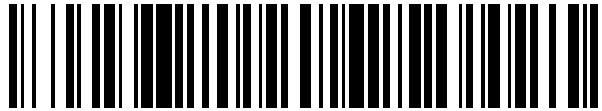


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 545 264**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **04.04.2007 E 07747276 (9)**

97 Fecha y número de publicación de la concesión europea: **10.06.2015 EP 2002017**

54 Título: **Detección de alto rendimiento de marcadores moleculares basada en fragmentos de restricción**

30 Prioridad:

04.04.2006 US 788706 P
12.01.2007 US 880052 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
09.09.2015

73 Titular/es:

KEYGENE N.V. (100.0%)
AGRO BUSINESS PARK 90
6708 PW WAGENINGEN, NL

72 Inventor/es:

VAN EIJK, MICHAEL JOSEPHUS THERESIA y
HOGERS, RENÉ CORNELIS JOSEPHUS

74 Agente/Representante:

SUGRAÑES MOLINÉ, Pedro

ES 2 545 264 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

Detección de alto rendimiento de marcadores moleculares basada en fragmentos de restricción

5 **Campo de la invención**

La presente invención se refiere al campo de la biología molecular y la biotecnología. En particular, la invención se refiere al campo de la identificación y detección de ácidos nucleicos. Más en particular, la invención se refiere a métodos para la detección y la identificación de marcadores, en particular marcadores moleculares. La invención se refiere a la provisión de métodos de alto rendimiento para la detección e identificación de marcadores moleculares. La invención también se refiere a la aplicación del método en la identificación de y/o detección de secuencias de nucleótidos que se relacionan con una gran diversidad de rasgos genéticos, genes, haplotipos y combinaciones de los mismos. La invención se puede usar en el campo de detección e identificación de alto rendimiento de marcadores moleculares de cualquier origen, ya sea vegetal, animal, humano, artificial o cualquier otro.

15

Antecedentes de la invención

La comunidad científica, en particular la comunidad médica, ha deseado durante mucho tiempo la exploración del ADN genómico. El ADN genómico tiene la clave para la identificación, diagnóstico y tratamiento de enfermedades tales como el cáncer y la enfermedad de Alzheimer. Además de la identificación y el tratamiento de la enfermedad, la exploración del ADN genómico puede proporcionar ventajas significativas en los esfuerzos para reproducción de plantas y animales, que puede proporcionar respuestas a los problemas de alimentación y nutrición en el mundo.

20

Se sabe que muchas enfermedades están asociadas con componentes genéticos específicos, en particular con polimorfismos en genes específicos. La identificación de polimorfismos en muestras grandes tales como genomas es en la actualidad una tarea laboriosa y que consume mucho tiempo. Sin embargo, tal identificación es de gran valor en áreas tales como investigación biomédica, desarrollo de productos farmacéuticos, tipificación de tejidos, genotipificación y estudios de población.

25

Los marcadores, es decir, marcadores genéticos, se han usado durante mucho tiempo como un método de tipificación genética, es decir, para conectar un rasgo fenotípico con la presencia, ausencia o cantidad de una parte en particular del ADN (gen). Una de las tecnologías de tipificación genética más versátiles es AFLP, desde hace ya muchos años y se puede aplicar ampliamente aplicable a cualquier organismo (para revisiones véase Savelkoul *et al.* J. Clin. Microbiol, 1999, 37 (10), 3083-3091; Bensch *et al.* Molecular Ecology, 2005, 14, 2899-2914).

30

La tecnología AFLP (Zabeau y Vos, 1993; Vos *et al.*, 1995) ha encontrado un amplio uso en la reproducción de plantas y otros campos desde su invención en la década de los noventa. Esto se debe a varias características de AFLP, de las cuales la más importante es que no se necesita información de la secuencia para generar un gran número de marcadores genéticos de una manera reproducible. Además, el principio de amplificación selectiva, una piedra angular de AFLP, asegura que el número de fragmentos amplificados se puede poner de acuerdo con la resolución del sistema de detección, independientemente del tamaño o el origen del genoma.

35

La detección de fragmentos de AFLP se realiza normalmente por electroforesis en bloques de gel (Vos *et al.*, 1995) o por electroforesis capilar (van der Meulen *et al.*, 2002). La mayoría de los marcadores de AFLP puntuados de esta forma representan polimorfismos (de un solo nucleótido) que se producen en cualquiera de los sitios de reconocimiento de enzimas de restricción usados para la preparación de molde de AFLP o sus nucleótidos de flanqueo cubiertos por cebadores de AFLP selectivos. El resto de los marcadores de AFLP son polimorfismos de inserción/supresión que se producen en las secuencias internas de los fragmentos de restricción y una fracción muy pequeña en sustituciones de un solo nucleótido que se producen en pequeños fragmentos de restricción (< aproximadamente 100 pb), que para estos fragmentos causan variaciones de movilidad reproducibles entre ambos alelos que se pueden observar en la electroforesis; estos marcadores de AFLP se pueden puntuar de forma codominante sin tener que depender de las intensidades de banda.

40

En una identificación genética de AFLP habitual, los marcadores de AFLP constituyen por lo tanto la minoría de los fragmentos amplificados (menos de un 50 por ciento, pero a menudo menos de un 20 por ciento), mientras que el resto se denomina normalmente fragmentos de AFLP constantes. Estos últimos, sin embargo, son útiles en el procedimiento de puntuación de gel ya que sirven como puntos de anclaje para calcular movilidades de fragmentos de marcadores de AFLP y ayudan en la cuantificación de los marcadores para la puntuación codominante. En la actualidad, la puntuación codominante (puntuación para homo o heterocigosidad) de marcadores de AFLP se limita al contexto de la identificación genética de una población segregada. En un panel de líneas no relacionadas, solamente es posible la puntuación dominante.

45

50

Aunque el rendimiento de AFLP es muy elevado debido a los niveles elevados de multiplexación en las etapas de amplificación y detección, la etapa limitante es el poder de resolución de la electroforesis. La electroforesis permite la identificación única de la mayoría de los fragmentos amplificados basándose en la combinación de combinaciones de enzimas de restricción (EC), combinaciones de cebadores (PC) y movilidad, pero la electroforesis solamente es

55

60

65

capaz de distinguir los fragmentos amplificados basándose en las diferencias de movilidad. Los fragmentos de movilidad similar se encuentran a menudo como las llamadas bandas apiladas' <[lambda]> y con electroforesis, con la electroforesis, no se puede proporcionar atención alguna a la información que se contiene en las denominadas bandas constantes', es decir, fragmentos de restricción amplificados que no parecen diferir entre especies comparadas. Además, en un sistema basado en gel habitual, o en un sistema de capilares tal como MegaBACE, las muestras se deben desarrollar en paralelo y solamente se pueden analizar de aproximadamente 100 a 150 bandas por calle en un gel o por capilaridad. Estas limitaciones también obstaculizan el rendimiento.

De forma ideal, el sistema de detección debería ser capaz de determinar toda la secuencia de los fragmentos amplificados para capturar todos los fragmentos de restricción amplificados. Sin embargo, la mayoría de las tecnologías de secuenciación de alto rendimiento aún no pueden proporcionar lecturas de secuenciación que incluyen la totalidad de los fragmentos de AFLP, que por lo general tienen 100-500 pb de longitud.

Hasta ahora, la detección de marcadores/secuencias de AFLP mediante secuenciación no ha sido económicamente factible debido a, entre otras limitaciones, limitaciones del coste de la tecnología de secuenciación dideoxi de Sanger y otras tecnologías de secuenciación convencionales.

La detección mediante secuenciación en lugar de la determinación de la movilidad aumentará el rendimiento porque:

- 1) se detectarán polimorfismos localizados en las secuencias internas en la mayoría (o todos) los fragmentos amplificados; esto aumentará considerablemente el número de marcadores por PC.
- 2) No habrá pérdida de marcadores de AFLP debido a la comigración de marcadores de AFLP y bandas constantes.
- 3) La puntuación codominante no se basa en la cuantificación de las intensidades de banda y es independiente de la relación de los individuos identificados genéticamente.

Sin embargo, la detección por secuenciación de todo el fragmento de restricción todavía es relativamente poco económico. Por otra parte, la tecnología de secuenciación de última generación actual tal como se describe en el presente documento en cualquier parte (de 454 Life Sciences, www.454.com y Solexa, www.solexa.com), a pesar de su poder secuenciación abrumador, solamente puede proporcionar fragmentos de secuenciación de longitud limitada. Además, los métodos actuales no permiten el procesamiento simultáneo de muchas muestras en un desarrollo.

Van Eijk *et al.*, presentaron un póster en la XIV Conferencia sobre Genomas de Plantas y Animales con el título "Complexity reduction of polymorphic sequences (Crops) : a novel approach for high throughput polymorphism discovery" [Online] enero de 2006. El póster desvela un método para el descubrimiento de SNP usando AFLP como un método para reducir la complejidad.

El documento de patente WO2006/137733 describe un método para la identificación de fragmentos de restricción basándose en la secuenciación de fragmentos de restricción ligados a adaptador amplificado.

Definiciones

En la descripción y ejemplos siguientes se usa una serie de términos. Con el fin de proporcionar una comprensión clara y coherente de la memoria descriptiva y las reivindicaciones, incluyendo al alcance que se da a tales términos, se proporcionan las siguientes definiciones. A menos que se defina de otro modo en el presente documento, todos los términos técnicos y científicos usados tienen el mismo significado como normalmente lo entiende un experto habitual en la materia a la que pertenece la presente invención. Las divulgaciones de todas las publicaciones, solicitudes de patentes, patentes y otras referencias se incorporan en el presente documento por referencia en su totalidad.

Ácido nucleico: un ácido nucleico de acuerdo con la presente invención puede incluir 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*, en las páginas 793-800 (Worth Pub. 1982). La presente invención contempla cualquier componente desoxirribonucleótido, ribonucleótido o ácido nucleico peptídico, y cualquier variante química de los mismos, tal como las formas metiladas, hidroximetiladas o glicosiladas de estas bases, y similares. Los polímeros u oligómeros pueden ser heterogéneos u homogéneos en su composición, y se pueden aislar a partir de fuentes de origen natural o se pueden producir de forma artificial o sintética. Además, los ácidos nucleicos pueden ser ADN o ARN, o una mezcla de los mismos, y pueden existir de forma permanente o transitoria en forma monocatenaria o bicatenaria, incluyendo homodúplex, heterodúplex y estados híbridos.

AFLP: AFLP se refiere a un método para la amplificación selectiva de ácidos nucleicos basado en digerir un ácido nucleico con una o más endonucleasas de restricción para producir fragmentos de restricción, ligar adaptadores a los fragmentos de restricción y amplificar los fragmentos de restricción ligados al adaptador con al menos un cebador que es (en parte) complementario al adaptador, (en parte) complementario a los restos de la endonucleasa de restricción, y que contiene adicionalmente al menos un nucleótido seleccionado aleatoriamente entre A, C, T, o G (o

U si este es el caso). El AFLP no requiere ninguna información previa de la secuencia y se puede realizar en cualquier ADN de partida. En general, el AFLP comprende las etapas de:

- 5 (a) digerir un ácido nucleico, en particular un ADN o ADNc, con una o más endonucleasas de restricción específicas, para fragmentar el ADN en una serie correspondiente de fragmentos de restricción;
- (b) ligar los fragmentos de restricción obtenidos de este modo con un adaptador de un oligonucleótido sintético bicatenario, siendo un extremo del mismo compatible con uno o ambos de los extremos de los fragmentos de restricción, para producir de este modo fragmentos de restricción ligados al adaptador, preferentemente marcados, del ADN de partida;
- 10 (c) poner en contacto los fragmentos de restricción ligados al adaptador, preferentemente marcados, en 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 marcado, hibridado con los cebadores por PCR o una técnica similar de tal modo que provoque la elongación adicional de los cebadores hibridados a lo largo de los fragmentos de restricción del ADN de partida con los que se hibridan los cebadores; y
- 15 (e) detectar, identificar o recuperar el fragmento de ADN amplificado o alargado obtenido de este modo.

Por lo tanto, el AFLP proporciona un subconjunto reproducible de fragmentos ligados al adaptador. El AFLP se describe en el documento de patente EP 534858, documento de patente US 6045994 y en Vos *et al.* Se hace referencia a estas publicaciones para detalles adicionales con respecto al AFLP. El AFLP se usa normalmente como una técnica de reducción de complejidad y una tecnología identificación genética del ADN. En el contexto del uso del AFLP como una tecnología de identificación genética, se ha desarrollado el concepto de un marcador del AFLP.

Marcador del AFLP: Un marcador del AFLP es un fragmento de restricción ligado al adaptador amplificado que es diferente entre dos muestras que se han amplificado usando el AFLP (identificación genética), usando el mismo conjunto de cebadores. Como tal, la presencia o ausencia de este fragmento de restricción ligado al adaptador amplificado se puede usar como un marcador que está unido a un rasgo o fenotipo. En la tecnología de gel convencional, un marcador de AFLP se muestra como una banda en el gel situada en una cierta movilidad. Otras técnicas de electroforesis tales como electroforesis capilar pueden no hacer referencia a esto como una banda, pero el concepto sigue siendo el mismo, es decir, un ácido nucleico con una cierta longitud y movilidad. La ausencia o presencia de la banda pueden ser indicativos de (o estar asociados con) la presencia o ausencia del fenotipo. Por lo general, los marcadores del AFLP implican SNP en el sitio de restricción de la endonucleasa o los nucleótidos selectivos. En ocasiones, los marcadores del AFLP pueden implicar indeles en el fragmento de restricción.

Banda constante: en la tecnología de AFLP, una banda constante es un fragmento de restricción ligado al adaptador amplificado que es relativamente invariable entre muestras. Por lo tanto, una banda constante en la tecnología de AFLP, en una serie de muestras, aparecerá aproximadamente en la misma posición en el gel, es decir, tiene la misma longitud/movilidad. En el AFLP convencional, estas se usan por lo general para anclar las calles correspondientes a las muestras en un gel o electroferogramas de múltiples muestras de AFLP detectadas por electroforesis capilar. Por lo general, una banda constante da menos información que un marcador de AFLP. Sin embargo, dado que los marcadores del AFLP implican habitualmente SNP en los nucleótidos selectivos o el sitio de restricción, las bandas constantes pueden comprender SNP en los fragmentos de restricción en sí mismos, haciendo que las bandas constantes sean una fuente alternativa de información genética interesante que es complementaria a los marcadores del AFLP.

Base selectiva: Situada en el extremo 3' del cebador que contiene una parte que es complementaria al adaptador y una parte que es complementaria a los restos del sitio de restricción, la base selectiva se selecciona aleatoriamente entre A, C, T o G. Al prolongar un cebador con una base selectiva, la amplificación posterior producirá solamente un subconjunto reproducible de los fragmentos de restricción ligados al adaptador s, es decir, solamente los fragmentos que se pueden amplificar usando el cebador que lleva la base selectiva. Se pueden añadir nucleótidos selectivos al extremo 3' del cebador en un número que varía entre 1 y 10. Por lo general es suficiente de 1 a 4. Ambos cebadores pueden contener un número variable de bases selectivas. Con cada base selectiva añadida, el subconjunto reduce la cantidad de fragmentos de restricción ligados al adaptador amplificados en el subconjunto en un factor de aproximadamente 4. Por lo general, el número de bases selectivas usadas en AFLP se indica por + N + M, en el que un cebador lleva N nucleótidos selectivos y los otros cebadores llevan M nucleótidos selectivos. Por lo tanto, un AFLP Eco/Mse +1/+2 es la abreviatura para la digestión del ADN de partida con EcoRI y MseI, ligación de los adaptadores apropiados y amplificación con un cebador dirigido a la posición limitada de EcoRI que lleva una base selectiva y el otro cebador dirigido al sitio limitado de MseI que lleva 2 nucleótidos selectivos. Un cebador usado en AFLP que lleva al menos un nucleótido selectivo en su extremo 3' también se representa como un cebador de AFLP. Los cebadores que no llevan un nucleótido selectivo en su extremo 3' y que, de hecho, son complementarios al adaptador y los restos del sitio de restricción se indican en ocasiones como cebadores de AFLP + 0.

Agrupamiento: con el término "agrupamiento" se entiende la comparación de dos o más secuencias de nucleótidos basada en la presencia de tramos cortos o largos de nucleótidos idénticos o similares. En la técnica se conocen varios métodos para el alineamiento de secuencias de nucleótidos, como se explicará adicionalmente a continuación. En ocasiones, los términos "ensamblaje" o "alineamiento" se usan como sinónimos.

- Identificador: una secuencia corta que se puede añadir a un adaptador o un cebador o incluir en su secuencia o usar de otro modo como una marca para proporcionar un identificador único. Tal identificador de secuencia puede ser una secuencia de bases única de longitud variable pero definida usada únicamente para identificar una muestra específica de ácido nucleico. Por ejemplo las marcas de 4 pb permiten $4^4 = 256$ marcas diferentes. Los ejemplos habituales son secuencias ZIP, conocidas en la técnica como marcas usadas normalmente para detección única por hibridación (Iannone *et al.* Cytometry 39: 131-140, 2000). Usando tan identificador, el origen de una muestra de PCR se puede determinar después de su procesamiento adicional. En caso de combinar productos transformados que se originan a partir de diferentes muestras de ácido nucleico, las diferentes muestras de ácido nucleico se identifican generalmente usando diferentes identificadores.
- Secuenciación: El término secuenciación se refiere a la determinación del orden de los nucleótidos (secuencias de bases) en una muestra de ácido nucleico, por ejemplo ADN o ARN.
- Identificación sistemática de alto rendimiento: La identificación sistemática de alto rendimiento, a menudo abreviado como HTS, es un método para experimentación científica especialmente relevante para los campos de la biología y la química. A través de una combinación de robótica moderna y otro hardware de laboratorio especializado, permite que un investigador identifique sistemáticamente de manera eficaz grandes cantidades de muestras de forma simultánea.
- Endonucleasa de restricción: una endonucleasa de restricción o enzima de restricción es una enzima que reconoce una secuencia de nucleótidos específica (sitio diana) en una molécula de ADN bicatenario, y escindirá las dos hebras de la molécula de ADN en o cerca de cada sitio diana.
- Fragmentos de restricción: las moléculas de ADN producidas por digestión con una endonucleasa de restricción se denominan fragmentos de restricción. Cualquier genoma dado (o ácido nucleico, independientemente de su origen) se digerirán con una endonucleasa de restricción particular en un conjunto discreto de fragmentos de restricción. Los fragmentos de ADN que resultan de la escisión de la endonucleasa de restricción se pueden usar adicionalmente en una diversidad de técnicas y se pueden detectar, por ejemplo, por electroforesis en gel.
- Electroforesis en gel: con el fin de detectar los fragmentos de restricción, puede ser necesario un método analítico para fraccionar moléculas de ADN sobre la base del tamaño. La técnica usará más habitualmente para conseguir tal fraccionamiento es la electroforesis en gel (capilar). La tasa a la que se mueven los fragmentos de ADN en tales geles depende de su peso molecular; por lo tanto, las distancias recorridas disminuyen a medida que aumentan las longitudes de los fragmentos. Los fragmentos de ADN fraccionados por electroforesis en gel se pueden visualizar directamente mediante un procedimiento de tinción, por ejemplo, tinción con plata o tinción usando bromuro de etidio, si el número de fragmentos incluidos en el patrón es lo suficientemente pequeño. Como alternativa, el tratamiento adicional de los fragmentos de ADN puede incorporar marcadas detectables en los fragmentos, tales como fluoróforos o marcas radiactivas, que se usan preferentemente para marcar una hebra del producto de AFLP.
- Ligación: la reacción enzimática catalizada por una enzima ligasa en la que se unen dos moléculas de ADN bicatenario de forma covalente entre sí se denomina ligación. En general, ambas hebras de ADN se unen de forma coherente entre sí, pero también es posible evitar la ligación de una de las dos hebras mediante la modificación química o enzimática de uno de los extremos de las hebras. En ese caso, la unión covalente se producirá solamente en una de las dos hebras de ADN.
- Oligonucleótido sintético: las moléculas de ADN monocatenario que tienen preferentemente de aproximadamente 10 a aproximadamente 50 bases, que se pueden sintetizar químicamente se denominan oligonucleótidos sintéticos. En general, estas moléculas de ADN sintéticas se diseñan para que tengan una secuencia de nucleótidos única o deseada, aunque es posible sintetizar familias de moléculas que tienen secuencias relacionadas y que tienen diferentes composiciones de nucleótidos en posiciones específicas dentro de la secuencia de nucleótidos. El término oligonucleótido sintético se usará para hacer referencia a moléculas de ADN que tienen una secuencia de nucleótidos diseñada o deseada.
- Adaptadores: moléculas de ADN bicatenario cortas con un número limitado de pares de bases, por ejemplo, de aproximadamente 10 a aproximadamente 30 pares de bases de longitud, que se diseñan de tal manera que se pueden ligar a los extremos de fragmentos de restricción. Los adaptadores están compuestos generalmente por dos oligonucleótidos sintéticos que tienen secuencias de nucleótidos que son parcialmente complementarias entre sí. Cuando se mezclan los dos oligonucleótidos sintéticos en solución en las condiciones apropiadas, se hibridarán entre sí formando una estructura bicatenaria. Después de la hibridación, un extremo de la molécula de adaptador se diseñan de modo que sea compatible con el extremo de un fragmento de restricción y se puede ligar al mismo; el otro extremo del adaptador se puede diseñar de modo que no se pueda ligar, pero no es necesario que este sea el caso (adaptadores dobles ligados).
- Fragmentos de restricción ligados al adaptador: fragmentos de restricción que se han protegido con adaptadores.
- Cebadores: en general, el término cebadores se refiere a hebras de ADN que pueden cebar la síntesis de ADN. La

ADN polimerasa no puede sintetizar ADN *de novo* sin cebadores: solamente puede prolongar una hebra de ADN existente en una reacción en la que la cadena complementaria se usa como molde para dirigir el orden de los nucleótidos a ensamblar. Los inventores harán referencia a las moléculas de oligonucleótidos sintéticos que se usan en una reacción en cadena de la polimerasa (PCR) como cebadores.

Amplificación de ADN: la expresión amplificación de ADN se usa por lo general para indicar la síntesis *in vitro* de moléculas de ADN bicatenario usando PCR. Se indica que existen otros métodos de amplificación y que se pueden usar en la presente invención sin apartarse de la esencia de la misma.

10 Sumario de la invención

Los presentes inventores han encontrado que los problemas que se han descrito anteriormente y otros problemas de la técnica se pueden superar mediante la elaboración de una forma genérica en la que la versatilidad y aplicabilidad de tecnología de marcadores (AFLP) se pueden combinar con esa tecnología de secuenciación de alto rendimiento de última generación.

Por lo tanto, los presentes inventores han encontrado que mediante la incorporación de un identificador específico de la muestra en el fragmento de restricción ligado al adaptador y/o la determinación solamente de una parte de la secuencia del fragmento de restricción proporciona una mejora muy eficaz y fiable de las tecnologías existentes. Se encontró que mediante la incorporación de un identificador específico de la muestra, se pueden secuenciar múltiples muestras en una única ejecución y mediante la secuenciación solamente de una parte del fragmento de restricción, se puede conseguir la identificación adecuada del fragmento de restricción. La presente invención se define en las reivindicaciones 1-18.

25 Breve descripción de las figuras

Figura 1: es una representación esquemática de la estructura del adaptador que se usa en un enfoque basado en AFLP regular para secuenciación de marca corta por detección con AFLP. Se muestra un fragmento habitual de AFLP derivado una digestión de una muestra de ADN con EcoRI y MseI y ligación del adaptador, seguido de un adaptador habitual para el sitio EcoRI. El adaptador comprende, desde el extremo 5' a 3', una secuencia del cebador 5', que es opcional, y que se puede usar para anclar cebadores de amplificación o para anclar el fragmento ligado al adaptador a una perla o superficie. Además se muestra un identificador (proporcionado como NNNNNN en una forma degenerada), seguido por restos de una secuencia de reconocimiento de una enzima de restricción (en este EcoRI, es decir, AATTC). El último nucleótido del identificador no comprende preferentemente una G para destruir el sitio de restricción EcoRI. Se proporciona un cebador adecuado que comprende la secuencia del cebador 5' opcional, un ejemplo de un cebador específico (ACTGAC), restos del sitio de reconocimiento y una sección que puede contener uno o más nucleótidos selectivos en el extremo 3'.

Figura 2: es una representación esquemática de la realización en la que una secuencia de reconocimiento para una endonucleasa de restricción de tipo IIs se incorpora en el adaptador. Después de la restricción con la enzima de tipo IIs, los adaptadores compatibles de tipo IIs se pueden ligar a uno o ambos de los fragmentos A y B de restricción. El adaptador de tipo IIs comprende una secuencia de unión cebadora opcional (o anclaje), un identificador y una sección que contiene nucleótidos (NN) (degenerados) para hibridarse con el saliente del sitio de restricción IIs. El cebador asociado puede contener uno o más nucleótidos selectivos (XYZ) en su extremo 3'.

45 Descripción detallada de la invención

En un aspecto, la presente divulgación se refiere a un método para la identificación de fragmentos de restricción en una muestra, que comprende las etapas de:

- 50 (a) proporcionar una muestra de ácido nucleico;
- (b) digerir el muestra de ácido nucleico con al menos una endonucleasa de restricción para obtener un conjunto de fragmentos de restricción;
- (c) proporcionar adaptadores sintéticos bicatenarios que comprenden una secuencia compatible con el cebador 5',
 - 55 - una sección identificadora específica de la muestra,
 - una sección que es complementaria con los restos de la secuencia de reconocimiento de la endonucleasa de restricción;
- 60 (d) ligación de los adaptadores sintéticos bicatenarios a los fragmentos de restricción en el conjunto, para proporcionar un conjunto de fragmentos de restricción ligados al adaptador;
- (e) amplificación del conjunto de fragmentos de restricción ligados al adaptador, con uno o más cebadores que son al menos complementarios a:
 - 65 - la sección identificadora específica de la muestra,
 - la sección que es complementaria a los restos de la secuencia de reconocimiento de la endonucleasa de restricción, para proporcionar fragmentos de restricción ligados al adaptador amplificados (amplicones);

(f) determinar la secuencia de al menos la sección identificadora específica de la muestra, los restos de la secuencia de reconocimiento de la endonucleasa de restricción y de parte de la secuencia del fragmento de restricción colocado adyacente a la misma (parte de) de los fragmentos de restricción ligados al adaptador amplificados.

(g) identificar la presencia o ausencia de fragmentos de restricción ligados al adaptador amplificados en la muestra.

Al tratar una muestra de ácido nucleico de esta manera, se obtiene un conjunto de fragmentos de restricción amplificados para cada muestra que se secuencian. Cada fragmento de restricción se puede identificar como procedente de una determinada muestra a través del identificador específico de la muestra que es diferente para cada muestra. La secuenciación de los fragmentos de restricción ligados a adaptadores amplificados proporciona información de la secuencia en al menos parte del fragmento de restricción ligado al adaptador. La información contenida en la parte que se deriva del adaptador contiene información sobre la muestra de la que se obtiene el fragmento, mientras que la información de la secuencia del fragmento de restricción en sí misma proporciona información sobre el fragmento de restricción y permite la identificación del fragmento de restricción. Esta información de la secuencia en el fragmento de restricción se usa para identificar el fragmento de restricción con una precisión que depende del número de nucleótidos que se determina y el número de fragmentos de restricción en el conjunto de fragmentos de restricción ligados a adaptadores amplificados.

Para proporcionar una solución al problema de la variación del muestreo que afecta a la precisión de la identificación de marcadores moleculares mediante secuenciación contenida en un conjunto de múltiples fragmentos, los presentes inventores también han encontrado que la detección de los marcadores a través de la secuenciación se realiza preferentemente con suficiente redundancia (profundidad) para muestrear todos los fragmentos amplificados al menos una vez y acompañado por medios estadísticos que abordan la cuestión de la variación del muestreo en relación con la precisión de los denominados genotipos. Además, simplemente como con la puntuación de AFLP, en el contexto de una población de segregación, la puntuación simultánea de los individuos precursores en un experimento, ayudará a determinar el umbral estadístico.

Por lo tanto, en ciertas realizaciones, la redundancia de los fragmentos de restricción ligados al adaptador amplificados marcados es al menos 6, preferentemente al menos 7, preferentemente al menos 8 y lo más preferentemente al menos 9. En ciertas realizaciones, la secuencia de cada fragmento de restricción ligado al adaptador se determina al menos 6, preferentemente al menos 7, más preferentemente al menos 8 y lo más preferentemente al menos 9 veces. En ciertas realizaciones, la redundancia se selecciona de modo que, suponiendo una posibilidad global de 50/50 de identificar el locus correctamente como homocigoto, la posibilidad de la identificación correcta del locus es superior a un 95 %, un 96 %, un 97 %, un 98 %, un 99 %, un 99,5 %.

A este respecto, el siguiente cálculo puede ser ilustrativo: La tecnología de secuenciación de Solexa como se describe en cualquier parte en el presente documento, proporciona 40.000.000 lecturas de aproximadamente 25 pb cada una, haciendo un total de un asombroso 1 billón de pb en una sola ejecución. Suponiendo una redundancia en el muestreo de 10 veces, se pueden evaluar 4.000.000 de fragmentos únicos en una ejecución. La combinación de 100 muestras permite secuencia 40.000 fragmentos para cada muestra. Visto desde la perspectiva del AFLP, esto equivale a 160 combinaciones de cebadores con 250 fragmentos cada uno.

Este método permite la identificación de fragmentos de restricción de manera que es diferente de la de la detección de marcadores convencionales basada en electroforesis.

En la primera etapa del método para la identificación de fragmentos de restricción se proporciona una muestra de ácido nucleico. Los ácidos nucleicos en la muestra generalmente se presentarán en forma de ADN. Sin embargo, la información de la secuencia de nucleótidos contenida en la muestra puede ser de cualquier fuente de ácidos nucleicos, incluyendo, por ejemplo, ARN, poliA + ARN, ADNc, ADN genómico, ADN organular tal como ADN mitocondrial o de cloroplastos, ácidos nucleicos sintéticos, bibliotecas de ADN (tales como bibliotecas de BAC/clones de BAC agrupados), bancos de clones o cualquier selección o combinaciones de los mismos. El ADN en la muestra de ácido nucleico puede ser bicatenario, monocatenario y ADN bicatenario desnaturalizado en el ADN monocatenario. La muestra de ADN puede ser de cualquier organismo, ya sea vegetal, animal, sintético o humano.

La muestra de ácido nucleico se restringe (o digiere) con al menos una endonucleasa de restricción para proporcionar un conjunto de fragmentos de restricción. En ciertas realizaciones, dos o más endonucleasas se pueden usar para obtener fragmentos de restricción. La endonucleasa puede ser un cortador frecuente (una secuencia de reconocimiento de 3 a 5 pb, tal como MseI) o un cortador raro (secuencia de reconocimiento de > 5 pb, tales como EcoRI). En ciertas realizaciones preferentes, se prefiere una combinación de un cortador raro y uno frecuente. En ciertas realizaciones, en particular cuando la muestra contiene o se deriva de un genoma relativamente grande, puede ser preferente usar una tercera enzima (cortador raro o frecuente) para obtener un conjunto más grande de fragmentos de restricción de tamaño más corto.

Como endonucleasas de restricción, será suficiente cualquier endonucleasa. Por lo general, son preferentes las

endonucleasas de tipo II tales como EcoRI, MseI, PstI, etcétera. En ciertas realizaciones, se puede usar una endonucleasa de tipo IIs, es decir, una endonucleasa cuya secuencia de reconocimiento se coloca alejada del sitio de restricción, es decir, tal como AceIII, AlwI, AlwXI, Alw26I, BbvI, BbvII, BbsI, BclI, Bce83I, BceI, BcgI, BlnI, BsaI, BsgI, BsmI, BsmfI, BspMI, EarI, EciI, Eco3II, Eco57I, Esp3I, FauI, FokI, GsuI, HgaI, HingII, HphI, Ksp632I, MboII, MmeI, MnlI, NgovIII, PfiI, Rleal, SapI, SfaNI, TaqI y ZthII. El uso de este tipo de endonucleasa de restricción conduce a ciertas adaptaciones al método como se describirá en cualquier parte en el presente documento.

Los fragmentos de restricción pueden tener extremos romos o extremos que sobresalen, dependiendo de la endonucleasa usada. Para estos fines, se pueden ligar adaptadores. Por lo general, los adaptadores usados en la presente invención tienen un diseño particular. Los adaptadores usados en la presente invención pueden comprender una secuencia compatible con el cebador 5', que puede ser opcional para proporcionar una longitud suficiente del adaptador para la posterior hibridación del cebador, seguido por una sección identificadora específica de la muestra que puede comprender de 4 a 16 nucleótidos. Preferentemente, el identificador específico de la muestra no contiene 2 o más bases idénticas consecutivas para evitar ensayos preliminares durante la etapa de secuenciación. Además, en el caso en que 2 o más muestras se combinan y se usan múltiples identificadores específicos de la muestra para distinguir el origen de las muestras, existe preferentemente una diferencia entre los identificadores específicos de la muestra de al menos 2, preferentemente 3 pb. Esto permite el aumento de la discriminación entre los diferentes identificadores específicos de la muestra dentro de un grupo combinado de muestras. En el extremo 3' del adaptador se localiza una sección, que es complementaria a los restos de la secuencia de reconocimiento de la endonucleasa de restricción. Por ejemplo, EcoRI reconoce 5'-GAATTC-3' y corta entre G y AATTC. Para EcoRI, la sección complementaria a los restos de la secuencia de reconocimiento de la endonucleasa de restricción es por lo tanto un C-nucleótido.

El adaptador se liga (se conecta de forma covalente) con uno o ambos lados del fragmento de restricción. Cuando la digestión se realiza con más de una endonucleasa, se pueden usar diferentes adaptadores, lo que dará lugar a diferentes conjuntos de fragmentos de restricción ligados al adaptador.

Los fragmentos de restricción ligados al adaptador se amplifican posteriormente con un conjunto de uno o más cebadores. El cebado puede ser complementario al adaptador solamente, es decir, amplificación no selectiva. El cebador contiene preferentemente una sección que es complementaria al identificador específico de la muestra y una sección que es complementaria con los restos de la secuencia de reconocimiento de la endonucleasa de restricción. En ciertas realizaciones, el cebador puede contener en su extremo 3' uno o más nucleótidos selectivos para proporcionar un subconjunto de fragmentos de restricción ligados al adaptador. El cebador también puede contener en su extremo 5' nucleótidos adicionales para ayudar en el anclaje del cebador a los fragmentos de restricción ligados al adaptador. En ciertas realizaciones, el cebado puede contener nucleótidos que expresan un aumento de las características de hibridación tales como LNAs o PNAs. Para amplificar fragmentos de restricción ligados al adaptador a partir de muestras combinadas en un grupo, es posible usar conjuntos de cebadores degenerados, es decir, conjuntos de cebadores en los que para cada muestra, el identificador de la muestra correspondiente se incorpora en el cebador. En ciertas realizaciones, es posible usar conjuntos de cebadores en los que la sección identificadora está totalmente degenerada (o al menos en una gran medida) es decir (casi) cada combinación de nucleótidos se proporciona en la sección identificadora específica de la muestra. En combinación con condiciones de hibridación rigurosas en la amplificación y el uso opcional de nucleótidos de tipo LNA o PNA para aumentar las características de la hibridación, esto puede conducir a una amplificación muy eficaz.

La amplificación de los fragmentos de restricción ligados al adaptador conduce a un conjunto de fragmentos de restricción ligados al adaptador amplificados, denominados en ocasiones amplicones.

Los amplicones (o al menos parte de los mismos) se someten a una etapa que comprende al menos la determinación de la secuencia del identificador específico de la muestra para determinar el origen del fragmento y de la parte de la secuencia del fragmento de restricción. En la práctica esto equivale también a la determinación de las secciones situadas entre ellos tales como los restos de la secuencia de reconocimiento de la endonucleasa de restricción. Al secuenciar el identificador específico de la muestra en combinación con parte del fragmento situado adyacente a la secuencia derivada del adaptador, es posible identificar fragmentos de restricción de forma única. Cuando se correlaciona con la presencia o ausencia de un fenotipo, estos fragmentos de restricción identificados de forma única se pueden usar como marcadores moleculares. Esto permite la definición de una nueva generación de marcadores y equivale, por lo tanto, a una nueva tecnología de marcadores con la versatilidad probada de la tecnología de AFLP, que sin embargo, es adecuada para tecnologías de alto rendimiento y por lo general se puede aplicar entre cualquier tipo de organismo o ácido nucleico. Los fragmentos de restricción identificados de forma única en una muestra mediante la determinación de una parte de su secuencia con este método se puede repetir para múltiples muestras. La presencia o ausencia de los fragmentos de restricción con la secuencia representada en la muestra es indicativo de la presencia o ausencia de un fenotipo.

Una ventaja adicional de la tecnología de marcadores inventada en el momento actual basada en la combinación de AFLP y secuenciación de alto rendimiento es la información adicional que se puede obtener en comparación con la tecnología convencional de AFLP. En el AFLP, los amplicones que se designan como marcadores de AFLP contienen por lo general polimorfismo en el sitio de reconocimiento, el sitio de restricción u opcionalmente, en los

nucleótidos selectivos. Los polimorfismos situados adicionalmente en el fragmento de restricción habitual no califican como marcadores de AFLP (quizá aparte de polimorfismos indel). Con la presente etapa de secuenciación, también se determinan los nucleótidos adyacentes a los nucleótidos selectivos opcionales y esto conduce a la identificación de un mayor número de marcadores moleculares y a una mejora de la tecnología de marcadores existentes.

La secuenciación de alto rendimiento que se usa en la presente invención es un método de experimentación científica especialmente adecuado para los campos de la biología y la química. Mediante una combinación de robótica moderna y otro hardware de laboratorio especializado, el investigador puede identificar sistemáticamente de forma eficaz grandes cantidades de muestras de forma simultánea.

Se prefiere que la secuenciación se realice usando métodos de secuenciación de alto rendimiento, tales como los métodos que se desvelan en los documentos de patente WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007, y WO 2005/003375 (todos en nombre de 454 Life Sciences), de Seo *et al.* (2004) Proc. Natl. Acad. Sci. USA 101: 5488- 93, y tecnologías de Helios, Solexa, US Genomics, etcétera.

Tecnología de 454 Life sciences

En ciertas realizaciones, es preferente que la secuenciación se realice usando el aparato y/o método que se desvela en los documentos de patente WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007, y WO2005/003375 (todos en nombre de 454 Life Sciences). La tecnología que se describe permite la secuenciación de 40 millones de bases en una sola ejecución y es 100 veces más rápida y económica que la tecnología alternativa. La tecnología de secuenciación consiste aproximadamente en 5 etapas: 1) fragmentación de ADN y ligación de adaptadores específicos para crear una biblioteca de ADN monocatenario (ADNss), 2) hibridación del ADNss con perlas, emulsión de las perlas en microrreactores de agua en aceite y realizar la PCR en emulsión para amplificar las moléculas individuales de ADNss en las perlas; 3) selección/enriquecimiento de las perlas que contienen moléculas de ADNss amplificado en su superficie, 4) sedimentación del ADN que porta las perlas en una Placa PicoTiter™; y 5) secuenciación simultánea en 100.000 pocillos mediante la generación de una señal óptica de pirofosfato. El método se explicará con más detalle a continuación.

En una realización preferente, la secuenciación comprende las etapas de:

- (a) hibridación de los fragmentos adaptados con perlas, hibridándose cada perla con un solo fragmento adaptado;
- (b) emulsión de las perlas en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una sola perla;
- (c) carga de las perlas en pocillos, comprendiendo cada pocillo una sola perla; y generación de una señal de pirofosfato.

En la primera etapa (a), los adaptadores de secuenciación se ligan a fragmentos dentro de la biblioteca de combinación. Dicho adaptador de secuenciación incluye al menos una región "clave" para hibridarse con una perla, una región del cebador de secuenciación y una región del cebador de PCR. Por lo tanto, se obtienen fragmentos adaptados. En una primera etapa, los fragmentos adaptados se hibridan con perlas, hibridándose cada perla con un solo fragmento adaptado. Al grupo de fragmentos adaptados, se añaden perlas en exceso con el fin de asegurar la hibridación de un solo fragmento adaptado por perla para la mayoría de las perlas (distribución de Poisson).

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

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

Después de la adición de de perlas que transportan enzimas, se determina la secuencia de los fragmentos usando pirosecuenciación. En etapas sucesivas, la Placa PicoTiter™ y las perlas así como las perlas de enzimas en las mismas se someten a diferentes desoxirribonucleótidos en presencia de reactivos de secuenciación convencionales y, después de 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 pirosecuenciación en sí misma se conoce en la técnica y se describe, entre otros, en www.biotagebio.com; sección de tecnología www.pyrosequencing.com / sección de tecnología. La tecnología se aplica adicionalmente, por ejemplo, en los documentos de patente WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007, y WO 2005/003375 (todos en nombre de 454 Life Sciences). En la presente invención, las perlas

están provistas preferentemente con secuencias cebadoras (unión) o partes de las mismas que son capaces de unirse a los amplicones, si este fuera el caso. En otras realizaciones, los cebadores usados en la amplificación están provistos con secuencias, por ejemplo en su extremo 5', que permiten la unión de los amplicones a las perlas con el fin de permitir la posterior polimerización en emulsión, seguido de secuenciación. Como alternativa, los amplicones se pueden ligar con adaptadores de secuenciación antes de la ligación a las perlas o la superficie. Los amplicones secuenciados revelarán la identidad del identificador y, por lo tanto, la presencia o ausencia del fragmento de restricción en la muestra.

Tecnologías Solexa

Uno de los métodos para secuenciación de alto rendimiento está disponible en Solexa, Reino Unido (www.solexa.co.uk) y se describe, entre otros, en los documentos de patente WO0006770, WO0027521, WO0058507, WO0123610, WO0157248, WO0157249, WO02061127, WO03016565, WO03048387, WO2004018497, WO2004018493, WO2004050915, WO2004076692, WO2005021786, WO2005047301, WO2005065814, WO2005068656, WO2005068089, WO2005078130. Básicamente, el método se inicia con fragmentos ligados al adaptador de ADN genómico. El ADN ligado al adaptador se une aleatoriamente a una capa densa de cebadores que se unen a una superficie sólida, por lo general en un flujo celular. El otro extremo del fragmento ligado al adaptador se hibrida con un cebador complementario en la superficie. Los cebadores se extienden en presencia de nucleótidos y polimerasas en la denominada amplificación por puente en fase sólida para proporcionar fragmentos bicatenarios. Dicha amplificación por puente en fase sólida puede ser una amplificación selectiva.

La desnaturalización y la repetición de la amplificación por puente en fase sólida dan como resultado formación de grupos densos de fragmentos amplificados distribuidos en la superficie. La secuenciación se inicia añadiendo cuatro nucleótidos terminadores reversibles marcados de forma diferente, cebadores y polimerasa al flujo celular. Después de la primera ronda de extensión del cebador, se detectan las marcas, se registra la identidad de las primeras bases incorporadas y el extremo 3' y el fluoróforo se retiran de la base incorporada. A continuación, se determina del mismo modo la identidad de la segunda base y se continúa con la secuenciación de este modo.

En la presente divulgación, los fragmentos de restricción ligados al adaptador o los amplicones se unen a la superficie mediante la secuencia de unión al cebador o la secuencia cebadora. La secuencia se determina tal como se ha descrito, incluyendo la secuencia identificadora y (parte de) el fragmento de restricción. La tecnología Solexa disponible en la actualidad permite la secuenciación de fragmentos de aproximadamente 25 pares de bases. Mediante el diseño económico de los adaptadores y los cebadores unidos a la superficie, la etapa de secuenciación lee, a través del identificador de la muestra, los restos de la secuencia de reconocimiento de la endonucleasa de restricción y cualquier base selectiva opcional. Cuando se usa un identificador de muestra con 6 pb, los restos son del cortador inusual EcoRI (AACCT), el uso de dos bases selectivas proporciona una secuencia interna del fragmento de restricción de 12 pb que se puede usar para identificar únicamente el fragmento de restricción en la muestra.

En una realización preferente basada en la tecnología de secuenciación Solexa mencionada anteriormente, la amplificación de los fragmentos de restricción ligados al adaptador se realiza con un cebador que contiene como máximo un nucleótido selectivo en su extremo 3', preferentemente ningún nucleótido selectivo en su extremo 3', es decir, el cebador solamente es complementario con el adaptador (un cebador +0).

En realizaciones alternativas dirigidas a los métodos de secuenciación que se describen en el presente documento, los cebadores usados en la amplificación pueden contener secciones específicas (como alternativa al cebador o secuencias de unión al cebador que se describen en el presente documento) que se usan en la etapa de secuenciación posterior para unir los fragmentos de restricción protegidos con el adaptador o los amplicones con la superficie. Estas se indican generalmente como la región clave o la secuencia compatible con el cebador 5'.

En una realización de la divulgación, la muestra de ácido nucleico se digiere con al menos una enzima de restricción y al menos se liga un adaptador que comprende una secuencia de reconocimiento para una endonucleasa de restricción de tipo IIs. La digestión posterior del fragmento de restricción ligado al adaptador con una endonucleasa de restricción de tipo IIs proporciona, ya que la distancia entre el sitio de reconocimiento y de restricción de una enzima de tipo IIs es relativamente corta (hasta aproximadamente 30 nucleótidos), un fragmento de restricción más corto y uno más largo, al que se puede ligar un adaptador compatible con un sitio de restricción de tipo IIs. Por lo general, se desconoce el saliente del sitio limitado por IIs de modo que se puede usar un conjunto de adaptadores que se degeneran en el saliente. Después de la amplificación (selectiva), los amplicones se pueden secuenciar. La secuencia adaptadora en esta realización por lo general es la que sigue: 5'-sitio de unión al cebador---secuencia identificadora de la muestra---secuencia del extremo cohesivo de tipo IIs degenerado- 3'. El cebador de PCR asociado generalmente es el que sigue: secuencia cebadora---secuencia identificadora de la muestra---secuencia del extremo cohesivo de tipo IIs degenerado---nucleótidos selectivos -3'. El cebador usado para iniciar la secuenciación por síntesis por lo general tiene la estructura: 5'-sitio de unión al cebador-3'. Puede ser preferente una etapa de selección del tamaño después de la digestión con la enzima IIs para retirar los fragmentos más pequeños. Al igual que en esta realización, los restos del sitio de restricción son para este tipo de enzima por lo general del

orden de 2 a 4 pb, esto resulta en combinación con un identificador de la muestra de 6 pb en la secuenciación de 15 a 17 pb de un fragmento de restricción. En un aspecto adicional, la divulgación se refiere a kits que comprenden uno o más cebadores, y/o uno o más adaptadores para uso del método, además de componentes convencionales para kits *per se*. Además, la presente invención encuentra aplicación, entre otros, en el uso del método para la identificación de marcadores moleculares, para genotipificación, análisis de segregación en masa, formación de mapas genéticos, retrocruzamiento asistido por marcador, formación de mapas de sitios de rasgos cuantitativos, formación de mapas de desequilibrio de unión.

Ejemplo

Se aisló ADN de 2 padres y 88 descendencias usando métodos convencionales. Los padres (2 x) y la descendencia (= 4 x) estaban en dúplex con diferentes índices para poner para someter al ensayo la reproducibilidad. Se usaron marcas para distinguir las muestras entre sí diferían por lo menos en 2 nucleótidos de cualquier otra marca usada en los experimentos. La calidad se está sometiendo a ensayo en las distintas etapas usando geles de agarosa y PAA.

Ejemplo 1

Para cada muestra de ADN, se realiza una etapa de restricción-ligación usando EcoRI y MseI como enzimas. Los adaptadores se basan en las secuencias de hibridación situadas en la superficie del sistema de secuenciación de alto rendimiento Solexa, más en particular, un adaptador de EcoRI contiene la secuencia P5 (parte del cebador de la secuencia) y el adaptador de MseI contiene la secuencia P7 (secuencia de cebador de PCR unido por puente). El adaptador de EcoRI contiene adicionalmente la muestra que identifica la marca. Se usan 96 adaptadores de EcoRI diferentes y un adaptador de MseI. Es posible usar un adaptador de EcoRI degenerado. La preparación del molde incluye una etapa de selección del tamaño mediante incubación de la mezcla durante 10 minutos a 80 grados Celsius después de la etapa de restricción (EcoRI + MseI) pero antes de la etapa de ligación al adaptador. Los fragmentos menores de 130 nt se retiran (en una muestra de maíz).

La complejidad de la mezcla se reduce mediante una amplificación previa selectiva usando +1 cebadores (es decir, que contiene un nucleótido aleatoriamente selectivo en el extremo 3', usando 96 cebadores de EcoRI+1 si un cebador de MseI+1 (o un cebador de EcoRI+1 con marca degenerada y un cebador de MseI +1).

Se realiza amplificación selectiva para reducir la complejidad de la mezcla hasta el tamaño deseado usando cebadores EcoRI+2 (= P5 lateral) y MseI+3 (= P7 lateral) que necesitan el uso de 96 cebadores de EcoRI+2 y un cebador de MseI+3. Se realiza PCR de cola usando un cebador de EcoRI con la secuencia del cebador de PCR con puente en P5 como la cola. Los productos se codifican usando columnas de Sephadex™. Las concentraciones se determinan y se normalizan y se crean grupos. Los grupos se someten a secuenciación masiva en paralelo basándose en tecnología de Solexa que comprende amplificación de PCR por puente y secuenciación seguido de análisis de datos para determinar los genotipos de los padres y la descendencia.

Un escenario alternativo no usa PCR de cola, pero usa cebadores de EcoRI+2 fosforilados. Debido a la falta de coincidencia con el adaptador original, la temperatura de hibridación en el perfil de pacificación se reduce en 3 grados Celsius para 13 ciclos de contacto hacia abajo desde 62-53 grados Celsius seguido de 23 ciclos a 53 grados Celsius. Después de la ligación del adaptador con la secuencia de PCR con puente de P5, se realiza PCR con cebadores de PCR con puente de P5 y P7.

Un segundo escenario alternativo se basa en la preparación de moldes convencionales como se ha descrito anteriormente en el presente documento, (pre)amplificación selectiva para reducir la complejidad. La amplificación selectiva se realiza con cebadores que contienen los sitios de restricción de EcoRI y MseI reconstituidos. Esto permite la retirada de las secuencias adaptadoras antes de la secuenciación, reduciendo de este modo la cantidad de datos a analizar. Se realiza purificación de los productos con columnas de Sephadex™ para retirar restos de ADN Taq polimerasa. Se preparan moldes en los que las secuencias adaptadoras (sitio reconstituido) se reemplazan con adaptadores de Solexa usando un aumento de diez veces de adaptador de EcoRI y enzima de EcoRI para compensar el aumento del número de sitios de EcoRI para el ADN genómico. Los adaptadores de EcoRI de Solexa también contienen las marcas, por lo tanto, se necesitan 96 adaptadores de EcoRI de Solexa marcados. La hebra de la parte final del adaptador se bloquea en el extremo 3' (en este caso con grupo amino 3') para bloquear la extensión con una polimerasa. Se realiza PCR con cebadores de PCR con puente en P5 y P7. Los productos se purifican con columnas de Qiagen.

Ejemplo 2

Se realizó detección de fragmentos de AFLP basada en secuencias usando tecnología de Matriz de Molécula Individual Clonal (CSMATM) de Solexa, una plataforma de Secuenciación por Síntesis capaz de analizar hasta 40 millones de fragmentos individuales en una sola ejecución de la secuencia. La secuencia experimental implica la preparación de moldes de AFLP, amplificación (AFLP) selectiva, amplificación por puente de una sola molécula y secuenciación de millones de marcas de la secuencia a partir de un extremo de la enzima de restricción de los fragmentos de AFLP. Se usaron las líneas precursoras B73 y Mo17 y 87 de maíz. Se usaron Líneas de la Misma

Estirpe Recombinantes (RILs) y se secuencian en 8,9 millones de extremos de fragmentos de EcoRI para AFLP para proporcionar una *prueba del principio* para detección con AFLP basada en secuencias.

5 Se seleccionaron líneas precursoras B73 y Mo17 y 87 RILs. Se prepararon moldes para AFLP usando la combinación de enzima de restricción *EcoRI/MseI*. Se realizó amplificación selectiva usando cebadores +2/+3 para AFLP. Se prepararon fragmentos de molde para amplificación por puente de CSMA de Solexa realizando una segunda restricción/ligación usando adaptadores de *EcoRI* que contienen secuencias únicas de marca de identificación (ID) de la muestra de 5 pb. Se incluyeron líneas precursoras y tres muestras de RIL dos veces usando diferentes marcas de ID de la muestra de 5 pb para medir la reproducibilidad dentro del experimento.

10 Se identificaron marcadores de AFLP basados en secuencia mediante extracción de marcas de secuencias de 27 pb observadas a diferentes frecuencias en B73 y Mo17, que se segregan en la descendencia de RIL.

15 Se compararon datos de marcadores de AFLP basados en secuencia con puntuaciones de marcadores de AFLP obtenidas con identificación genética de AFLP convencional usando detección basada en la longitud de las cuatro combinaciones de *EcoRI/MseI* +3/+3 cebador correspondiente.

Estadística de la ejecución de la secuencia en 5 flujos celulares

# marcas de la secuencia generadas	8.941.407
# marcas de la secuencia con IDs de la muestra desconocidas	8.029.595
# marcas de la secuencia diferentes con IDs de la muestra conocidas	206.758
# datos de la secuencia en Mbp generados por el total del intervalo de la frecuencia	241,4
# marcas de la secuencia por muestra	55.374 - 112.527
# marcadores de con marca en la secuencia	125
intervalo de frecuencia de los marcadores de AFLP con marca en la secuencia en la puntuación precursora presente	90 - 17.218

20

Definición y puntuación de marcador de AFLP con marca en la secuencia

- tabular la representación de marcas en secuencia por muestra
- retirar las marcas de la secuencia con muestra de IDs desconocida
- 25 - normalizar la representación de la muestra basándose en las marcas totales en la secuencia por muestra
- retirar las marcas de la secuencia con una diferencia en la frecuencia > 2 veces en duplos precursores
- frecuencias medias de la marca de duplos precursores
- definir el marcador de AFLP con marca en la secuencia si la frecuencia de P1/P2 supera el valor del umbral
- 30 - puntuar la presencia / ausencia de de marcadores con marca en la secuencia en la descendencia de RIL

30

AFLP distribución del marcador de AFLP +3/+3; secuencia / basado en gel

	total	EcoRI + 3 base			
		+A	+C	+G	+T
# marcadores de AFLP con marca en la secuencia	125	34	37	37	17
# marcadores de AFLP basados en gel	82	29	18	17	18

Reproducibilidad de descendencia de 3 RIL de duplos marcadores de AFLP con marca en la secuencia

35

# marcadores de AFLP con marca en la secuencia puntuados	125
# número de puntos de datos en comparación	375
# puntos de datos idénticos para duplos	372
% de concordancia con los duplos del experimento	99,2 %

Detección de bloque de gel convencional:

40

	B73	Mo17	1	2	3	4	5	6	7	8	9	10	11	12
Marcador de AFLP														
E36/M50-175.9	-	+	+	-	-	-	-	+	-	+	-	-	-	+
E36/M50-280	+	-	+	-	-	+	-	+	+	-	+	-	-	-
E36/M50-405.8	-	+	+	-	-	-	+	+	+	+	+	+	-	+
E36/M50-243.7	+	-	+	-	-	-	-	-	+	+	+	+	+	+
E36/M50-124.02	+	-	+	-	+	+	+	+	-	-	-	-	+	+
E36/M60-379	+	-	+	-	-	+	+	+	+	+	+	+	-	+
E36/M50-468.9	+	-	+	-	+	+	-	+	-	+	+	+	+	+

Detección basada en Solexa

	B73	Mo17	1	2	3	4	5	6	7	8	9	10	11	12
Marcador de AFLP														
CGGCGACGTACCGC	-	+	+	-	-	-	-	+	-	+	-	-	-	+
CTAGTAATTATTCC	+	-	+	-	-	+	-	+	+	-	+	-	-	-
CAGCGCCTTCTCCT	-	+	+	-	-	-	+	+	+	+	-	+	-	+
CAGAACTCTGACTT	+	-	+	-	-	-	-	-	+	+	+	+	+	+
CAAATCTGTTAGAT	+	-	+	-	+	+	+	+	-	+	-	-	+	+
CATGAAGGATTTAT	+	-	+	-	-	+	+	+	+	+	+	+	-	+
CAAACAGACAACCG	+	-	+	-	+	+	-	+	-	+	+	+	+	+

- 5 El marcador de AFLP basado en la viabilidad de la secuencia se generó usando tecnología CSMA de Solexa, mediante la cual un número mayor de marcadores de AFLP se puntúa usando detección basada en secuencias que en bloques de gel convencionales, debido supuestamente al aumento de la resolución (tamaño del fragmento) y secuenciación profunda que también captura fragmentos de baja abundancia. Las comparaciones de vectores de datos del marcador revelan patrones de segregación similares entre detección basada en secuencias y detección de bloque de gel: la prueba de concordancia esperada la secuenciación de marcadores de AFLP basados en gel.
- 10

REIVINDICACIONES

1. Método para identificar la presencia o ausencia de fragmentos de restricción en una muestra, que comprende las etapas de:
- 5
- (a) proporcionar dos o más muestras de ácidos nucleicos;
 - (b) digerir cada muestra de ácido nucleico con al menos una endonucleasa de restricción para obtener un conjunto de fragmentos de restricción;
 - (c) proporcionar adaptadores sintéticos bicatenarios que comprenden
- 10
- una secuencia complementaria del cebador 5',
 - una sección identificadora específica de la muestra,
 - al menos un extremo
- 15
- que se puede ligar al extremo como o que sobresale de un fragmento de restricción;
- (d) ligar los adaptadores sintéticos bicatenarios a los fragmentos de restricción en el conjunto, para proporcionar un conjunto de fragmentos de restricción ligados al adaptador;
 - (e) opcionalmente, amplificar el conjunto de fragmentos de restricción ligados al adaptador, con uno o más cebadores que son al menos complementarios a:
- 20
- la secuencia complementaria del cebador 5' del adaptador,
 - la sección identificadora específica de la muestra del adaptador, y,
 - opcionalmente, una sección del adaptador que es complementaria a los restos que sobresalen de la secuencia de reconocimiento de al menos una endonucleasa de restricción,
- 25
- para proporcionar fragmentos de restricción ligados al adaptador amplificados (amplicones);
- (f) determinar la secuencia de al menos la sección identificadora específica de la muestra, los restos de la secuencia de reconocimiento de al menos una endonucleasa de restricción y parte de la secuencia del fragmento de restricción situada adyacente a la secuencia derivada del adaptador o adyacente a los restos de la secuencia de reconocimiento de al menos una endonucleasa de restricción,
 - (g) comparar dos o más muestras para la presencia o ausencia de fragmentos de restricción ligados al adaptador;
 - (h) identificar la presencia o ausencia de fragmentos de restricción ligados al adaptador en la muestra.
- 30
2. Método de acuerdo con la reivindicación 1, en el que los fragmentos de restricción son marcadores moleculares.
- 35
3. Método de acuerdo con la reivindicación 2, en el que los marcadores moleculares son marcadores de AFLP.
4. Método de acuerdo con la reivindicación 1, en el que dos o más muestras se combinan en un grupo después de la etapa de ligación de los adaptadores.
- 40
5. Método de acuerdo con la reivindicación 4, en el que, para cada muestra en el grupo, se usa un identificador específico de la muestra que se diferencia de los otros identificadores específicos de la muestra en el grupo.
- 45
6. Método de acuerdo con la reivindicación 1, en el que los cebadores contienen uno o más nucleótidos selectivos en el extremo 3'.
7. Método de acuerdo con la reivindicación 1, en el que la endonucleasa de restricción es una endonucleasa de restricción de tipo II.
- 50
8. Método de acuerdo con la reivindicación 1, en el que la endonucleasa de restricción es una endonucleasa de restricción de tipo IIs.
9. Método de acuerdo con la reivindicación 1, en el que se usan dos o más endonucleasas de restricción.
- 55
10. Método de acuerdo con la reivindicación 1, en el que la secuenciación se realiza por medio de secuenciación de alto rendimiento.
- 60
11. Método de acuerdo con la reivindicación 10, en el que la secuenciación de alto rendimiento se realiza en un soporte sólido.
- 60
12. Método de acuerdo con la reivindicación 10, en el que la secuenciación de alto rendimiento se basa en Secuenciación por Síntesis.
- 65
13. Método de acuerdo con la reivindicación 10, en el que la secuenciación de alto rendimiento comprende las etapas de:

ES 2 545 264 T3

- hibridar los amplicones o fragmentos de restricción ligados al adaptador con perlas, hibridándose cada perla con un solo fragmento de restricción ligado al adaptador o amplicón;
 - emulsionar las perlas en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una sola perla;
- 5
- realizar PCR por emulsión para amplificar el fragmento de restricción ligado al adaptador o amplicones;
 - opcionalmente, seleccionar / enriquecer las perlas que contienen amplicones amplificados;
 - cargar las perlas en pocillos, comprendiendo cada pocillo una sola perla; y
 - generar una señal de pirofosfato.
- 10
14. Método de acuerdo con la reivindicación 10, en el que la secuenciación de alto rendimiento comprende las etapas de:
- hibridar los fragmentos de restricción ligados al adaptador o amplicones con una superficie que contiene el primer y segundo cebadores o la primera y segunda secuencias de unión al cebador respectivamente,
- 15
- realizar amplificación por puente para proporcionar fragmentos de restricción ligados al adaptador amplificados o amplicones amplificados,
 - determinar la secuencia de nucleótidos de los fragmentos de restricción ligados al adaptador amplificados o amplicones amplificados usando nucleótidos terminadores reversibles marcados.
- 20
15. Método de acuerdo con la reivindicación 1, en el que el identificador tiene de 4 a 16 pb, preferentemente de 4 a 10 pb, más preferentemente de 4 a 8 pb, lo más preferentemente de 4 a 6 pb.
- 25
16. Método de acuerdo con la reivindicación 15, en el que el identificador no contiene 2 o más bases consecutivas idénticas.
- 30
17. Método de acuerdo con la reivindicación 15, en el que, para dos o más muestras, los identificadores correspondientes contienen al menos dos nucleótidos diferentes.
18. Uso del método como se define en las reivindicaciones 1-17 para la identificación de marcadores moleculares, para genotipificación, análisis de segregación en masa, formación de mapas genéticos, retrocruzamiento asistido por marcador, formación de mapas de sitios de rasgos cuantitativos, formación de mapas de desequilibrio de unión.

FIG 1

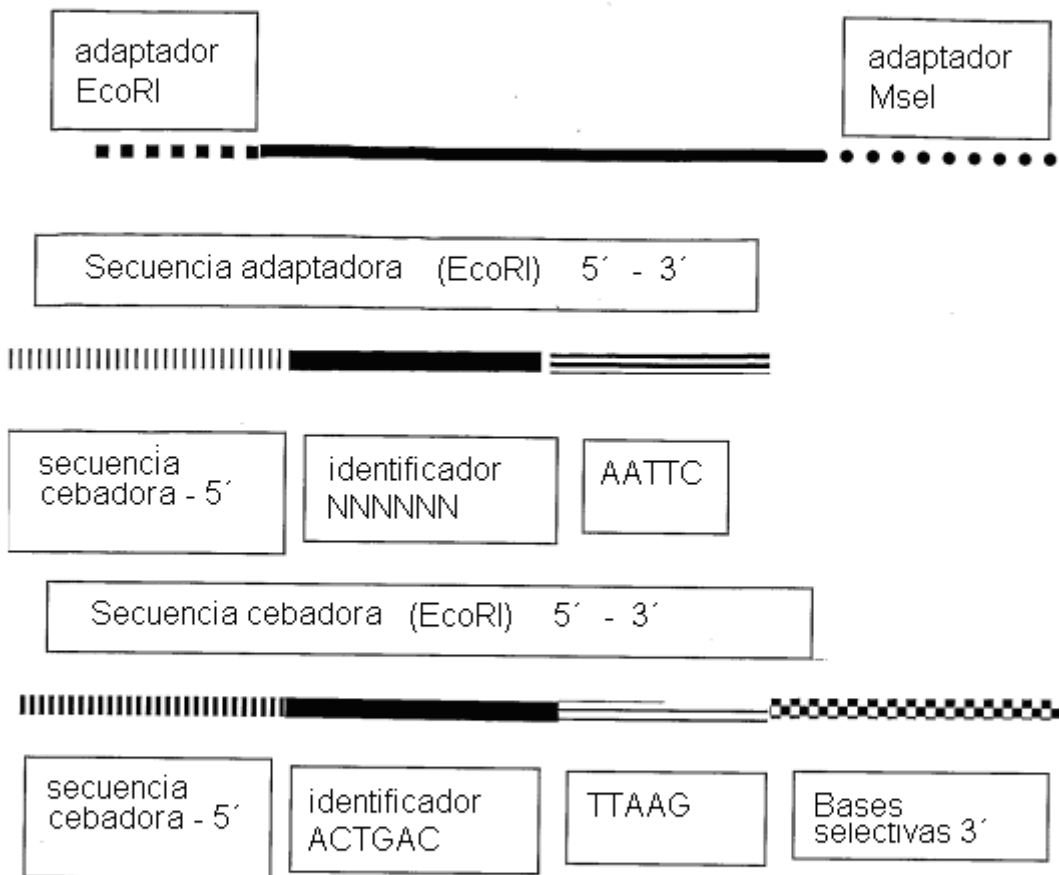
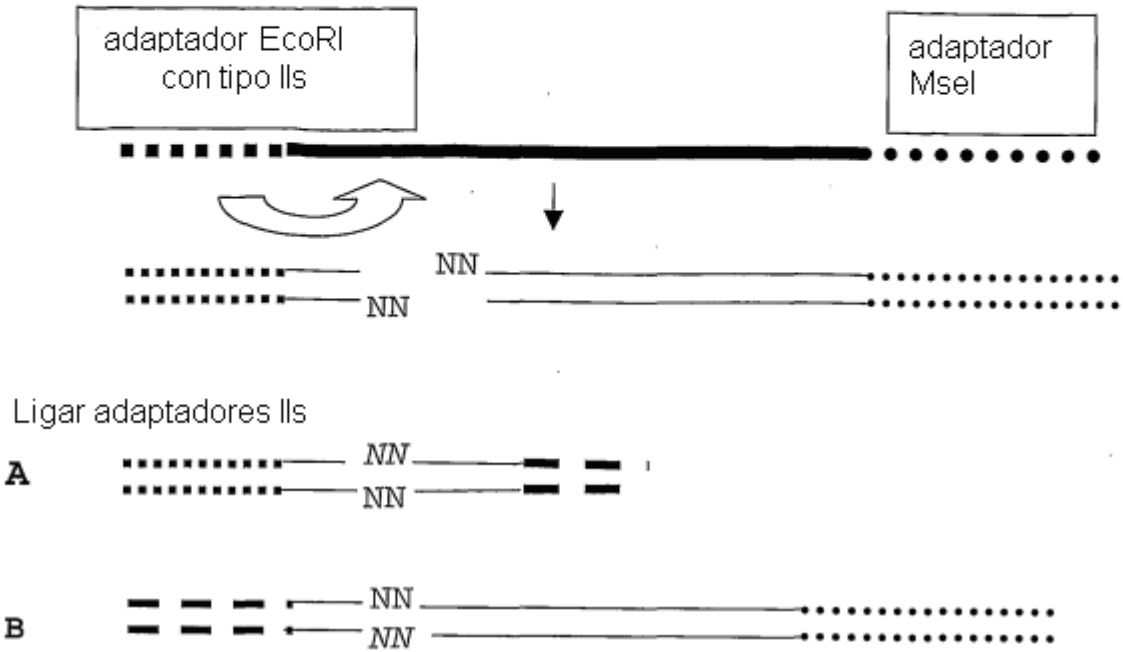
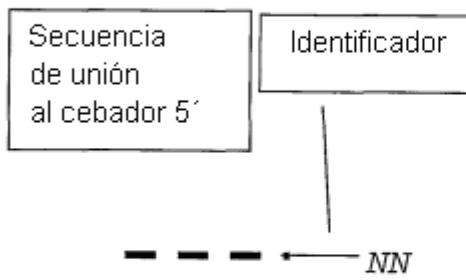


FIG 2



Estructura del adaptador:



Estructura del cebador:

