



MINISTERIO
DE INDUSTRIA, TURISMO
Y COMERCIO



Oficina Española
de Patentes y Marcas

Justificante de presentación electrónica de solicitud de patente

Este documento es un justificante de que se ha recibido una solicitud española de patente por vía electrónica, utilizando la conexión segura de la O.E.P.M. Asimismo, se le ha asignado de forma automática un número de solicitud y una fecha de recepción, conforme al artículo 14.3 del Reglamento para la ejecución de la Ley 11/1986, de 20 de marzo, de Patentes. La fecha de presentación de la solicitud de acuerdo con el art. 22 de la Ley de Patentes, le será comunicada posteriormente.

Número de solicitud:	P201430012	
Fecha de recepción:	08 enero 2014, 13:14 (CET)	
Oficina receptora:	OEPM Madrid	
Su referencia:	2013-0377	
Solicitante:	CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC)	
Número de solicitantes:	1	
País:	ES	
Título:	METODO PARA LA IDENTIFICACION DE DIANAS PARA INTRONES DEL GRUPO II EN CUALQUIER SECUENCIA DE ADN Y SU INSERCIÓN EFICIENTE	
Documentos enviados:	Descripcion.pdf (14 p.) Reivindicaciones.pdf (3 p.) Resumen.pdf (1 p.) Dibujos.pdf (6 p.) OLF-ARCHIVE.zip OTRO-1.pdf (1 p.) SEQLPDF.pdf (4 p.)	package-data.xml es-request.xml application-body.xml es-fee-sheet.xml feesheet.pdf request.pdf
Enviados por:	CN=NOMBRE UNGRIA LOPEZ JAVIER - NIF 05211582N,OU=500050022,OU=FNMT Clase 2 CA,O=FNMT,C=ES	
Fecha y hora de recepción:	08 enero 2014, 13:14 (CET)	
Codificación del envío:	4A:11:2B:1F:40:A0:6E:03:EC:72:05:36:92:37:30:7D:95:6B:C1:FC	

/Madrid, Oficina Receptora/



(1) MODALIDAD:	PATENTE DE INVENCION MODELO DE UTILIDAD	<input checked="" type="checkbox"/> <input type="checkbox"/>
(2) TIPO DE SOLICITUD:	PRIMERA PRESENTACION ADICION A LA PATENTE EUROPEA ADICION A LA PATENTE ESPAÑOLA SOLICITUD DIVISIONAL CAMBIO DE MODALIDAD TRANSFORMACION SOLICITUD PATENTE EUROPEA PCT: ENTRADA FASE NACIONAL	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
(3) EXP. PRINCIPAL O DE ORIGEN:	MODALIDAD: N.º SOLICITUD: FECHA SOLICITUD:	
4) LUGAR DE PRESENTACION:		OEPM, Presentación Electrónica
(5-1) SOLICITANTE 1:	DENOMINACION SOCIAL: UNIVERSIDAD PÚBLICA NACIONALIDAD: CÓDIGO PAÍS: NIF/NIE/PASAPORTE: CNAE: PYME: DOMICILIO: LOCALIDAD: PROVINCIA: CÓDIGO POSTAL: PAÍS RESIDENCIA: CÓDIGO PAÍS: TELÉFONO: FAX: CORREO ELECTRÓNICO: PERSONA DE CONTACTO: MODO DE OBTENCIÓN DEL DERECHO: PORCENTAJE DE TITULARIDAD:	CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC) [] España ES Q2818002D Serrano nº 117 MADRID 28 Madrid 28006 España ES <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> 100,00 %
(6-1) INVENTOR 1:	APELLIDOS: NOMBRE: NACIONALIDAD: CÓDIGO PAÍS: NIF/NIE/PASAPORTE:	GARCIA RODRIGUEZ FERNANDO MANUEL España ES
(6-2) INVENTOR 2:	APELLIDOS: NOMBRE: NACIONALIDAD: CÓDIGO PAÍS: NIF/NIE/PASAPORTE:	TORO GARCIA NICOLAS España ES
(7) TÍTULO DE LA INVENCION:		METODO PARA LA IDENTIFICACION DE

(18) NOTAS:	
(19) FIRMA:	FIRMA DEL SOLICITANTE O REPRESENTANTE: LUGAR DE FIRMA: FECHA DE FIRMA:

NOMBRE UNGRIA LOPEZ
JAVIER - NIF 05211582N
MADRID
08 Enero 2014

MÉTODO PARA LA IDENTIFICACIÓN DE DIANAS PARA INTRONES DEL GRUPO II EN CUALQUIER SECUENCIA DE ADN Y SU INSERCIÓN EFICIENTE.

RESUMEN

5

La presente invención se refiere a un método para rediseñar un intrón del grupo II para su inserción eficiente en un ADN que, a partir de un intrón inicial del grupo II que codifica una proteína IEP carente del dominio endonucleasa, comprende:

- seleccionar una diana de al menos 25 nucleótidos consecutivos en la secuencia de dicho ADN según su eficiencia de invasión;
 - modificar total o parcialmente las secuencias de hibridación EBS1, EBS2 y EBS3 del intrón inicial para ser totalmente complementarias con las secuencias IBS1, IBS2 e IBS3 de la secuencia diana.
- 15 Son asimismo objeto de la presente invención, el intrón rediseñado por el método anterior y el vector que lo comprende. Además, la invención también hace referencia al uso de estos intrones rediseñados en diversas aplicaciones biotecnológicas.

MÉTODO PARA LA IDENTIFICACIÓN DE DIANAS PARA INTRONES DEL GRUPO II EN CUALQUIER SECUENCIA DE ADN Y SU INSERCIÓN EFICIENTE.

5

DESCRIPCIÓN

Sector de la técnica

El método de la invención se encuadra en el sector biotecnológico, siendo de aplicación tanto *in vivo* como *in vitro* en el marcaje de genes, control de la expresión génica, mutagénesis, reparación génica etc.

Objeto de la invención

El método objeto de la presente invención permite la identificación en cualquier secuencia de ADN de la mejor diana posible para ser invadida eficientemente por intrones del grupo II. El método consiste en el análisis de una secuencia de ADN (mediante un algoritmo informático) en grupos de 25 nucleótidos (diana), este grupo se va desplazando a lo largo de la secuencia de nucleótido en nucleótido identificando las mejores dianas. Posteriormente se modifican, en el ARN del intrón, aquellas regiones mediante las cuales el intrón reconoce su diana, denominadas EBS, de forma que reconozcan a la diana seleccionada y se aporta una proteína codificada por el intrón cuya forma silvestre carece de dominio endonucleasa.

Estado de la técnica

25

Los intrones del grupo II son ARN catalíticos que maduran a partir de su ARN mensajero a un intermediario en forma de lazo por un mecanismo similar al descrito para los intrones nucleares. Los intrones del grupo II se describieron en un principio en las mitocondrias de levaduras y en cloroplastos y posteriormente en bacterias y arqueas (Michel y Ferat 1995 Annu. Rev. Biochem. 64: 435-461; Toro 2003 Environ. Microbiol. 5: 143-151).

El ARN de los intrones del grupo II posee una estructura tridimensional organizada en seis dominios del DI al DVI. En el DI se encuentran las regiones EBS que intervienen en el reconocimiento de las dianas de estos elementos. En el dominio DIV la mayoría de los intrones bacterianos poseen un ORF (Open ReadinG Frame) que codifica una proteína (IEP). Esta IEP posee en su parte amino-terminal un dominio reverso transcriptasa (RT)

seguido por un posible dominio de unión al ARN con actividad maturasa (dominio X) y en la parte carboxi-terminal un dominio de unión a ADN (D) seguido por un dominio endonucleasa (En). Algunos intrones del grupo II son elementos genéticos móviles que se insertan en posiciones determinadas de otro alelo carente del intrón, proceso denominado “retrohoming”. El movimiento de los intrones del grupo II esta mediado por un complejo ribonucleoproteico (RNP) formado por la IEP codificado por el ORF y el ARN del intrón en forma de lazo que permanece asociado a la IEP. El complejo ribonucleoproteico reconoce la diana del intrón a través tanto de la proteína IEP como del ARN en forma de lazo del intrón. La parte central de la diana es reconocida por apareamiento del ARN a través de las EBSs (Exon Binding Site) (EBS1, 2 y 3) que hibridan con la región complementaria en el ADN diana denominadas IBSs (Intron Binding Site) (IBS1, 2 y 3) (Michel y Ferat 1995 Annu. Rev. Biochem. 64: 435-461). Los extremos de la diana son reconocidos por la IEP (Singh y Lambowitz 2001 J. Mol. Biol. 309:361-386). El ARN del intrón corta el ADN en la cadena codificante justo en el sitio donde se va a insertar a través del mecanismo conocido como “splicing reverso”, mientras que la IEP corta la cadena no codificante en la posición +9 o +10 del exón 2. Este extremo 3’ del sitio de corte es usado como cebador de la reacción de reversotranscripción sobre el ARN del intrón que ha sufrido “splicing reverso” en la cadena codificante del ADN diana.

Muchas de las IEPs codificadas por intrones del grupo II bacterianos carecen de dominio endonucleasa responsable del corte en la cadena no codificante en la posición +9 o +10 del exón 2. Este es el caso de la IEP del intrón Rmlnt1 de *Sinorhizobium meliloti* que aun careciendo del dominio En se ha demostrado que es un intrón móvil (Martínez-Abarca y col. 2000 Mol. Microbiol. 35:1405-1412). El “retrohoming” de Rmlnt1 es independiente de los mecanismos de recombinación homóloga. La vía mayoritaria de retrohoming de Rmlnt1 es mediante “splicing reverso” en ADN de cadena sencilla en la horquilla de replicación y usa la cadena “lagging” como cebador para la reverso transcripción (Martínez-Abarca y col. 2004, Nucleic Acids Res. 32:2880-2888). Esta preferencia hacia la cadena de ADN que sirve de molde para la síntesis de la cadena “lagging” durante la replicación es la contraria a la que muestra un mutante del intrón LI.LtrB sin dominio En (Zhong y Lambowitz 2003, EMBO J 22:4555-4565). El mecanismo que Rmlnt1 utiliza para el retrohoming es diferente al que presentan los intrones cuyas IEPs poseen dominios endonucleasa (En) y las reglas por las que Rmlnt1 reconoce a su diana son también distintas indicando características específicas en el reconocimiento de su diana por el ARN o la proteína.

35

Esta especificidad intrínseca que depende de secuencia convierte a los intrones del grupo II en herramientas moleculares útiles para su uso en biotecnología, para el análisis funcional de genomas y en terapia génica (Guo y col. 2000 Science 289:452-457).

5 Los intrones del grupo II están siendo desarrollados como herramientas de mutagénesis porque pueden ser dirigidos a secuencias específicas de ADN disminuyendo la aparición de mutaciones inespecíficas que suceden cuando se utilizan otros métodos como los basados en transposones. Como la actividad de estos intrones no depende de los mecanismos de recombinación celulares pueden ser herramientas muy útiles en organismos y líneas
10 celulares en los que no están presentes estos mecanismos de recombinación.

Aunque es conocido que, en la diana para los intrones del grupo II, las posiciones -15 y +4 con respecto al sitio de inserción del intrón tienen que estar ocupadas por T y G respectivamente para que el intrón se pudiera insertar, otras posiciones de la parte distal de los exones y las interacciones EBS-IBS están vinculadas a la eficiencia de la inserción del intrón y no son conocidas. La presente invención determina cuáles son esas posiciones y que nucleótido debe ocuparlas para que se dé una invasión eficiente (figura 5). Mediante un algoritmo informático y utilizando las frecuencias de nucleótidos de la figura 4 se obtiene una puntuación para cada una de las posibles dianas del intrón. Cuanto mayor sea la puntuación
20 obtenida mayor será el porcentaje de invasión.

Aunque hay algunos métodos de mutagénesis basados en el intrón LI.ItrB cuya IEP contiene el dominio de unión al ADN (D) y el dominio endonucleasa (En), y cuya diana está formada por 35 nucleótidos (patentes US7.592.161; US6.027.895; US5.869.634; US5.698.421), el desarrollo de nuevos métodos para identificar mejores dianas para otros intrones del grupo II carentes del dominio En en su IEP, y cuya diana mínima está compuesta por sólo 25 nucleótidos, aumenta las posibilidades de encontrar dianas para estos intrones en cualquier ADN ampliando además el rango de herramientas biotecnológicas que pueden ser usadas en diferentes ADNs.

30

Explicación de la invención

El objeto de la presente invención lo constituye un método para la identificación de las mejores dianas para intrones del grupo II dentro de cualquier secuencia de ADN y la modificación de los intrones del grupo II para su inserción de una manera eficiente en estas dianas que comprende los siguientes pasos:

35

- 5 a) Selección de una secuencia de 25 nucleótidos dentro de cualquier ADN mediante la aplicación de un algoritmo informático para un intrón del grupo II silvestre que codifica en su dominio IV una proteína (IEP) cuya forma silvestre carece de dominio endonucleasa. La secuencia seleccionada comprende por una parte, tres regiones, IBS1, IBS2 e IBS3 que hibridan con las secuencias EBS1, EBS2 y EBS3 respectivamente, contenidas en el dominio I del ARN de un intrón silvestre del grupo II. Por otra parte dos regiones distales, la del exón 1 y la del exón 2 que son reconocidas por la IEP del intrón del grupo II.
- 10 b) Modificar total o parcialmente dichas secuencias de hibridación, EBS1, EBS2 y EBS3 por separado o conjuntamente para que aparezcan totalmente con las secuencias IBS1, IBS2 e IBS3 de la secuencia diana seleccionada en el apartado anterior.

15 Los intrones del grupo II utilizados en el presente método codifican proteínas cuyas cadenas de aminoácidos carecen en su forma silvestre del dominio endonucleasa (En).

20 En particular, el ARN del intrón del grupo II es un ARN del intrón Rmlnt1 silvestre o modificado y la proteína codificada por el intrón del grupo II es una proteína codificada por dicho intrón Rmlnt1. La modificación del intrón puede consistir entre otras en la delección del ORF que codifica la IEP en el dominio IV del intrón o la inserción en ese dominio IV sin ORF de secuencias heterólogas como genes de resistencia, promotores inducibles o promotores constitutivos, etc.

Breve descripción de las figuras

25

FIG. 1. La figura 1 es una representación de los diferentes dominios del ARN del intrón del grupo II Rmlnt1. Se señala también la EBS1, EBS2 y EBS3 y la interacción que establecen con la IBS1, IBS2 e IBS3 de la diana. Se señala también la posición “ δ ” que hibrida con la posición “ δ ” y no con la IBS3. Se muestran también el ORF de la proteína codificada por el intrón en el DIV, este ORF sería eliminado y/o sustituido por otro ADN heterólogo en los intrones modificados. Así mismo se muestran las interacciones que mantienen el plegamiento del ARN del intrón.

30

FIG. 2. La figura 2 muestra una representación de los diferentes dominios de la proteína de Rmlnt1 y la de Ll.LtrB.

35

FIG. 3. En la figura 3 se representa la diana silvestre de Rmlnt1 y las interacciones entre las IBSs y las EBSs. Muestra también la parte distal del exón 1 y 2.

FIG. 4. La figura 4 muestra las frecuencias obtenidas en experimentos de selección para cada uno de los nucleótidos en cada una de las posiciones de la diana de Rmlnt1 de 25 nucleótidos (seleccionados), también se muestra las frecuencias de partida antes de la selección (sin seleccionar).

FIG. 5. La figura 5 es una representación de los requerimientos de secuencia en cada una de las posiciones de la diana de 25 nucleótidos de Rmlnt1. Los nucleótidos necesarios no siempre son los que están presentes en la diana silvestre de Rmlnt1.

FIG. 6. La figura 6 muestra la diana silvestre (wt) del intrón Rmlnt1 y las interacciones EBSs/IBSs. Se muestran también las dianas identificadas por el algoritmo en la secuencia del gen lacZ y las modificaciones en las EBSs para que el intrón reconozca las dianas identificadas. Se muestra la puntuación dada por el algoritmo a cada una de las dianas identificadas (S) y el porcentaje de invasión de cada una de las dianas identificadas.

FIG. 7. La figura 7 muestra un diagrama de flujo que esquematiza el funcionamiento del algoritmo.

Descripción de la invención

Un primer aspecto de la invención hace referencia a un método para rediseñar un intrón del grupo II para su inserción eficiente en un ADN cualquiera partiendo de un intrón inicial del grupo II que codifica una proteína IEP carente del dominio endonucleasa, la cual además carece de la región de unión al ADN (D). Dicho método comprende las siguientes etapas:

- seleccionar al menos una diana que comprende al menos 25 nucleótidos consecutivos en la secuencia de dicho ADN, preferentemente una diana de entre 25 y 35 nucleótidos, y más preferentemente de 25 nucleótidos, mediante la aplicación de un algoritmo que determina la mejor diana para dicho intrón inicial según su eficiencia de invasión, donde la diana seleccionada comprende:
 - a. un sitio de inserción para el intrón situado entre dos nucleótidos consecutivos de la diana seleccionada, preferentemente comprendidos entre las posiciones 20 y 26 respecto al extremo 5' en dianas de 25-35 nucleótidos, que preferentemente,

cuando la diana tiene 25 nucleótidos, dicho sitio de inserción está situado entre los nucleótidos en posiciones 20 y 21 respecto al extremo 5' de la diana seleccionada;

b. dos regiones distales que son reconocidas por la IEP del intrón del grupo II, una primera región, la región distal del exón 1 que comprende los nucleótidos en posiciones -14 a -20 respecto al sitio de inserción de la secuencia diana seleccionada, y una segunda región, la región distal del exón 2 que comprende los nucleótidos en posiciones +2 a +5 respecto al sitio de inserción de la secuencia diana seleccionada, que preferentemente, cuando la diana tiene 25 nucleótidos, la región distal del exón 1 comprende los nucleótidos en posiciones 1 a 7 respecto al extremo 5' de la secuencia seleccionada, y la región distal del exón 2 comprende los nucleótidos en posiciones 22 a 25 respecto al extremo 5' de la secuencia diana seleccionada;

c. tres regiones IBS1, IBS2 e IBS3, que hibridan con las secuencias EBS1, EBS2 y EBS3 del intrón del grupo II en su forma silvestre, respectivamente, donde la región IBS1 comprende los nucleótidos en posiciones -1 a -7 respecto al sitio de inserción de la secuencia diana seleccionada, la región IBS2 comprende los nucleótidos en posiciones -9 a -13 respecto al sitio de inserción de la secuencia diana seleccionada, y la región IBS3 consiste en el nucleótido en posición +1 respecto al sitio de inserción de la secuencia diana seleccionada, que preferentemente, cuando la diana tiene 25 nucleótidos, la región IBS1 comprende los nucleótidos en posiciones 14 a 20 respecto al extremo 5' de la secuencia seleccionada, la región IBS2 comprende los nucleótidos en posiciones 8 a 12 respecto al extremo 5' de la secuencia seleccionada, y la región IBS3 consiste en el nucleótido en posición 21 respecto al extremo 5' de la secuencia seleccionada;

- modificar total o parcialmente al menos una de las secuencias de hibridación EBS1, EBS2 y EBS3 del intrón del grupo II inicial para que sean totalmente complementarias con las respectivas secuencias IBS1, IBS2 e IBS3 de la secuencia diana seleccionada en la etapa anterior, y así obtener un intrón rediseñado para la inserción eficiente en dicha secuencia diana.

Los intrones del grupo II que se encuentran en la naturaleza, a partir de ahora también referidos como intrones de tipo silvestre, se transcriben en una molécula de ARN desde cualquier promotor situado corriente arriba. El ARN del intrón presenta una actividad catalítica que media la escisión de su ARN mensajero precursor generando un intermediario en forma de lazo. Este ARN del intrón del grupo II escindido tiene una estructura secundaria y terciaria características y está estructurada en 6 dominios. En el dominio I se encuentran

las secuencias capaces de hibridar con una de las cadenas de la secuencia de ADN diana. Estas secuencias presentes en el ARN del intrón son tres y se denominan EBS1, EBS2 y EBS3 e hibridan con las secuencias IBS1, IBS2 e IBS3 del ADN diana. En el dominio IV del ARN del intrón existe un marco abierto de lectura que codifica la proteína del intrón del grupo II "IEP". Esta IEP posee el dominio Reversotranscriptasa (RT) y el dominio maturasa (X). En el caso de intrones del grupo II a los que se refiere la presente invención, esta IEP se caracteriza además porque su forma silvestre carece de la región de unión al ADN (D) y/o el dominio endonucleasa (En) presentes en otros intrones del grupo II que se ha demostrado que son móviles (LI.LtrB de *Lactococcus lactis* y Ecl5 de *Escherichia coli*).

10

Según la presente invención, el intrón inicial o de partida del grupo II al que se refiere el método descrito anteriormente puede tener una secuencia de tipo silvestre, es decir, una secuencia que es idéntica a la secuencia del intrón del grupo II que se encuentra en la naturaleza, o bien, poseer una secuencia modificada, es decir, una secuencia que sea diferente de la secuencia encontrada en la naturaleza, preferentemente modificada por comprender al menos una delección del ORF que codifica la IEP en el dominio IV del intrón, pudiendo comprender además una inserción de al menos una secuencia heteróloga en dicho dominio IV sin ORF, tal como son por ejemplo, un gen de resistencia, un promotor inducible, un promotor constitutivo, etc. Asimismo, la secuencia modificada puede comprender una o más inserciones de secuencias heterólogas en el dominio IV del intrón, tanto si existe o no la delección del ORF.

15

20

Lógicamente y de manera similar, el intrón rediseñado y obtenido con el método descrito en la presente invención, puede tener una secuencia que, a excepción de las secuencias de hibridación EBS1, EBS2 y EBS3 que se modifican por dicho método, sea idéntica a la secuencia del intrón del grupo II que se encuentra en la naturaleza, pudiendo comprender una modificación por al menos una delección de dicho ORF, y opcionalmente además por inserción de secuencias heterólogas.

25

30 Cuando el intrón rediseñado presenta una secuencia modificada, el método puede además comprender al menos una etapa previa o posterior a la modificación de las secuencias de hibridación, para deleccionar la región del intrón del grupo II inicial que codifica la proteína IEP y/o para insertar al menos una secuencia de ADN heterólogo como pueden ser un promotor constitutivo, un promotor inducible y/o un marcador de resistencia, entre otros. Preferentemente, en estas realizaciones del método de la invención, el intrón de partida es un intrón del grupo II de tipo silvestre.

35

A partir de los intrones del grupo II conocidos en el estado de la técnica actual, y según la presente descripción, el experto medio puede identificar los intrones del grupo II de partida que son susceptibles de ser rediseñados por el método de la presente invención, como puede ser entre otros, y sin que sirva de limitación, el intrón del grupo II de *Sinorhizobium meliloti*, Rmlnt1.

5

El intrón del grupo II de *Sinorhizobium meliloti*, Rmlnt1, tiene la EBS1 formada por siete nucleótidos, posiciones 267-273 y tiene la secuencia 5'-UUUCGUC-3' (Fig.1). La EBS2 del ARN de Rmlnt1 está localizada en la posición 236-240 y tiene la secuencia 5'-AUGAA-3' (Fig.1). La posición 329 del ARN de Rmlnt1 corresponde con la EBS3 y es una "G" (Fig.1). Las posiciones referidas de las EBSs están dadas según las posiciones de la SEQ ID NO:1.

10

SEQ ID NO: 1 corresponde a la secuencia de ADN del intrón del grupo II Rmlnt1 de *Sinorhizobium meliloti*, en el cual los nucleótidos comprendidos entre las posiciones 547 y 1803 definen la secuencia nucleotídica que codifica la proteína IEP de tipo silvestre de Rmlnt1 SEQ ID NO: 2.

15

La IEP de tipo silvestre de Rmlnt1, SEQ ID NO:2, está codificada por un marco abierto de lectura situado en el dominio IV del ARN y tiene una longitud de 419 aminoácidos. Posee un dominio reverso transcriptasa (RT) entre las posiciones 17-291 de SEQ ID NO:2, y un dominio maturasa (X) entre las posiciones 307-399 de SEQ ID NO:2, careciendo del dominio de unión a ADN (D) y el dominio endonucleasa (En) que poseen otras IEP codificadas por otros intrones como Ll.ItrB y Ecl5 (Fig.2).

20

Dicho intrón Rmlnt1 de partida puede ser un intrón del tipo silvestre, es decir, una secuencia que es idéntica a la secuencia del intrón del grupo II que se encuentra en la naturaleza o bien un intrón Rmlnt1 modificado, es decir, una secuencia que sea diferente de la secuencia del intrón del grupo II Rmlnt1 encontrada en la naturaleza, preferentemente modificada por comprender al menos una delección del ORF que codifica la IEP en el dominio IV del intrón, pudiendo comprender además una inserción de al menos una secuencia heteróloga en dicho dominio IV sin ORF, tal como son por ejemplo, un gen de resistencia, un promotor inducible, un promotor constitutivo, etc. Asimismo, la secuencia del intrón Rmlnt1 modificado puede comprender una o más inserciones de secuencias heterólogas en el dominio IV del intrón, tanto si existe o no la delección del ORF.

30

35

Cuando el intrón Rmlnt1 rediseñado presenta una secuencia modificada por delección del ORF y/o inserción una secuencia heteróloga, el método puede además comprender al menos una etapa previa o posterior a la modificación de las secuencias de hibridación, para eliminar la región del intrón Rmlnt1 inicial que codifica la proteína IEP, y/o para insertar al menos una secuencia de ADN heterólogo como pueden ser un promotor constitutivo, un promotor inducible y/o un marcador de resistencia, entre otros. Preferentemente, en estas realizaciones del método de la invención, el intrón Rmlnt1 de partida es de tipo silvestre, y más preferentemente consiste en la secuencia SEQ ID NO: 1.

La diana que los intrones del grupo II reconocen está formada por 35-25 nucleótidos de los que 25-20 están localizados corriente arriba del sitio de inserción del intrón denominado exón 1 y 5-10 nucleótidos localizados corriente abajo del sitio de inserción del intrón denominado exón 2. En el exón 1 se encuentra una región denominada IBS1 que hibrida con la secuencia EBS1 del ARN del intrón. Separada un nucleótido de la IBS1 se encuentra otra secuencia denominada EBS2 que hibrida con la región EBS2 del ARN del intrón. El resto de nucleótidos del exón 1 no hibridan con el ARN del intrón pero son reconocidos por la proteína que este codifica en el dominio IV y se denominan parte distal del exón 1. El nucleótido +1, con respecto al sitio de inserción del intrón, del exón 2 es el que hibrida con la EBS3 del ARN del intrón y se denomina IBS3. El resto de nucleótidos del exón 2 no hibridan con el ARN del intrón pero son reconocidos por la proteína que codifica el intrón y se denominan parte distal del exón 2.

La diana natural del intrón Rmlnt1 de tipo silvestre, comprende una secuencia específica de 25 nucleótidos (SEQ ID NO: 3, Fig.3), de ellos, 20 están localizados corriente arriba del sitio de inserción de Rmlnt1 y se denominan exón 1. Los otros 5 nucleótidos están localizados corriente abajo del sitio de inserción del intrón y se denominan exón 2. En el exón 1 se encuentra la secuencia IBS1 cuya secuencia silvestre es 5'-GATGAGA-3', a continuación se encuentra la región IBS2 separada un nucleótido "C" de la IBS1. La secuencia de la IBS2 en la diana silvestre es 5'-TTCAT-3'. A continuación se encuentran los nucleótidos de la región distal del exón 1 cuya secuencia en la diana silvestre es 5'-CCTCGTT-3' esta región no es reconocida por apareamiento con el ARN del intrón Rmlnt1 pero si es reconocida por la proteína IEP que este codifica. En el exón 2 se encuentra la EBS3 que en la diana silvestre corresponde a una "C". El resto de las posiciones del exón 2 forman la región denominada parte distal del exón 2 y su secuencia en la diana silvestre es 5'-CTGG-3'. Esta parte distal del exón 2 tampoco es reconocida por apareamiento del ARN del intrón Rmlnt1 pero si por la proteína IEP que este codifica.

Para determinar la mejor diana dentro de una secuencia de ADN para el intrón del grupo II se puede analizar la secuencia con un algoritmo informático que permita clasificar las dianas según su eficiencia de invasión, preferentemente mediante datos de frecuencias obtenidas en experimentos de selección para cada uno de los nucleótidos en cada una de las posiciones de la diana y de las frecuencias de partida antes de la selección, y más preferentemente utilizando los datos presentados en la figura 4. Este algoritmo permite analizar la totalidad de la secuencia del ADN en el que se quiere determinar una diana para el intrón del grupo II, en grupos de nucleótidos según el tamaño de la diana a seleccionar, preferentemente en grupos de 25 a 35 nucleótidos, y más preferentemente de 25 nucleótidos, que los va desplazando de nucleótido en nucleótido a lo largo de toda la secuencia. El programa presenta una lista de posibles dianas ordenadas según una puntuación. Preferentemente, la puntuación (S) se obtiene según un modelo de Markov de orden cero aplicando la ecuación:

$$S = \sum_p \log_2 \frac{f_{n_p(s)}}{f_{n_p}}$$

donde $f_{n_p(s)}$ es la frecuencia con que aparece el nucleótido n en la posición p de las frecuencias obtenidas en experimentos de selección para cada uno de los nucleótidos en cada una de las posiciones de la diana (seleccionados), tal como son las frecuencias presentadas en la figura 4 marcadas como seleccionadas. f_{n_p} es la frecuencia con que aparece un nucleótido n en la posición p de las frecuencias de partida antes de la selección (sin seleccionar), tal como son las frecuencias presentadas en la figura 4 marcadas como sin seleccionar.

A mayor puntuación la diana se invadirá con más eficiencia. Preferentemente, según la presente invención, un porcentaje de invasión eficiente se encuentra con valores de S por encima de 5,35 y que permite detectar la inserción del intrón en su diana mediante métodos sencillos como PCR sin el uso de selección. Este método mejora los métodos utilizados hasta ahora basados sólo en homología con el ADN diana natural de Rmlnt1 y basados sólo en dianas de 20 nucleótidos (patente ES2190866) porque tiene en cuenta los requerimientos de secuencia de los 25 nucleótidos de la diana. Una vez seleccionada la mejor diana de la lista que el algoritmo genera, las EBSs del intrón del grupo II se modificarán para que reconozcan a las IBS de la diana seleccionada.

Un diagrama de flujo del algoritmo informático usado se encuentra en la figura 7.

Para la selección de las dianas para el intrón del grupo II Rmlnt1 en un ADN determinado, se analiza la secuencia de ese ADN utilizando un algoritmo informático que preferentemente utiliza los datos presentados en la figura 4. Estos datos corresponden a la frecuencia con que aparece cada nucleótido en cada una de las posiciones de la diana obtenidas por experimentos de selección usando dianas randomizadas en cada una de las posiciones. Estos experimentos de selección se pueden realizar utilizando por ejemplo dos librerías de plásmidos, una contiene el intrón del grupo II con las secuencias EBS1, 2 y 3 randomizadas y además tiene insertado el promotor del fago T7 en el dominio IV. La otra librería tiene clonada la diana del intrón de 25 nucleótidos con cada una de las 25 posiciones también randomizadas, delante de un gen que confiere resistencia a tetraciclina aunque este gen no tiene promotor y en condiciones normales no proporciona resistencia a tetraciclina. Las dos librerías se transforman en HMS174 (DE3), una cepa de *E. coli* que expresa la RNA polimerasa del fago T7. Las bacterias transformadas se siembran en medio que contenga tetraciclina y en medio que contenga ampicilina. En el medio con tetraciclina crecerán las bacterias en la que se haya producido la invasión de la diana por el intrón que al contener el promotor del fago T7 producirá la transcripción del gen de resistencia a tetraciclina y la bacteria se convierte en resistente a este antibiótico. De estas colonias se extraen los plásmidos y se secuencian tanto la diana que ha sido invadida como el intrón que la invadió. Estos forman los datos recogidos en la figura 4 y marcados como seleccionadas. Las colonias que crecen en el medio con ampicilina son aquellas que tienen los plásmidos con las dianas del intrón randomizadas en las que no ha habido salto, estas colonias también poseen los plásmidos donadores del intrón con las EBSs randomizadas aunque estos no hayan saltado. A estas colonias también se les extraen los plásmidos y se secuencian tanto la diana del intrón como el intrón. Estos datos están recogidos en la figura 4 y marcados como no seleccionadas. Estos experimentos nos determinan que nucleótido es el óptimo, para que el intrón invada su diana, en cada posición independientemente del nucleótido presente en la diana silvestre que no siempre es el óptimo. El algoritmo analiza la totalidad de la secuencia del ADN en el que se quiere determinar una diana para el intrón del grupo II, Rmlnt1, en grupos de 25 que los va desplazando de nucleótido en nucleótido a lo largo de toda la secuencia. El programa presenta una lista de posibles dianas ordenadas según una puntuación. A mayor puntuación la diana se invadirá con más eficiencia. Este método mejora a los métodos utilizados hasta ahora en los que la selección de la diana se basaba sólo en homología con el ADN diana natural del intrón del grupo II, utilizaba dianas de sólo 20 nucleótidos (patente ES2190866).

En los métodos anteriores, las IBS de las dianas seleccionadas podían tener cualquier secuencia, sin tener en cuenta que hay ciertas posiciones de las IBS que deben estar ocupadas por nucleótidos determinados (Fig. 5), este requerimiento si es tenido en cuenta en este nuevo método que discrimina entre las dianas que poseen esos nucleótidos críticos en las posiciones específicas de todas las posibles dianas. Así mismo, el presente método también analiza los nucleótidos críticos de la parte distal de los exones que en el anterior método no se realizaba. Además los métodos anteriores modifican la posición “ δ ” del intrón localizada justo antes de la EBS1 (Fig.1) para que interaccione con la posición +1 de la diana. Esta posición no interacciona con el nucleótido +1 de la diana sino que es la posición EBS3 la que interacciona con la diana (Fig.1 y Fig.3). En el método definido en este aspecto de la invención, esta posición si es tomada en cuenta a la hora de la selección de la diana. Una vez seleccionada la mejor diana de la lista que el algoritmo genera las EBSs del intrón del grupo II se modificarán para que reconozcan a las IBSs de la diana seleccionada.

15 Modificación del ARN del intrón del grupo II:

Según el método definido en este aspecto de la invención, si las secuencias EBS1, EBS2 y EBS3 del ARN del intrón del grupo II silvestre o modificado, preferentemente del intrón Rmlnt1 silvestre o modificado, no son o son sólo parcialmente complementarias a las secuencia IBS1, IBS2 e IBS3 de la diana seleccionada las secuencias EBSs serán modificadas para hacerlas totalmente complementarias a las secuencias IBSs de la diana seleccionada. Dicha modificación de las secuencias de hibridación puede comprender modificar una secuencia de hibridación seleccionada entre EBS1, EBS2, EBS3, o una combinación de dos o más de las anteriores, ya sea de forma individual, por parejas o las tres conjuntamente. Las moléculas susceptibles de esta modificación son las correspondientes al intrón del grupo II, preferentemente Rmlnt1, silvestre o modificado clonado en vectores plasmídicos bajo cualquier promotor, a partir de ahora denominados plásmidos donadores, y las modificaciones se llevan a cabo mediante cualquier protocolo conocido de mutagénesis de ADN o utilizando sistemas comerciales. La modificación de las secuencias EBSs implica modificaciones complementarias en las secuencias IBSs de los exones contenidos en el plásmido donador para asegurar la maduración eficiente del ARN del intrón así modificado. El plásmido donador que contiene el intrón del grupo II Rmlnt1 modificado a partir del silvestre o de un intrón del grupo II Rmlnt1 previamente modificado es replicado en cualquier organismo vivo silvestre, mutante o modificado mediante ingeniería genética.

Vehículo para el intrón modificado:

5 El intrón modificado puede ser utilizado en cualquier aplicación de ingeniería genética, terapia génica, análisis funcional de genomas, mutagénesis, control de la expresión génica, marcaje de genes, reparación de genes no funcionales, contención biológica, etc. El intrón modificado puede ser aportado *in vivo* expresado a partir de cualquier vector conocido como plásmidos, T-DNA, adenovirus, etc., o directamente como RNPs (Ribonucleopartículas) obtenidas a partir del intrón correspondiente.

10 Ejemplos

Los siguientes ejemplos específicos sirven para ilustrar la naturaleza de la presente invención. Estos ejemplos se incluyen solamente con fines ilustrativos y no han de ser interpretados como limitaciones a la invención que aquí se reivindica. Por tanto, los ejemplos descritos más adelante ilustran la invención sin limitar el campo de aplicación de la misma.

15 El uso del método descrito en la presente memoria para la identificación de la mejor diana en cualquier secuencia de ADN, para la inserción eficiente de un intrón del grupo II mejorará su uso en diferentes aplicaciones como mutagénesis, estudio de la función génica, control de la expresión génica, marcado de genes, reparación de genes no funcionales, contención biológica entre otros.

Ejemplo 1. Selección de las mejores dianas en un gen para su mutagénesis por intrones del grupo II. Uso de los intrones del grupo II para mutar el gen *lacZ* en el cromosoma de RMO17

25 El gen *lacZ* cuando está presente en las bacterias y en el medio se encuentra x-gal las colonias de las bacterias se vuelven de color azul. El primer paso para mutar el gen *lacZ* mediante el uso de intrones del grupo II es identificar las mejores dianas para ser invadidas por el intrón del grupo II. La identificación de estas dianas se hace aplicando a la secuencia del gen el algoritmo informático anteriormente descrito. Este algoritmo nos muestra unas posibles dianas ordenadas en función de una puntuación (S) calculada por el algoritmo utilizando las frecuencias presentadas en la figura 4. La lista de las dianas seleccionadas aparece en la figura 6.

35 A continuación se modifican las regiones EBSs del intrón de forma que estas aparezcan totalmente con las regiones IBS de las dianas seleccionadas como aparece en la figura 6.

La modificación de las EBSs se realiza mediante mutagénesis puntual mediante PCR solapantes. En un primer PCR se introducen las modificaciones necesarias en la IBS2, IBS1 y EBS2. El segundo PCR modifica la EBS1. Los productos purificados de estos PCRs se utilizan como molde para el segundo PCR. El fragmento de PCR resultante contiene las EBSs e IBSs modificadas y se utiliza para sustituir las EBSs e IBSs silvestres del intrón mediante las enzimas de restricción *SacI* y *XhoI*.

El plásmido que contiene el intrón se introduce en la bacteria RMO17 mediante conjugación. Tras el crecimiento de los transconjugantes en medio Ty líquido estos se siembran en medio sólido conteniendo x-gal. Las bacterias donde se ha producido la invasión darán lugar a colonias blancas mientras que las colonias en las que el gen *lacZ* no se ha invadido producirán colonias azules. Calculando el porcentaje de colonias blancas con respecto al de azules se obtiene el porcentaje de invasión del gen *lacZ*. En la figura 6 se muestran los resultados observándose que las dianas con mayor puntuación se invaden mejor que las de menor puntuación.

REIVINDICACIONES

1. Un método para rediseñar un intrón del grupo II para su inserción eficiente en un ADN caracterizado por que, a partir de un intrón inicial del grupo II que codifica una proteína IEP carente del dominio endonucleasa y/o de unión al ADN, comprende:
- seleccionar al menos una diana que comprende al menos 25 nucleótidos consecutivos en la secuencia de dicho ADN, mediante la aplicación de un algoritmo que determina la mejor diana para dicho intrón inicial según su eficiencia de invasión, donde la diana seleccionada comprende:
 - a. un sitio de inserción para el intrón situado entre dos nucleótidos consecutivos comprendidos en la diana seleccionada,
 - b. dos regiones distales que son reconocidas por la IEP del intrón del grupo II, una primera región, la región distal del exón 1 que comprende los nucleótidos en posiciones -14 a -20 respecto al sitio de inserción de la secuencia diana seleccionada, y una segunda región, la región distal del exón 2 que comprende los nucleótidos en posiciones +2 a +5 respecto al sitio de inserción de la secuencia diana seleccionada,
 - c. tres regiones IBS1, IBS2 e IBS3, que hibridan con las secuencias EBS1, EBS2 y EBS3 del intrón del grupo II en su forma silvestre, respectivamente, donde la región IBS1 comprende los nucleótidos en posiciones -1 a -7 respecto al sitio de inserción de la secuencia diana seleccionada, la región IBS2 comprende los nucleótidos en posiciones -9 a -13 respecto al sitio de inserción de la secuencia diana seleccionada, y la región IBS3 consiste en el nucleótido en posición +1 respecto al sitio de inserción de la secuencia diana seleccionada;
 - modificar total o parcialmente al menos una de las secuencias de hibridación EBS1, EBS2 y EBS3 del intrón del grupo II inicial para que sean totalmente complementarias con las respectivas secuencias IBS1, IBS2 e IBS3 de la secuencia diana seleccionada en la etapa anterior, y así obtener un intrón rediseñado para la inserción eficiente en dicha secuencia diana.
2. Un método según la reivindicación 1, donde la diana seleccionada consiste en 25 nucleótidos.
3. Un método según la reivindicación 2, donde:
- el sitio de inserción para el intrón está situado entre los nucleótidos en posiciones 20 y 21 respecto al extremo 5' de la secuencia diana seleccionada,

- la región distal del exón 1 comprende los nucleótidos en posiciones 1 a 7 respecto al extremo 5' de la secuencia diana seleccionada,
- la región distal del exón 2 comprende los nucleótidos en posiciones 22 a 25 respecto al extremo 5' de la secuencia diana seleccionada,
- 5 - la región IBS1 comprende los nucleótidos en posiciones 14 a 20 respecto al extremo 5' de la secuencia diana seleccionada,
- la región IBS2 comprende los nucleótidos en posiciones 8 a 12 respecto al extremo 5' de la secuencia diana seleccionada, y
- la región IBS3 consiste en el nucleótido en posición 21 respecto al extremo 5' de la
10 secuencia diana seleccionada.

4. Un método según una cualquiera de las reivindicaciones 1 a 3, donde la modificación de las secuencias de hibridación comprende modificar total o parcialmente al menos una secuencia de hibridación seleccionada entre EBS1, EBS2, EBS3 y una combinación de dos
15 o más de las anteriores, de forma individual, por parejas o las tres conjuntamente.

5. Un método según una cualquiera de las reivindicaciones 1 a 4 que, previa o posteriormente a la modificación de las secuencias de hibridación, comprende además:
- delecionar la región del intrón del grupo II inicial que codifica la proteína IEP.

20

6. Un método según una cualquiera de las reivindicaciones 1 a 5 que además comprende:

- insertar al menos una secuencia de ADN heterólogo seleccionado entre un promotor constitutivo, un promotor inducible y un marcador de resistencia.

25

7. Un método según una cualquiera de las reivindicaciones 1 a 6, en el que la proteína IEP carente de dominio endonucleasa es la proteína IEP que codifica el intrón RmInt1.

8. Un método según la reivindicación 7 donde la proteína IEP que codifica el intrón
30 RmInt1 es SEQ ID NO: 2.

9. Un método según una cualquiera de las reivindicaciones 1 a 6 donde el intrón del grupo II inicial es un intrón RmInt1 silvestre o modificado al menos por una delección en su dominio ORF.

35

10. Un método según la reivindicación 9 donde el intrón del grupo II inicial es el intrón RmInt1 silvestre SEQ ID NO: 1.
- 5 11. Un intrón del grupo II rediseñado para su inserción eficiente en una secuencia de ADN, caracterizado por que se obtiene por un método definido en una cualquiera de las reivindicaciones 1 a 10.
12. Uso de al menos un intrón del grupo II rediseñado según se define en la reivindicación 11 en una aplicación biotecnológica *in vivo* o *in vitro*.
- 10 13. Uso de al menos un intrón del grupo II rediseñado según se define en la reivindicación 11 para el marcaje de un gen.
14. Uso de al menos un intrón del grupo II rediseñado según se define en la
15 reivindicación 11 para su empleo en mutagénesis.
15. Uso de al menos un intrón del grupo II rediseñado según se define en la reivindicación 11 para el control de la expresión génica.
- 20 16. Uso de al menos un intrón del grupo II rediseñado según se define en la reivindicación 11 para la reparación de genes no funcionales.
17. Uso de al menos un intrón del grupo II rediseñado según se define en la reivindicación 11 para contención biológica.
- 25 18. Un vector de expresión que comprende un intrón del grupo II rediseñado según se define en la reivindicación 11.

FIGURAS

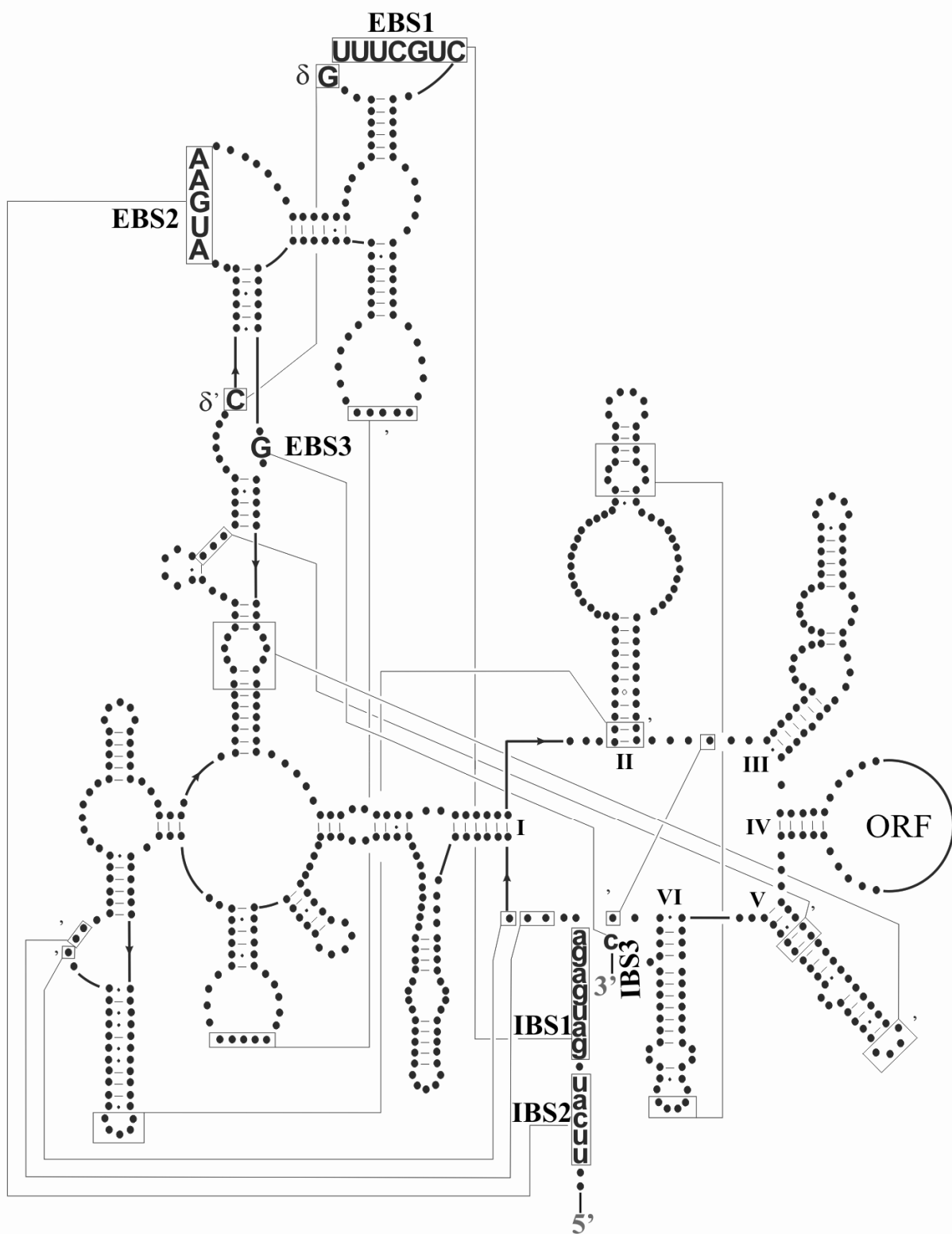


FIG. 1

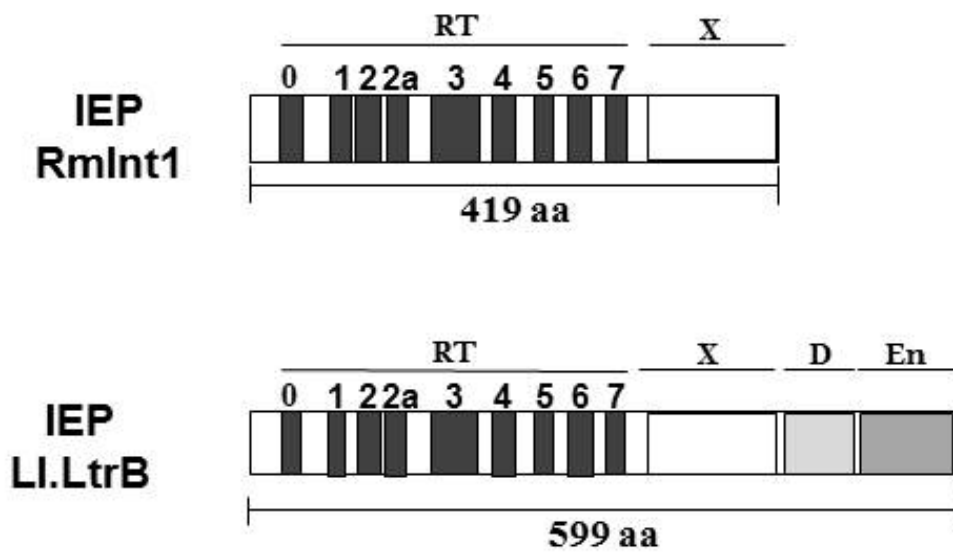


FIG. 2

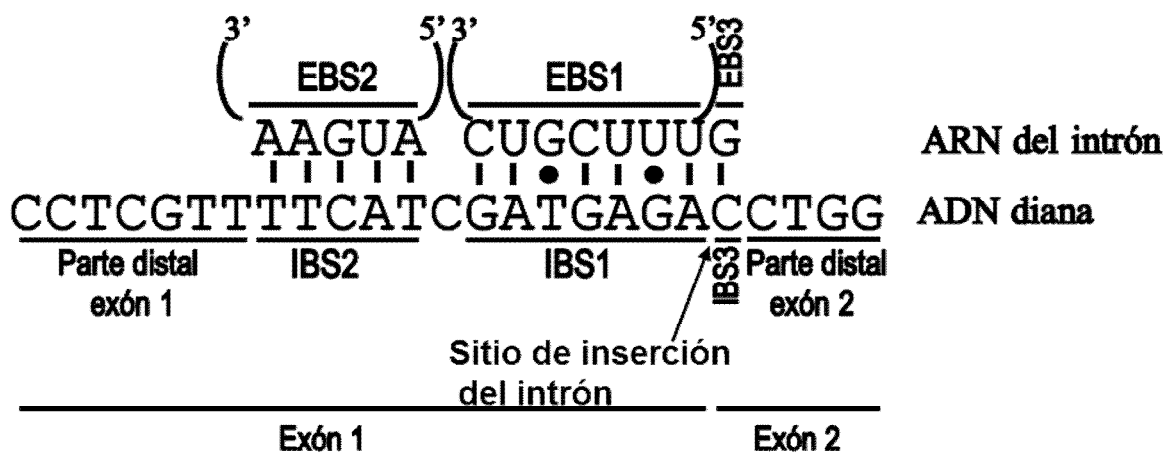


FIG. 3

Parte distal exón 1																IBS2											IBS1											Parte distal exón 2																						
C C T C G T T T T C A T C																-20 -19 -18 -17 -16 -15 -14 -13 -12 -11 -10 -9											G A T G A G A C G A C G											-7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5																						
A	2	24	26	30	32	32	25	16	8	9	17	14	21	30	19	27	19	30	23	18	26	34	39	20	18	11	11	29	24	31	30	22	36	27	29	30	32	23	30	11	31	17	21	27	15	21	13	18	30	37	40	33	26	21	26	43	23	18	19	31
C	19	19	23	12	14	19	16	13	19	21	17	16	7	16	7	23	25	36	33	1	44	29	4	42	7	12	9	17	42	20	10	84	30	25	8	9	75	41	47	29	7	13	32	8	12	6	77	25	12	32	37	31	26	31	25	7	14	39	11	25
G	25	29	20	26	30	23	30	30	16	20	20	24	7	12	9	17	42	20	10	84	30	25	8	9	75	41	47	29	7	13	32	8	12	6	77	25	12	32	37	31	26	31	25	7	14	39	11	25												
T	34	29	31	31	23	26	30	41	57	50	47	46	7	16	7	23	25	36	33	1	44	29	4	42	7	12	9	17	42	20	10	84	30	25	8	9	75	41	47	29	7	13	32	8	12	6	77	25	12	32	37	31	26	31	25	7	14	39	11	25
A	21	28	44	36	48	1	32	2	0	0	5	2	7	16	7	23	25	36	33	1	44	29	4	42	7	12	9	17	42	20	10	84	30	25	8	9	75	41	47	29	7	13	32	8	12	6	77	25	12	32	37	31	26	31	25	7	14	39	11	25
C	4	11	19	15	1	1	14	3	18	22	64	0	7	12	9	17	42	20	10	84	30	25	8	9	75	41	47	29	7	13	32	8	12	6	77	25	12	32	37	31	26	31	25	7	14	39	11	25												
G	15	11	13	21	43	0	18	12	6	2	9	0	75	41	47	29	7	13	32	8	12	6	77	25	75	41	47	29	7	13	32	8	12	6	77	25	75	41	47	29	7	13	32	8	12	6	77	25	75	41	47	29	7	13	32	8	12	6	77	25
T	60	50	24	28	9	98	35	83	76	76	21	98	12	32	37	31	26	31	25	7	14	39	11	25	12	32	37	31	26	31	25	7	14	39	11	25	12	32	37	31	26	31	25	7	14	39	11	25												

FIG. 4

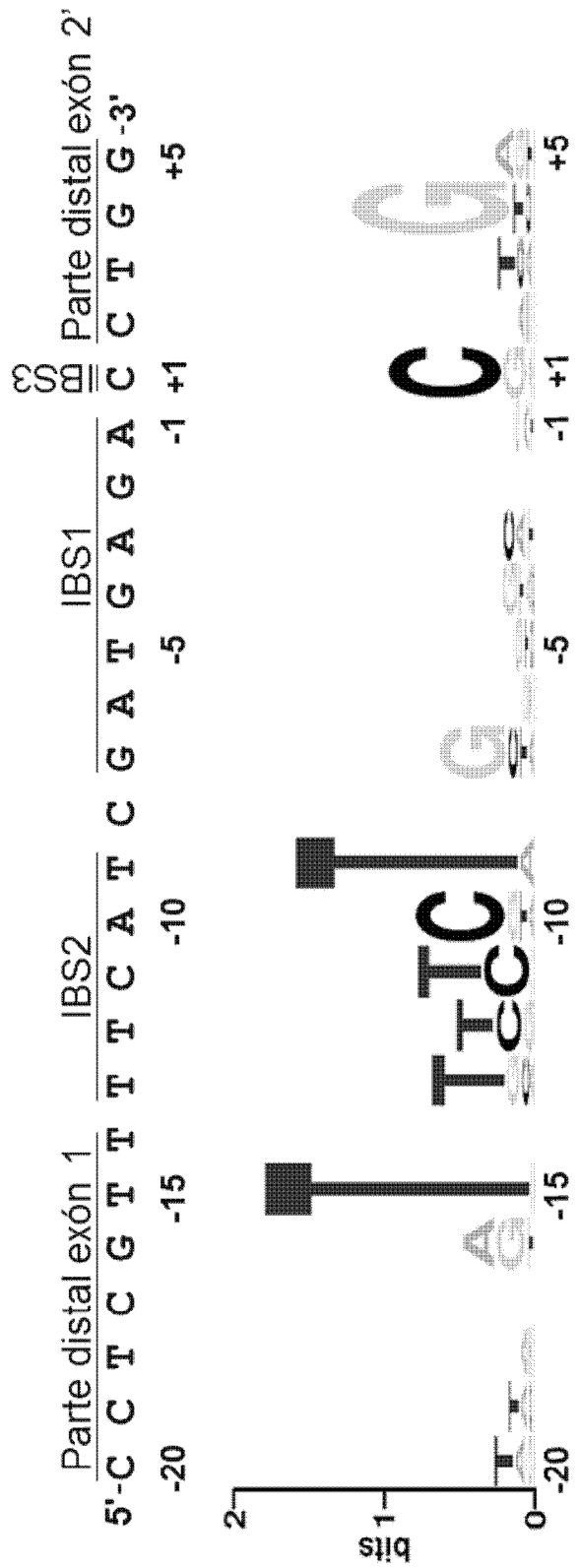


FIG. 5

			Puntuación (S)	% Invasión
WT	<pre> EBS2 EBS1 EBS3 AAGUA CUGC UUUG CCTCGT-15TTCATCGATGAGACCTGG+4 IBS2 IBS1 IBS3 </pre>			
2222as	<pre> AGCGA CACCAGUC ATCCAT-15TTCGCTGGTGGTCAGATGC+4 </pre>	7.89	38.28 ± 12.22	
2095as	<pre> CGUGA GCGCAUGA TTCGG-15TTCGACTACGCGTACTGTGA+4 </pre>	7.15	36.60 ± 2.86	
2021s	<pre> GGAGA CTACAGCG TGAAG-15TGCCTCTGGATGTCGCTCCA+4 </pre>	6.78	25.97 ± 1.23	
1496as	<pre> AGCCA CGGCACCA AATAA-15TATCGGTGGCCGTGGTGTTCG+4 </pre>	5.35	2.79 ± 0.79	
1585s	<pre> CGAAA CGATGGAC AAAAA-15TGGCTTTCGCTACCTGGAGA+4 </pre>	4.62	0.52 ± 0.57	

FIG. 6

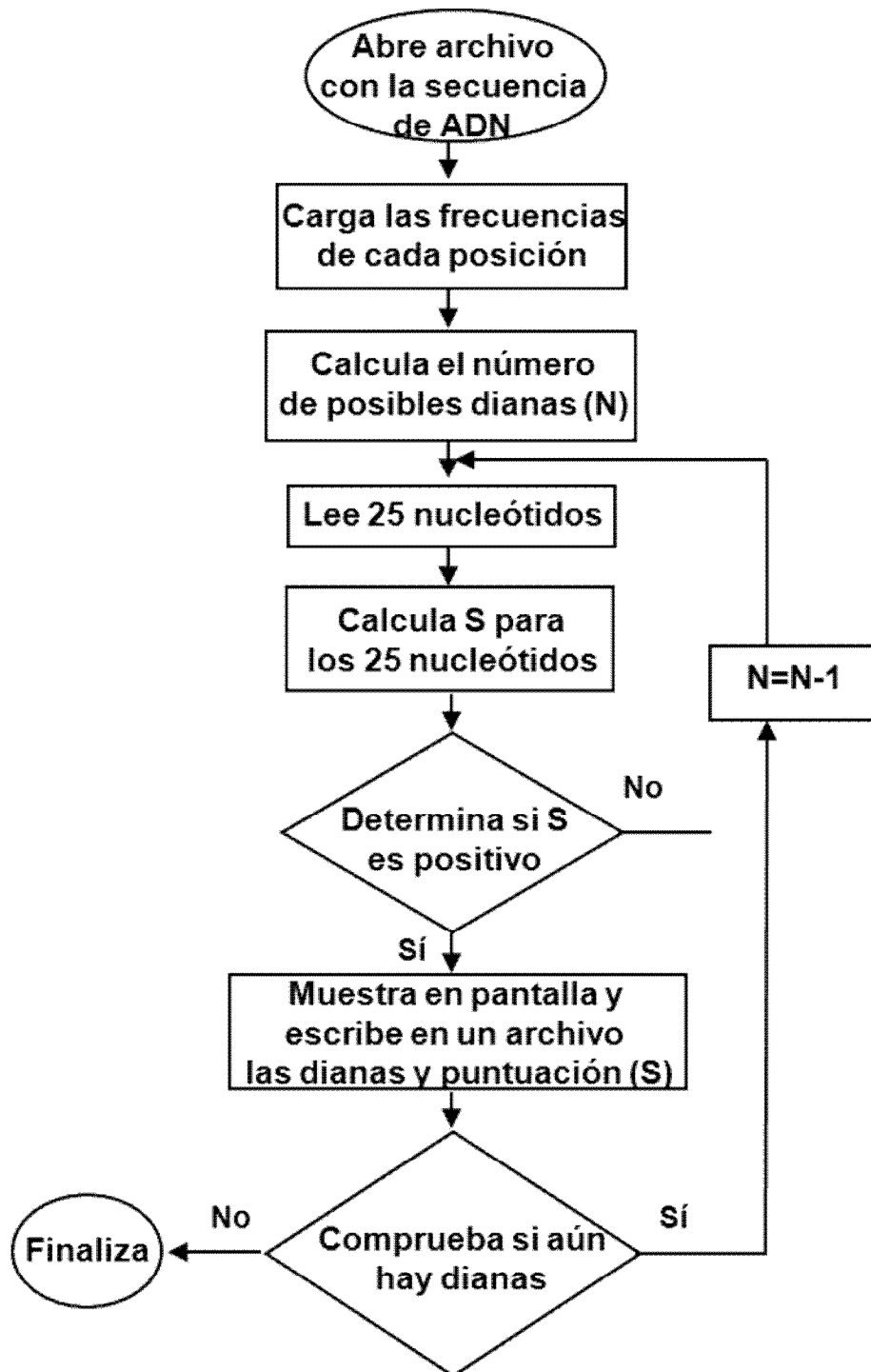


FIG. 7

LISTA DE SECUENCIAS

- <110> Consejo Superior de Investigaciones Científicas
- <120> Método para la identificación de dianas para intrones del grupo II en cualquier secuencia de ADN y su inserción eficiente
- <130> 2013-0377
- <160> 3
- <170> PatentIn version 3.5
- <210> 1
- <211> 1884
- <212> DNA
- <213> Sinorhizobium meliloti

- <220>
- <221> CDS
- <222> (547)..(1803)

```

<400> 1
gtgtgctgca gaggcacgga aggagttaa catgaactaa gaccgtggcg taaagctgcg 60
tgaatgatgg gggacggccc tccgggatcg gctttcagga gcggttctca aaccagtccg 120
agctgctgcg gtaaagagcc gtggtggtga gcgtcggatg aaacgttcgg acgagatccg 180
agcaggtgca tgtccaaaag acgaacgaaa gtgaaccctc cgaggacgcg tcgttatgaa 240
cgtaagtgtc gtcgaaacca ggaccgtttc gtcattcctg gacaagtccg ccagatgcct 300
gatgaccggg cgggcggcga ccggcgtaga gggggcgtga gttggacata ggctttcacg 360
cggaactgca ggaaccaggc tcctgatgtc aagggagaag ctcaagcggc gcaaaccgca 420
aggcgagagt accgatgcag gagactgggg cggatcggcc cgtatgagcg tcgaggacct 480
tgtaatgggg tcggagcaaa gggggcggat caggccgtcg tattgtttga aacaactgga 540
aacagg atg act tcg gaa agt acg aca gac aag ccg ttt cga att gag      588
      Met Thr Ser Glu Ser Thr Thr Asp Lys Pro Phe Arg Ile Glu
      1          5          10

aaa cgt cga gtg tac gaa gct tac aaa gcg gtc aaa gcc aac cgt ggc      636
Lys Arg Arg Val Tyr Glu Ala Tyr Lys Ala Val Lys Ala Asn Arg Gly
15          20          25          30

gcg gcc ggg gtg gac ggg cag acg ctg gag ata ttt gag aaa gac ctt      684
Ala Ala Gly Val Asp Gly Gln Thr Leu Glu Ile Phe Glu Lys Asp Leu
          35          40          45

gca gca aac ctc tac aag atc tgg aat cgg atg tcc tcg gga acc tac      732
Ala Ala Asn Leu Tyr Lys Ile Trp Asn Arg Met Ser Ser Gly Thr Tyr
          50          55          60

ttt ccg ccg ccg gtg cgc gcc gtc tcc att ccg aag aag gct gga ggc      780
Phe Pro Pro Pro Val Arg Ala Val Ser Ile Pro Lys Lys Ala Gly Gly
          65          70          75

gaa agg gtt ttg ggt gtg ccc acg gtc agc gat cgg atc gcg cag atg      828
Glu Arg Val Leu Gly Val Pro Thr Val Ser Asp Arg Ile Ala Gln Met
          80          85          90

gtg gtc aag cag atg atc gag ccg gat ttg gac tcc ctc ttt ctt ccg      876
Val Val Lys Gln Met Ile Glu Pro Asp Leu Asp Ser Leu Phe Leu Pro
          95          100          105          110

```

gac Asp	tcc Ser	tac Tyr	ggt Gly	tac Tyr 115	agg Arg	ccg Pro	gga Gly	aaa Lys	tcg Ser 120	gcc Ala	ctg Leu	gat Asp	gct Ala	gtc Val 125	gga Gly	924
gtg Val	acg Thr	cgt Arg	cag Gln 130	cgg Arg	tgc Cys	tgg Trp	aag Lys	tat Tyr 135	gat Asp	tgg Trp	ggt Val	ttg Leu	gaa Glu 140	ttc Phe	gac Asp	972
atc Ile	aaa Lys	ggg Gly 145	ctg Leu	ttt Phe	gac Asp	aat Asn	ctt Leu 150	ccg Pro	cat His	gat Asp	ctc Leu	ttg Leu 155	ctg Leu	aag Lys	gcg Ala	1020
gtc Val	aga Arg 160	aaa Lys	gac Asp	gtc Val	aaa Lys	tgc Cys 165	aac Asn	tgg Trp	gct Ala	ctg Leu	ctc Leu 170	tac Tyr	atc Ile	gaa Glu	aga Arg	1068
tgg Trp 175	ctg Leu	act Thr	gcg Ala	cct Pro	atg Met 180	gaa Glu	aag Lys	aac Asn	gga Gly	gaa Glu 185	gtc Val	atc Ile	gag Glu	cgg Arg	tca Ser 190	1116
cgc Arg	ggt Gly	acc Thr	cca Pro	cag Gln 195	gga Gly	ggc Gly	gtg Val	gtt Val	agc Ser 200	ccg Pro	atc Ile	ttg Leu	gcg Ala	aat Asn 205	ctc Leu	1164
ttt Phe	ctg Leu	cac His	tat Tyr 210	gca Ala	ttt Phe	gat Asp	ctc Leu	tgg Trp 215	atg Met	acg Thr	cgg Arg	acg Thr	cat His 220	ccc Pro	gac Asp	1212
ctt Leu	cca Pro	tgg Trp 225	tgt Cys	cga Arg	tat Tyr	gcc Ala	gac Asp 230	gat Asp	ggt Gly	ctt Leu	gtt Val	cac His 235	tgc Cys	cag Gln	agc Ser	1260
gag Glu	caa Gln 240	caa Gln	gcc Ala	gaa Glu	gcc Ala	ctc Leu 245	agg Arg	gtg Val	gag Glu	ctg Leu	agt Ser 250	tct Ser	cgg Arg	ctg Leu	gca Ala	1308
gcg Ala 255	tgc Cys	gga Gly	ctt Leu	cag Gln 260	atg Met	cat His	ccg Pro	aca Thr	aag Lys	acc Thr 265	aag Lys	att Ile	gtc Val	tac Tyr	tgc Cys 270	1356
aag Lys	gat Asp	caa Gln	cgg Arg	cgc Arg 275	agg Arg	gag Glu	gcg Ala	tat Tyr	ccg Pro 280	aat Asn	gtc Val	acg Thr	ttc Phe	gac Asp 285	ttt Phe	1404
ctc Leu	ggg Gly	tat Tyr	cag Gln 290	ttc Phe	cgg Arg	ccg Pro	cga Arg	cgg Arg 295	gtg Val	gcg Ala	aac Asn	aca Thr	cag Gln 300	cgg Arg	gac Asp	1452
gag Glu	ttc Phe	ttc Phe 305	tgt Cys	ggc Gly	tac Tyr	acg Thr	cct Pro 310	gcg Ala	gtc Val	agt Ser	ccg Pro	acg Thr 315	gcg Ala	ctc Leu	aag Lys	1500
tcg Ser	atg Met 320	cgg Arg	gca Ala	acg Thr	atc Ile	aaa Lys 325	agt Ser	ttg Leu	aac Asn	atc Ile	ccg Pro 330	cgg Arg	cag Gln	acg Thr	ccg Pro	1548
ggg Gly 335	acg Thr	ctg Leu	gcc Ala	gaa Glu	ata Ile 340	gcc Ala	aaa Lys	cag Gln	ctc Leu	aat Asn 345	cca Pro	ctc Leu	ctt Leu	cgg Arg	gga Gly 350	1596
tgg Trp	att Ile	gcc Ala	tac Tyr	tat Tyr 355	gga Gly	cgg Arg	tac Tyr	agt Ser	cgt Arg 360	tcg Ser	gcc Ala	ctg Leu	tcc Ser	act Thr 365	ctg Leu	1644
gct Ala	gat Asp	tac Tyr 370	gtt Val	aat Asn	cag Gln	aaa Lys	ctc Leu	agg Arg 375	gct Ala	tgg Trp	atc Ile	agg Arg	cga Arg 380	aag Lys	ttc Phe	1692
aaa Lys	cgc Arg	ttt Phe	cag Gln	tcc Ser	cat His	aag Lys	aca Thr	cgc Arg	gcc Ala	agc Ser	ctc Leu	ttc Phe	ttg Leu	cga Arg	aag Lys	1740

	385		390		395																
ctg	gcg	cgg	gaa	aat	ccg	ggg	ctg	ttc	gtg	cat	tgg	aag	gcg	ttc	gga						1788
Leu	Ala	Arg	Glu	Asn	Pro	Gly	Leu	Phe	Val	His	Trp	Lys	Ala	Phe	Gly						
	400					405					410										
acg	aac	acg	ttt	acc	tgatgggagc	ggtgtgaatc	gagaggttca	cgaccgttc													1843
Thr	Asn	Thr	Phe	Thr																	
	415																				
tgcgagaggc	cggctggtga	aactcctccg	gcctactcac	c																	1884

<210> 2
 <211> 419
 <212> PRT
 <213> Sinorhizobium meliloti

<400> 2

Met Thr Ser Glu Ser Thr Thr Asp Lys Pro Phe Arg Ile Glu Lys Arg
1 5 10 15

Arg Val Tyr Glu Ala Tyr Lys Ala Val Lys Ala Asn Arg Gly Ala Ala
20 25 30

Gly Val Asp Gly Gln Thr Leu Glu Ile Phe Glu Lys Asp Leu Ala Ala
35 40 45

Asn Leu Tyr Lys Ile Trp Asn Arg Met Ser Ser Gly Thr Tyr Phe Pro
50 55 60

Pro Pro Val Arg Ala Val Ser Ile Pro Lys Lys Ala Gly Gly Glu Arg
65 70 75 80

Val Leu Gly Val Pro Thr Val Ser Asp Arg Ile Ala Gln Met Val Val
85 90 95

Lys Gln Met Ile Glu Pro Asp Leu Asp Ser Leu Phe Leu Pro Asp Ser
100 105 110

Tyr Gly Tyr Arg Pro Gly Lys Ser Ala Leu Asp Ala Val Gly Val Thr
115 120 125

Arg Gln Arg Cys Trp Lys Tyr Asp Trp Val Leu Glu Phe Asp Ile Lys
130 135 140

Gly Leu Phe Asp Asn Leu Pro His Asp Leu Leu Leu Lys Ala Val Arg
145 150 155 160

Lys Asp Val Lys Cys Asn Trp Ala Leu Leu Tyr Ile Glu Arg Trp Leu
165 170 175

Thr Ala Pro Met Glu Lys Asn Gly Glu Val Ile Glu Arg Ser Arg Gly
180 185 190

Thr Pro Gln Gly Gly Val Val Ser Pro Ile Leu Ala Asn Leu Phe Leu
195 200 205

His Tyr Ala Phe Asp Leu Trp Met Thr Arg Thr His Pro Asp Leu Pro
210 215 220

Trp Cys Arg Tyr Ala Asp Asp Gly Leu Val His Cys Gln Ser Glu Gln
225 230 235 240

Gln Ala Glu Ala Leu Arg Val Glu Leu Ser Ser Arg Leu Ala Ala Cys
245 250 255

Gly Leu Gln Met His Pro Thr Lys Thr Lys Ile Val Tyr Cys Lys Asp
260 265 270

Gln Arg Arg Arg Glu Ala Tyr Pro Asn Val Thr Phe Asp Phe Leu Gly
275 280 285

Tyr Gln Phe Arg Pro Arg Arg Val Ala Asn Thr Gln Arg Asp Glu Phe
290 295 300

Phe Cys Gly Tyr Thr Pro Ala Val Ser Pro Thr Ala Leu Lys Ser Met
305 310 315 320

Arg Ala Thr Ile Lys Ser Leu Asn Ile Pro Arg Gln Thr Pro Gly Thr
325 330 335

Leu Ala Glu Ile Ala Lys Gln Leu Asn Pro Leu Leu Arg Gly Trp Ile
340 345 350

Ala Tyr Tyr Gly Arg Tyr Ser Arg Ser Ala Leu Ser Thr Leu Ala Asp
355 360 365

Tyr Val Asn Gln Lys Leu Arg Ala Trp Ile Arg Arg Lys Phe Lys Arg
370 375 380

Phe Gln Ser His Lys Thr Arg Ala Ser Leu Phe Leu Arg Lys Leu Ala
385 390 395 400

Arg Glu Asn Pro Gly Leu Phe Val His Trp Lys Ala Phe Gly Thr Asn
405 410 415

Thr Phe Thr

<210> 3
<211> 25
<212> DNA
<213> Sinorhizobium meliloti

<400> 3
cctcgttttc atcgatgaga cctgg

**DECLARACIÓN SOBRE EL CONTENIDO DEL SOPORTE ELECTRÓNICO Y EL
DOCUMENTO IMPRESO CON LAS SECUENCIAS**

Ref: Nueva solicitud de patente en España

Enunciado: “Método para la identificación de dianas para intrones del grupo II en cualquier secuencia de ADN y su inserción eficiente”

Titulares: CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

N/Ref: 2013-0377

El abajo firmante, Agente de la Propiedad Industrial, D. Javier UNGRÍA, como representante autorizado del solicitante en relación con la solicitud mencionada, declara que la información sobre la secuencia contenida en el soporte electrónico en formato TXT es idéntica a la secuencia que consta en el documento PDF que se aporta.

Madrid, 07 de enero de 2014.

Javier UNGRIA

