



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 344 802**

51 Int. Cl.:
C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **06757809 .6**

96 Fecha de presentación : **23.06.2006**

97 Número de publicación de la solicitud: **1910563**

97 Fecha de publicación de la solicitud: **16.04.2008**

54 Título: **Estrategias mejoradas para la secuenciación de genomas complejos utilizando tecnologías de secuenciación de alto rendimiento.**

30 Prioridad: **23.06.2005 US 693053 P**
08.09.2005 US 714897 P
16.01.2006 EP 06075104
17.01.2006 US 759034 P

45 Fecha de publicación de la mención BOPI:
07.09.2010

45 Fecha de la publicación del folleto de la patente:
07.09.2010

73 Titular/es: **KEYGENE N.V.**
Agro Business Park 90
6708 PW Wageningen, NL

72 Inventor/es:
Van Eijk, Michael, Josephus, Theresia;
Sørensen, Anker, Preben y
Hogers, René, Cornelis, Josephus

74 Agente: **Manresa Val, Manuel**

ES 2 344 802 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

Estrategias mejoradas para la secuenciación de genomas complejos utilizando tecnologías de secuenciación de alto rendimiento.

Campo técnico

La presente invención se refiere a los campos de la biología molecular y de la genética. La presente invención se refiere a unas estrategias mejoradas para determinar la secuencia de genomas preferentemente complejos (es decir, grandes), basándose en la utilización de técnicas de secuenciación.

Antecedentes de la invención

El ensamblaje secuencias aleatorias del genoma completas en genomas grandes (de 100 Mbp o superiores) con secuencias borrador del genoma es un tema complicado. Muchas plantas y animales contienen además un gran número de secuencias repetidas, con lo que se complica aún más el problema. Dicho problema informático aún es mayor con la aparición de las técnicas de alta capacidad de secuenciación, tales como las técnicas de 454 Life Science. Dichas técnicas ya no se basan en el método de didesoxisecuenciación de Sanger, sino fundamentalmente en la secuenciación por síntesis (pírosecuenciación), que resulta más sencilla de realizar en una superficie sólida. La secuenciación por síntesis proporciona una gran cantidad de secuencias, aunque de una longitud relativamente corta (aproximadamente 100 bp [pares de bases]) en comparación con la longitud relativamente grande de entre 500 y 1000 bp tal como resulta habitual en el método de didesoxisecuenciación de Sanger.

Una de las desventajas de las que adolecen dichos fragmentos cortos es que el ensamblaje de los cóntigos para determinar la secuencia genómica requiere una gran potencia informática, lo que hace que los procedimientos actuales de secuenciación constituyan una tarea relativamente costosa y lenta. Por consiguiente, existe la necesidad de procedimientos de secuenciación compleja que sean económicos, fiables y rápidos, es decir, grandes genomas, para reforzar la tecnología que a veces se ha denominado del “genoma de 1000 \$”, es decir, un procedimiento que permita la determinación de la secuencia entera de un genoma complejo (en particular humano) que no cueste más de 1000 \$. Ello permitiría, entre otras cosas, el desarrollo de medicaciones personalizadas.

Sumario de la invención

Los presentes inventores han descubierto ahora que dicho problema se puede solucionar con una estrategia distinta y que las técnicas de alta capacidad de secuenciación se pueden utilizar eficientemente en el ensamblaje del genoma.

La presente invención comprende la utilización de una técnica que divide el genoma en partes complementarias y reproducibles mediante la restricción del genoma con una o más endonucleasas de restricción para obtener un conjunto de fragmentos de restricción y proporcionar posteriormente un subconjunto de fragmentos de restricción mediante amplificación selectiva. Se secuencian y ensambla el subconjunto a un cóntigo. Mediante la repetición de dicha etapa para uno o más conjuntos distintos de endonucleasas de restricción, se obtienen distintos cóntigos. Dichos cóntigos distintos se utilizan para ensamblar la secuencia borrador del genoma. La presente invención no requiere conocimiento alguno de la secuencia y se puede aplicar a genomas de cualquier tamaño y complejidad. La presente invención se puede ajustar para cualquier tipo y tamaño de genoma. La presente invención proporciona un acceso más rápido, fiable y veloz a cualquier genoma de interés y permite, por lo tanto, acelerar el análisis del genoma.

Definiciones

En la descripción y ejemplos siguientes, se utiliza un cierto número de términos. A fin de proporcionar una comprensión clara y consistente de la presente memoria y reivindicaciones, comprendiendo el alcance a proporcionar a dichos términos, se proporcionan las siguientes definiciones. Excepto si se define de otro modo en la presente memoria, todos los términos técnicos y científicos tienen el mismo significado que les atribuyen habitualmente los expertos en la materia a la que pertenece la presente invención.

Ácido nucleico: un ácido nucleico según la presente invención puede comprender cualquier polímero u oligómero de bases de pirimidina y purina, preferentemente citosina, timina y uracilo, y adenina y guanina, respectivamente (véase Albert L. Lehninger, *Principles of Biochemistry* (“Principios de Bioquímica”), en 793-800 (Worth Pub. 1982)). La presente invención contempla cualquier desoxirribonucleótido, ribonucleótido o componente peptídico de un ácido nucleico, y cualquier variante química de los mismos, tales como las formas metiladas, hidroximetiladas o glucosiladas de los mismos. Los polímeros u oligómeros pueden presentar una composición heterogénea u homogénea, y se pueden aislar a partir de fuentes naturales o se pueden producir artificialmente o sintéticamente. Además, los ácidos nucleicos pueden ser ADN o ARN, o una mezcla de los mismos, y pueden existir de un modo permanente o transitorio en forma monocatenaria o bicatenaria, comprendiendo los estados homodúplex, heterodúplex e híbrido.

Reducción de la complejidad: el término reducción de la complejidad se utiliza para indicar un procedimiento en el que la complejidad de una muestra de ácido nucleico, tal como un ADN genómico, se reduce al generar un subconjunto de la muestra. Dicho subconjunto puede ser representativo de la muestra entera (es decir, complejo) y es

preferentemente un subconjunto reproducible. Reproducible significa en el presente contexto que cuando se reduce la complejidad de la misma muestra utilizando el mismo procedimiento, se obtiene el mismo subconjunto, o por lo menos uno de comparable. El procedimiento utilizado en la reducción de la complejidad puede ser cualquier procedimiento de reducción de la complejidad conocido en la técnica. Los ejemplos de procedimientos de reducción de la complejidad comprenden, por ejemplo, el AFLP® (Keygene N.V., Países Bajos; véase, por ejemplo, el documento EP 0 534 858), los procedimientos descritos por Dong (véanse, por ejemplo, los documentos WO 03/012118, WO 00/24939, EP 1 124 990) o Barany (por ejemplo el documento US 6 534 293), ligamiento indexado (Unrau *et al.*, véase posteriormente), etc. Los procedimientos de reducción de la complejidad utilizados en la presente invención tienen en común que son reproducibles. Reproducibles en el sentido de que cuando se reduce la complejidad de la misma muestra del mismo modo, se obtiene el mismo subconjunto de la muestra, al contrario de lo que sucede con una reducción de la complejidad más aleatoria tal como la microdissección o la utilización de ARNm (ADNc) que representa una parte del genoma transcrito en un tejido seleccionado y que su reproducibilidad depende del tejido, el período de aislamiento, etc.

Etiquetado: el término etiquetado se refiere a la adición de una etiqueta a una muestra de ácido nucleico a fin de poder distinguirla de una segunda u otra muestra de ácido nucleico. El etiquetado se puede realizar, por ejemplo, mediante la adición de un identificador de la secuencia durante la reducción de la complejidad o mediante cualquier otro medio conocido en la técnica. Dicho identificador de la secuencia puede, por ejemplo, ser una secuencia de bases única con una longitud variable pero definida que se utiliza para identificar una muestra específica de ácido nucleico. Los ejemplos habituales de las mismas son, por ejemplo secuencias ZIP. Al utilizar dicha etiqueta, se puede determinar el origen de una con un procesamiento adicional. En el caso de combinación de productos procesador procedentes de muestras de ácido nucleico distintas, se tendrían que identificar las distintas mezclas de ácidos nucleicos utilizando distintas etiquetas.

Genoteca etiquetada: el término genoteca etiquetada se refiere a una genoteca de ácido nucleico etiquetado.

Secuenciación: el término secuenciación se refiere a la determinación del orden de nucleótidos (secuencias de bases) en a muestra de ácido nucleico, por ejemplo ADN o ARN.

Alinear y alineación: los términos “alinear” y “alineación” se refieren a la comparación de dos o más secuencias de nucleótidos basándose en la presencia de extensiones cortas o largas de nucleótidos idénticos o similares. En la técnica se conocen diversos procedimientos de alineación de secuencias de nucleótidos, tal como se describirá con mayor detalle posteriormente. A veces se utiliza ensamblaje como sinónimo.

Técnicas de alta capacidad de identificación: las técnicas de alta capacidad de identificación, con frecuencia abreviadas como HTS, consisten en un procedimiento de experimentación científica especialmente adecuado para los campos de la biología y la química. Mediante una combinación de robótica actual y otro hardware especializado para laboratorios, permite al investigador identificar simultáneamente con efectividad grandes cantidades de muestras.

Endonucleasa de restricción: una endonucleasa de restricción o enzima de restricción es un enzima que reconoce unas secuencias específicas de nucleótidos (sitio seleccionado) en una molécula de ADN bicatenario y separará ambas cadenas de la molécula de ADN en cada sitio seleccionado.

Fragmentos de restricción: se hace referencia a las moléculas de ADN producidas mediante la digestión con una endonucleasa de restricción como fragmentos de restricción. Cualquier genoma determinado (o ácido nucleico, con independencia de su origen) se digerirá mediante una endonucleasa de restricción particular en un conjunto descrito de fragmentos de restricción. Los fragmentos de ADN que se obtienen a partir de la escisión mediante la endonucleasa de restricción se pueden continuar utilizando en diversas técnicas y, por ejemplo, se pueden detectar mediante electroforesis en gel.

Electroforesis en gel: a fin de detectar los fragmentos de restricción, se puede requerir un procedimiento analítico para fraccionar las moléculas de ADN bicatenario que se basa en el tamaño. La técnica que se utiliza más habitualmente para alcanzar dicho fraccionamiento es la electroforesis (capilar) en gel. La velocidad a la que dichos fragmentos de ADN se desplazan en dichos geles depende de su peso molecular; de este modo, la distancia de desplazamiento disminuye al aumentar la longitud del fragmento. Los fragmentos de ADN fraccionados mediante electroforesis en gel se pueden visualizar directamente mediante un procedimiento de tinción, por ejemplo, tinción con plata o una tinción que utilice bromuro de etidio, si el número de fragmentos que se encuentran en la banda es suficientemente pequeño. Alternativamente, un tratamiento adicional de los fragmentos de ADN puede incorporar etiquetas detectables en los fragmentos, tales como fluoróforos o etiquetas radiactivas.

Ligación: se hace referencia a la reacción enzimática catalizada por un enzima ligasa en la que dos moléculas de ADN bicatenario se unen entre sí con un enlace covalente como ligación. En general, ambas cadenas de ADN se unen entre sí con un enlace covalente, pero es asimismo posible evitar la ligación de una de las dos cadenas mediante la modificación química o enzimática de uno de los extremos de las cadenas. En dicho caso, el enlace covalente se producirá únicamente en una de las dos cadenas de ADN.

Oligonucleótido sintético: se hace referencia a las moléculas de ADN monocatenario que presentan preferentemente entre aproximadamente 10 y aproximadamente 50 bases, que se pueden sintetizar químicamente como oligo-

nucleótidos sintéticos. En general, dichas moléculas sintéticas de ADN se diseñan para que presenten una secuencia de nucleótidos única o pretendida, aunque resulta posible sintetizar familias de moléculas que presenten secuencias relacionadas y que presenten unas composiciones distintas en los nucleótidos en unas posiciones específicas de la secuencia de nucleótidos. El término oligonucleótido sintético se utilizará para hacer referencia a moléculas de ADN que presenten una secuencia de nucleótidos diseñada o pretendida.

Adaptadores: moléculas cortas de ADN bicatenario con un número limitado de pares de bases, por ejemplo entre aproximadamente 10 y aproximadamente 30 pares de bases en longitud, que se diseñan de tal modo que se puedan ligar a los extremos de fragmentos de restricción. Los adaptadores están compuestos generalmente de dos oligonucleótidos sintéticos que presentan unas secuencias de nucleótidos que son parcialmente complementarias entre sí. Cuando se mezclan dos oligonucleótidos sintéticos en disolución con unas condiciones apropiadas, se aparearán entre sí formando una estructura bicatenaria. Tras el apareamiento, un extremo de la molécula adaptadora se diseña de tal modo que sea compatible con el extremo de un fragmento de restricción y se pueda ligar al mismo; el otro extremo del adaptador se puede diseñar de tal modo que no se pueda ligar, pero esta necesidad no constituye el caso (adaptadores con doble ligación).

Fragmentos de restricción ligados al adaptador: fragmentos de restricción que se han rematado con adaptadores como resultado de la ligación.

Cebadores: en general, el término cebadores se refiere a una cadena de ADN que puede iniciar la síntesis de ADN. La ADN polimerasa no puede sintetizar ADN *de novo* sin cebadores: únicamente puede extender una cadena existente de ADN en una reacción en la que la cadena complementaria se utiliza como molde para dirigir el orden de nucleótidos a ensamblar. Nos referiremos a las moléculas de oligonucleótidos sintéticos que se utilizan en la reacción en cadena de la polimerasa (PCR) como cebadores.

Amplificación del ADN: el término amplificación del ADN se utilizará habitualmente para indicar la síntesis *in vitro* de moléculas de ADN bicatenario utilizando la PCR. Se ha de indicar que existen otros procedimientos de amplificación y que se pueden utilizar en la presente invención sin apartarse de la esencia de la misma.

30 Descripción detallada de la invención

La presente invención proporciona un procedimiento para determinar una secuencia genómica que comprende las etapas de:

- 35 (a) proporcionar un primer subconjunto del genoma mediante la digestión del genoma con por lo menos una primera endonucleasa de restricción para proporcionar fragmentos de restricción;
- (b) realizar la ligación de por lo menos un adaptador con los fragmentos de restricción del primer subconjunto para proporcionar un primer conjunto de fragmentos de restricción ligados al adaptador;
- 40 (c) amplificar selectivamente el primer conjunto de fragmentos de restricción ligados al adaptador utilizando una primera combinación de cebador en la que por lo menos un primer cebador contiene una sección complementaria al adaptador y a una parte de la secuencia de reconocimiento de la endonucleasa de restricción y que contiene además una primera secuencia seleccionada en el extremo 3' de la secuencia del cebador, comprendiendo la primera secuencia seleccionada entre 1 y 10 nucleótidos selectivos, para proporcionar 45 un primer subconjunto de fragmentos de restricción amplificados ligados al adaptador;
- (d) repetir la etapa (c) con por lo menos un segundo y/o una o más combinaciones adicionales del cebador en las que el cebador contiene una distinta secuencia seleccionada segunda y/o adicionales en su extremo 3' 50 que contiene el mismo número de nucleótidos selectivos, para proporcionar unos subconjuntos segundo y/o adicionales de fragmentos de restricción amplificados ligados al adaptador;
- (e) fragmentar cada uno de los subconjuntos, primero, segundo y/o adicionales de fragmentos de restricción amplificados ligados al adaptador, seguido opcionalmente por la selección del tamaño de los fragmentos 55 en el intervalo óptimo de tamaños, para generar unas genotecas de secuenciación primera, segunda y/o adicionales, y a continuación mezclar opcionalmente las genotecas;
- (f) determinar (por lo menos una parte de) las secuencias de nucleótidos de (por lo menos una parte de) los fragmentos contenidos en cada una de las genotecas de secuenciación primera, segunda y/o adicionales;
- 60 (g) alinear la secuencia de los fragmentos en cada una de las genotecas primera, segunda y/o adicionales para generar cóntigos de los fragmentos de restricción amplificados ligados al adaptador obtenidos a partir del/de los subconjunto(s) del genoma;
- 65 (h) repetir las etapas (a)-(g) para por lo menos unas endonucleasas de restricción segunda y/o adicionales;
- (i) alinear los cóntigos obtenidos en las etapas (g) y (h) para cada una de las endonucleasas de restricción segunda y/o adicionales para proporcionar una secuencia del genoma.

En la etapa (a) del procedimiento, el genoma de interés se somete a una o más endonucleasas de restricción. En determinadas formas de realización, se utilizan por lo menos dos endonucleasas de restricción. En determinadas formas de realización, en particular con genomas grandes, se pueden utilizar tres o más endonucleasas de restricción. La digestión del genoma proporciona un primer subconjunto del genoma. Las endonucleasas de restricción pueden ser cortadoras frecuentes (es decir, habitualmente cortadoras 4 y 5, es decir, endonucleasas de restricción que presentan una secuencia de reconocimiento de 4 ó 5 nucleótidos, respectivamente) o pueden ser cortadoras raras, (es decir, habitualmente cortadoras 6 y superiores (7, 8, ... etc., es decir, endonucleasas de restricción que presentan una secuencia de reconocimiento de 6 o más nucleótidos, respectivamente)), o combinaciones de las mismas. En determinadas formas de realización se utiliza una combinación de una cortadora poco frecuente y una frecuente. En determinadas formas de realización se pueden utilizar dos cortadoras poco frecuentes. Las endonucleasas de restricción pueden ser de cualquier tipo, comprendiendo los tipos IIs y IIsa que cortan el ADN fuera de su secuencia de reconocimiento, tanto en una como en ambas caras de la secuencia de reconocimiento.

En etapa (b) del procedimiento, por lo menos un adaptador se liga a los fragmentos de restricción obtenidos en la etapa (a). Preferentemente, los adaptadores son de tal modo que el sitio de restricción no se restablece mediante ligación del adaptador. Resulta asimismo posible, por ejemplo, en el caso de dos o más endonucleasas de restricción utilizar dos o más adaptadores distintos. Dicha etapa de ligación produce fragmentos de restricción ligados al adaptador. Los adaptadores, en función de la endonucleasa de restricción, pueden presentar un extremo romo o pueden contener un extremo saliente.

En determinadas formas de realización, el adaptador puede ser un conjunto de adaptadores conocidos como ligadores de indización (Unrau *et al.*, 1994, *Gene*, 145: 163-169).

En etapa (c), el primer conjunto de fragmentos de restricción ligados al adaptador se amplifica utilizando una primera combinación del cebador. La combinación del cebador comprende por lo menos un primer cebador que contiene una sección complementaria a (por lo menos una parte de) el adaptador y a una parte de la secuencia de reconocimiento de la endonucleasa de restricción utilizada en la restricción del genoma. Habitualmente, la parte de la secuencia de reconocimiento es aquella parte que permanece tras la restricción de la secuencia con la endonucleasa de restricción. En su extremo 3', el cebador contiene una primera secuencia seleccionada. La primera secuencia seleccionada comprende a conjunto seleccionado previamente de 1 a 10 nucleótidos, preferentemente entre 1 y 8 nucleótidos seleccionados, preferentemente entre 1 y 5, más preferentemente entre 1 y 3. Dicho cebador puede presentar la siguiente estructura ilustrativa (para 2 nucleótidos selectivos (AC)) "5' - región específica del adaptador - región específica de la secuencia de restricción - AC - 3'". Este primer cebador presentado a título de ejemplo contiene 2 nucleótidos selectivos AC que amplificarán únicamente fragmentos ligados al adaptador que comprendan la secuencia complementaria TG como los dos primeros nucleótidos obtenidos a partir de la secuencia del fragmento de restricción. Ello proporciona el primer subconjunto de fragmentos de restricción amplificados ligados al adaptador.

La primera combinación del cebador puede comprender asimismo dos cebadores selectivos, transportando cada uno de ellos una secuencia seleccionada en su extremo 3'. Los cebadores se pueden etiquetar para permitir estrategias de agrupamiento.

La amplificación se realiza preferentemente utilizando la PCR larga. En determinadas formas de realización se prefiere la utilización de la PCR.

En la etapa (d), se repite la amplificación selectiva con las combinaciones del cebador segunda y adicional. Por lo menos uno de los cebadores de cada una de las combinaciones adicionales del cebador contiene una secuencia seleccionada distinta en su extremo 3'. La selección de las secuencias seleccionadas se realiza de tal modo que, dado el número de nucleótidos seleccionados, se utilizan todas las permutaciones posibles de los nucleótidos selectivos. En el ejemplo anterior ello significa AT, AG, AA, CA, CT, CG, CA, etc. En la práctica ello significa que todos los fragmentos de restricción ligados al adaptador en el subconjunto del genoma (es decir, en el conjunto de fragmentos de restricción obtenidos utilizando una o más endonucleasas de restricción) se han amplificado.

En una forma de realización preferida de la presente invención, la reducción de la complejidad del genoma mediante amplificación selectiva se realiza mediante la técnica AFLP® (Keygene N. V., Países Bajos; véase por ejemplo el documento EP 0 534 858 y Vos *et al.* (1995). AFLP: una nueva técnica para la obtención de la huella genética del ADN, *Nucleic Acids Research*, vol. 23, n.º 21, 4407-4414).

La AFLP consiste en un procedimiento de amplificación de un fragmento de restricción selectivo. La AFLP no requiere información previa alguna de la secuencia y se puede realizar en cualquier ADN inicial. En general, la AFLP comprende las etapas de:

- (a) digerir un ácido nucleico, en particular a ADN, con una o más endonucleasas de restricción específicas, para fragmentar el ADN en la serie correspondiente de fragmentos de restricción;
- (b) ligar los fragmentos de restricción obtenidos de este modo con adaptador de un oligonucleótido sintético bicatenario, siendo un extremo del mismo compatible con uno o ambos extremos de los fragmentos de restricción, para producir de este modo fragmentos de restricción del ADN inicial ligados al adaptador, preferentemente etiquetados;

(c) poner en contacto los fragmentos de restricción ligados al adaptador, preferentemente etiquetados, en unas condiciones de hibridación con uno o más cebadores de oligonucleótidos que contienen nucleótidos selectivos en su extremo 3';

(d) amplificar el fragmento de restricción ligado al adaptador, preferentemente etiquetado hibridado con los cebadores mediante la PCR o una técnica similar de tal modo que provoque una elongación adicional de los cebadores hibridados a lo largo de los fragmentos de restricción del ADN inicial con los que se hibridan los cebadores; y

(e) detectar, identificar o recuperar el fragmento amplificado o alargado de ADN obtenido de este modo.

El AFLP proporciona de este modo un subconjunto reproducible de fragmentos ligados al adaptador. Una variante útil de la técnica AFLP no utiliza nucleótidos selectivos (es decir, cebadores +0/+0) y a veces se denomina PCR de enlace. Ello proporciona asimismo una reducción de la complejidad muy adecuada, en particular para los genomas más pequeños.

Para una descripción adicional de la AFLP, sus ventajas, sus formas de realización, así como las técnicas, enzimas, adaptadores, cebadores y compuestos y herramientas adicionales utilizados en la misma, se hace referencia a los documentos US 6.045.994, EP-B-O 534 858, EP 976835 y EP 974672, WO 01/88189 y Vos *et al. Nucleic Acids Research*, 1995, 23, 4407-4414.

De este modo, en una forma de realización preferida del procedimiento de la presente invención, se reduce la complejidad del genoma mediante las siguientes etapas:

(a) digerir la muestra de ácido nucleico con por lo menos una endonucleasa de restricción para fragmentar la misma en fragmentos de restricción;

(b) ligar los fragmentos de restricción obtenidos con por lo menos un adaptador de un oligonucleótido sintético bicatenario que presente un extremo compatible con uno o ambos extremos de los fragmentos de restricción para producir fragmentos de restricción ligados al adaptador;

(c) poner en contacto dichos fragmentos de restricción ligados al adaptador con uno o más cebadores de oligonucleótidos en unas condiciones de hibridación; y

(d) amplificar dichos fragmentos de restricción adaptados mediante la elongación de uno o más cebadores de oligonucleótidos,

(e) comprendiendo por lo menos uno de los uno o más cebadores de oligonucleótidos una secuencia de nucleótidos que presente una secuencia de nucleótidos como parte terminal de las cadenas en los extremos de dichos fragmentos de restricción adaptados, comprendiendo los nucleótidos implicados en la formación de la secuencia seleccionada para dicha endonucleasa de restricción y comprendiendo por lo menos una parte de los nucleótidos presentes en los adaptadores, en los que, opcionalmente, por lo menos uno de dichos cebadores comprende en su extremo 3' una secuencia seleccionada que comprende por lo menos un nucleótido dispuesto inmediatamente adyacente a los nucleótidos implicados en la formación de la secuencia seleccionada para dicha endonucleasa de restricción.

La AFLP constituye un procedimiento muy reproducible para la reducción de la complejidad y resulta, por lo tanto, particularmente adecuado como método según la presente invención.

Hasta la fecha, en la técnica de la secuenciación, no se ha descrito o propuesto la utilización de dichas amplificación selectiva en la determinación de la secuencia de genomas enteros y, en particular, en genomas complejos. La técnica AFLP se conoce en la especialidad como técnica de obtención de la huella genética y no se ha identificado todavía como una solución para ayudar en la secuenciación de genomas complejos. En particular, la utilización de un conjunto de combinaciones del cebador que cubren todas o la mayoría de las permutaciones de nucleótidos para un número determinado de nucleótidos selectivos (por ejemplo, 16 combinaciones del cebador en el caso de dos nucleótidos selectivos) proporciona un método fiable y rápido para proporcionar subconjuntos complementarios y reproducibles de un genoma que se puede secuenciar. En determinadas formas de realización, los cebadores utilizados para reducir la complejidad contienen uno o más enlaces tioato para aumentar su selectividad y/o rendimiento.

En ciertas formas de realización alternativas, la reducción de la complejidad comprende el método CHIP. Otros métodos aptos para la reducción de la complejidad son la inmunoprecipitación de la cromatina (ChIP). Ello significa que se aísla el ADN del núcleo, mientras que las proteínas tales como los factores de transcripción se enlazan con el ADN. Con la técnica Chip, en primer lugar se utiliza un anticuerpo contra la proteína, lo que tiene como resultado un complejo Ab-proteína-ADN. Al purificar dicho complejo y precipitarlo, se selecciona el ADN al que se enlaza dicha proteína. Posteriormente, se puede utilizar el ADN en la construcción y la secuenciación de la genoteca. Es decir, este es un método para realizar una reducción de la complejidad de un modo no aleatorio dirigido a unas áreas funcionales espe-

cíficas; en el presente ejemplo, factores de transcripción específicos. Unas formas de realización alternativas pueden utilizar el diseño de los cebadores de la PCR dirigidos contra motivos conservados tales como las regiones SSRs, NBS (regiones de enlace de nucleótidos), secuencias promotor/potenciador, secuencias consenso de telómeros, secuencias de genes MADS, familias de genes de la familia de genes de la adenosinatrifosfatasa y otras familias de genes.

En etapa (e), se generan las genotecas de secuenciación primera segunda y adicional para cada subconjunto de fragmentos de restricción amplificados ligados al adaptador. Las genotecas se generan habitualmente mediante la fragmentación de los fragmentos de restricción amplificados ligados al adaptador. Se puede realizar la fragmentación mediante técnicas físicas, es decir, corte, sonicación u otros procedimientos de fragmentación. En la etapa (f), se determina por lo menos una parte, pero preferentemente la totalidad, de la secuencia de nucleótidos de por lo menos una parte de, pero preferentemente de todos los fragmentos contenidos en las genoteca.

La secuenciación se puede realizar en principio utilizando cualquier medio conocido en la técnica, tal como el procedimiento de terminación de la cadena de didesoxirribonucleótidos. Se prefiere, sin embargo, que la secuenciación se realice utilizando procedimientos de secuenciación de alto rendimiento, tales como los procedimientos descritos en los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007 y WO 2005/003375 (todos ellos a nombre de 454 Life Sciences), por Seo *et al.* (2004) *Proc. Natl. Acad. Sci. EE. UU.* 101: 5488-93, y las técnicas de Helios, Solexa, US Genomics. Se prefiere que la secuenciación se realice utilizando el aparato y/o el procedimiento descrito en los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007 y WO 2005/003375 (todos ellos a nombre de 454 Life Sciences). La técnica descrita permite la secuenciación de 40 millones bases en una serie simple y es 100 veces más rápida y económica que las técnicas alternativas. La técnica de secuenciación consiste aproximadamente en 5 etapas: 1) fragmentación de ADN y ligación de un adaptador específico para crear una genoteca de ADN monocatenario (ADNmc); 2) apareamiento del ADNmc con perlas, emulsificación de las perlas en microrreactores de agua en aceite y realizar la PCR en emulsión para amplificar las moléculas individuales de ADNmc en las perlas; 3) selección de/enriquecimiento de las perlas que contienen moléculas ADNmc amplificado en su superficie 4) sedimentación del ADN que transporta las en una PicoTiterPlate®; y 5) secuenciación simultánea en 100,000 pocillos mediante la generación de una señal óptica de pirofosfato. El procedimiento se explicará posteriormente con mayor detalle.

En una forma de realización preferida, la secuenciación comprende las etapas de:

- (a) aparear los fragmentos adaptados a las perlas, apareándose cada perla con un fragmento adaptado simple;
- (b) emulsionar las perlas en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una perla simple;
- (c) cargar las perlas en pocillos, comprendiendo cada pocillo una perla simple; y generar una señal de pirofosfato.

En la primera etapa (a), los adaptadores de la secuenciación se enlazan con fragmentos en la genoteca de la combinación. Dicho adaptador de la secuenciación comprende por lo menos una región "clave" para aparearse con una perla, una región del cebador de la secuenciación y una región del cebador de la PCR. De este modo se obtienen los fragmentos adaptados.

En una primera etapa, los fragmentos adaptados se aparean con las perlas, apareándose cada perla con un fragmento adaptado simple. Al grupo de fragmentos adaptados, se añaden las perlas en exceso para garantizar el apareamiento de un fragmento adaptado simple por perla en el caso de la mayoría de las perlas (distribución de Poisson).

En una etapa posterior, se emulsionan las perlas en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una perla simple. Los reactivos de la PCR se encuentran presentes en los microrreactores de agua en aceite permitiendo que se produzca una reacción de la PCR en los microrreactores. Posteriormente, se rompen los microrreactores y se enriquecen las perlas que comprenden ADN (perlas positivas en ADN).

En una etapa posterior, se cargan las perlas en pocillos, comprendiendo cada pocillo una perla simple. Los pocillos forman parte preferentemente de una PicoTiter™Plate que permite la secuenciación simultánea de una gran cantidad de fragmentos.

Tras la adición de perlas que transportan enzimas, se determina la secuencia de los fragmentos utilizando la pirosecuenciación. En etapas sucesivas, la PicoTiter™Plate y las perlas así como las los enzimas de las mismas se someten a distintos desoxirribonucleótidos en presencia de reactivos convencionales de la secuenciación y, mediante la incorporación de un desoxirribonucleótido, se genera una señal óptica que se registra. La incorporación del nucleótido correcto generará una señal de pirosecuenciación que se puede detectar.

La propia pirosecuenciación resulta conocida en la técnica y se describe, entre otros, en www.biotagebio.com; www.pyrosequencing.com/sección_técnica. La técnica se aplica además en, por ejemplo, los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007 y WO 2005/003375 (todos ellos a nombre de 454 Life Sciences).

En la etapa (g) del procedimiento de la presente invención, se alinean las secuencias determinadas de los fragmentos de las genotecas primera, segunda y/o adicionales. La alineación proporciona cóntigos de los fragmentos en los subconjuntos de los fragmentos de restricción amplificados ligados al adaptador. De este modo para cada fragmento de restricción amplificado ligado al adaptador, se genera un cóntigo a partir de los fragmentos secuenciados, es decir, el cóntigo de un fragmento de restricción amplificado ligado al adaptador, se construye a partir de la alineación de la secuencia de los diversos fragmentos obtenidos a partir de la fragmentación de la etapa (e). Mediante la construcción de cóntigos a partir de las secuencias que representan los fragmentos de restricción dispersos de una pequeña parte del genoma, disminuyen considerablemente los problemas de la construcción de cóntigos que son consecuencia de las abundantes secuencias repetidas obteniéndose secuencia borrador del genoma de una calidad superior que contiene menos errores debidos a uniones falsas de secuencias repetidas. Además, el proceso de ensamblaje resultará menos complejo desde el punto de vista informático y, por lo tanto, más rápido de realizar. Al alinear las secuencias de las distintas genotecas, se pueden construir los cóntigos de cada fragmento de restricción del conjunto de fragmentos de restricción para cada combinación del cebador. Ello tiene como resultado un conjunto de cóntigos, correspondiendo cada uno a un fragmento de restricción particular. Como resultado de ello, cada fragmento de restricción obtenido a partir de la restricción del genoma con por lo menos una endonucleasa de restricción presenta ahora una secuencia determinada (cóntigo). El procedimiento de la presente invención se ilustra en las figuras 1 y 2.

Los procedimientos de alineación de secuencias con un propósito comparativo resultan muy conocidos en la técnica. Se describen diversos programas y algoritmos de alineación en: Smith & Waterman (1981) *Adv. Appl. Math.* 2: 482; Needleman & Wunsch (1970) *J. Mol. Biol.* 48: 443; Pearson & Lipman (1988) *Proc. Natl. Acad. Sci. EE. UU.* 85: 2444; Higgins & Sharp (1988) *Gene* 73: 237-244; Higgins & Sharp (1989) *CABIOS* 5: 151-153; Corpet *et al.* (1988) *Nucl. Acids Res.* 16: 10881-90; Huang *et al.* (1992) *Computer Appl. in the Biosci.* 8: 155-65; y Pearson *et al.* (1994) *Meth. Mol. Biol.* 24: 307-31, que se incorporan en la presente memoria como referencia. Altschul *et al.* (1994) *Nature Genet.* 6: 119-29 (que se incorpora en la presente memoria como referencia) presenta una consideración detallada de procedimientos de alineación de secuencias y cálculos de analogías.

El programa básico de búsqueda de alineación local (BLAST) del NCBI (Altschul *et al.*, 1990) se encuentra disponible a partir de fuentes diversas, entre ellas el Centro Nacional de Información Biológica (NCBI, Bethesda, Md.) y en Internet, para utilizar junto con los programas de análisis de secuencias blastp, blastn, blastx, tblastn y tblastx. Se puede conseguir en <<http://www.ncbi.nlm.nih.gov/BLAST/>>. Una descripción de como determinar la identidad de la secuencia utilizando dicho programa se encuentra disponible en <http://www.ncbi.nlm.nih.gov/BLAST/blast_help.html>. Una aplicación adicional se puede encontrar en la extracción de microsatélites (véase Varshney *et al.* (2005) *Trends in Biotechn.* 23(1): 48-55).

Habitualmente, la alineación se realiza en los datos de la secuencia que se han extraído para los adaptadores/cebador y/o identificadores pero con secuencias de reconocimiento reconstruidas de enzimas de restricción, es decir, utilizando únicamente los datos de la secuencia procedentes de los fragmentos que se originan en la muestra de ácido nucleico. Habitualmente, los datos de la secuencia obtenidos se utilizan para identificar el origen del fragmento (es decir, de qué muestra), las secuencias obtenidas a partir del adaptador y/o identificador se eliminan de los datos y la alineación se realiza en este conjunto extraído.

En la etapa (h), se repite el procedimiento completo por lo menos una vez con una o más endonucleasas de restricción distintas, es decir, una endonucleasa de restricción que contiene, preferentemente, un sitio de reconocimiento distinto al de la primera endonucleasa, para proporcionar un conjunto de fragmentos de restricción segundo o incluso adicionales que posteriormente se enlazan con el adaptador, se amplifican selectivamente utilizando combinaciones del cebador con una secuencia seleccionada que se selecciona independientemente, es decir, no tiene relación alguna con la secuencia seleccionada, (tanto en el número como en el tipo de nucleótidos) con las que se han utilizado con el mismo objetivo con la primera endonucleasa de restricción. Por ejemplo, el primer subconjunto se puede obtener mediante la restricción con MseI/PstI y amplificar selectivamente utilizando un cebador selectivo para los restos MseI del sitio de reconocimiento y que transporta 2 nucleótidos selectivos en su extremo 3'. El segundo subconjunto se puede obtener mediante la digestión de EcoRI/HindIII y la amplificación selectiva con un cebador selectivo para los restos EcoRI del sitio de reconocimiento con 1 nucleótido selectivo.

Por lo tanto, se puede obtener de este modo un segundo (y/o adicionales) conjunto de cóntigos para todos los fragmentos de restricción, de un modo similar a tal como se ha descrito anteriormente en la presente memoria. Ello resulta necesario, ya que para una endonucleasa de restricción determinada, las fracciones del genoma que se están secuenciando son complementarias, no se superponen. Los cóntigos obtenidos con distintas combinaciones enzimáticas se superponen y, por lo tanto, permiten la generación de un cóntigo a partir de los mismos y, de este modo, permiten la formación de una secuencia genómica (borrador).

En etapa (i) del procedimiento, los cóntigos obtenidos a partir de las etapas anteriores del procedimiento para cada fragmento se alinean para que formen una secuencia del genoma.

En determinadas formas de realización, se puede contribuir a la construcción del cóntigo de tanto los fragmentos de restricción como de la secuencia genómica mediante la utilización de secuencias de nucleótidos del genoma que se obtienen a partir de otras fuentes, entre ellas, pero sin limitarse a las mismas, secuencias del extremo BAC, secuencias aleatorias BAC, secuencias EST o secuencias aleatorias del genoma entero.

El procedimiento de la presente invención es independiente de la fuente de ADN, es decir, se puede aplicar a todos los organismos ya que no requiere información previa de la secuencia. Mediante la selección apropiada de enzimas, adaptadores, cebadores y del número de nucleótidos selectivos, se presenta la técnica graduable que se puede aplicar a genomas de todos los tamaños y complejidades. Además, las fracciones del genoma que se obtienen con las distintas combinaciones del cebador y/o con los cebadores selectivos que difieren entre sí en las secuencias selectivas específicas de nucleótidos del extremo 3', son complementarias. Ello significa que cuando, para un número determinado de nucleótidos selectivos, se utilizan todas las permutaciones (1 nucleótido selectivo equivale a 4 variantes (A, C, T, G), 2 nucleótidos selectivos 16, 3 nucleótidos selectivos 64 variantes, etcétera), los fragmentos de restricción combinados constituyen el genoma restringido.

Breve descripción de las figuras

Figura 1: Partiendo del ADN genómico, se realiza una digestión con una combinación de endonucleasas de restricción (combinación enzimática 1, EC1), lo que tiene como resultado un conjunto de fragmentos de restricción. Los fragmentos de restricción se enlazan con adaptadores tras los que los fragmentos de restricción enlazados con adaptadores se amplifican con una primera combinación selectiva del cebador (PC1) para obtener n fragmentos. Cada fragmento se divide para secuenciación de alto rendimiento y se somete a secuenciación y alineación para generar los cóntigos de los fragmentos de restricción. De este modo, se determina la secuencia n de todos o de la mayoría de fragmentos de restricción enlazados con adaptadores para una combinación del cebador.

Figura 2: Para cada combinación del cebador posible de una combinación enzimática determinada (EC1), se repiten las etapas de fragmentación de los fragmentos de restricción amplificados enlazados con adaptadores, secuenciación, alineación y construcción de cóntigos. Ello significa que cuando, por ejemplo, se realiza la amplificación selectiva con cebadores que transportan cada uno un nucleótido selectivo en su extremo 3' (es decir, cebadores +1/+1), 16 combinaciones del cebador (PC1 ... PCm) cubren todas las permutaciones y con 16 combinaciones del cebador todos los fragmentos de restricción enlazados con adaptadores se han amplificado y posteriormente secuenciado. A partir de los cóntigos generados con la EC1, es decir, a partir de EC1/PC1 ... EC1/PCm, un ensamblaje cubrirá una gran parte del genoma, pero necesitará fijarse a fin de proporcionar una secuencia genómica. Con este objetivo, se utiliza una segunda combinación enzimática (y, si es necesario una tercera y una cuarta, etc.). Las etapas de las figuras 1 y 2 se repiten con la combinación enzimática 2 (EC2), es decir, restricción, ligación con el adaptador, etc. La amplificación selectiva se realiza con un conjunto de cebadores selectivos que habitualmente pueden ser distintos (secuencia y selectividad) de los cebadores utilizados con la EC1. Los fragmentos de restricción enlazados con adaptadores se amplifican de nuevo con todas las permutaciones posibles de un conjunto de cebadores selectivos obteniéndose unos subconjuntos distintos y complementarios. La fragmentación de cada subconjunto de fragmentos de restricción enlazados con adaptadores amplificados selectivamente, y la posterior secuenciación de alto rendimiento, la construcción de cóntigos, etc. conduce a un segundo ensamblaje, que de nuevo cubre una gran parte del genoma. A partir de dichos dos ensamblajes (y la combinación enzimática opcional tercera, cuarta, etc.), que se superponen entre sí en grandes áreas, se genera la secuencia borrador del genoma que se investiga.

Figura 3. Secuencia de 662 pares de bases del cóntigo 606 prevista informáticamente, que contiene los fragmentos de restricción que se superponen EcoRI/HindIII +C/+CT y BamHI/XbaI +C+G.

Figura 4. Cóntigos de la secuencia observada del cóntigo 606 previsto informáticamente basándose en la secuenciación por AFLP de las genotecas de fragmentos EcoRI/HindIII +C/+CT (r1_9_35974-36087) y BamHI/XbaI +C/+G (r2_9_36138-36200). Se ha de indicar que la superposición de 42 pares de bases entre (r1_9_35974-36087) y (r2_9_36138-36200) se cubre completamente mediante las secuencias obtenidas a partir de ambas genotecas de fragmentos.

La presente invención se puede ilustrar mediante los ejemplos siguientes que no pretenden limitar la presente invención en modo alguno, si no que se presentan a título simplemente ilustrativo.

Secuenciación del genoma entero utilizando la PCR larga

Etapas 1

El ADN se somete a restricción utilizando dos 6 cortadoras A y B (por ejemplo EcoRI e HindIII). Ello genera tres tipos de fragmentos de restricción: A-A (25%), B-B (25%) y A-B (50%) con una longitud media de aproximadamente 3-4 kb, en función del contenido en GC del genoma de interés, y la selección de las endonucleasas de restricción. Tras la ligación de los adaptadores, se realiza la PCR larga con cebadores +X/+Y (es decir, uno de los cebadores contiene X nucleótidos selectivos y el otro Y), hasta una secuencia de 1 Mb por combinación del cebador. En el caso de que X = 2 e Y = 3, ello se repite para todas las 1024 combinaciones del cebador. En el caso de que X = 1 e Y = 2, ello se repite para todas las 64 combinaciones del cebador.

A: Tamaño del genoma maíz de 2700 Mbp: fragmentos de tipo A-B = 1350 Mbp. Mediante una amplificación selectiva +2/+3 (1024 combinaciones del cebador) el producto de la amplificación de cada una de las combinaciones del cebador contiene, en promedio, una secuencia de $1350/1024 = 1,32$ Mbp. Con una longitud media de aproximadamente 3000 de pares de bases por fragmento A-B, ello produce $1320000/3000 = 440$ fragmentos A-B.

ES 2 344 802 T3

- B: 130 Mbp de *Arabidopsis*: fragmentos del tipo A-B de 65 Mbp. Con $X = 1$ e $Y = 2$, cada combinación del cebador (PC) contiene una secuencia de aproximadamente 1 Mbp. Con una longitud media de 3000 de pares de bases por fragmento, es decir, fragmentos A-B de $1000000/3000 = 330$.

Etapa 2

Construcción de genotecas cortando cada conjunto de fragmentos de restricción amplificados ligados al adaptador y secuenciación utilizando la PCR en emulsión junto con la pirosecuenciación de 454 Technologies tal como se describe en otras partes de la presente memoria, para cada combinación del cebador (1024 ó 64 veces). La secuenciación utilizando esta tecnología proporciona unos datos de 40 Mbp de la secuencia por genoteca, lo que significa que cada genoteca presenta un nivel de repeticiones en la secuencia de 40 veces. Variando el número de nucleótidos, se amplifica una cantidad distinta de fragmentos A-B y se obtiene un nivel distinto de repeticiones. Ello se puede determinar en su puesta en práctica.

Etapa 3

Ensamblaje de las secuencias por genoteca de la secuencia (por PC)

Se realiza el ensamblaje para generar los cóntigos de todos los fragmentos A-B que se han amplificado por PC. Con ello se producen aproximadamente entre 300 y 500 cóntigos por PC cuya longitud media se corresponderá con la longitud media del fragmento A-B. La secuenciación de todos los PC tiene como resultado un número de cóntigos que varía desde varias decenas de miles hasta varios centenares de miles (*Arabidopsis* 21000, maíz 450000).

Etapa 4

Repetición de las etapas 1 a 3 para por lo menos otra combinación enzimática (EC), por ejemplo, A-C. Ello resulta imprescindible ya que todos los PC del EC A-B proporcionan únicamente cóntigos complementarios y no cóntigos que se superpongan, con lo que se cubre únicamente el 50% del genoma. Se puede aumentar asimismo la cobertura del genoma procesando todos los fragmentos A-A y B-B. Al utilizar EC adicionales, se consigue la superposición entre los cóntigos de AB y de AC y aumenta la cobertura del genoma.

Etapa 5

Ensamblaje de todas los cóntigos de las combinaciones del cebador A-B (opcionalmente también A-A, B-B) y A-C hasta una secuencia genómica (borrador)

Una de las ventajas de dicho procedimiento radica en el hecho de que uno de los problemas del ensamblaje del genoma y la posibilidad de formación de cóntigos erróneos debido a que se minimiza la presencia de múltiples secuencias repetidas mediante la formación de cóntigos dispersos pequeños (es decir, no adyacentes) de 1-10 kb en una fracción del genoma de 1-5 Mbp en vez del genoma entero. Los cóntigos con longitudes mucho más largas se pueden etiquetar en las primeras etapas como producto de un enlace falso y descartarlos. Una ventaja adicional es que el ensamblaje resulta menos complejo desde el punto de vista informático debido al ensamblaje inicial (etapa 3), menos secuencias se encuentran implicadas que cuando se va a ensamblar la secuencia genómica entera en una etapa. Una ventaja adicional es que el proceso de amplificación selectiva permite graduar el procedimiento entero a un genoma de cualquier tamaño y es de aplicación universal.

Secuenciación del genoma entero utilizando una cortadora poco frecuente y otra frecuente

Etapa 1

Tal como en el caso anterior, con una cortadora 6 (EcoRI) y una cortadora 4 (MseI). La longitud media del fragmento es aproximadamente de 250 pares de bases. Los fragmentos A-B representan aproximadamente entre el 8 y el 15% del genoma. En comparación con la digestión con la enzima de restricción utilizando dos enzimas de restricción cortadoras 6, se necesita en promedio aproximadamente 1 nucleótido selectivo menos para obtener un nivel de complejidad de secuencia por PC de aproximadamente 1-5 Mbp.

- A: Tamaño del genoma de maíz de 2700 Mbp: fragmentos de tipo A-B = 270 Mbp, (10%) mediante una amplificación selectiva +2/+2 (256 combinaciones del cebador) contiene el producto de la amplificación de cada una de las combinaciones del cebador, en promedio, secuencia de $270/256 = 1,05$ Mbp. Con una longitud media por fragmento A-B de aproximadamente 250 pares de bases, se obtienen $1050000/250 = 4200$ fragmentos A-B/cóntigos.

- B: 130 Mbp de *Arabidopsis*: fragmentos de tipo A-B de 13 Mbp (10%). Con $X = 1$ e $Y = 1$, cada combinación del cebador (PC) contiene aproximadamente una secuencia de 1 Mbp. Con una longitud media de 250 pares de bases por fragmento, es decir, $1000000/3000 = 4000$ fragmentos A-B.

ES 2 344 802 T3

Etapa 2

Tal como en el caso anterior. Para evitar el error sistemático de los fragmentos demasiado cortos, se puede utilizar una selección del tamaño para eliminar los fragmentos inferiores a 100-150 pares de bases.

Etapa 3

Ensamblaje de las secuencias por genoteca de secuencias (por PC)

Se realiza el ensamblaje para generar cóntigos de todos los fragmentos A-B que se han amplificado por PC. Ello produce varios miles de cóntigos por PC cuya longitud media se corresponderá con la longitud del fragmento A-B (250 pares de bases). La secuenciación de todos los PC tiene como resultado un número de cóntigos que varía desde varias decenas de miles hasta aproximadamente un millón (*Arabidopsis* 64000, maíz 1000000).

Etapa 4

Repetición de las etapas 1-3 con diversos EC (A-C, B-C, C-C, C-D, A-D, etc.). Ello es necesario ya que los PC de combinación enzimática A-B no cubren más del 8-15% del genoma y, tal como en el caso anterior, los cóntigos de los PC no se superponen.

Etapa 5

Tal como en el caso anterior.

Secuenciación del genoma entero utilizando una endonucleasa de restricción

Etapa 1

Digestión del ADN con una endonucleasa de restricción A (EcoRI, por ejemplo). Fragmentos de restricción de aproximadamente 3-4 kb, en función del contenido en GC y la elección del enzima. Se enlaza la mezcla al adaptador y se realiza la PCR larga (véase anteriormente) con cebadores selectivos que reducen la cantidad de secuencia por PC hasta aproximadamente 1 Mb. En el caso de $X = 2$ e $Y = 3$ se repite para los 1024 PC. En el caso $(X,Y) = (+1/+2)$ se repite para todos los 64 PC.

A: Tamaño del genoma de maíz de 2700 Mbp: fragmentos de tipo A-A = 2700 Mbp, mediante una amplificación selectiva +2/+3 (1024 combinaciones del cebador) contiene el producto de la amplificación de cada una de las combinaciones del cebador, en promedio, secuencia de $2700/1024 = 2,64$ Mb. Con una longitud media por fragmento A-A de aproximadamente 3000 pares de bases, se obtienen $2640000/3000 = 880$ fragmentos A-A/cóntigos.

B: 130 Mbp de *Arabidopsis*: fragmentos de tipo A-A de 130 Mb. Con $X = 1$ e $Y = 2$, cada combinación del cebador (PC) contiene aproximadamente una secuencia de 2 Mb. Con una longitud media de 3000 pares de bases por fragmento, es decir, $2000000/3000 = 660$ fragmentos A-A.

Etapa 2

Construcción de genotecas cortando cada conjunto de fragmentos de restricción amplificados ligados al adaptador y secuenciación utilizando la PCR en emulsión junto con la pirosecuenciación de 454 Technologies tal como se describe en otras partes de la presente memoria, para cada combinación del cebador (1024 ó 64 veces). La secuenciación utilizando esta tecnología proporciona unos datos de 40 Mbp de la secuencia por genoteca, lo que significa que cada genoteca presenta un nivel de repeticiones en la secuencia de 20 veces.

Etapa 3

Ensamblaje de las secuencias por genoteca de secuencias (por PC)

Se realiza el ensamblaje para generar cóntigos de todos los fragmentos A-A que se han amplificado por PC. Ello produce, teóricamente, entre 600 y 900 cóntigos por PC cuya longitud media se corresponderá con la longitud del fragmento A-A (3000 pares de bases). La secuenciación de todos los PC tiene como resultado un número de cóntigos que varía desde varias decenas de miles hasta aproximadamente unos centenares de miles (*Arabidopsis* 42000, maíz 900000).

Etapa 4

Repetir las etapas 1-3 con por lo menos otro EC (B-B). Ello es necesario ya que los PC de combinación enzimática A-B no cubren más del 8-15% del genoma y, tal como en el caso anterior, los cóntigos de los PC no se superponen.

Etapa 5

Tal como en el caso anterior.

Ejemplo 1

El presente ejemplo describe la capacidad de utilizar la secuenciación de alto rendimiento de fragmentos AFLP obtenidos a partir 2 combinaciones de enzimas de restricción para determinar la secuencia genómica de un genoma vegetal complejo.

En el presente ejemplo se realizaron las siguientes etapas:

A) *Predicción informática de fragmentos de restricción por AFLP de la secuencia genómica de Arabidopsis (Genbank), utilizando el programa RECOMB, descrito en el documento WO 0044937 (Keygene N. V)*

La secuencia genómica entera del genoma de Arabidopsis (ecotipo Colombia) se descargó de Genbank. Se realizó la predicción informática de los fragmentos de AFLP +1/+1 para la combinación de enzimas de restricción BamHI/XbaI utilizando los nucleótidos selectivos +C y +G, respectivamente, utilizando RECOMB. De un modo similar, se realizó la predicción de los fragmentos AFLP +1/+2 para la combinación de enzimas de restricción EcoRI/HindIII utilizando los nucleótidos selectivos +C y +CT. La recogida de los fragmentos AFLP obtenidos a partir de dos digestiones informáticas tuvo como resultado diversas secuencias de fragmentos AFLP (de aproximadamente 14) que se superponían entre las combinaciones enzimáticas BamHI/XbaI y EcoRI/HindIII. Uno de los fragmentos de restricción superpuestos forma un cóntigo denominado cóntigo "606", que presenta una longitud total de 662 pares de bases. La secuencia de dicho cóntigo se representa en la figura 3.

El fragmento previsto EcoRI/HindIII AFLP +C/+CT en dicho cóntigo presenta 218 pares de bases de longitud, lo que supone el 32,9 % de la longitud total del cóntigo de 606 pares de bases. El fragmento BamHI/XbaI AFLP +C/+G presenta una longitud de 486 pares de bases, lo que equivale al 73,4% de la longitud total del cóntigo. Ambos fragmentos se superponen en 42 pares de bases, tal como se indica en la figura 3.

B) *Preparación y amplificación del molde AFLP*

Se prepararon los moldes del ADN genómico del ecotipo de Arabidopsis Colombia y AFLP moldes para las combinaciones de enzimas de restricción EcoRI/HindIII y BamHI/XbaI basándose en los protocolos descritos por Zabeau & Vos, 1993: *Selective restriction fragment amplification; a general method for DNA fingerprinting* (documento EP 0534858-A1, B1; la patente US 6045994) y Vos *et al.* (Vos, P., Hogers, R., Bleeker, M., Reijmans, M., van de Lee, T., Homes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. *et al.* (1995) AFLP: *a new technique for DNA fingerprinting. Nucl. Acids Res.*, 21, 4407-4414).

Se utilizaron las siguientes secuencias (5'-3') del adaptador:

BamHI:

91M35: CTCGTAGACTGCGTACC [SEC ID 1]

93U01: GATCGGTACGCAGTC [SEC ID 2]

XbaI:

90K02: CTCGTAGACTGCGTACA [SEC ID 3]

92A16: CTAGTGTACGCAGTCT [SEC ID 4]

ES 2 344 802 T3

EcoRI:

91M35: CTCGTAGACTGCGTACC [SEC ID 5]

91M36: AATTGGTACGCAGTCTAC [SEC ID 6]

HindIII:

91M35: CTCGTAGACTGCGTACC [SEC ID 7]

91M37: AGCTGGTACGCAGTCTAC [SEC ID 8]

Se realizaron las amplificaciones (+1/+1) selectivas (E/H y B/X) utilizando los siguientes cebadores de fosforotioato (5'-3'):

BamHI+C-thio: 96R22thio: GACTGCGTACCGATCsCsC [SEC ID 9]

XbaI+G: 96X03thio: GACTGCGTACACTAGsAsG [SEC ID 10]

EcoRI+C-thio: 93Tl4thio: GACTGCGTACCAATTsCsC [SEC ID 11]

HindIII+C-thio: 95H18thio GACTGCGTACCAGCTTsCsT [SEC ID 12]

con una "s" indicando la posición de los enlaces fosforotioato de los oligonucleótidos.

Las mezclas de las reacciones por AFLP presentaron la composición siguiente:

5 µl 1/10 en molde AFLP diluido con MQ

10 µl de solución amortiguadora para PCR de herculasa II

0,5 µl de dNTP (20 mM)

1,5 µl de cebador AFLP 1 (50 ng/µl)

1.5 µl de cebador AFLP 2 (50 ng/µl)

1 µl de la ADN-polimerasa Herculase® II Fusion

30,5 µl de MQ

Las condiciones del ciclo de la PCR fueron las siguientes:

Desnaturalización inicial 94°C 2 min

Desnaturalización 94°C 10 seg

Apareamiento 56°C 30 seg 10 ciclos

Elongación 68°C 2 min

Desnaturalización 94°C 15 seg

Apareamiento 56°C 30 seg 20 ciclos

Elongación 68°C 2 min *

* Ajuste: 20 seg por ciclo

ES 2 344 802 T3

Tras la amplificación AFLP, se purificaron los productos de las reacciones utilizando columnas Qiagen siguiendo los protocolos de los fabricantes.

C) Preparación de la genoteca de secuencias 454

Se prepararon dos genotecas de secuencias 454 utilizando fragmentos purificados BamHI/XbaI AFLP y fragmentos EcoRI/HindIII AFLP como ADN inicial, respectivamente, tal como describen Margulies y colaboradores, empezando con la nebulización (fragmentación) de los productos purificados de la reacción AFLP. Se realizó una serie simple de la secuencia 454 secuencia utilizando un instrumento de secuenciación GS20 (Roche Molecular Diagnostics), aplicando cada una de las genotecas de fragmentos de las dos combinaciones enzimáticas AFLP a una mitad de una PicoTiterPlate GS20.

D) Procesamiento de los datos

Tras completar la serie de la secuencia, se procesaron los datos originales utilizando el programa RUNASSEMBLY del GS20. Los datos obtenidos de las genotecas de fragmentos EcoRI/HindIII y BamHI/XbaI AFLP se procesaron por separado y juntos, obteniéndose los cóntigos de las lecturas de las secuencias de superposición.

A continuación, se cartografiaron los cóntigos obtenidos del RUNASSEMBLY con respecto al genoma de referencia (cóntigo 606 previsto en la etapa a) anterior utilizando RUNMAPPING, a fin de determinar hasta qué grado se habían secuenciado los fragmentos previstos informáticamente BamHI/XbaI +C/+G y EcoRI/HindIII +C/+CT AFLP contenidos en el cóntigo 606. Los porcentajes de cobertura obtenidos a partir de las genotecas respectivas se representan en la tabla siguiente.

TABLA

Resumen de la cobertura de la secuencia del cóntigo 606 de Arabidopsis

	% Esperado	Número Observado de cóntigos en el cóntigo 606	% Cobertura de la secuencia observada en el cóntigo 606	Longitud de la superposición (pares de bases)	Superposición de cobertura (pares de bases)	Superposición de cobertura (%)
EcoRI/HindIII	32,9	3	35,2	42	42	100
BamHI/XbaI	71,4	2	59,0	42	42	100
EcoRI/HindIII + BamHI/XbaI	100	4	84,8	42	42	100

Las secuencias de los cóntigos resultantes se representan en la figura 4.

Dichos resultados demuestran la viabilidad determinar la secuencia genómica de genomas vegetales complejos mediante la digestión del ADN genómico total con múltiples combinaciones de enzimas de restricción por AFLP, seguido por el ensamblaje de cóntigos por genoteca de fragmentos y el posterior ensamblaje de los cóntigos en la secuencia genómica del vegetal.

REIVINDICACIONES

1. Procedimiento para determinar una secuencia genómica que comprende las etapas de:

- (a) proporcionar un primer subconjunto del genoma mediante la digestión del genoma con por lo menos una primera endonucleasa de restricción para proporcionar fragmentos de restricción;
- (b) realizar la ligación de por lo menos un adaptador con los fragmentos de restricción del primer subconjunto para proporcionar un primer conjunto de fragmentos de restricción ligados al adaptador;
- (c) amplificar selectivamente el primer conjunto de fragmentos de restricción ligados al adaptador utilizando una primera combinación de cebador en la que por lo menos un primer cebador contiene una sección complementaria al adaptador y a una parte de la secuencia de reconocimiento de la endonucleasa de restricción y que contiene además una primera secuencia seleccionada en el extremo 3' de la secuencia del cebador, comprendiendo la primera secuencia seleccionada entre 0 y 10 nucleótidos selectivos, para proporcionar un primer subconjunto de fragmentos de restricción amplificados ligados al adaptador;
- (d) repetir la etapa (c) con por lo menos un segundo y/o una o más combinaciones adicionales del cebador en las que el cebador contiene una distinta secuencia seleccionada segunda y/o adicionales en su extremo 3' que contiene el mismo número de nucleótidos selectivos, para proporcionar unos subconjuntos segundo y/o adicionales de fragmentos de restricción amplificados ligados al adaptador;
- (e) fragmentar cada uno de los subconjuntos primero, segundo y/o adicionales de fragmentos de restricción amplificados ligados al adaptador para generar unas genotecas de secuenciación primera, segunda y/o adicionales;
- (f) determinar por lo menos una parte de las secuencias de nucleótidos de por lo menos una parte de los fragmentos contenidos en cada una de las genotecas primera, segunda y/o adicionales;
- (g) alinear la secuencia de los fragmentos en cada una de las genotecas primera, segunda y/o adicionales para generar cóntigos de los fragmentos de restricción amplificados ligados al adaptador obtenidos que representan fracciones dispersas del genoma;
- (h) repetir las etapas (a)-(g) para por lo menos unas endonucleasas de restricción segunda y/o adicionales;
- (i) alinear los cóntigos obtenidos en las etapas (g) y (h) para cada una de las endonucleasas de restricción segunda y/o adicionales para proporcionar una secuencia del genoma.

2. Procedimiento según la reivindicación 1, en el que por lo menos una de las endonucleasas de restricción primera, segunda y/o adicionales es una cortadora poco frecuente.

3. Procedimiento según cualquiera de las reivindicaciones 1 ó 2, en el que por lo menos una de las endonucleasas de restricción primera, segunda y/o adicionales es una cortadora frecuente.

4. Procedimiento según cualquiera de las reivindicaciones 1, 2 ó 3, en el que se utilizan por lo menos dos cortadoras poco frecuentes.

5. Procedimiento según cualquiera de las reivindicaciones 1 a 4, en el que se utilizan una cortadora poco frecuente y una cortadora frecuente.

6. Procedimiento según cualquiera de las reivindicaciones 1 a 5, en el que el procedimiento de amplificación es la PCR.

7. Procedimiento según cualquiera de las reivindicaciones 1 a 6 en el que la secuencia seleccionada en el extremo 3' del cebador contiene entre 1 y 8 nucleótidos seleccionados.

8. Procedimiento según cualquiera de las reivindicaciones 1 a 7, en el que las secuencias seleccionadas primera, segunda y adicionales presentan el mismo número de nucleótidos pero difieren en las secuencias de nucleótidos entre sí de la secuencia selectiva que se encuentra en el extremo 3' del cebador.

9. Procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que secuenciación se realiza mediante la didesoxisecuenciación de Sanger.

10. Procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que secuenciación se realiza en un soporte sólido tal como una perla.

ES 2 344 802 T3

11. Procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que la secuenciación se basa en la secuenciación de alto rendimiento.

5 12. Procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que la secuenciación se basa en la secuenciación la secuenciación por síntesis.

13. Procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que la secuenciación se basa en la pirosecuenciación.

10 14. Procedimiento según cualquiera de las reivindicaciones 1 a 8, en el que la secuenciación comprende las etapas de:

(f1) enlazar los adaptadores de la secuenciación con los fragmentos;

15 (f2) aparear los fragmentos de la secuenciación enlazados con el adaptador con las perlas, apareándose cada perla con un fragmento simple,

(f3) emulsionar las perlas en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una perla simple;

20 (f4) realizar la PCR en emulsión para amplificar los fragmentos enlazados con el adaptador sobre la superficie de perlas

(f5) seleccionar/enriquecer las perlas que contienen fragmentos amplificados enlazados con el adaptador

25 (f6) cargar las perlas en pocillos, comprendiendo cada pocillo una perla simple; y

(f7) generar una señal de pirofosfato.

30 15. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que la construcción de contig se ve ayudada además por la utilización de secuencias de nucleótidos obtenidas a partir de otras fuentes, que comprenden, pero sin limitarse a las mismas, secuencias del extremo BAC, secuencias aleatorias BAC, secuencias EST o secuencias aleatorias del genoma entero.

35 16. Procedimiento según cualquiera de las reivindicaciones anteriores, en el que el procedimiento para reducir la complejidad de la mezcla se basa en ligadores de indización, la inmunoprecipitación de la cromatina (CHIP) o cebadores de la PCR dirigidos contra motivos conservados.

40

45

50

55

60

65

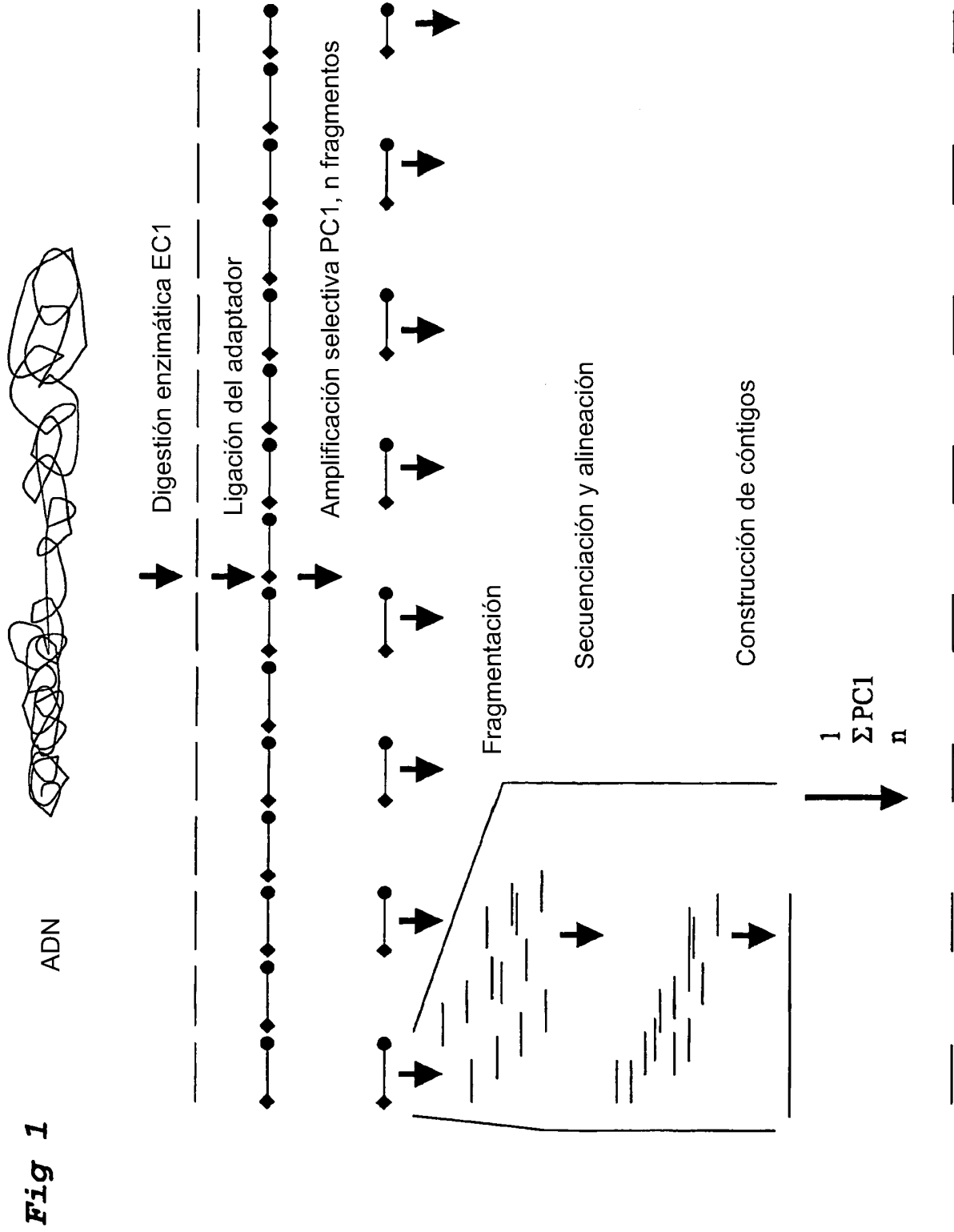


Fig 2

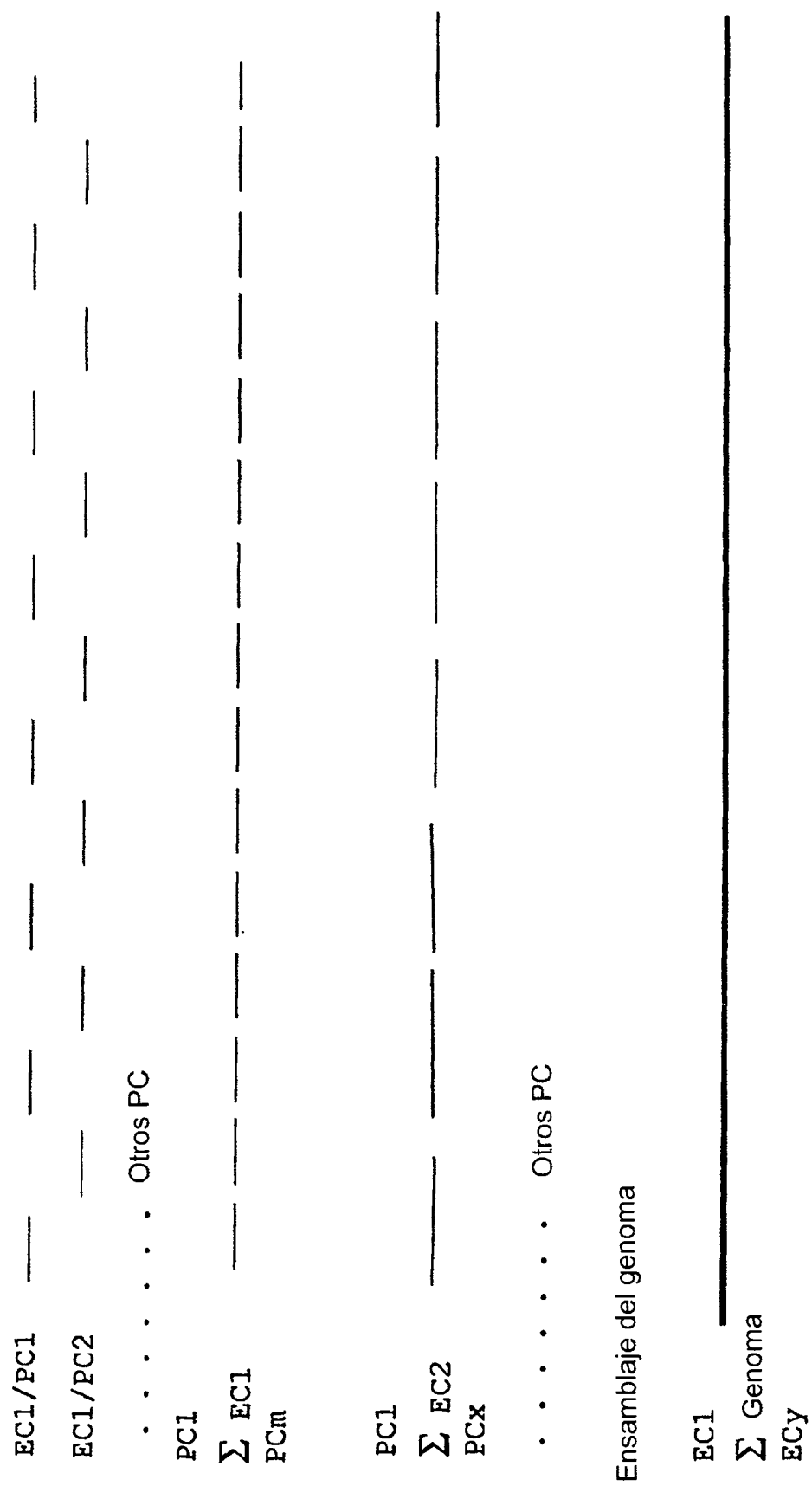


FIG 3

Proyecto: Secuenciación cóntigo E_C-H_CT+B_C-x_G Cóntigos
seleccionado NweRecomb 090606, Cóntigo 606 secuenciados

Página 1

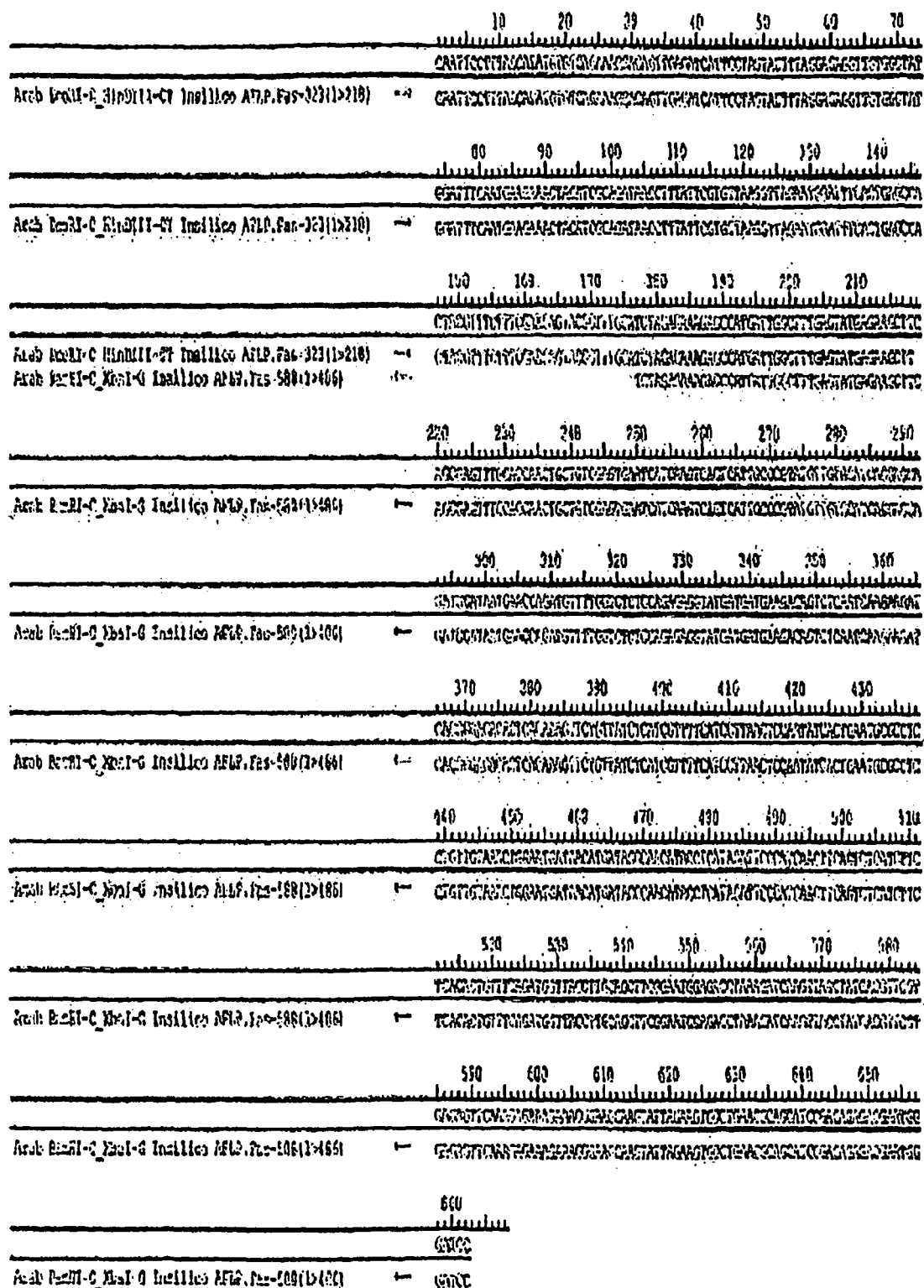


FIG 4

```

      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u gaattcctttaccatattgtgtcagaaggacagttgagatcattgctagtactttaggagaggttgtggctatggatttc
r1_9_35974-36087 u aattcctttaccatattgtgtcagaaggacagttgagatcattgctagtactttaggagaggttgtggctatggatttc

consenso u      gaattcctttaccatattgtgtcagaaggacagttgagatcattgctagtactttaggagaggttgtggctatggatttc
      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u aatgaagaaactacatcgcagataaacctttattcgtgttaagggttagaatggatttcactgagccactacgtttcttttcg
r1_9_35974-36087 u aatgaagaaactacatcgcagataaacctttattcgt
r1_9_36094-36200 u      ggttagaatggatttcactgagccactacgtttcttttcg

consenso u      aatgaagaaactacatcgcagataaacctttattcgtgttaagggttagaatggatttcactgagccactacgtttcttttcg
      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u aagagttacgggtttgcattctagagaaagagccatgattggctttgagttatgagaagcttcagcggagtttgcaccaactgct
r1_9_36094-36200 u aagagttacgggtttgcattctagagaaagagccatgattggctttgagttatgagaagct
r2_9_36138-36227 u      cttagagaaagagccatgattggctttgagttatgagaagcttcagcggagtttgcaccaactgct

consenso u      aagagttacgggtttgcattctagagaaagagccatgattggctttgagttatgagaagcttcagcggagtttgcaccaactgct
      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u gtcgagtgaaatcatcaagtcagtcattgccggtatgttggcatcaggaggagatggataatgaaccagatgttttggctc
r2_9_36138-36227 u gtcgagtgaaatcatc
r2_9_36236-36536 u      cattgccggtatgttggcatcaggaggagatggataatgaaccagatgttttggctc

consenso u      gtcgagtgaaatcatcaagtcagtcattgccggtatgttggcatcaggaggagatggataatgaaccagatgttttggctc
      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u tctccagagaggttatgatgatgaagacagttctcaatcaagaagatcacgggagacactcacaaagttctgttatctcatc
r2_9_36236-36536 u tctccagagaggttatgatgatgaagacagttctcaatcaagaagatcacgggagacactcacaaagttctgttatctcatc

consenso u      tctccagagaggttatgatgatgaagacagttctcaatcaagaagatcacgggagacactcacaaagttctgttatctcatc
      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u gttttcatccttaactccaatatcactgaatggcgcctcctgttgtaaactgggaatgataacatgataggcaacataccctc
r2_9_36236-36536 u gttttcatccttaactccaatatcactgaatggcgcctcctgttgtaaactgggaatgataacatgataggcaacataccctc

consenso u      gttttcatccttaactccaatatcactgaatggcgcctcctgttgtaaactgggaatgataacatgataggcaacataccctc
      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u atagattcccatcaacttcagttctcatcttctcacactgtttcagatggttaccttgcagcttcgggaatggagacctaaa
r2_9_36236-36536 u atagattcccatcaacttcagttctcatcttctcacactgtttcagatggttaccttgcagcttcgggaatggagacctaaa

consenso u      atagattcccatcaacttcagttctcatcttctcacactgtttcagatggttaccttgcagcttcgggaatggagacctaaa
      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u gatcaagtttagctatgaggttgggtgagagttcaaaagagaaaggaagcaagatttagaagtgccctgaacgcagcat
r2_9_36236-36536 u gatc

consenso      gatcaagtttagctatgaggttgggtgagagttcaaaagagaaaggaagcaagatttagaagtgccctgaacgcagcat
      .   :   .   :   .   :   .   :   .   :   .   :   .   :   .   :
cóntigo_606 u ccgacagagacgaatgggatcc

consenso u      ccgacagagacgaatgggatcc

```

ES 2 344 802 T3

LISTA DE SECUENCIAS

<110> Keygene NV

5 <120> Estrategias mejoradas para la secuenciación de genomas complejos utilizando técnicas de secuenciación de alto rendimiento.

<130> P27925PC00

10 <150> US60/693,053
<151> 2005-06-23

15 <150> US60/714,897
<151> 2005-09-08

<150> EP06075104.7

20 <151> 2006-01-16

<150> US60/759,034
<151> 2006-01-17

25 <160> 12

<170> PatentIn version 3.3

30 <210> 1
<211> 17
<212> ADN

35 <213> Artificial

<220>
<223> adaptador

40 <400> 1

45 ctcgtagact gcgtacc 17

<210> 2
<211> 15
<212> ADN

50 <213> Artificial

<220>
<223> adaptador

55 <400> 2

60 gatcggtacg cagtc 15

<210> 3
<211> 17
<212> ADN

65 <213> Artificial

ES 2 344 802 T3

	<220>		
	<223> adaptador		
5	<400> 3		
	ctcgtagact gegtaca		17
10	<210> 4		
	<211> 16		
	<212> ADN		
	<213> Artificial		
15	<220>		
	<223> adaptador		
20	<400> 4		
	ctagtgtacg cagtct		16
25	<210> 5		
	<211> 17		
	<212> ADN		
	<213> Artificial		
30	<220>		
	<223> adaptador		
35	<400> 5		
	ctcgtagact gcgtacc		17
40	<210> 6		
	<211> 18		
	<212> ADN		
	<213> Artificial		
45	<220>		
	<223> adaptador		
50	<400> 6		
	aattggtacg cagtctac		18
55	<210> 7		
	<211> 17		
	<212> ADN		
	<213> Artificial		
60	<220>		
	<223> adaptador		
65	<400> 7		
	ctcgtagact gcgtacc		17

ES 2 344 802 T3

	<210> 8	
	<211> 18	
	<212> ADN	
5	<213> Artificial	
	<220>	
	<223> adaptador	
10	<400> 8	
	agctgggtacg cagtctac	18
15	<210> 9	
	<211> 17	
	<212> ADN	
20	<213> Artificial	
	<220>	
	<223> cebador	
25	<220>	
	<221> base modificada	
30	<222> (16)..(17)	
	<223> enlaces fosforotioato	
	<400> 9	
35	gactgcgtac cgatecc	17
	<210> 10	
40	<211> 17	
	<212> ADN	
	<213> Artificial	
45	<220>	
	<223> cebador	
	<220>	
50	<221> base modificada	
	<222> (16)..(17)	
	<223> enlaces fosforotioato	
55	<400> 10	
	gactgcgtac cgatecc	17
60	<210> 11	
	<211> 17	
	<212> ADN	
	<213> Artificial	
65	<220>	
	<223> cebador	

ES 2 344 802 T3

	<220>	
	<221> base modificada	
	<222> (16)..(17)	
5	<223> enlaces fosforotioato	
	<400> 11	
10	gactgcgtac cgatccc	17
	<210> 12	
	<211> 19	
15	<212> ADN	
	<213> Artificial	
	<220>	
20	<223> cebador	
	<220>	
	<221> base modificada	
25	<222> (16)..(17)	
	<223> enlaces fosforotioato	
30	<400> 12	
	gactgcgtac cgatcscsc	19
35		
40		
45		
50		
55		
60		
65		