



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS  
ESPAÑA



⑪ Número de publicación: **2 829 295**

⑮ Int. Cl.:  
**C12Q 1/6853** (2008.01)

⑫

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑯ Fecha de presentación y número de la solicitud europea: **04.04.2007 E 17163116 (1)**

⑯ Fecha y número de publicación de la concesión europea: **19.08.2020 EP 3239304**

⑮ Título: **Detección de alto rendimiento de marcadores moleculares basados en AFLP y secuenciación de alto rendimiento**

⑯ Prioridad:

**04.04.2006 US 788706 P  
12.01.2007 US 880052 P**

⑮ Fecha de publicación y mención en BOPI de la traducción de la patente:  
**31.05.2021**

⑮ Titular/es:

**KEYGENE N.V. (100.0%)  
P.O. Box 216  
6700 AE Wageningen, NL**

⑯ Inventor/es:

**VAN EJK, MICHAEL JOSEPHUS THERESIA y  
HOGERS, RENÉ CORNELIS JOSEPHUS**

⑯ Agente/Representante:

**TOMAS GIL, Tesifonte Enrique**

**ES 2 829 295 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 basados en AFLP y secuenciación de alto rendimiento

Campo de la invención

5 [0001] 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 la 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 y la identificación de marcadores moleculares. La invención se refiere además a la aplicación del método en la identificación y/o la detección de secuencias de nucleótidos que están relacionadas con una amplia variedad de rasgos genéticos, 10 genes, haplotipos y combinaciones de los mismos. La invención se puede usar en el campo de la detección de alto rendimiento y la identificación de marcadores moleculares de cualquier origen, sea vegetal, animal, humano, artificial u otro.

Antecedentes de la invención

15 [0002] La comunidad científica, médica en particular, ha deseado durante mucho tiempo la exploración del ADN genómico. El ADN genómico es la clave para la identificación, el diagnóstico y el tratamiento de las enfermedades tales como el cáncer y la enfermedad de Alzheimer. Además de la identificación y el tratamiento de las enfermedades, la exploración del ADN genómico puede proporcionar ventajas significativas a los esfuerzos en mejora vegetal y animal, que pueden proporcionar respuestas a los problemas de alimentación y nutrición en el mundo.

20 [0003] Muchas enfermedades son conocidas por estar asociadas a componentes genéticos específicos, en particular a polimorfismos en genes específicos. La identificación de polimorfismos en muestras grandes tales como los genomas es en la actualidad una tarea laboriosa y que conlleva mucho tiempo. Sin embargo, dicha identificación es de gran valor para áreas tales como la investigación biomédica, el desarrollo de productos farmacéuticos, la tipificación de tejidos, la genotipificación y los estudios poblacionales.

25 [0004] Los marcadores, es decir, los marcadores genéticos, han sido usados durante mucho tiempo como un método de tipificación genética, es decir, para conectar un rasgo fenotípico con la presencia, la ausencia o la cantidad de una parte particular del ADN (gen). Una de las tecnologías de tipificación genética más versátil es la de AFLP, ya existente durante muchos años y ampliamente aplicable a cualquier organismo (para revisiones véanse Savelkoul *et al.* *J. Clin. Microbiol.*, 1999, 37 (10), 3083-3091; Bensch *et al.* *Molecular Ecology*, 2005, 14, 30 2899-2914).

35 [0005] La tecnología de AFLP (Zabeau & Vos, 1993; Vos *et al.*, 1995) ha encontrado un uso amplio en la mejora vegetal y otros campos desde su invención a principios de la década de los noventa. Esto se debe a varias características de AFLP, de las cuales la más importante es que no se necesita ninguna información de secuencia previa para generar un gran número de marcadores genéticos r/*l* de una forma reproducible. Además, el principio de amplificación selectiva, un principio básico de AFLP, asegura que el número de fragmentos amplificados se puede ajustar a la resolución del sistema de detección, sin tener en cuenta el origen o el tamaño del genoma.

40 [0006] La detección de fragmentos AFLP se realiza comúnmente por electroforesis en geles planos (Vos *et al.*, 1995) o electroforesis capilar (van der Meulen *et al.*, 2002). La mayoría de marcadores de AFLP clasificados de esta manera representan polimorfismos (de nucleótido único) que ocurren o en los sitios de reconocimiento de enzimas de restricción usadas para la preparación del modelo de AFLP o en sus nucleótidos flanqueantes cubiertos por cebadores de AFLP selectivos. El resto de los marcadores de AFLP son polimorfismos de inserción/deletión que ocurren en las secuencias internas de los fragmentos de restricción y una fracción muy pequeña en sustituciones de nucleótido único que ocurren en fragmentos de restricción pequeños (< 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 clasificar codominantemente sin tener que depender de las intensidades de banda.

45 [0007] Por lo tanto, en una huella de AFLP típica, los marcadores de AFLP constituyen la minoría de fragmentos amplificados (menos de un 50 por ciento, pero a menudo menos de un 20 por ciento), mientras que el resto se denominan comúnmente fragmentos AFLP constantes. Sin embargo, estos últimos son útiles en el procedimiento de clasificación en gel, ya que sirven como puntos de anclaje para calcular movilidades de fragmentos de los marcadores de AFLP y ayudar en la cuantificación de los marcadores para la clasificación codominante. La clasificación codominante (clasificación según la homocigosidad o la heterocigosidad) de los marcadores de AFLP

está actualmente restringida al contexto de obtención de huellas de una población segregante. En un panel de líneas no relacionadas, solo es posible la clasificación dominante.

[0008] Aunque el rendimiento de AFLP es muy alto debido a altos niveles de multiplexado en los pasos de amplificación y detección, el paso limitante de la velocidad es la potencia de resolución de la electroforesis. La electroforesis permite la identificación única de la mayoría de fragmentos amplificados basándose en la combinación de combinaciones de enzimas de restricción (EC), combinaciones de cebadores (PC) y movilidad, pero la electroforesis solo es capaz de distinguir los fragmentos amplificados en función de las diferencias de movilidad. Los fragmentos de movilidad similar se encuentran a menudo como las denominadas 'bandas apiladas' y, con la electroforesis, no se le puede prestar ninguna atención a la información que está contenida en las denominadas 'bandas constantes', es decir, fragmentos de restricción amplificados que no parecen diferir entre las especies comparadas. Además, en un sistema basado en gel típico, o en un sistema capilar tal como un MegaBACE, las muestras deben correrse en paralelo y solo pueden analizarse aproximadamente 100-150 bandas por carril en un gel o por capilar. Estas limitaciones obstaculizan también el rendimiento.

[0009] Idealmente, 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 tecnologías de secuenciación de alto rendimiento no pueden proporcionar aún lecturas de secuenciación que abarquen fragmentos AFLP enteros, que tienen típicamente 100-500 pb de longitud.

[0010] Hasta ahora, la detección de marcadores/secuencias AFLP por secuenciación no ha sido económicamente realizable debido a, entre otras limitaciones, las limitaciones de los costes de la tecnología de secuenciación didesoxi de Sanger y otras tecnologías de secuenciación convencionales.

[0011] La detección por secuenciación en vez de la determinación de la movilidad aumentará el rendimiento porque:

- 1) Los polimorfismos situados en las secuencias internas se detectarán en la mayoría de (o en todos) los fragmentos amplificados; esto aumentará el número de marcadores por PC considerablemente.
- 2) No habrá ninguna pérdida de marcadores de AFLP debida a la comigración de marcadores de AFLP y bandas constantes.
- 3) La clasificación codominante no depende de la cuantificación de las intensidades de bandas y es independiente de la relación de los individuos de los que se han obtenido las huellas.

[0012] Sin embargo, la detección por secuenciación de todo el fragmento de restricción sigue siendo relativamente no económica. Además, el estado de la técnica actual de la tecnología de secuenciación tal como se describe aquí en otra parte (de 454 Life Sciences, [www.454.com](http://www.454.com) y Solexa, [www.solexa.com](http://www.solexa.com)), a pesar de su potencia de secuenciación sobrecogedora, solo puede proporcionar fragmentos de secuenciación de longitud limitada. Además, los métodos actuales no permiten el procesamiento simultáneo de muchas muestras en una ronda.

#### Definiciones

[0013] En la descripción y los ejemplos siguientes se utilizan una serie de términos. Para proporcionar una comprensión clara y consistente de la especificación y las reivindicaciones, incluido el alcance que ha de darse a tales términos, se proporcionan las definiciones siguientes. A menos que se defina de otro modo aquí, todos los términos técnicos y científicos usados tienen el mismo significado que entiende comúnmente una persona experta en la materia a la que pertenece esta invención.

[0014] Ácido nucleico: un ácido nucleico conforme a la presente invención puede incluir cualquier polímero u oligómero de las bases pirimidínicas y púricas, preferiblemente citosina, timina y uracilo, y adenina y guanina, respectivamente (véase Albert L. Lehninger, *Principles of Biochemistry*, en 793-800 (Worth Pub. 1982)). La presente invención contempla cualquier desoxirribonucleótido, ribonucleótido o componente de ácido nucleico péptido, y cualquier variante química de los mismos, tales como las formas metiladas; hidroximetiladas o glicosiladas de estas bases, y similares. Los polímeros u oligómeros pueden tener una composición heterogénea u homogénea, y se pueden aislar de fuentes de origen natural o se pueden producir artificial o sintéticamente. Además, los ácidos nucleicos pueden ser de ADN o de ARN, o una mezcla de los mismos, y pueden existir de manera permanente o transitoria en la forma monocatenaria o bicatenaria, incluidos homodúplex, heterodúplex y estados híbridos.

[0015] AFLP: AFLP se refiere a un método para la amplificación selectiva de ácidos nucleicos que se basa en la digestión de un ácido nucleico con una o más endonucleasas de restricción para producir fragmentos de restricción, el ligamiento de adaptadores a los fragmentos de restricción y la amplificación de los fragmentos de restricción

ligados a los adaptadores con al menos un cebador que sea complementario (en parte) al adaptador, complementario (en parte) a los restos de la endonucleasa de restricción, y que además contenga al menos un nucleótido seleccionado de forma aleatoria de entre A, C, T o G (o U, según el caso). AFLP no requiere ninguna información de secuencia previa y se puede realizar en cualquier ADN de partida. En general, AFLP comprende los pasos 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;
- 10 (b) ligar los fragmentos de restricción así obtenidos con un adaptador oligonucleotídico sintético bicitenario, un extremo del cual es compatible con uno o los dos extremos de los fragmentos de restricción, para producir así, fragmentos de restricción ligados a adaptadores, preferiblemente etiquetados, del ADN de partida;
- 15 (c) poner en contacto los fragmentos de restricción ligados a adaptadores, preferiblemente etiquetados, bajo condiciones de hibridación con uno o más cebadores oligonucleotídicos que contienen nucleótidos selectivos en sus extremos 3';
- (d) amplificar el fragmento de restricción ligado a un adaptador, preferiblemente etiquetado, hibridado con los cebadores por PCR o una técnica similar para causar un alargamiento adicional de los cebadores hibridados a lo largo de los fragmentos de restricción del ADN de partida con los que los cebadores han hibridado; y
- 15 (e) detectar, identificar o recuperar el fragmento de ADN amplificado o alargado así obtenido.

[0016] AFLP proporciona así un subconjunto reproducible de fragmentos ligados a adaptadores. AFLP se describe en la EP 534858, la US 6045994 y en Vos *et al.* Se hace referencia a estas publicaciones para detalles adicionales con respecto a AFLP. El AFLP se usa comúnmente como una técnica de reducción de la complejidad y una tecnología de obtención de huellas genéticas. En el contexto del uso de AFLP como una tecnología de obtención de huellas, se ha desarrollado el concepto de un marcador de AFLP.

[0017] Marcador de AFLP: un marcador de AFLP es un fragmento de restricción ligado a un adaptador amplificado que es diferente entre dos muestras que han sido amplificadas usando AFLP (se han obtenido sus huellas), usando el mismo conjunto de cebadores. Como tal, la presencia o la ausencia de este fragmento de restricción ligado a un adaptador amplificado se puede usar como un marcador que está enlazado a un rasgo o fenotipo. En la tecnología de gel convencional, un marcador de AFLP aparece como una banda en el gel localizada con una determinada movilidad. En otras técnicas electroforéticas como la electroforesis capilar puede que no se haga referencia a esto como una banda, pero el concepto es el mismo, es decir, un ácido nucleico con una longitud y una movilidad determinadas. La ausencia o la presencia de la banda puede ser indicativa de (o estar asociada a) la presencia o la ausencia del fenotipo. Los marcadores de AFLP implican típicamente SNPs en el sitio de restricción de la endonucleasa o los nucleótidos selectivos. Ocasionalmente, los marcadores de AFLP pueden implicar indels en el fragmento de restricción.

[0018] Banda constante: una banda constante en la tecnología de AFLP es un fragmento de restricción ligado a un adaptador amplificado que es relativamente invariable entre muestras. Así, una banda constante en la tecnología de AFLP aparecerá, en un rango de muestras, en aproximadamente la misma posición en el gel, es decir, tiene la misma longitud/movilidad. En AFLP convencional estas se usan típicamente para anclar los cariles correspondientes a las muestras en un gel o los electroferogramas de múltiples muestras AFLP detectadas por electroforesis capilar. Típicamente, una banda constante es menos informativa que un marcador de AFLP. Sin embargo, como los marcadores de AFLP habituales implican SNPs en los nucleótidos selectivos o el sitio de restricción, las bandas constantes pueden comprender SNPs en los propios fragmentos de restricción, convirtiendo a las bandas constantes en una fuente alternativa interesante de información genética que es complementaria a los marcadores de AFLP.

[0019] 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 de forma aleatoria de entre A, C, T o G. Extendiendo un cebador con una base selectiva, la amplificación posterior producirá solo un subconjunto reproducible de los fragmentos de restricción ligados a adaptadores, es decir, solo 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 variable entre 1 y 10. Típicamente 1-4 bastan. 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 a adaptadores amplificados en el subconjunto por un factor de aproximadamente 4. Típicamente, el número de bases selectivas usadas en AFLP se indica por  $+N+M$ , donde un cebador lleva N nucleótidos selectivos y los otros cebadores llevan M nucleótidos selectivos. Así, un AFLP Eco/Mse +1/+2 es la forma abreviada para la digestión del ADN de partida con EcoRI y MseI, el ligamiento de adaptadores apropiados y la amplificación con un cebador dirigido a la posición restringida por EcoRI que lleva una base selectiva y el otro cebador dirigido al sitio restringido por 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 AFLP-cebador. Los cebadores que no llevan un nucleótido selectivo en su extremo 3' y que son, de hecho, complementarios al adaptador y los restos del sitio de restricción se indican a veces como cebadores de AFLP+0.

[0020] Agrupamiento: el término "agrupamiento" se refiere a la comparación de dos o más secuencias de nucleótidos en función de la presencia de extensiones cortas o largas de nucleótidos idénticos o similares. Se conocen en la técnica varios métodos para el alineamiento de secuencias de nucleótidos, como se explicará adicionalmente a continuación. A veces, los términos "ensamblaje" o "alineamiento" se usan como sinónimos.

5 [0021] Identificador: una secuencia corta que se puede añadir a un adaptador o un cebador o incluir en su secuencia o, en su defecto, usarse como un marcador 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 de ácido nucleico específica. Por ejemplo, las etiquetas de 4 pb permiten  $4^{(exp4)} = 256$  etiquetas diferentes. Ejemplos típicos son las secuencias ZIP, conocidas en la técnica como etiquetas comúnmente 10 usadas para la detección única por hibridación (Iannone *et al.* Cytometry 39:131-140, 2000). Usando dicho identificador, se puede determinar el origen de una muestra de PCR tras un procesamiento adicional. En caso de combinar productos procesados que se originan a partir de muestras de ácido nucleico diferentes, las muestras de ácido nucleico diferentes se identifican generalmente usando diferentes identificadores.

15 [0022] 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.

[0023] Cribado de alto rendimiento: el cribado de alto rendimiento, a menudo abreviado como HTS, es un método para la experimentación científica especialmente pertinente 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 crie eficazmente grandes cantidades de muestras simultáneamente.

20 [0024] 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 bícatenario, y escindirá ambas cadenas de la molécula de ADN en o cerca de cada sitio diana.

25 [0025] 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) será digerido por una endonucleasa de restricción particular en un conjunto discreto de fragmentos de restricción. Los fragmentos de ADN que se originan a partir de la escisión por endonucleasas de restricción se pueden usar adicionalmente en una variedad de técnicas y pueden, por ejemplo, ser detectados por electroforesis en gel.

30 [0026] Electroforesis en gel: para detectar fragmentos de restricción, puede requerirse un método analítico para fraccionar moléculas de ADN basándose en el tamaño. La técnica más frecuentemente usada para conseguir dicho fraccionamiento es la electroforesis en gel (capilar). La velocidad a la que se mueven los fragmentos de ADN en dichos geles depende de su peso molecular; así, las distancias recorridas se reducen a medida que las longitudes de los fragmentos aumentan. Los fragmentos de ADN fraccionados por electroforesis en gel se pueden visualizar directamente por un procedimiento de tinción, por ejemplo tinción de plata o tinción usando bromuro de etidio, si el número de fragmentos incluidos en el patrón es suficientemente pequeño. Alternativamente el tratamiento adicional de los fragmentos de ADN puede incorporar marcadores detectables en los fragmentos, tales como fluoróforos o marcadores radiactivos, que se usan preferiblemente para marcar una cadena del producto AFLP.

35 [0027] Ligamiento: la reacción enzimática catalizada por una enzima ligasa en la que dos moléculas de ADN bícatenario se unen de manera covalente se denomina ligamiento. En general, ambas cadenas de ADN se unen de manera covalente, pero también es posible evitar el ligamiento de una de las dos cadenas a través de la modificación química o enzimática de uno de los extremos de las cadenas. En ese caso la unión covalente ocurrirá en solo una de las dos cadenas de ADN.

40 [0028] Oligonucleótido sintético: las moléculas de ADN monocatenario con preferiblemente de aproximadamente 10 a aproximadamente 50 bases que se pueden sintetizar químicamente se denominan oligonucleótidos sintéticos. 45 En general, estas moléculas de ADN sintéticas están diseñadas para tener 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 de la secuencia de nucleótidos. El término oligonucleótido sintético se usará para referirse a moléculas de ADN con una secuencia diseñada o deseada de nucleótidos.

50 [0029] Adaptadores: moléculas de ADN bícatenario cortas con un número limitado de pares de bases, por ejemplo de aproximadamente 10 a aproximadamente 30 pares de bases de longitud, que están diseñadas de manera que se puedan ligar a los extremos de los fragmentos de restricción. Los adaptadores están generalmente compuestos por dos oligonucleótidos sintéticos que tienen secuencias de nucleótidos que son parcialmente complementarias entre sí. Cuando se mezclan los dos oligonucleótidos sintéticos en solución bajo condiciones apropiadas, se

aparearán entre sí formando una estructura bicatenaria. Después del apareamiento, un extremo de la molécula de adaptador está diseñada 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 pueda ligarse, pero no es necesario que sea así (adaptadores ligados doblemente).

5 [0030] Fragmentos de restricción ligados a adaptadores: fragmentos de restricción que han sido cubiertos por adaptadores.

10 [0031] Cebadores: en general, el término cebadores se refiere a cadenas de ADN que pueden promover la síntesis de ADN. La ADN-polimerasa no puede sintetizar ADN *de novo* sin cebadores: solo puede extender una cadena de ADN existente en una reacción en la que la cadena complementaria se usa como un modelo para dirigir el orden de los nucleótidos que se van a ensamblar. Nos referiremos a las moléculas de oligonucleótido sintéticas que se usan en una reacción en cadena de la polimerasa (PCR) como cebadores.

[0032] Amplificación de ADN: el término amplificación de ADN se usará típicamente para indicar la síntesis *in vitro* de moléculas de ADN bicatenario usando una PCR. Cabe señalar que existen otros métodos de amplificación y que se pueden usar en la presente invención sin apartarse de lo esencial.

15 Resumen de la invención

[0033] La presente invención se define en las reivindicaciones anexas.

20 [0034] Los presentes inventores han descubierto que los problemas descritos anteriormente y otros problemas en la técnica se pueden superar concibiendo una manera genérica donde la versatilidad y la aplicabilidad de la tecnología de marcadores (AFLP) se pueda combinar con la de la tecnología de secuenciación de alto rendimiento del estado de la técnica.

25 [0035] Así, los presentes inventores han descubierto que la incorporación de un identificador específico de muestra en el fragmento de restricción ligado a un adaptador y/o la determinación de solo parte de la secuencia del fragmento de restricción proporciona una mejora muy eficiente y fiable de las tecnologías existentes. Se ha observado que, mediante la incorporación de un identificador específico de muestra, se pueden secuenciar múltiples muestras en una única ronda y, mediante la secuenciación de solo parte del fragmento de restricción, se puede conseguir una identificación adecuada del fragmento de restricción.

Breve descripción de los dibujos

[0036]

30 **Figura 1:** es una representación esquemática de la estructura del adaptador que se usa en un método basado en AFLP habitual para la secuenciación de etiquetas cortas para la detección de AFLP. Se muestra un fragmento AFLP típico derivado de un producto de digestión de una muestra de ADN con EcoRI y MseI y un ligamiento del adaptador posterior, seguido de un adaptador típico para el sitio EcoRI. El adaptador comprende, del extremo 5' al 3', una secuencia de cebador 5', que es opcional, y se puede usar para anclar cebadores de amplificación o para anclar el fragmento ligado a un adaptador a una esfera o superficie.

35 Además, se muestra un identificador (dado como NNNNNNN en una forma degenerada), seguido de restos de una secuencia de reconocimiento de una enzima de restricción (en esta EcoRI, es decir, AATTC). El último nucleótido del identificador preferiblemente no comprende una G para destruir el sitio de restricción EcoRI. Se proporciona un cebador adecuado que comprende la secuencia de 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'.

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

Descripción detallada de la invención

50 [0037] En un aspecto, la presente descripción se refiere a un método para la identificación de fragmentos de restricción en una muestra, que comprende los pasos de:

(a) proporcionar un ácido nucleico de muestra;  
 (b) digerir el ácido nucleico de muestra con al menos una endonucleasa de restricción para obtener un conjunto de fragmentos de restricción;  
 (c) proporcionar adaptadores sintéticos bicatenarios que comprenden

5 - una secuencia compatible con el cebador 5',  
 - una sección identificadora específica de muestra,  
 - una sección que sea complementaria a los restos de la secuencia de reconocimiento de la endonucleasa de restricción;

10 (d) ligar los adaptadores sintéticos bicatenarios a los fragmentos de restricción en el conjunto, para proporcionar un conjunto de fragmentos de restricción ligados a adaptadores;

(e) amplificar el conjunto de fragmentos de restricción ligados a adaptadores, con uno o más cebadores que son al menos complementarios a:

15 - la sección identificadora específica de 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 a adaptadores amplificados (amplicones);

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

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

25 [0038] Tratando un ácido nucleico de muestra de esta manera, se obtiene un conjunto de fragmentos de restricción amplificados por cada muestra que se secuencie. Cada fragmento de restricción se puede identificar como originado a partir de cierta muestra a través del identificador específico de muestra que es diferente para cada muestra. La secuenciación de los fragmentos de restricción ligados a adaptadores amplificados proporciona información de secuencia en al menos parte del fragmento de restricción ligado a un adaptador. La información contenida en la parte derivada del adaptador contiene información acerca de la muestra de la que se obtiene el fragmento, mientras que la información de secuencia del fragmento de restricción mismo proporciona información acerca del fragmento de restricción y permite la identificación del fragmento de restricción. Esta información de secuencia en el fragmento de restricción se utiliza para identificar el fragmento de restricción con una exactitud que depende del número de nucleótidos que se determine y el número de fragmentos de restricción en el conjunto de fragmentos de restricción ligados a adaptadores amplificados.

35 [0039] Para proporcionar una solución al problema de variación de muestreo que afecta a la exactitud de la identificación de marcadores moleculares por secuenciación contenidos en un conjunto de múltiples fragmentos, los presentes inventores también han descubierto que la detección de marcadores a través de la secuenciación se realiza preferiblemente con suficiente redundancia (profundidad) para muestrear todos los fragmentos amplificados al menos una vez y acompañada por medios estadísticos que abordan el problema de la variación de muestreo en relación a la exactitud de los genotipos llamados. Además, al igual que con la clasificación AFLP, en el contexto 40 de una población segregante, la clasificación simultánea de los individuos parentales en un experimento, ayudará a determinar el umbral estadístico.

45 [0040] Así, en determinadas formas de realización, la redundancia de los fragmentos de restricción ligados a adaptadores amplificados y etiquetados es al menos 6, preferiblemente al menos 7, más preferiblemente al menos 8 y más preferiblemente al menos 9. En determinadas formas de realización, la secuencia de cada fragmento de restricción ligado a un adaptador se determina al menos 6, preferiblemente al menos 7, más preferiblemente al menos 8 y más preferiblemente al menos 9 veces. En determinadas formas de realización, la redundancia se selecciona de modo que, suponiendo una probabilidad general de 50/50 de identificar el locus correctamente como homocigótico, la probabilidad de identificar correctamente el locus sea más de un 95%, 96%, 97%, 98%, 99%, 99,5%.

50 [0041] A este respecto puede ser ilustrativo el cálculo siguiente: la tecnología de secuenciación de Solexa como se describe en este caso en otra parte, proporciona 40.000.000 de lecturas de aproximadamente 25 pb cada una, alcanzando el asombroso total de 1 mil millones pb en una única ronda. Suponiendo una redundancia en el muestreo de 10 veces, se pueden evaluar 4.000.000 de fragmentos únicos en una ronda. La combinación de 100 muestras permite secuenciar 40.000 fragmentos por cada muestra. Visto desde la perspectiva de AFLP, esto equivale a 160 combinaciones de cebadores con 250 fragmentos cada uno.

[0042] Este método permite la identificación de fragmentos de restricción de manera que es diferente de la detección de marcadores convencional a base de electroforesis.

[0043] En el primer paso del método para la identificación de fragmentos de restricción se proporciona un ácido nucleico de muestra. Los ácidos nucleicos en la muestra estarán usualmente en forma de ADN. Sin embargo, la 5 información de secuencia de nucleótidos contenida en la muestra puede ser de cualquier fuente de ácidos nucleicos, incluidos, por ejemplo, ARN, ARN poliA+, ADNc, ADN genómico, ADN organular tal como ADN mitocondrial o de cloroplastos, ácidos nucleicos sintéticos, genotecas de ADN (tal como genotecas de BAC/clones de BAC agrupados), bancos de clones o cualquier selección o combinación de los mismos. El ADN en la muestra de ácido nucleico puede ser bicatenario, monocatenario y ADN bicatenario desnaturalizado en ADN 10 monocatenario. La muestra de ADN puede ser de cualquier organismo, ya sea vegetal, animal, sintético o humano.

[0044] 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 determinadas formas de realización, se pueden usar dos o más endonucleasas para obtener fragmentos de restricción. La endonucleasa puede ser un cortador 15 frecuente (una secuencia de reconocimiento de 3-5 pb, tal como Msel) o un cortador raro (secuencia de reconocimiento de >5 pb, tal como EcoRI). En determinadas formas de realización preferidas, se prefiere una combinación de un cortador raro y uno frecuente. En determinadas formas de realización, en particular cuando la muestra contiene o deriva de un genoma relativamente grande, se puede preferir usar una tercera enzima (cortador raro o frecuente) para obtener un conjunto mayor de fragmentos de restricción de tamaño más corto.

[0045] Como endonucleasas de restricción, cualquier endonucleasa bastará. Típicamente, se prefieren 20 endonucleasas de tipo II tales como EcoRI, Msel, PstI, etc. En determinadas formas de realización, se puede usar una endonucleasa de tipo IIs, es decir, una endonucleasa cuya secuencia de reconocimiento está localizada distante del sitio de restricción, es decir, tales como AcellI, AlwI, Alw26I, BbvI, BbvII, BbsI, BccI, Bce83I, Bcefl, Bcgl, BinI, Bsal, BsgI, BsmAI, BsmFI, BspMI, Earl, Ecil, Eco31I, Eco57I, Esp3I, Faul, FokI, Gsul, Hgal, HinGII, HphI, Ksp632I, Mboll, Mmel, MnII, NgoVIII, PleI, RleAI, SapI, SfaNI, TaqJI y ZthII III. El uso de este tipo 25 de endonucleasa de restricción lleva a ciertas adaptaciones al método, como se describirá aquí en otra parte.

[0046] Los fragmentos de restricción pueden tener extremos romos o extremos salientes, dependiendo de la endonucleasa usada. A estos extremos, pueden ligarse adaptadores. Típicamente, los adaptadores usados en la 30 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 el posterior apareamiento del cebador, seguido de una sección identificadora 35 específica de muestra que puede comprender de 4 a 16 nucleótidos. Preferiblemente el identificador específico de muestra no contiene 2 o más bases idénticas consecutivas para evitar ultralecturas durante el paso de secuenciación. Además, en el caso de que se combinen 2 o más muestras y se utilicen múltiples identificadores específicos de muestra para distinguir el origen de las muestras, hay preferiblemente una diferencia entre los 40 identificadores específicos de muestra de al menos 2, preferiblemente 3 pb. Esto permite una discriminación mejorada entre los diferentes identificadores específicos de 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 AATT. Por lo tanto, para EcoRI, la sección complementaria a los restos de la secuencia de reconocimiento de la endonucleasa de restricción es un nucleótido C.

[0047] El adaptador se liga (se conecta de manera covalente) con uno o ambos lados del fragmento de restricción. Cuando se realiza una digestión con más de una endonucleasa, se pueden usar diferentes adaptadores que darán lugar a diferentes conjuntos de fragmentos de restricción ligados a adaptadores.

[0048] Los fragmentos de restricción ligados a adaptadores se amplifican posteriormente con un conjunto de uno 45 o más cebadores. El cebador puede ser complementario solo al adaptador, es decir, amplificación no selectiva. El cebador contiene preferiblemente una sección que es complementaria al identificador específico de muestra y una sección que es complementaria a los restos de la secuencia de reconocimiento de la endonucleasa de restricción. En determinadas formas de realización, el cebador puede contener en su extremo 3' uno o más nucleótidos selectivos para proporcionar un subconjunto de fragmentos de restricción ligados a adaptadores amplificados. El 50 cebador puede, en su extremo 5', contener también nucleótidos adicionales para ayudar al anclaje del cebador a los fragmentos de restricción ligados a adaptadores. En determinadas formas de realización, el cebador puede contener nucleótidos que expresan características de hibridación mejoradas tales como LNAs o PNAs. Para amplificar fragmentos de restricción ligados a adaptadores de muestras combinadas en un grupo es posible usar 55 conjuntos de cebadores degenerados, es decir, conjuntos de cebadores donde, para cada muestra, el identificador de muestra correspondiente se incorpora en el cebador. En determinadas formas de realización, es posible usar conjuntos de cebadores donde la sección identificadora es completamente degenerada (o al menos en gran parte), es decir, (casi) cada combinación de nucleótidos se proporciona en la sección identificadora específica de 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 hibridación, esto puede conducir a una amplificación muy eficiente.

[0049] La amplificación de los fragmentos de restricción ligados a adaptadores llevan a un conjunto de fragmentos de restricción ligados a adaptadores amplificados, denominados a veces amplicones.

5 [0050] Los amplicones (o al menos parte de los mismos) se someten a un paso que comprende al menos la determinación de la secuencia del identificador específico de muestra para determinar el origen del fragmento y de parte de la secuencia del fragmento de restricción. En la práctica esto equivale también a la determinación de las secciones situadas entremedias tales como los restos de la secuencia de reconocimiento de la endonucleasa de restricción. Secuenciando el identificador específico de muestra en combinación con parte del fragmento situado 10 adyacente a la secuencia derivada del adaptador, es posible identificar únicamente fragmentos de restricción. Cuando se correlaciona con la presencia o la ausencia de un fenotipo, estos fragmentos de restricción identificados únicamente se pueden usar como marcadores moleculares. Esto permite la definición de una nueva generación 15 de marcadores y equivale por lo tanto a una tecnología de marcadores novedosa con la versatilidad probada de la tecnología de AFLP, pero que es adecuada para las tecnologías de alto rendimiento y es generalmente aplicable entre cualquier tipo de organismo o ácido nucleico. La identificación de manera única de fragmentos de restricción en una muestra por determinación de parte de su secuencia mediante este método se puede repetir para múltiples muestras. La presencia o la ausencia de los fragmentos de restricción con la secuencia representada en la muestra es indicativa de la presencia o la ausencia de un fenotipo.

20 [0051] Una ventaja adicional de la tecnología de marcadores actualmente inventada basada en la combinación de AFLP y la secuenciación de alto rendimiento es la información adicional que se puede obtener en comparación con la tecnología de AFLP convencional. En AFLP, los amplicones que se designan como marcadores de AFLP contienen típicamente un polimorfismo en el sitio de reconocimiento, el sitio de restricción u, opcionalmente, en los 25 nucleótidos selectivos. Los polimorfismos situados más alejados en el fragmento de restricción típicamente no se califican como marcadores de AFLP (aparte de quizás los polimorfismos de indels). Con el presente paso de secuenciación, también se determinan los nucleótidos adyacentes a los nucleótidos selectivos opcionales y esto lleva a la identificación de un mayor número de marcadores moleculares y a una mejora en la tecnología de 30 marcadores existente.

[0052] La secuenciación de alto rendimiento usada en la presente invención es un método de experimentación científica especialmente pertinente para los campos de la biología y la química. A través de una combinación de 35 robótica moderna y otro *hardware* de laboratorio especializado, permite que un investigador críbe eficazmente grandes cantidades de muestras simultáneamente. Se prefiere que la secuenciación se realice usando métodos de secuenciación de alto rendimiento, tales como los métodos descritos en la WO 03/004690, la WO 03/054142, la WO 2004/069849, la WO 2004/070005, la WO 2004/070007 y la WO 2005/003375 (todas en nombre de 454 Life Sciences), por Seo *et al.* (2004) Proc. Natl. Acad. 40 Sci. USA 101:5488-93, y las tecnologías de Helios, Solexa, US Genomics, etc.

#### Tecnología de 454 Life Sciences

[0053] En determinadas formas de realización, se prefiere que la secuenciación se realice usando el aparato y/o el método descrito en la WO 03/004690, la WO 03/054142, la WO 2004/069849, la WO 2004/070005, la WO 40 2004/070007 y la WO 2005/003375 (todas en nombre de 454 Life Sciences). La tecnología descrita permite la secuenciación de 40 millones de bases en una única ronda y es 100 veces más rápida y menos costosa que la tecnología competidora. La tecnología de secuenciación consiste en aproximadamente 5 pasos: 1) fragmentación del ADN y ligamiento de adaptadores específicos para crear una genoteca de ADN monocatenario (ssADN); 2) apareamiento de ssADN a esferas, emulsión de las esferas en microrreactores de agua en aceite y realización de 45 PCR en emulsión para amplificar las moléculas de ssADN individuales en esferas; 3) selección de enriquecimiento para las esferas que contienen moléculas de ssADN amplificado en su superficie, 4) deposición de esferas que llevan ADN en una placa PicoTiter™; y 5) secuenciación simultánea en 100.000 pocillos por generación de una señal luminosa de pirofosfato. El método se explicará con más detalle a continuación.

En una forma de realización preferida, la secuenciación comprende los pasos de:

50 (a) aparear fragmentos adaptados a esferas, donde cada esfera se aparea con un único fragmento adaptado; (b) emulsionar las esferas en microrreactores de agua en aceite, donde cada microrreactor de agua en aceite comprende una única esfera; (c) cargando las esferas en pocillos, donde cada pocillo comprende una única esfera; y generar una señal de pirofosfato.

55 En el primer paso (a), los adaptadores de secuenciación se ligan a fragmentos de la genoteca de combinación. Dicho adaptador de secuenciación incluye al menos una región "clave" para aparearse a una esfera, una región

de cebador de secuenciación y una región de cebador de PCR. Así, se obtienen fragmentos adaptados. En un primer paso, los fragmentos adaptados se aparean a esferas, donde cada esfera se aparea con un único fragmento adaptado. Al grupo de fragmentos adaptados, se añaden esferas en exceso para asegurar el apareamiento de un único fragmento adaptado por esfera para la mayoría de las esferas (distribución de Poisson).

- 5 En un paso siguiente, las esferas se emulsionan en microrreactores de agua en aceite, donde cada microrreactor de agua en aceite comprende una única esfera. Los reactivos de la PCR están presentes en los microrreactores de agua en aceite, lo que permite que se produzca una reacción por PCR en los microrreactores. Posteriormente, los microrreactores se rompen, y las esferas que comprenden ADN (esferas positivas para ADN) se enriquecen.
- 10 En un paso siguiente, las esferas se cargan en pocillos, donde cada pocillo comprende una única esfera. Los pocillos son preferiblemente parte de una placa PicoTiter™, lo que permite una secuenciación simultánea de una gran cantidad de fragmentos.

Después de la adición de las esferas que llevan enzimas, la secuencia de los fragmentos se determina usando una pirosecuenciación. En pasos sucesivos, la placa PicoTiter™ y las esferas al igual que las esferas enzimáticas incluidas se someten a diferentes desoxirribonucleótidos en presencia de reactivos de secuenciación convencionales y, tras la incorporación de un desoxirribonucleótido, se genera una señal lumínica que se registra.

15 La incorporación del nucleótido correcto generará una señal de pirosecuenciación que puede ser detectada.

La pirosecuenciación misma se conoce en la técnica y se describe entre otros en [www.bioteagebio.com](http://www.bioteagebio.com); [www.pyrosequencing.com](http://www.pyrosequencing.com) / sección de tecnología. La tecnología se aplica posteriormente, por ejemplo, en la WO 03/004690, la WO 03/054142, la WO 2004/069849, la WO 2004/070005, la WO 2004/070007 y la WO 2005/003375 (todas en nombre de 454 Life Sciences).

20 En la presente invención, las esferas están equipadas preferiblemente con secuencias (de unión) de cebadores o partes de las mismas que son capaces de unirse a los amplicones, según el caso. En otras formas de realización, los cebadores usados en la amplificación están equipados con secuencias, por ejemplo en su extremo 5', que permiten la unión de los amplicones a las esferas para permitir la polimerización en emulsión posterior seguida de 25 la secuenciación. Alternativamente los amplicones se pueden ligar con adaptadores de secuenciación antes del ligamiento a las esferas o la superficie. Los amplicones secuenciados revelarán la identidad del identificador y así de la presencia o la ausencia del fragmento de restricción en la muestra.

#### Tecnologías de Solexa

30 [0054] Uno de los métodos para la secuenciación de alto rendimiento está disponible de Solexa, Reino Unido ([www.solexa.co.uk](http://www.solexa.co.uk)) y se describe, entre otras, en la WO0006770, la WO0027521, la WO058507, la WO0123610, la WO0157248, la WO0157249, la WO02061127, la WO03016565, la WO03048387, la WO2004018497, la WO2004018493, la WO2004050915, la WO2004076692, la WO2005021786, la WO2005047301, la WO2005065814, la WO2005068656, la WO2005068089 y la WO2005078130. En esencia, el método empieza con

35 fragmentos ligados a adaptadores de ADN genómico. El ADN ligado al adaptador se une de forma aleatoria a un manto denso de cebadores que se fijan a una superficie sólida, típicamente en una célula de flujo. El otro extremo del fragmento ligado a un adaptador hibrida con un cebador complementario en la superficie. Los cebadores se extienden en presencia de nucleótidos y polimerasas en una amplificación denominada de puente en fase sólida para proporcionar fragmentos bicatenarios. Esta amplificación de puente en fase sólida puede ser una amplificación selectiva. La desnaturización y la repetición de la amplificación de puente en fase sólida da lugar a

40 agrupaciones densas de fragmentos amplificados distribuidos sobre la superficie. La secuenciación se inicia por adición de cuatro nucleótidos de terminación reversibles marcados de manera diferente, cebadores y polimerasa a la célula de flujo. Después de la primera ronda de extensión de cebadores, se detectan los marcadores, se registra la identidad de las primeras bases incorporadas y se eliminan el extremo 3' bloqueado y el fluoróforo de la base incorporada. Luego la identidad de la segunda base se determina de la misma manera y así continúa la 45 secuenciación.

50 [0055] En la presente descripción, los fragmentos de restricción ligados a adaptadores o los amplicones se unen a la superficie a través de la secuencia de unión de cebadores o la secuencia de cebador. La secuencia se determina como se ha resumido, incluida la secuencia identificadora y (parte de) el fragmento de restricción. La tecnología de Solexa disponible actualmente 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, el paso

55 de secuenciación lee a través del identificador de 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 de 6 pb, los restos son del cortador raro EcoRI (AACCT), el uso de dos bases selectivas produce una secuencia interna del fragmento de restricción de 12 pb que se puede usar para identificar únicamente el fragmento de restricción en la muestra.

[0056] En una forma de realización preferida basada en la tecnología de secuenciación de Solexa anterior, la amplificación de los fragmentos de restricción ligados a adaptadores se realiza con un cebador que contiene como mucho un nucleótido selectivo en su extremo 3', preferiblemente ningún nucleótido selectivo en su extremo 3', es decir, el cebador solo es complementario al adaptador (un cebador +0).

[0057] En formas de realización alternativas dirigidas a los métodos de secuenciación descritos aquí, los cebadores usados en la amplificación pueden contener secciones específicas (como alternativa a las secuencias de cebador o de unión de cebadores aquí descritas) que se usan en el paso de secuenciación posterior para unir los fragmentos de restricción cubiertos por un adaptador o amplicones a la superficie. Estos se representan generalmente como la región clave o la secuencia compatible con el cebador 5'.

[0058] En una forma de realización de la presente descripción, la muestra de ácido nucleico se digiere con al menos una enzima de restricción y se liga al menos un adaptador que comprende una secuencia de reconocimiento para una endonucleasa de restricción de tipo IIs. La digestión posterior del fragmento de restricción ligado a un adaptador con una endonucleasa de restricción de tipo IIs produce, ya que la distancia entre el sitio de reconocimiento y el 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, a los que puede ligarse un adaptador compatible con el sitio de restricción IIs. Típicamente, se desconoce la proyección del sitio restringido por IIs, de modo que se puede usar un conjunto de adaptadores que están degenerados en la proyección. Después de la amplificación (selectiva), los amplicones se pueden secuenciar. La secuencia de adaptador en esta forma de realización sigue generalmente: 5'-sitio de unión del cebador---secuencia identificadora de muestra---secuencia final cohesiva de tipo IIs degenerada-3'. El cebador de PCR asociado sigue generalmente: secuencia de cebador---secuencia identificadora de muestra---secuencia final cohesiva de tipo IIs degenerada---nucleótidos selectivos-3'. El cebador usado para iniciar la secuenciación por síntesis tiene entonces generalmente la estructura: 5'-sitio de unión del cebador-3'. Se puede preferir un paso de selección de tamaño después de digerir con la enzima IIs para eliminar los fragmentos más pequeños. Como en esta forma de realización los restos del sitio de restricción para este tipo de enzima tienen típicamente del orden de 2-4 pb, esto da lugar a la combinación con un identificador de muestra de 6 pb en la secuenciación de 15-17 pb de un fragmento de restricción.

[0059] En otro aspecto, la presente descripción se refiere a kits que comprenden uno o más cebadores, y/o uno o más adaptadores para usar en el método, aparte de componentes convencionales para los kits *per se*. Además, la presente invención encuentra aplicación en, entre otros, el uso del método para la identificación de marcadores moleculares, para la genotipificación, el análisis de segregantes en masa, el mapeo genético, el retrocruce asistido por marcadores, el mapeo de locus de rasgos cuantitativos y el mapeo de desequilibrio de enlace.

### Ejemplo

[0060] Se aisló ADN de 2 progenitores y 88 descendientes usando métodos convencionales. Los progenitores (2x) y los descendientes (= 4x) estaban en dúplex con diferentes índices para probar la reproducibilidad. Las etiquetas usadas para distinguir las muestras entre sí diferían al menos en 2 nucleótidos de cualquier otra etiqueta usada en los experimentos. La calidad se está evaluando a lo largo de todos los varios pasos usando geles de agarosa y PAA.

### Ejemplo 1

[0061] Para cada muestra de ADN se realiza un paso de restricción-ligamiento 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 de Solexa, más en particular el adaptador de EcoRI contiene la secuencia P5 (parte del cebador de la secuencia) y el adaptador de MseI contiene la secuencia P7 (secuencia del cebador de la PCR de puente). El adaptador de EcoRI contiene además la etiqueta de identificación de muestra. Se usan 96 adaptadores diferentes de EcoRI y un adaptador de MseI. Es posible usar un adaptador de EcoRI degenerado. La preparación del modelo incluye un paso de selección de tamaño por incubación de la mezcla durante 10 minutos a 80 grados Celsius después del paso de restricción (EcoRI+ MseI) pero antes del paso de ligamiento de adaptadores. Los fragmentos más pequeños de 130 nt se eliminan (en una muestra de maíz). La complejidad de la mezcla se reduce por una preamplificación selectiva usando cebadores +1 (es decir, con un nucleótido selectivo de forma aleatoria en el extremo 3', usando 96 cebadores EcoRI+1 y un cebador MseI+1 (o un cebador EcoRI+1 con etiqueta degenerada y un cebador MseI+1). La amplificación selectiva para reducir la complejidad de la mezcla al tamaño deseado se realiza usando cebadores EcoRI+2 (= lado P5) y MseI+3 (= lado P7) que necesitan el uso de 96 cebadores EcoRI+2 y un cebador MseI+3. La PCR de cola se realiza usando un cebador de EcoRI con la secuencia del cebador de PCR de puente P5 como la cola. Los productos se purifican usando columnas Sephadex™. Se determinan y se normalizan las concentraciones y se crean grupos. Los grupos se someten a la secuenciación masiva en paralelo basándose en la tecnología de Solexa que comprende la amplificación por PCR de puente y la secuenciación seguida de un análisis de datos para determinar los genotipos de los progenitores y los descendientes.

[0062] Un escenario alternativo no usa la PCR de cola, sino que emplea cebadores EcoRI+2 fosforilados. Debido a la falta de coincidencia con el adaptador original, la temperatura de apareamiento en el perfil de amplificación se baja 3 grados Celsius a 13 ciclos de *touch-down* de 62-53 grados Celsius seguidos de 23 ciclos a 53 grados

Celsius. Despues del ligamiento del adaptador con la secuencia de la PCR de puente P5, se realiza la PCR con los cebadores de la PCR de puente P5 y 'P7.

[0063] Un segundo escenario alternativo se basa en la preparación del modelo estándar como se ha resumido aquí antes, la (pre)amplificación selectiva para reducir la complejidad. La amplificación selectiva se realiza con cebadores que contienen los sitios de restricción de EcoRI y Msel reconstituidos. Esto permite la eliminación de las secuencias adaptadoras antes de la secuenciación, reduciendo así la cantidad de datos que hay que analizar. Purificación de los productos por columnas Sephadex™ para eliminar los restos de la Taq ADN polimerasa. Preparación del modelo donde se sustituyen secuencias adaptadoras (de sitio reconstituido) por adaptadores de Solexa usando un adaptador de EcoRI y la enzima EcoRI aumentados diez veces para compensar el mayor número de sitios de EcoRI en comparación con el ADN genómico. Los adaptadores de EcoRI de Solexa también contienen las etiquetas, por lo tanto, se necesitan 96 adaptadores de EcoRI de Solexa etiquetados. La cadena inferior del adaptador está bloqueada en el extremo 3' (en este caso por un amino 3') para bloquear la extensión por una polimerasa. La PCR se realiza con los cebadores de PCR de puente P5 y P7. Los productos se purifican por columnas Qiagen.

15 Ejemplo 2:

[0064] La detección basada en secuencias de fragmentos AFLP se realizó usando la tecnología de matriz de moléculas únicas clonal de Solexa (CSMATM), una plataforma de secuenciación por síntesis capaz de analizar hasta 40 millones de fragmentos individuales en una única ronda de secuenciación. La secuencia experimental implica la preparación del modelo AFLP, la amplificación selectiva (AFLP), la amplificación de puente de moléculas únicas y la secuenciación de millones de etiquetas de secuencias de un extremo de una enzima de restricción de los fragmentos AFLP. Las líneas parentales de maíz B73 y Mo17 y 87 líneas endogámicas recombinantes (RILs) se usaron y secuenciaron sobre 8,9 millones de terminales de fragmentos AFLP de EcoRI se secuenciaron para proporcionar una *prueba de principio* para la detección de AFLP basada en secuencias. Se seleccionaron las líneas parentales B73 y Mo17 y 87 RILs. Se prepararon modelos de AFLP usando la combinación de enzimas de restricción EcoRI/Msel. La amplificación selectiva se realizó usando cebadores de AFLP +2/+3. Se prepararon fragmentos modelo para la amplificación de puente CSMA de Solexa realizando una segunda restricción/ligamiento usando adaptadores de EcoRI que contienen secuencias de etiqueta de identificación (ID) de muestra de 5 pb únicas. Las líneas parentales y tres muestras RIL se incluyeron dos veces usando diferentes etiquetas de ID de muestra de 5 pb para medir la reproducibilidad intraexperimental.

20 Los marcadores de AFLP basados en secuencias fueron identificados por extracción de etiquetas de secuencias de 27 pb observadas con diferentes frecuencias en B73 y Mo17, segregando en los descendientes RIL.

25 Los datos de marcadores de AFLP basados en secuencias se compararon con puntuaciones de marcadores de AFLP obtenidas por obtención de huellas de AFLP convencional usando la detección basada en la longitud de las cuatro combinaciones de cebadores EcoRI/Msel +3/+3 correspondientes.

30

35 **Estadísticas de ronda de secuencia 5 células de flujo**

[0065]

# etiquetas de secuencia generadas	8.941.407
# etiquetas de secuencia con IDs de muestra conocidas	8.029.595
# etiquetas de secuencia diferentes con IDs de muestra conocidas	206.758
# MpB de datos de secuencia generados	241,4
rango de frecuencia de # total etiquetas de secuencia por muestra	55.374-112.527
# marcadores de AFLP de etiquetas de secuencia	125

rango de frecuencia de marcadores de AFLP de etiquetas de secuencia en la clasificación de 90-17.218 progenitores presente

#### Definición y clasificación de marcadores de AFLP de etiquetas de secuencia

[0066]

40 - tabular la representación por muestra de etiquetas de secuencia

- eliminar las etiquetas de secuencia con IDs de muestra desconocidas
  - normalizar la representación de las muestras en función de las etiquetas de secuencia totales por muestra
  - eliminar las etiquetas de secuencia con una diferencia de frecuencia > 2 veces en los duplicados parentales
  - promediar las frecuencias de etiqueta de los duplicados parentales
- 5 - definir marcador de AFLP de etiqueta de secuencia si la frecuencia P1/P2 excede el valor umbral
- clasificar la presencia/ausencia de los marcadores de etiqueta de secuencia en los descendientes RIL

#### Distribución de marcadores de AFLP AFLP +3/+3: secuencia/basados en gel

[0067]

		Base EcoRI+3			
		total	+A	+C	+G
# marcadores de AFLP de etiquetas de secuencia		125	34	37	37
# marcadores de AFLP basados en gel		82	29	18	17

#### Reproducibilidad de duplicados de marcadores de AFLP de etiqueta secuencia de 3 descendientes RIL

10 [0068]

# marcadores de AFLP de etiquetas de secuencia clasificados	125
# número de puntos de datos en comparación	375
# puntos de datos idénticos para los duplicados	372
% concordancia de duplicados intraexperimentales	99,2%

Detección en gel plano convencional:

	873	Mo17	1	2	3	4	5	6	7	8	9	10	11	12
<b>Marcador de AFLP</b>														
E361M50-175.9	-	+	+	-	-	-		+	-	+	-	-	-	+
E36/M50-280	+	-	+	-	-	+	-	+	+	-	+	-	-	-
36/M50-405.8	-	+	+	-	-	-	+	+	+	+	-	+	-	-
E36/M50-243.7	+	-	+	-	-	-	-	-	+	+	+	+	+	+
E36IM50-124.02	+	-	+	-	+	+	+	+	-	-	-	-	-	+
E36/M50-379	+	-	+	-	-	-	+	+	+	+	+	-	-	+
E36/M50-468.9	+	-	+	-	+	+	-	+	-	+	+	+	+	+

Detección basada en Solexa

	B73	Mo17	1	2	3	4	5	6	7	8	9	10	11	12
<b>Marcador de AFLP</b>														
CGGCGACGTACCGC	-	+	+	-	-	-	-	+	-	+	-	-	+	+
CTAGTAATTATTCC	+	-	+	-	-	+	-	+	+	-	+	-	-	-
CAGCGCCTTCTCCT	+	+	+	-	-		+	+	+	+	-	+	-	+
CAGAACTCTGACTT	+	-	+	-	-	-	-	-	+	+	+	+	+	+
CAAATCTGTTAGAT	+	-	+	-	+	+	+	+	-	+	-	-	+	+
CATGAAGGATTAT	+	-	+	-	-	+	+	+	+	+	+	+	-	+
CAAACAGACAAACCG	+	-	+	-	+	+	-	+	-	+	+	+	+	+

- [0069] La viabilidad de la detección de marcadores de AFLP basada en secuencias fue generada (usando la tecnología CSMA de Solexa, por la cual un mayor número de marcadores de AFLP se clasifica usando la detección basada en secuencias que en los geles planos convencionales, supuestamente debido a una resolución mejorada (tamaño de los fragmentos), y la secuenciación profunda, que captura también los fragmentos poco abundantes.
- 5 Las comparaciones de vectores de datos de marcadores revelan patrones de segregación similares entre la detección basada en secuencias y la detección en gel plano: la prueba de concordancia espera la secuenciación de marcadores de AFLP basados en geles.

## Listado de secuencias

[0070]

- 10 < 110> Keygene NV  
 <120> Detección de alto rendimiento de marcadores moleculares basados en fragmentos de restricción  
 <130> p28122EP01  
 <140> pct/n12007/000094  
 <141> 2007-04-04
- 15 <150> US 60/788, 706  
 <151> 2006-04-04  
 <150> US 60/872, 999  
 <151> 2007-01-12
- 20 <160> 7  
 <170> versión de PatentIn 3.3  
 <210> 1  
 <211> 14  
 <212> ADN  
 <213> Zea mays
- 25 <400> 1  
 cggcgacgta ccgc 14  
 <210> 2  
 <211> 14  
 <212> ADN  
 30 <213> Zea mays  
 <400> 2  
 ctagtaattattcc 14  
 <210> 3  
 <211> 14  
 35 <212> ADN  
 <213> Zea mays  
 <400> 3  
 <cagcgcccttc tcct 14  
 <210> 4  
 <211> 14  
 <212> ADN  
 40 <213> Zea mays  
 <400> 4

cagaactctg act 14  
<210> 5  
<211> 14  
<212> ADN  
5 <213> Zea mays  
  
<400> 5  
caaatctgtt agat 14  
  
<210> 6  
<211> 14  
10 <212> ADN  
<213> Zea mays  
  
<400> 6  
catgaaggat ttat 14  
  
<210> 7  
15 <211> 14  
<212> ADN  
<213> Zea mays  
  
<400> 7  
caaacagaca accg 14

## REIVINDICACIONES

1. Método para detectar polimorfismos en fragmentos de restricción de dos o más muestras, que comprende los pasos de:
  - (a) proporcionar dos o más muestras de ácido nucleico;
  - (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
    - una secuencia compatible con el cebador 5', seguida de una sección identificadora específica de muestra, y
    - un extremo 3' que se puede ligar al extremo romo o saliente 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 a adaptadores;
  - (e) opcionalmente amplificar el conjunto de fragmentos de restricción ligados a adaptadores, con uno o más cebadores que son al menos complementarios a:
    - la sección identificadora específica de muestra,
    - una sección que es complementaria a los restos de la secuencia de reconocimiento de la endonucleasa de restricción,
 para proporcionar fragmentos de restricción ligados a adaptadores amplificados (amplicones);
  - (f) determinar la secuencia de al menos
    - la sección identificadora específica de muestra;
    - los restos de la secuencia de reconocimiento de la endonucleasa de restricción; y,
    - parte de la secuencia del fragmento de restricción situada adyacente a los restos de la secuencia de reconocimiento de la endonucleasa de restricción; y
  - (g) detectar polimorfismos en la secuencia de los fragmentos de restricción ligados a adaptadores en las dos o más muestras obtenidas en el paso (f).
2. Método según la reivindicación 1, donde los fragmentos de restricción son marcadores moleculares.
3. Método según la reivindicación 2, donde los marcadores moleculares son marcadores de AFLP.
4. Método según la reivindicación 1, donde dos o más muestras se comparan para la presencia o la ausencia de fragmentos de restricción y/o marcadores moleculares.
5. Método según la reivindicación 1, donde dos o más muestras se combinan en un grupo después del paso de ligar los adaptadores.
6. Método según la reivindicación 5, donde para cada muestra en el grupo se usa un identificador específico de muestra que difiere de los otros identificadores específicos de muestra en el grupo.
7. Método según la reivindicación 1, donde los cebadores contienen uno o más nucleótidos selectivos en el extremo 3'.
8. Método según la reivindicación 1, donde la endonucleasa de restricción es una endonucleasa de restricción de tipo II.
9. Método según la reivindicación 1, donde la endonucleasa de restricción es una endonucleasa de restricción de tipo IIs.
10. Método según la reivindicación 1, donde se usan dos o más endonucleasas de restricción.
11. Método según la reivindicación 1, donde se realiza la secuenciación por medio de secuenciación de alto rendimiento.

12. Método según la reivindicación 11, donde la secuenciación de alto rendimiento se realiza en un soporte sólido.

13. Método según la reivindicación 11, donde la secuenciación de alto rendimiento se basa en la secuenciación por síntesis.

14. Método según la reivindicación 11, donde la secuenciación de alto rendimiento comprende los pasos de:

- 5      - aparear los amplicones o fragmentos de restricción ligados a adaptadores a esferas, donde cada esfera se aparea con un único fragmento de restricción ligado a un adaptador o amplicón;
- 10     - emulsionar las esferas en microrreactores de agua en aceite, donde cada microrreactor de agua en aceite comprende una única esfera;
- 15     - realizar una PCR en emulsión para amplificar fragmentos de restricción ligados a adaptadores o amplicones en la superficie de las esferas;
- 20     - opcionalmente, seleccionar/enriquecer las esferas que contienen amplicones amplificados;
- 25     - cargar las esferas en pocillos, donde cada pocillo comprende una única esfera; y
- 30     - determinar la secuencia de nucleótidos de los fragmentos de restricción ligados a adaptadores amplificados o amplicones amplificados usando la generación de una señal de pirofosfato.

15     15. Método según la reivindicación 11, donde la secuenciación de alto rendimiento comprende los pasos de:

- 20     - aparear los fragmentos de restricción ligados a adaptadores o amplicones a una superficie que contiene primeros y segundos cebadores o primeras y segundas secuencias de unión de cebadores respectivamente,
- 25     - realizar una amplificación de puente para proporcionar agrupaciones de fragmentos de restricción ligados a adaptadores amplificados o amplicones amplificados,
- 30     - determinar la secuencia de nucleótidos de los fragmentos de restricción ligados a adaptadores amplificados o amplicones amplificados usando nucleótidos de terminación reversibles marcados.

16. Método según la reivindicación 1, donde el identificador tiene de 4 a 16 pb, preferiblemente de 4 a 10, más preferiblemente de 4 a 8, más preferiblemente de 4 a 6 pb.

17. Método según la reivindicación 1, donde el identificador no contiene 2 o más bases consecutivas idénticas.

25     18. Método según la reivindicación 1, donde, para dos o más muestras, los identificadores correspondientes contienen al menos dos nucleótidos diferentes.

19. Uso del método de cualquiera de las reivindicaciones 1-18 para la identificación de marcadores moleculares, para la genotipificación, el análisis de segregantes en masa, el mapeo genético, el retrocruce asistido por marcadores, el mapeo de locus de rasgos cuantitativos, el mapeo de desequilibrio de enlace.

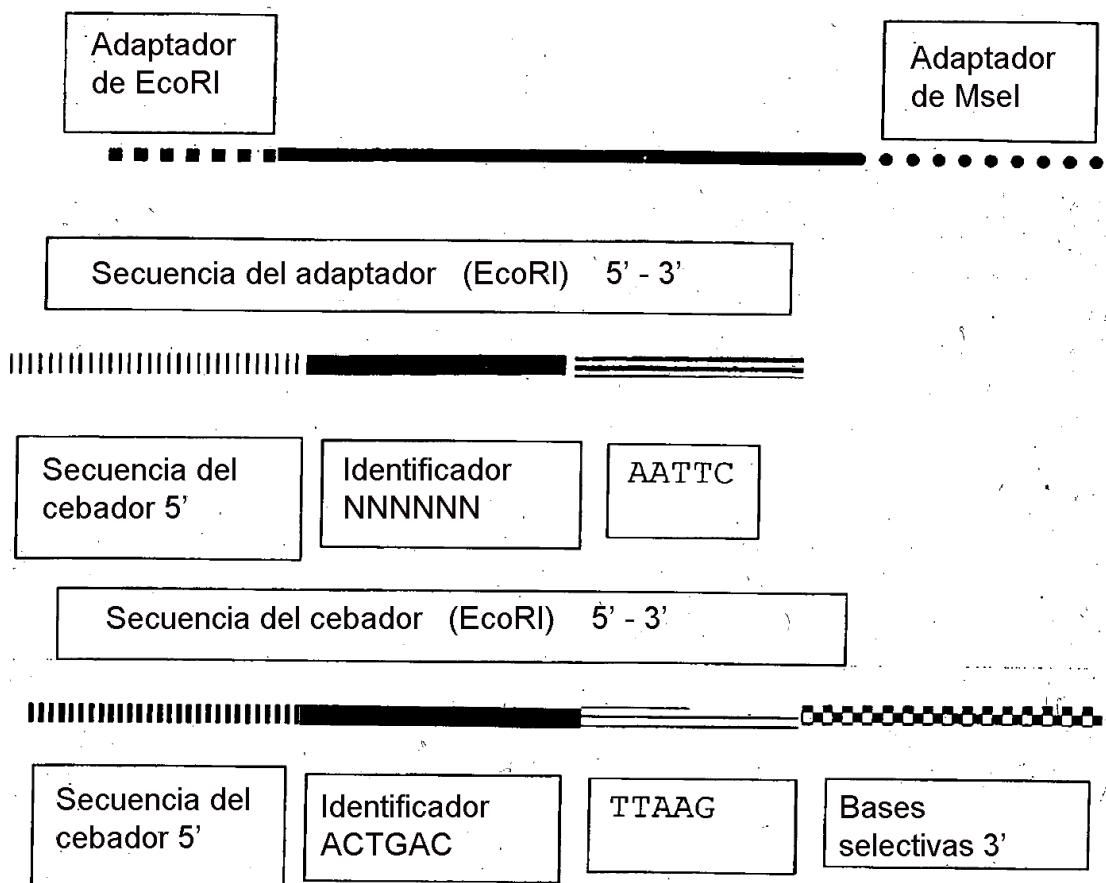
30     20. Kit para usar en el método según cualquiera de las reivindicaciones 1-18 que comprende:

- 35     - uno o más adaptadores que comprenden una secuencia compatible con el cebador 5', seguida de una sección identificadora específica de muestra, y un extremo 3' de que se puede ligar al extremo romo o saliente de un fragmento de restricción para proporcionar fragmentos de restricción ligados a adaptadores, y
- 40     - uno o más cebadores capaces de hibridar con los fragmentos de restricción ligados a adaptadores, que comprenden una parte que es complementaria al uno o más adaptadores y que comprenden en el extremo 3' uno o más nucleótidos selectivos para proporcionar un subconjunto de fragmentos de restricción ligados a adaptadores amplificados por amplificación selectiva.

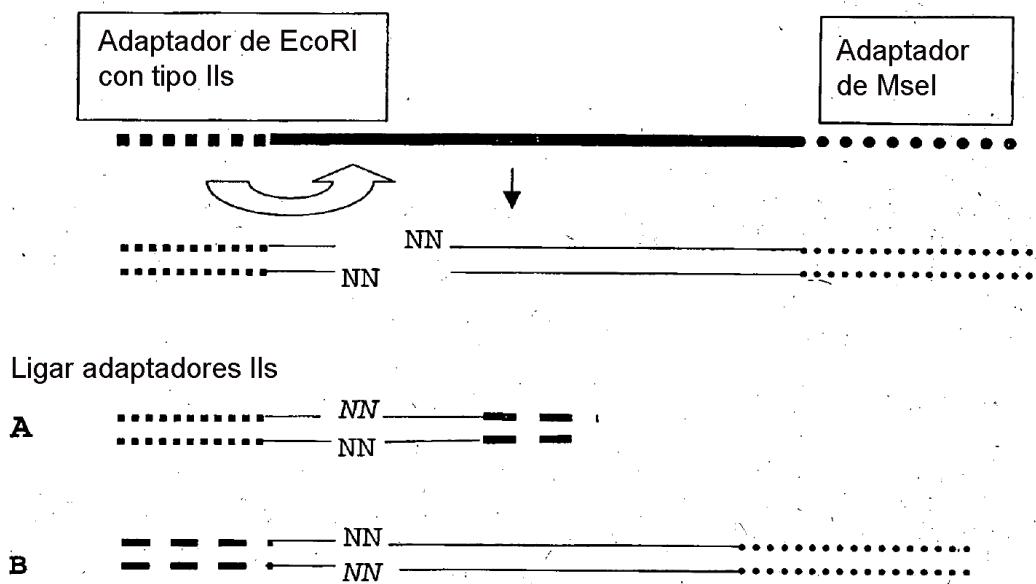
21. Kit para usar en el método según cualquiera de las reivindicaciones 1-18 que comprende:

- 45     - uno o más adaptadores que comprenden una secuencia compatible con el cebador 5', seguida de una sección identificadora específica de muestra, y un extremo 3' que se puede ligar al extremo romo o saliente de un fragmento de restricción para proporcionar fragmentos de restricción ligados a adaptadores, y
- 50     - uno o más cebadores capaces de hibridar con los fragmentos de restricción ligados a adaptadores, que comprenden una parte que es complementaria al uno o más adaptadores y que comprenden en el extremo 3' uno o más nucleótidos selectivos para proporcionar un subconjunto de fragmentos de restricción ligados a adaptadores amplificados por amplificación selectiva, y
- 55     - uno o más cebadores para la amplificación no selectiva que son complementarios al uno o más adaptadores solo.

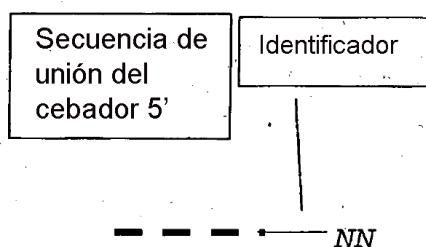
**FIG 1**



**FIG 2**



**Estructura del adaptador**



**Estructura del cebador**

