro de un plásmido pUC-19 y se secuenciaron. El 100 % de los clones secuenciados tenían una delección precisa de la región entre los dos sitios cortados por BRP2 (como se muestra en el diagrama en la Fig. 4b). Estos resultados indican que una meganucleasa de cadena sencilla modificada por ingeniería es capaz de escindir un sitio de reconocimiento de meganucleasa *in vivo* siempre y cuando tenga una secuencia central optimizada.

**Tabla 7. Escisión *in vivo* de secuencias de reconocimiento de meganucleasa optimizada por un homodímero de meganucleasa modificado por ingeniería.**

Construcción	Antes del choque térmico		Después del choque térmico	
	Cortada	No cortada	Cortada	No cortada
GTAC	0	23	8	15
TAGA	0	59	1	58

**55 EJEMPLO 3**

Escisión de un sitio de reconocimiento de meganucleasa optimizado por un homodímero de meganucleasa derivada de I-Crel racionalmente diseñada *in vitro*

60 La meganucleasa BRP2 descrita en el Ejemplo 1 (SEQ ID NO: 8) se expresó en *E. coli* y se purificó como en el

Ejemplo 1 del documento WO 2007/047859. A continuación, se añadió la meganucleasa purificada a concentraciones variantes a reacciones que contenían plásmidos que portaban secuencias de reconocimiento de BRP2 con o bien una secuencia central GTAC o TAGA (0,25 picomoles de sustrato plásmido en 25 microlitros de tampón SA: Tris-HCL 25 mM, pH 8.0, NaCl 100 mM, MgCl<sub>2</sub> 5 mM, EDTA 5 mM). Las reacciones se incubaron a 37 °C durante 2 horas y, a continuación, se visualizaron por electroforesis en gel y el porcentaje de cada sustrato plásmido escindido por la meganucleasa se trazó como una función de la concentración de meganucleasa (Figura 5). Se encontró que el sustrato plásmido con la secuencia central TAGA se escindió por la meganucleasa *in vitro*, pero que la escisión de este sustrato requirió una concentración mucho mayor de meganucleasa BRP2 que hizo escisión del sustrato GTAC.

**LISTADO DE SECUENCIAS**

SEQ ID NO: 1 (I-Crel de tipo natural, N.º de referencia de Genbank PO5725)

1 MNTKYNKEFL LYLAFVVDGD GSIIAQIKPN QSYKFKHQLS LAFQVTQKTQ RRWFLDKLVD  
 61 EIGVGYVRDR GSVSDYILSE IKPLHNFLTQ LQPFLKQK QANLVLKIIW RLPSAKESPD  
 121 KFLEVCTWVD QIAALNDSKT RKTTSETVRA VLDSLSEKKK SSP

SEQ ID NO: 2 (Secuencia de reconocimiento de I-Crel de tipo natural)  
 1 GAAACTGTCT CACGACGTTT TG

SEQ ID NO: 3 (Secuencia de reconocimiento de I-Crel de tipo natural)  
 1 GAAAACGTCTG TGAGACAGTT TC

SEQ ID NO: 4 (secuencia de aminoácidos de DJ1)

1 MNTKYNKEFL LYLAFVVDGD GSIIAQIDPR QNYKFKHELRLRFQVTQKTQ RRWFLDKLVD  
 61 EIGVGYVRDR GSVSDYILSE IKPLHNFLTQ LQPFLKQK QANLVLKIIIE QLPSAKESPD  
 121 KFLEVCTWVD QIAALNDSKT RKTTSETVRA VLDSLSEKKK SSP

SEQ ID NO: 5 (secuencia de reconocimiento de DJ1- secuencia central GTGA)  
 1 AACGGTGTCTG TGAGACACCG TT

SEQ ID NO: 6 (secuencia de reconocimiento de DJ1- secuencia central GTGA)  
 1 AACGGTGTCT CACGACACCG TT

SEQ ID NO: 7 (medio sitio DJ1)  
 1 AACGGTGTC

SEQ ID NO: 8 (secuencia de aminoácidos de BRP2)

1 MGPKKRKRVI MNTKYNKEFL LYLAFVVDGD GSIIASIRPR QSKFKHELE  
 51 LRFQVTQKTQ RRWFLDKLVD EIGVGYVRDR GSVSDYRLSQ IKPLHNFLTQ  
 101 LQPFLKQK QANLVLKIIIE QLPSAKESPD KFLEVCTWVD QIAALNDSKT  
 151 RKTTSETVRA VLDSLSEKKK SSP

SEQ ID NO: 9 (secuencia de reconocimiento de BRP2 – secuencia central GTAC)  
 1 CTCCGGGTCTG TACGACCCGG AG

SEQ ID NO: 10 (secuencia de reconocimiento de BRP2 – secuencia central GTAC)  
 1 CTCCGGGTCTG TACGACCCGG AG

SEQ ID NO: 11 (secuencia de reconocimiento de BRP2 – secuencia central TAGA)  
 1CTCCGGGTCTAGAGACCCGG AG

SEQ ID NO: 12 (secuencia de reconocimiento de BRP2 – secuencia central TAGA)  
 1 CTCCGGGTCT CTAGACCCGG AG

SEQ ID NO: 13 (secuencia de aminoácidos de la señal de localización nuclear de SV40)  
 1 MAPKKRKY

# ES 2 625 941 T3

SEQ ID NO: 14 (secuencia de ADN de BRP2 optimizada con codones)

```
1 ATGGGCCCCGA AGAAGAAGCG CAAGGTCATC ATGAACACCA AGTACAACAA
51 GGAGTTCCTG CTCTACCTGG CGGGCTTCGT GGACGGGGAC GGCTCCATCA
101 TCGCCTCCAT CCGCCCCTGG CAGTCCTGCA AGTTCAAGCA TGAGCTGGAA
151 CTCCGGTTCC AGGTCACGCA GAAGACACAG CGCCGTTGGT TCCTCGACAA
201 GCTGGTGGAC GAGATCGGGG TGGGCTACGT GCGCGACCGC GGCAGCGTCT
251 CCGACTACCG CCTGAGCCAG ATCAAGCCTC TGCACAACCT CCTGACCCAG
301 CTCCAGCCCT TCCTGAAGCT CAAGCAGAAG CAGGCCAACC TCGTGTGAA
351 GATCATCGAG CAGCTGCCCT CCGCCAAGGA ATCCCCGGAC AAGTTCCTGG
401 AGGTGTGCAC CTGGGTGGAC CAGATCGCCG CTCTGAACGA CTCCAAGACC
451 CGCAAGACCA CTTCCGAGAC CGTCCGCGCC GTGCTGGACA GTCTCTCCGA
501 GAAGAAGAAG TCGTCCCCCT AG
```

SEQ ID NO: 15 (secuencia de aminoácidos de BRP12-SC)

5

```
1 MGPKKKRVKVI MNTKYNKEFL LYLAFVDGD GSIKAQIRPR QSRKFKHELE
51 LTFQVTQKTQ RRWFLDKLVD EIGVGKVDYDR GSVSDYELSQ IKPLHNFLTQ
101 LQPFLLKQK QANLVLKIE QLPSAKESPD KFLEVCTWVD QIAALNDSKT
151 RKTTSETVRA VLDSLPGSVG GLSPSQASSA ASSASSSPGS GISEALRAGA
201 TKSKEFLLYL AGFVDGDGSI IASIRPRQSC KFKHELELRF QVTQKTQRRW
251 FLDKLVDEIG VGYVRDRGSV SDYRLSQIKP LHNFLTQLQP FLKQKQAN
301 LVLKIEQLP SAKESPKFL EVCTWVDQIA ALNDSKTRKT TSETVRAVL
351 SLSEKKKSSP
```

SEQ ID NO: 16 (secuencia de reconocimiento de BRP12-SC – secuencia central GTAC)

1 TGCCTCCTCG TACGACCCGG AG

10

SEQ ID NO: 17 (secuencia de reconocimiento de BRP12-SC – secuencia central GTAC)

1 CTCCGGGTCTG TACGAGGAGG CA

SEQ ID NO: 18 (secuencia de reconocimiento de BRP12-SC – secuencia central TAGA)

1 TGCCTCCTCT AGAGACCCGG AG

15

SEQ ID NO: 19 (secuencia de reconocimiento de BRP12-SC – secuencia central TAGA)

1 CTCCGGGTCT CTAGAGGAGG CA

SEQ ID NO: 20 (secuencia de aminoácidos enlazadora de BRP12-SC)

1 PGVGLSPS QASSAASSAS SSPGSGISEA LLAGATKS

20

SEQ ID NO: 21 (secuencia de ADN de BRP12-SC optimizada con codones)

ES 2 625 941 T3

```

1   ATGGGCCCCGA AGAAGAAGCG CAAGGTCATC ATGAACACCA AGTACAACAA
51  GGAGTTCCTG CTCTACCTGG CCGGCTTCGT GGACGGCGAC GGCTCCATCA
101 AGGCGCAGAT CCGTCCGCGG CAGAGCCGGA AGTTCAAGCA CGAGCTCGAG
151 CTGACCTTCC AGGTGACCCA GAAGACGCAG AGGCGCTGGT TCCTCGACAA
201 GCTGGTGGAC GAGATCGGGG TGGCAAGGT CTACGACCGC GGGTCGGTGT
251 CCGACTACGA GCTCTCCCAG ATCAAGCCCC TGCACAACCT CCTCACCCAG
301 CTCCAGCCGT TCCTGAAGCT CAAGCAGAAG CAGGCCAACC TCGTGCTGAA
351 GATCATCGAG CAGCTGCCCT CCGCCAAGGA ATCCCCGGAC AAGTTCCTGG
401 AGGTGTGCAC GTGGGTGGAC CAGATCGCGG CCCTCAACGA CAGCAAGACC
451 CGCAAGACGA CCTCGGAGAC GGTGCGGGCG GTCCTGGACT CCCTCCCAGG
501 ATCCGTGGGA GGTCTATCGC CATCTCAGGC ATCCAGCGCC GCATCCTCGG
551 CTTCCCTCAAG CCCGGGTTC AAGCACTCAG AGCTGGAGCA
601 ACTAAGTCCA AGGAATTCCT GCTCTACCTG GCGGGCTTCG TGGACGGGGA
651 CGGCTCCATC ATCGCCTCCA TCCGCCCCG TCAGTCTG CAGTTCAAGC
701 ATGAGCTGGA ACTCCGGTTC CAGGTCACGC AGAAGACACA GCGCCGTTGG
751 TTCCTCGACA AGCTGGTGG AAGATCGGG GTGGGCTACG TCGCGGACCG
801 CGGCAGCGTC TCCGACTACC GCCTGAGCCA GATCAAGCCT CTGCACAAC
851 TCCTGACCCA GCTCCAGCCC TTCCTGAAGC TCAAGCAGAA GCAGGCCAAC
901 CTCGTGCTGA AGATCATCGA GCAGCTGCCC TCCGCCAAGG AATCCCCGGA
951 CAAGTTCCTG GAGGTGTGCA CCTGGGTGGA CCAGATCGCC GCTCTGAACG
1001 ACTCCAAGAC CCGCAAGACC ACTTCCGAGA CCGTCCGCGC CGTGCTGGAC
1051 AGTCTCTCCG AGAAGAAGAA GTCGTCCCC TAG

```

**REIVINDICACIONES**

1. Un método *ex vivo* para escindir un ADN de doble cadena, que comprende:

- 5 (a) identificar en dicho ADN al menos un sitio de reconocimiento para una meganucleasa derivada de I-Crel racionalmente diseñada con especificidad alterada en relación a I-Crel, en donde dicho sitio de reconocimiento no se escinde mediante una I-Crel de origen natural, en donde dicho sitio de reconocimiento se identifica basándose en la presencia de una secuencia central de cuatro pares de bases seleccionada entre el grupo que  
10 consiste en TTGT, TTAT, TTTC, TTCC, TTAG, TTAC y TTGC;  
(b) proporcionar dicha meganucleasa racionalmente diseñada; y  
(c) poner en contacto dicho ADN con dicha meganucleasa racionalmente diseñada; por lo cual dicha meganucleasa racionalmente diseñada escinde dicho ADN.

2. Un método *ex vivo* para escindir un ADN de doble cadena que comprende:

- 15 (a) incorporar una secuencia central de cuatro pares de bases seleccionada entre el grupo que consiste en TTGT, TTAT, TTTC, TTCC, TTAG, TTAC y TTGC en un sitio de reconocimiento para una meganucleasa derivada de I-Crel racionalmente diseñada con especificidad alterada en relación a I-Crel, en donde dicho sitio de reconocimiento no se escinde mediante una I-Crel de origen natural,  
20 (b) introducir dicho sitio de reconocimiento en dicho ADN,  
(c) proporcionar dicha meganucleasa racionalmente diseñada; y  
(d) poner en contacto dicho ADN con dicha meganucleasa racionalmente diseñada; por lo cual dicha meganucleasa racionalmente diseñada escinde dicho ADN.

25 3. El método *ex vivo* de una cualquiera de las reivindicaciones 1 a 2, en el que dicha escisión de ADN es *in vitro*, opcionalmente en el que dicho ADN se selecciona entre el grupo que consiste en un producto de PCR; un cromosoma artificial; ADN genómico aislado de bacterias, de hongos, de plantas o de células animales; y ADN vírico.

30 4. El método *ex vivo* de una cualquiera de las reivindicaciones 1 a 3, que comprende además diseñar racionalmente dicha meganucleasa derivada de I-Crel para reconocer dicho sitio de reconocimiento.

5. El método *ex vivo* de una cualquiera de las reivindicaciones 1 a 4, que comprende además producir dicha meganucleasa derivada de I-Crel racionalmente diseñada.

35

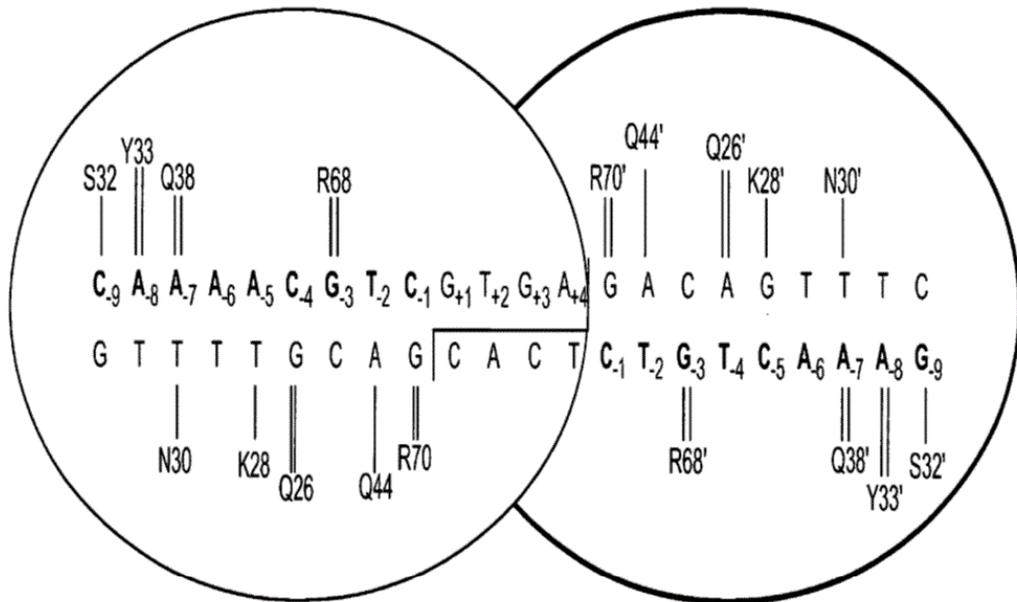


FIG. 1A

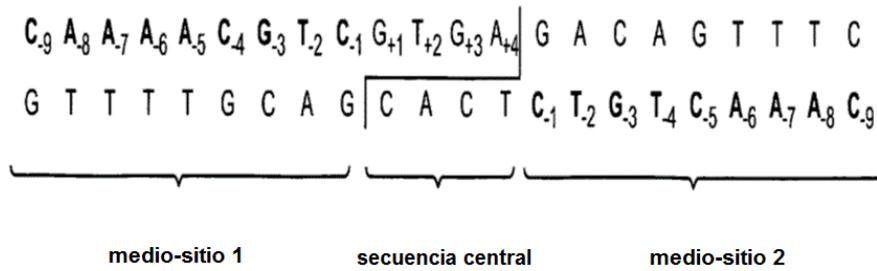


FIG. 1B

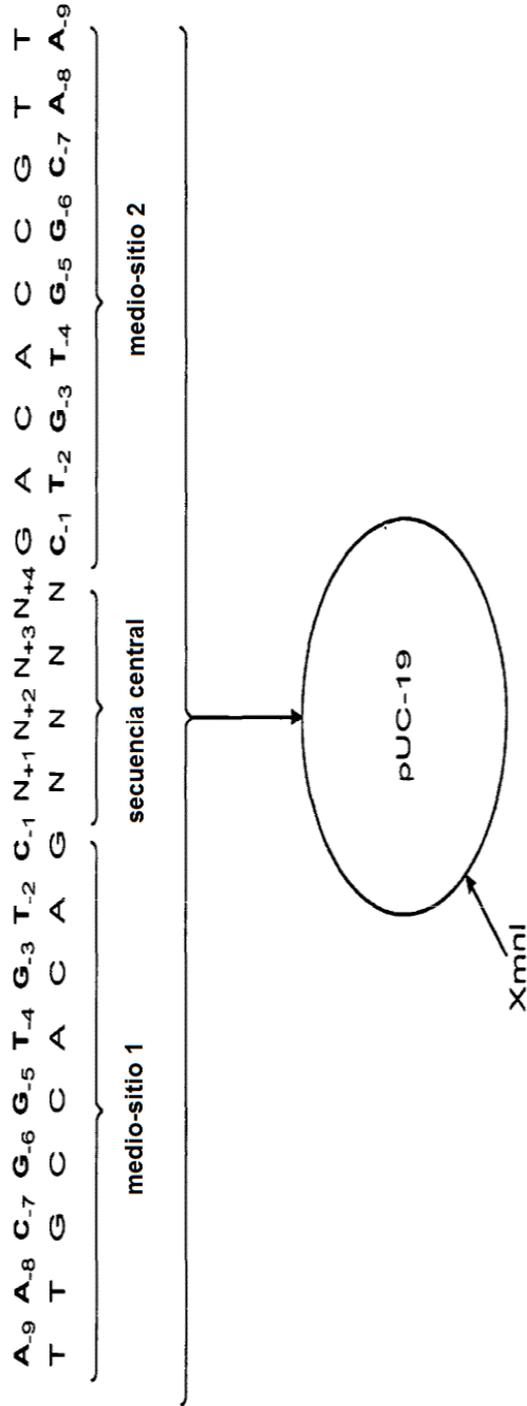


FIG. 2A

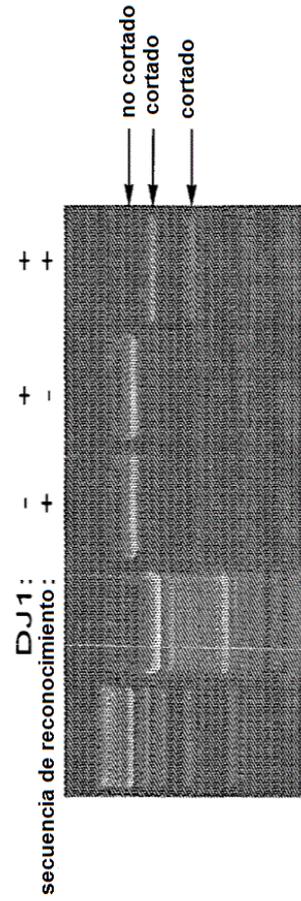
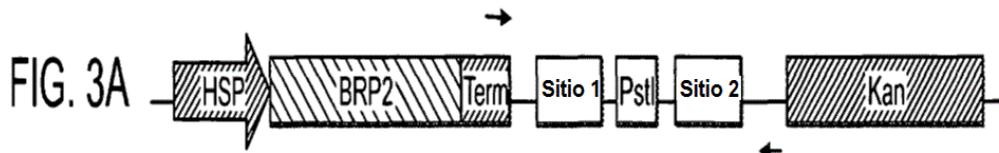


FIG. 2B



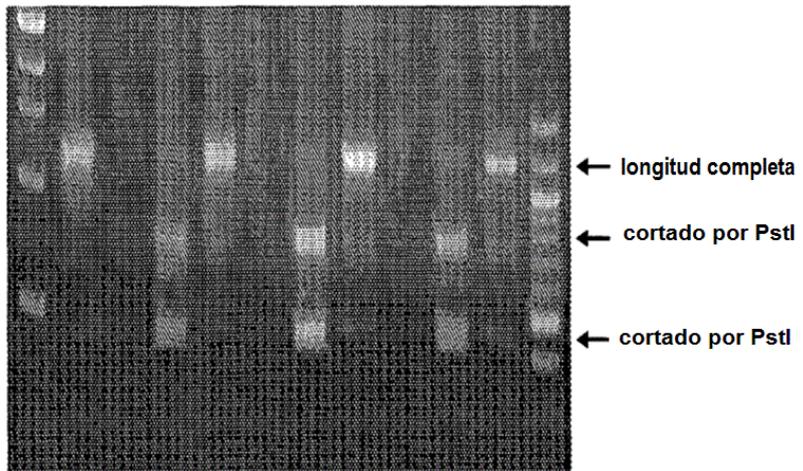
Escisión Meganucleasa



FIG. 3C  
 GTAC: CTCCGGGTCGTACGACCCGGAG  
 TAGA: CTCCGGGTCTAGAGACCCGGAG

FIG. 3D

Planta Nº:	C	44	45	46
Choque térmico:		- +	- +	- +



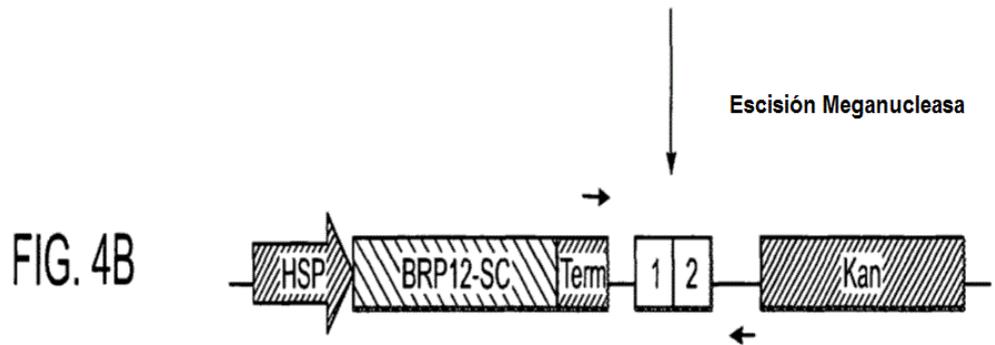
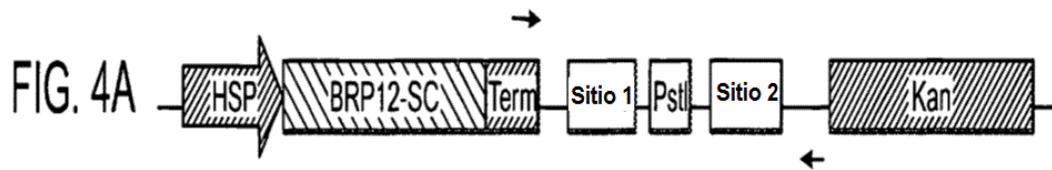


FIG. 4C

GTAC: TGCCTCCTCGTACGCCGGAG  
 TAGA: TGCCTCCTCTAGAGGCCGGAG

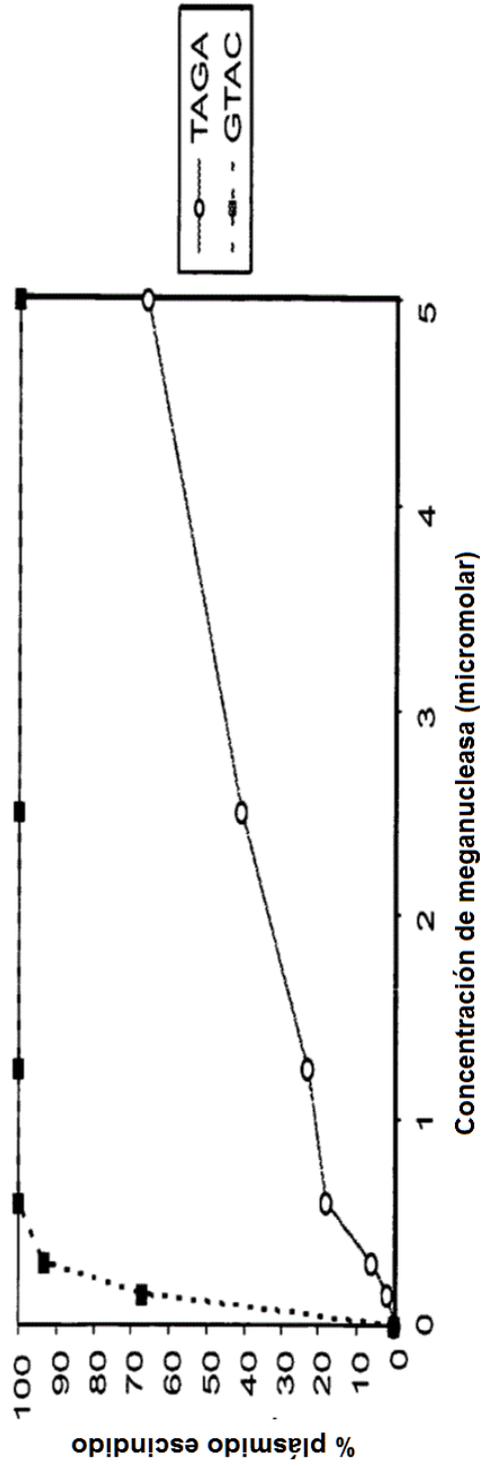


FIG. 5