

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 556 580**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **09.07.2012 E 12738652 (2)**

97 Fecha y número de publicación de la concesión europea: **16.09.2015 EP 2729580**

54 Título: **Genotipado a base de secuencias en base a ensayos de ligación a oligonucleótidos**

30 Prioridad:

**08.07.2011 US 201161505787 P**

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

**19.01.2016**

73 Titular/es:

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

72 Inventor/es:

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

74 Agente/Representante:

**PONTI SALES, Adelaida**

**ES 2 556 580 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

Genotipado a base de secuencias en base a ensayos de ligación a oligonucleótidos

5 Campo de la invención

10 [0001] La presente invención se refiere al campo de la biología molecular y la biotecnología. En particular, la presente invención se refiere al campo de la detección de ácido nucleico, más en particular al diseño y composición de (colecciones) de sondas que se pueden utilizar para la detección de alto rendimiento de ácidos nucleicos. La presente invención también se refiere a procedimientos para la detección de ácidos nucleicos usando las sondas y composiciones. La presente invención proporciona además sondas que son capaces de hibridarse a una secuencia diana de interés, cebadores para la amplificación de sondas ligadas, uso de estas sondas y cebadores en la identificación y/o detección de secuencias de nucleótidos que pueden estar relacionadas con una amplia variedad de características genéticas y genes. La presente invención proporciona además kits de cebadores y/o sondas adecuados para su uso en el procedimiento según la presente invención. La presente invención es aplicable en el campo de la detección de alto rendimiento de las secuencias de nucleótidos diana en muestras, ya sean de origen artificial, vegetal, animal o humano o combinaciones de los mismos. La presente invención encuentra una aplicación particular en el campo de genotipado de alto rendimiento.

20 Antecedentes de la invención

25 [0002] Con el incremento exponencial cerca de la disponibilidad de información genética debido al desarrollo de tecnologías avanzadas para la obtención de información sobre rasgos, alelos y secuenciación, existe una creciente necesidad de ensayos eficientes, fiables y escalables para ensayar muestras y en muchos casos múltiples muestras de una manera rápida, a menudo en paralelo. En particular, los polimorfismos de un solo nucleótido (SNP) contienen información valiosa sobre la composición genética de los organismos y la detección de los mismos es un campo que ha atraído un gran interés y actividad innovadora.

30 [0003] Uno de los procedimientos principales usados para el análisis de los ácidos nucleicos de una secuencia conocida se basa en la hibridación de dos sondas a una secuencia diana y, cuando las sondas se hibridan de forma adyacente a la secuencia diana, la ligación de las sondas. La detección de un evento de ligación con éxito es entonces indicativo de la presencia de la secuencia diana en la muestra. En Ensayo de ligación de oligonucleótidos (OLA) es una tecnología que se ha encontrado como adecuada para la detección de tales polimorfismos de nucleótido único y se ha descrito a lo largo de los años en muchas variaciones en un conjunto de solicitudes de patentes y artículos científicos.

35 [0004] El principio OLA (Ensayo de ligación de oligonucleótidos) se ha descrito, entre otros, en el documento US 4.988.617 (Landegren et al.). Esta publicación da a conocer un procedimiento para determinar la secuencia de ácido nucleico en una región de una secuencia de ácido nucleico conocida que tiene una mutación o polimorfismo conocido posible. Para detectar la mutación, se seleccionan oligonucleótidos para hibridarse a segmentos inmediatamente adyacentes de la secuencia a determinar. Una de las sondas de oligonucleótidos seleccionadas tiene una región terminal en la que uno de los nucleótidos de la región terminal es complementario al nucleótido normal o mutado en la posición correspondiente en la secuencia de ácido nucleico conocida. Se proporciona una ligasa que conecta covalentemente las dos sondas cuando están correctamente emparejadas de bases y están situadas inmediatamente adyacentes entre sí. La presencia, ausencia o cantidad de las sondas unidas es una indicación de la presencia de la secuencia y/o mutación conocida. Otras variantes de técnicas basadas en OLA se han descrito, entre otros, en Nilsson et al. Human mutation, 2002, 19, 410-415; Science 1994, 265: 2085-2088; US 5.876.924; WO 98/04745; WO 98/04746; US 6.221.603; US 5.521.065; US 5.962.223; EP 185494B1; US 6.027.889; US 4.988.617; EP 246864B1; US 6.156.178; EP 745140 B1; EP 964704 B1; WO 03/054511; US 40 2003/0119004; US 2003/190646; EP 1313880; US 2003/0032016; EP 912.761; EP 956359; US 2003/108913; EP 45 1255871; EP 1194770; EP 1252334; WO 96/15271; WO 97/45559; US 2003/0119004A1; US 5.470.705.

50 [0005] Otros avances en las técnicas de OLA se han descrito por KeyGene, Wageningen, Países Bajos. En los documentos WO 2004/111271, WO 2005/021794, WO 2005/118847 y WO 03/052142, se han descrito varios procedimientos y diseños de sondas que mejoraban la fiabilidad de los ensayos de ligación de oligonucleótidos. Estas aplicaciones describen además la mejora significativa en los niveles multiplex que se pueden conseguir. También "SNPWave: a flexible multiplexed SNP genotyping technology", van Eijk MJ, Broekhof JL, van der Poel HJ, Högers R.C., Geerlings H, Buntjer JB, van Oeveren AJ, Vos P Nucleic Acids Res. 2004; 32 (4): e47 describe las mejoras realizadas en este campo.

55 [0006] Con el inicio de las tecnologías de Next Generation Sequencing (NGS), tal como se describe en Janitz Ed. Next Generation Genome sequencing, Wiley VCH, 2008 y disponible en el mercado en plataformas proporcionadas por Roche (GS FLX y sistemas relacionados) y Illumina (Genome Analyzer y sistemas relacionados), surgió la necesidad de adaptar el ensayo de OLA a la secuenciación como plataforma de detección. Las mejoras en este campo se han descrito, entre otros, en el documento WO 2007100243 de Keygene NV. En WO2007100243, se ha descrito la aplicación de la tecnología de secuenciación de la siguiente generación en los resultados de ensayos

de ligación de oligonucleótidos.

[0007] Sigue existiendo la necesidad de nuevas mejoras en este campo, no sólo desde el punto de vista de la fiabilidad y precisión, sino también de los motores económicos para reducir aún más los costes mediante el aumento de escala.

[0008] Existe una continua necesidad de sondas de oligonucleótidos que combinan las ventajas y evitan las desventajas específicas de los diversos tipos de sondas de ligación y procedimientos de detección conocidos en la técnica. También existe la necesidad de una nueva mejora de la tecnología proporcionando sondas que tienen ventajas adicionales. Uno de los objetivos de la presente invención es proporcionar tales sondas y procedimientos. Otro objetivo de la presente invención es evitar las desventajas de las sondas comúnmente conocidas mencionadas anteriormente en el presente documento. Un objetivo adicional de la invención es proporcionar sondas que sean adecuados para los procedimientos de detección de alto rendimiento. También un objetivo de la presente invención es proporcionar un procedimiento eficiente, fiable y/o de alto rendimiento para la detección de secuencias de nucleótidos diana mediante la realización de ensayos de ligación de oligonucleótidos.

[0009] Los presentes inventores han establecido para eliminar o al menos disminuir los problemas existentes en la técnica, mientras que al mismo tiempo se trata de mantener los muchos aspectos ventajosos de la misma, y para mejorar adicionalmente la tecnología. Otros problemas en la técnica y soluciones aportadas a la misma por la presente invención se pondrán de manifiesto a lo largo de la descripción, las figuras y las diversas realizaciones descritas en el presente documento.

#### Breve descripción de los dibujos

[0010] La presente invención se ilustra mediante las siguientes figuras:

**Figura 1:** En la Figura 1, los diferentes tipos de sondas (Figura 1A, Figura 1B, Figura 1C) se ilustran esquemáticamente vis-a-vis una secuencia de nucleótidos diana (T) de interés. Se han representado diversos componentes de las sondas, utilizando representaciones idénticas a lo largo de la figura.

La **Figura 1A** ilustra un ensayo de ligación de oligonucleótidos general basado en un tipo de sonda lineal dirigido a una secuencia diana (T), en donde una primera sonda (P1) comprende una primera sección específica de diana (TS1) y una primera sección etiqueta (TAG1) que comprende un primer identificador (ID1) y una primera secuencia de unión a cebador (PBS1), capaz de hibridar con un primer cebador (PR1). Una segunda sonda (P2) comprende una segunda sección específica de diana (TS2) y una segunda sección etiqueta (TAG2) que comprende un segundo identificador opcional (ID2) y una segunda secuencia de unión a cebador (PBS2), capaz de hibridar con un segundo cebador (PR2). En realizaciones para la detección específica de alelo, TS2 puede contener, preferiblemente en su extremo 3', un nucleótido específico de alelo, preferiblemente junto con un identificador diferente (ID2) en la sección de etiqueta. En otras realizaciones para la detección específica de alelo, TS1 puede contener, preferiblemente en su extremo 5', un nucleótido específico de alelo, preferiblemente junto con un identificador diferente (ID1) en la sección de etiqueta. La combinación de locus-alelo puede entonces determinarse (genotipado) mediante la detección de la presencia o ausencia de ID1 y/o ID2. De forma similar, todas las variantes alélicas de un polimorfismo se pueden genotipar (por ejemplo, 2 alelos de un polimorfismo bialélico utilizando dos sondas con cada sección diana específica de alelo o para 4 alelos, utilizando 4 secciones diana específicas de alelo). Cuando la detección se basa en la secuenciación, la detección de la presencia, ausencia o cantidad también puede basarse en la información de secuencia de las sondas ligadas mediante una combinación de identificadores y la información de secuencia de la sección específica de diana. Así, el alelo se puede determinar a través de la secuenciación de (parte de la propia sección diana, mientras que el locus se puede determinar por secuenciación del identificador y viceversa).

La **Figura 1B** ilustra un ensayo de ligación de oligonucleótido de acuerdo con la invención basada en un tipo de sonda lineal dirigida a una secuencia diana (T). Las sondas se han equipado con una secuencia de reconocimiento para una endonucleasa de restricción (RE1, RE2).

La **Figura 1C** ilustra un ensayo de ligación de oligonucleótido de acuerdo con la invención basada en un tipo de sonda lineal dirigida a una secuencia diana (T). Las sondas se han equipado con una secuencia de reconocimiento para una endonucleasa de restricción (RE1, RE2) e identificadores (ID1, ID2).

La **Figura 2** ilustra, en base a la configuración de la sonda en la Figura 1B, que las dos sondas se hibridan con la secuencia diana y se ligan cuando la hibridación se realiza correctamente. Dos rutas ahora abiertas, una en la que las sondas ligadas (LP) están restringidas por una endonucleasa de restricción, cuando sea necesario ayudada por el uso de adaptadores de horquilla u otros oligonucleótidos que proporcionan localmente una cadena ds que puede restringirse esencialmente como se describe en el presente documento en otro punto y se ilustra en la figura 5C. El resultado es una sonda ligada restringida (RLP). La otra ruta amplifica las sondas ligadas utilizando uno o más cebadores (PR1, PR2) para producir amplicones (A) que pueden restringirse para producir amplicones restringidos (RA). A RLP y/o RA, se pueden ligar adaptadores que pueden contener identificadores (específicos de muestra). Tanto RLP como RA se pueden someter a secuenciación, lo que da lugar a la identificación de la presencia, ausencia y/o cantidad de la secuencia diana en la muestra. El genotipado (codominante) de la muestra se puede basar entonces en la identificación de la secuencia diana en la muestra a través de la información de secuencia a partir de (parte de) la sección o secciones diana y/o identificadores proporcionados en las secciones de etiqueta.

La **Figura 2A** ilustra una realización en la que después de la restricción de los amplicones o sondas ligadas, se liga

un adaptador (AD1, AD2) que puede contener un identificador (AD1 ID1, AD2 ID2), cuyo identificador puede servir por ejemplo para identificar un origen de la muestra.

La **Figura 3** ilustra una representación esquemática de una serie de elementos que están presentes en varias realizaciones de la invención. Para facilitar la consulta se han relacionado con las indicaciones usadas anteriormente. Por lo tanto, TS representa una sección específica de diana, ID indica un identificador, RE una secuencia de reconocimiento para una endonucleasa de restricción. Un sitio de unión del cebador se indica como PBS. Un adaptador que se liga a un fragmento restringido se representa como AD. Los cebadores utilizados en la etapa de amplificación (ya sea para la amplificación como parte de la preparación de la biblioteca o como parte de la etapa de secuenciación) se indican como PR. Los amplicones restringidos/sondas ligadas restringidas se indican como RA/RLA. Para indicar que ciertos elementos se utilizan en la secuenciación del procedimiento de la invención, esto puede indicarse por el prefijo SEQ. Por lo tanto, un adaptador utilizado en la etapa de secuenciación se puede indicar como "adaptador de secuenciación" o SEQ AD. Cuando dos o más elementos están presentes al mismo tiempo, esto se indica con un sufijo numérico. Por lo tanto, ID1 es el primer identificador, ID2 es el segundo y así sucesivamente.

La **Figura 3A** ilustra una sonda ligada típica de acuerdo con la invención que comprende sitios de unión a cebadores (PBS1, PBS2), sitios de reconocimiento para endonucleasas de restricción (RE1, RE2), identificadores (ID1, ID2), secciones específicas de diana (TS1, TS2) y cebadores para amplificación (PR1, PR2)

La **Figura 3B** ilustra una serie de las diversas posibilidades donde identificadores pueden situarse en las sondas, amplicones y fragmentos ligados a adaptadores de la invención. Los identificadores pueden estar presentes independientemente, o combinados. La secuencia específica diana también puede servir como un identificador en el procedimiento para la genotipificación basado en la secuencia, el locus (L) y el alelo (AI) también se indican aquí como posibles identificadores. Por ejemplo, el alelo puede representarse por un identificador (1) y el locus por la secuencia de la sección diana (1). Tanto el alelo como el locus pueden representarse por un identificador (2). O el locus está representado por un identificador y el alelo por una sección específica de diana (3). La muestra puede representarse por un identificador, independientemente de la representación del alelo y/o locus (5). Las variaciones 6, 7, 8 y 9 muestran varias combinaciones de identificadores que pueden servir para fines distintos o idénticos. Otras variaciones son posibles y son equivalentes a las posibilidades que se muestran actualmente.

La **Figura 4** ilustra varias realizaciones para la secuenciación de los fragmentos de la invención. Como las sondas ligadas (amplificadas) se han restringido utilizando las endonucleasas de restricción, los amplicones restringidos/sondas ligadas restringidas se indican como RA/RLP (y por lo tanto contienen la información del locus y el alelo, L y AI, respectivamente). Después de la restricción, se han añadido adaptadores (AD1, AD2) que puede ser adaptadores de secuenciación (SEQ AD1, SEQ AD2) que se pueden utilizar en la etapa de secuenciación (se conocen ejemplos en la técnica como cebadores P5/P7 (Illumina)). Los cebadores utilizados en la etapa de secuenciación (cebadores de secuenciación) se indican como SEQ PR1, SEQ PR2 y son complementarios a las secciones (secuencias de unión a cebadores) en los adaptadores de secuenciación.

La **Figura 4A** ilustra la secuenciación de lectura única (I) desde un extremo o desde el extremo opuesto y la secuenciación de extremos emparejado (I y II combinadas). El fragmento de secuenciación, que comprende RA/RLP y que comprende además uno o más identificadores (ID, L, AI) está secuenciado en una dirección (flecha rayada), usando un cebador de secuenciación que conduce a (dependiendo de la longitud de lectura generada por la plataforma de secuenciación) una lectura que identifica la presencia/ausencia o cantidad de la secuencia diana proporcionando la secuencia del identificador y, opcionalmente, (parte de) RA/RLP. La lectura producida mediante I proporciona principalmente información de la secuencia del adaptador (incluyendo el ID), II principalmente de la secuencia diana y III de ambos, siempre que la lectura sea lo suficientemente larga, dependiendo de la plataforma.

La **Figura 4B** ilustra la secuenciación de doble etiqueta unidireccional. La sonda ligada restringida o el amplicón restringido (RA/RLP) se han ligado a dos adaptadores de secuenciación y se han secuenciado en una dirección pero con dos cebadores (SEQ PR1, PR2 SEQ). Los dos cebadores dan lugar a dos lecturas diferentes, una lectura corta y una lectura larga. La secuenciación proporciona al menos información de la secuencia de los identificadores (tanto para la lectura corta como para la lectura larga) y la secuencia de (parte de) RLP/AR, posiblemente incluyendo L y AI.

La **Figura 4C** ilustra una realización en la que se realiza una etapa de reagrupamiento. Los fragmentos de secuenciación se secuencian en una dirección que da lugar a la primera lectura (Long1). El fragmento de secuenciación se reagrupado mediante hibridación del otro adaptador de secuenciación al portador en el que se realiza la secuenciación. El fragmento es secuenciado desde el otro extremo, lo que da lugar a otra lectura larga (Long 2).

La **Figura 4D** ilustra una realización que es una combinación de la realización descrita en las figuras 4B y 4C. Por lo tanto, primero un procedimiento de secuenciación de doble etiqueta unidireccional es seguido por el reagrupamiento y la secuenciación del fragmento de la otra dirección, dando lugar a la combinación de dos lecturas largas y una lectura corta.

La **Figura 5** ilustra el uso de un adaptador en forma de Y.

La **Figura 5A** muestra un adaptador en forma de Y en el que el adaptador en los brazos de la Y puede contener diferentes elementos, tales como identificadores, adaptadores de secuenciación, sitios de unión a cebadores, etc. Uno de los brazos de la Y es por lo tanto diferente del otro brazo de la Y. La parte inferior de la forma Y tiene doble cadena (es decir, contiene cadenas complementarias) y ambas cadenas son capaces de ligarse al fragmento de restricción. Esta realización se puede utilizar cuando sólo un sitio de reconocimiento ha sido introducido en las dos secciones etiqueta de las dos sondas, se necesita sólo un adaptador para ligar los adaptadores a ambos lados de un fragmento de restricción, tal como una RLA o un RA (ver I). La autoligación del adaptador en forma de Y se muestra

en la Figura 5A II.

La **Figura 5B** muestra cómo evitar la autoligación de los adaptadores en forma de Y representados en la Figura 5A (II). El extremo de hibridación del adaptador se puede diseñar para ligar una sola cadena. La otra cadena tiene un hueco, evitando la ligación. Esto se ilustra por un extremo escalonado en el fragmento de CTA, combinado con un saliente en el Y de sólo el GA. Esto evita la autoligación.

La Figura 5C ilustra algunas de las realizaciones de la invención en las que la cadena única de la sonda ligada se restringe utilizando un oligonucleótido adicional o mediante la ligación de una sonda de horquilla. Ambas realizaciones proporcionan cadenas que se pueden cortar por una endonucleasa de restricción.

La **Figura 6** ilustra el procedimiento de secuenciación de pares acoplados ("mate pair"), que comprende una etapa de circularización y una etapa en la que los dos extremos del fragmento están unidos y, después de la fragmentación y la ligación de adaptadores, son posteriormente secuenciados juntos.

#### Características de la invención

**[0011]** Los presentes inventores han sido capaces de combinar nuevas tecnologías de secuenciación de alto rendimiento con la versatilidad de ensayos basados en la ligación de oligonucleótidos. En particular, la invención se refiere a un procedimiento para la detección de alto rendimiento de secuencias de nucleótidos diana en base a ensayos de ligación de oligonucleótidos, en los que las sondas utilizadas en los ensayos de ligación se modifican de tal manera que se puede utilizar un procedimiento de secuenciación de alto rendimiento para revelar inequívocamente la presencia/ausencia de la cantidad de dichas una o más secuencias de nucleótidos diana. Los inventores han descubierto que se pueden hacer mejoras mediante la adaptación de los procedimientos existentes para centrarse en la detección de las partes del producto de ligación que son relevantes para una detección adecuada de las secuencias diana en una pluralidad de muestras. El procedimiento se basa en el uso de (combinaciones de) identificadores basados en la secuencia en combinación con una etapa que reduce la cantidad de datos de secuencia no relevantes mediante la eliminación (recorte) de una parte de las sondas (ligadas) antes de la secuenciación. El uso de adaptadores (que pueden contener identificadores) que están conectados a las sondas ligadas restringidas o sondas ligadas amplificadas permite el uso de conjuntos genéricos de adaptadores e identificadores y cebadores en combinación con sondas deliberadamente diseñadas para secuencias diana. La estrategia modular, que separa los elementos de la sonda que están conectados a la propia diana y los elementos que están conectados a muestras multiplex y la detección basada en secuencias permite una flexibilidad ventajosa en combinación con una fiabilidad probada. La presente invención conduce a procedimientos ventajosos para el genotipado de alto rendimiento de grandes cantidades de muestras para un gran número de genotipos con alta precisión y bajo coste por punto de dato. La presente invención también permite adaptar la tecnología OLA demostrada a las nuevas plataformas de detección basadas en la secuenciación.

#### Descripción detallada de la invención

**[0012]** La presente invención en su forma más amplia se refiere a un procedimiento para la detección de una secuencia de nucleótidos diana en una muestra basado en un ensayo de ligación de oligonucleótidos en el que se usan sondas que se proporcionan con o contiene (una combinación de) identificadores basados en la secuencia que pueden identificar la muestra y/o la secuencia diana (es decir, combinación de locus y/o alelo) en el que después de la etapa de ligación, las sondas ligadas, o después de la amplificación, las sondas ligadas amplificadas, se restringen usando enzimas de restricción para cortar parte de las sondas, en caso necesario, se ligan al identificador que contiene adaptadores y continúan con aquellas partes (identificadores y/o secuencia diana) que contienen la información relevante en la etapa de secuenciación para la identificación adecuada de la muestra y/o genotipo basado en la presencia y/o ausencia de identificador o identificadores.

**[0013]** Por lo tanto, con más detalle, la invención se refiere a un procedimiento para la determinación de una secuencia de nucleótidos diana en una muestra que comprende las etapas de

(a) proporcionar para cada secuencia de nucleótidos diana (T) una primera sonda (P1) y una segunda sonda (P2), en el que la primera sonda comprende una primera sección específica de diana (TS1) y una primera sección etiqueta (TAG1) que es no complementaria a la secuencia de nucleótidos diana y que comprende opcionalmente una primera secuencia de unión a cebador (PBS1), en el que la primera sección etiqueta comprende una primera secuencia de reconocimiento (RE1) para una primera endonucleasa de restricción; en el que la segunda sonda comprende una segunda sección específica de diana (TS2) y una segunda sección etiqueta (TAG2) que es no complementaria a la secuencia de nucleótidos diana y que comprende una segunda secuencia de unión a cebador (PBS2) opcional, en el que la segunda sección etiqueta comprende una segunda secuencia de reconocimiento (RE2) para una segunda endonucleasa de restricción;

(b) permitir que la primera y segunda sección específica de diana de la respectiva primera y segunda sonda se hibriden con la secuencia diana;

(c) ligar la primera y segunda sonda cuando las secciones específicas de diana respectivas de las sondas se hibridan con secciones esencialmente adyacentes en la secuencia diana para proporcionar sondas ligadas (LP);

(d) opcionalmente amplificar las sondas ligadas con un primer cebador opcional y/o un segundo cebador opcional para proporcionar amplicones (A);

(e) restringir las sondas ligadas o amplicones con la primera y/o segunda endonucleasa de restricción para proporcionar sondas ligadas restringidas (RLP) o amplicones restringidos (RA), ligar un primer y/o un segundo

adaptador que contiene un identificador de base adaptador (AD ID1, AD ID2) a las sondas ligadas restringidas (RLP) o amplicones restringidos (RA)

(f) someter las sondas ligadas restringidas (RLP) ligadas a adaptador o amplicones restringidos (RA) ligados a adaptador a tecnología de secuenciación de alto rendimiento para determinar al menos parte de la secuencia de nucleótidos de las sondas ligadas restringidas o amplicones restringidos

(g) identificar la presencia, ausencia o cantidad de la secuencia de nucleótidos diana en la muestra.

**[0014]** El procedimiento comienza con la disposición de una o más muestras (que pueden estar combinadas o agrupadas) que pueden contener (o se sospecha que contienen) la secuencia de nucleótidos diana (secuencia de interés). A esta muestra, se añade el conjunto (de una primera y una segunda sonda) de sondas (para cada secuencia diana, se pueden proporcionar diferentes conjuntos de sondas) y se deja que las secciones específicas de diana de las sondas se hibriden a la secuencia diana bajo condiciones adecuadas. Después de la hibridación, cualquier sonda hibridada esencialmente adyacente en la secuencia diana se ligan para dar lugar a sondas ligadas. Las sondas ligadas se pueden amplificar o, alternativamente, someter directamente a secuenciación usando procedimientos de secuenciación de alto rendimiento. Con la etapa de secuenciación, se determina la presencia de la secuencia diana (específico de alelo) en la muestra y se pueden determinar los genotipos.

**[0015]** Un aspecto de la presente invención se refiere al diseño ventajoso de las sondas utilizadas en la presente invención. Estas sondas se discutirán en más detalle en este documento a continuación. Otro aspecto ventajoso de la invención reside en la conexión entre las tecnologías de secuenciación de alto rendimiento del estado de la técnica como una plataforma de detección para los ensayos de ligación de oligonucleótidos y el poder discriminatorio de los ensayos basados en OLA. Los presentes inventores han observado que, aparte de innovaciones en el diseño de la sonda, también los procedimientos para realizar ensayos de OLA en combinación con secuenciación de alto rendimiento requieren modificaciones considerables para sondas y protocolos.

**[0016]** En la etapa (a) del procedimiento para cada secuencia de nucleótidos diana (T) en la muestra (S) se proporciona un conjunto de sondas. El conjunto de sondas puede comprender una primera sonda (P1) y una segunda sonda (P2).

**[0017]** La primera sonda comprende una primera sección específica de diana (TS1) y una primera sección etiqueta (TAG1). La primera sección etiqueta no es complementaria a la secuencia de nucleótidos diana, es decir, está compuesta de una secuencia de nucleótidos que no se hibrida o une a la secuencia diana bajo condiciones de rigurosidad empleadas para la hibridación de la sección específica de secuencia diana. En ciertas realizaciones, la primera sonda comprende una sección específica de diana en su extremo 3'. La primera sección etiqueta puede comprender además una primera secuencia de unión a cebador (PBS1). La primera secuencia de unión a cebador es capaz de unirse a un cebador (PR1).

**[0018]** La segunda sonda comprende una segunda sección específica de diana (TS2) y una segunda sección etiqueta (TAG2). La segunda sección etiqueta no es complementaria a la secuencia de nucleótidos diana, es decir, está compuesta de una secuencia de nucleótidos que no se hibrida o une a la secuencia diana bajo condiciones de rigurosidad empleadas para la hibridación de la sección específica de secuencia diana. En ciertas realizaciones, la segunda sonda comprende una segunda sección específica de diana en su extremo 5'. La segunda sección etiqueta puede comprender además una segunda secuencia de unión a cebador (PBS2). La segunda secuencia de unión a cebador (si está presente) es capaz de unirse a un cebador (PR2).

**[0019]** Al menos una de las secciones etiqueta contiene una secuencia de reconocimiento para una endonucleasa de restricción. La primera y/o la segunda sección etiqueta puede comprender, independientemente, una primera y/o una segunda secuencia de reconocimiento (RE1, RE2) para una primera y/o segunda endonucleasa de restricción. La primera y la segunda secuencia de reconocimiento pueden ser la misma o diferente (es decir, RE1 = RE2 o RE1 ≠ RE2) entre sí. Existe una preferencia para las endonucleasas de restricción que tienen dos secuencias de reconocimiento diferentes (RE1 ≠ RE2). La secuencia de reconocimiento se encuentra entre el sitio de unión a cebador (si está presente) y la sección específica de diana. La primera secuencia de reconocimiento puede estar situada entre la primera secuencia de unión a cebador opcional y la primera sección específica de diana. La segunda secuencia de reconocimiento puede estar situada entre la segunda secuencia de unión a cebador (si está presente) y la segunda sección específica de diana.

**[0020]** Las primera y segunda secciones específicas de diana respectivas de las sondas se dejan hibridar con secciones preferiblemente esencialmente adyacentes en la secuencia diana, aunque en algunas realizaciones puede estar presente un hueco de uno o más nucleótidos entre las dos secciones (ligación de huecos, véase, por ejemplo WO2007/100243, WO00/77260, US5185243, EP439182 y más a continuación).

**[0021]** En ciertas realizaciones, la primera y segunda sondas están ligadas, es decir, están conectadas entre sí. Las sondas se ligan entre sí esencialmente cuando se hibridan (o unen) la respectiva sección diana (primera, segunda) a secciones esencialmente adyacentes en la secuencia diana. La ligación de la primera y segunda sonda proporciona sondas ligadas (LP).

**[0022]** Las sondas ligadas (LP) ahora se:

i). restringen con la primera y/o segunda endonucleasa de restricción que es capaz de reconocer la primera y/o segunda secuencia de reconocimiento de la endonucleasa de restricción para proporcionar sondas ligadas restringidas (RLP) (esto puede requerir el uso de oligonucleótidos complementarios y/o sondas en horquilla o el uso de endonucleasas de ssDNA, tal como se describe aquí en otros puntos); o

ii). amplifican (lineal o exponencial) con un primer cebador y/o un segundo cebador opcional para dar lugar a amplicones (A) y a continuación se restringen con la primera y/o segunda endonucleasa de restricción que es capaz de reconocer la secuencia de reconocimiento de la endonucleasa de restricción para proporcionar amplicones restringidos (RA).

**[0023]** Mediante el tratamiento de las sondas ligadas con una o más endonucleasas de restricción, las sondas ligadas restringidas se han liberado de partes de la sección etiqueta. Esto reduce significativamente la longitud de las sondas ligadas y, por consiguiente, la cantidad de datos producidos mejora la adaptabilidad de la tecnología de ensayos de ligación de oligonucleótidos para el uso de estrategias de secuenciación de alto rendimiento que prefieren lecturas más cortas. Las sondas ligadas restringidas comprenden los restos de la primera y/o la segunda secuencia de reconocimiento de la endonucleasa de restricción, la primera y la segunda sección complementaria diana. Las sondas ligadas restringidas pueden contener además uno o más identificadores basados en la secuencia (ID), esencialmente tal como se explica en este documento a continuación. Para ayudar en el tratamiento de las sondas ligadas con endonucleasas de restricción, los oligonucleótidos adicionales se pueden proporcionar de manera que pueden hibridarse con las sondas ligadas en las posiciones de los sitios de restricción y/o reconocimiento para proporcionar sitios de restricción y/o reconocimiento de doble cadena. Alternativamente, las sondas de forma de horquilla se pueden ligar a las sondas ligadas que cubren las posiciones de los sitios de restricción y/o de reconocimiento para proporcionar sitios de restricción y/o reconocimiento de doble cadena que posteriormente se pueden restringir. Alternativamente, la primera y/o las propias sondas pueden contener una estructura de horquilla a tal efecto.

**[0024]** Los amplicones restringidos (o sondas ligadas restringidas) se someten ahora a la tecnología de secuenciación de alto rendimiento, esencialmente tal como se describe en el presente documento a continuación para determinar al menos parte de la secuencia de nucleótidos de los amplicones restringidos o sondas ligadas restringidas. En ciertas realizaciones, se determina al menos parte de la sección específica de diana. En ciertas realizaciones, cuando se han incorporado identificadores en las sondas, tal como se explica en más detalle a continuación, y por lo tanto, en las sondas ligadas restringidas o amplicones restringidos, se determina al menos la secuencia de uno o más de los identificadores. En ciertas realizaciones preferidas, se determina una combinación de la sección específica de diana (información de alelo y/o locus) y dichos uno o más identificadores.

**[0025]** Mediante la determinación de la secuencia de los identificadores y/o al menos parte de la sección específica de diana, se identifica la presencia, ausencia o cantidad de la secuencia diana en la muestra.

#### Secuencia de nucleótidos diana

**[0026]** En su definición más amplia, la secuencia diana puede ser cualquier secuencia de nucleótidos de interés. La secuencia diana puede ser cualquier secuencia de la que se desea la determinación/detección, por ejemplo porque es indicativa, está asociada o es representativa de una determinada dolencia o constitución o trastorno genético. La secuencia diana preferiblemente es una secuencia de nucleótidos que contiene, representa o está asociada con un polimorfismo.

**[0027]** Tal como se utiliza en el presente documento, el término "polimorfismo" se refiere a la presencia de dos o más variantes de una secuencia de nucleótidos en una población. Un polimorfismo puede comprender uno o más cambios de bases, una inserción, una repetición, o una delección. Un polimorfismo incluye, por ejemplo, una repetición de secuencia simple (SSR) y un polimorfismo de un solo nucleótido (SNP), que es una variación, que se producen cuando se altera un solo nucleótido: adenina (A), timina (T), citosina (C) o guanina (G). Una variación generalmente debe ocurrir en al menos el 1% de la población para ser considerado un SNP. Los SNP componen, por ejemplo, el 90% de todas las variaciones genéticas humanas, y se producen cada 100 a 300 bases a lo largo del genoma humano. Dos de cada tres SNP sustituyen citosina (C) por Timina (T). Las variaciones en las secuencias de ADN de, por ejemplo, los seres humanos o plantas pueden afectar a la manera en que se tratan enfermedades, bacterias, virus, productos químicos, fármacos, etc.

**[0028]** Un marcador o sitio polimórfico es el locus en el que se produce la divergencia de secuencia. Los marcadores preferidos tienen al menos dos alelos, produciéndose cada uno a una frecuencia mayor del 1%, y más preferiblemente mayor del 10% o 20% de una población seleccionada. Un locus polimórfico puede ser tan pequeño como un par de bases. Los marcadores polimórficos incluyen polimorfismos de la longitud del fragmento de restricción, número variable de repeticiones en tándem (VNTR), regiones hipervariables, minisatélites, repeticiones de dinucleótidos, repeticiones de trinucleótidos, repeticiones de tetranucleótidos, repeticiones de secuencia simple, (elementos de) Quantitative Trait Loci, y elementos de inserción, tales como Alu. La primera forma alélica identificada se designa arbitrariamente como la forma de referencia (tipo salvaje) y otras formas alélicas son designadas como alelos alternativos o variantes. La forma alélica que se produce con mayor frecuencia en una

población seleccionada se refiere a veces como la forma tipo salvaje. Los organismos diploides (y tetraploides/hexaploides) pueden ser homocigotos o heterocigotos para las formas alélicas. Un polimorfismo dialélico tiene dos formas. Un polimorfismo trialélico tiene tres formas. Un polimorfismo de un solo nucleótido se produce en un sitio polimórfico ocupado por un único nucleótido, que es el sitio de variación entre secuencias alélicas. El sitio está normalmente precedido de y seguido por secuencias altamente conservadas del alelo (por ejemplo, secuencias que varían en menos de 1/100 o 1/1000 miembros de las poblaciones). Un polimorfismo de un solo nucleótido habitualmente surge debido a la sustitución de un nucleótido por otro en el sitio polimórfico. Los polimorfismos de un solo nucleótido también pueden surgir de una delección de un nucleótido o una inserción de un nucleótido con relación a un alelo de referencia. Otros polimorfismos incluyen (pequeñas) delecciones o inserciones de varios nucleótidos, conocidos como indels. El proceso de análisis de las variaciones genéticas particulares (polimorfismos) existentes en una muestra de ADN individual usando los procedimientos actualmente descritos se refiere a veces en esta solicitud como genotipado o genotipado de SNP en el caso de polimorfismos de un solo nucleótido. El procedimiento de la presente invención permite el genotipado codominante utilizando un conjunto de sondas para cada alelo. Esta realización es ventajosa en el caso de muestras heterocigóticas.

**[0029]** Tal como se utiliza en el presente documento, el término "alelo" o "alelos" significa cualquiera de una o más formas alternativas de un gen en un locus particular. En una célula diploide de un organismo, los alelos de un gen determinado se encuentran en una ubicación específica, o locus (loci en plural) en un cromosoma. Un alelo está presente en cada cromosoma del par de cromosomas homólogos. Un diploide, o especie de plantas, pueden comprender un gran número de diferentes alelos en un locus particular. El locus de adhesiones de tipo salvaje puede, por lo tanto, comprender diversos alelos, que pueden variar ligeramente en la secuencia de nucleótidos y/o la secuencia de aminoácidos codificada.

**[0030]** Tal como se utiliza en el presente documento, el término "locus" (loci en plural) significa un lugar o lugares específicos o un sitio en un cromosoma donde por ejemplo se encuentra un gen o marcador genético. Por ejemplo, el "locus" se refiere a la posición en el genoma donde se encuentran el gen (y los correspondientes alelos).

Muestra

**[0031]** Una muestra puede contener al menos una secuencia diana y, en principio, el procedimiento de la presente invención se puede ejecutar en una muestra que contiene una secuencia diana ("muestra única – monoplex"). Se prefiere que una muestra contenga dos o más secuencias diana diferentes ("muestra única – multiplex"), es decir, dos o más se refiere a la identidad en lugar de la cantidad de las secuencias diana en la muestra. En particular, la muestra comprende al menos dos secuencias diana diferentes, en particular al menos 100, preferiblemente al menos 250, más preferiblemente al menos 500, más en particular al menos 1000, preferiblemente al menos 2.500, más preferiblemente al menos 5.000 y lo más preferiblemente al menos 10.000 secuencias diana adicionales. En la práctica, el número de secuencias diana en una muestra que puede ser analizada está limitado, entre otros, por el número de amplicones o sondas ligadas que pueden ser detectados. Los procedimientos de detección actualmente empleados permiten grupos relativamente grandes de secuencias diana. La muestra se puede aislar directamente de un individuo o grupo de individuos o puede derivarse de los mismos, tales como ADNc, plásmidos, YAC, BAC, cósmidos, bibliotecas de cromosomas artificiales, etc.

Pluralidad de muestras

**[0032]** En ciertas realizaciones, se pueden analizar una pluralidad de muestras usando el procedimiento de la invención. Cada muestra se puede derivar de un origen diferente, por ejemplo, diferentes pacientes que tienen que ser examinados para la presencia o ausencia de ciertas disposiciones genéticas para ciertas dolencias. O las muestras se pueden derivar de la progenie de un cruce para cribar diferentes polimorfismos. El presente procedimiento se puede utilizar para analizar al menos dos muestras diferentes, en particular al menos 100, preferiblemente al menos 250, más preferiblemente al menos 500, más en particular al menos 1.000, preferiblemente al menos 2.500, más preferiblemente al menos 5.000 y lo más preferiblemente al menos 10.000 muestras para la ausencia o presencia de una ("monoplex-monoplex") o más o una pluralidad ("multiplex-multiplex") de secuencias diana. Las muestras se pueden distinguir en el procesamiento adicional del procedimiento utilizando uno o más (combinaciones de) identificadores tal como se describe en el presente documento en otros lugar.

ADN

**[0033]** En la muestra (ácido nucleico), los ácidos nucleicos que comprenden la secuencia de nucleótidos diana puede ser cualquier ácido nucleico de interés. A pesar de que los ácidos nucleicos en la muestra estarán habitualmente en forma de ADN, la información de la secuencia de nucleótidos contenida en la muestra puede ser de cualquier fuente de ácidos nucleicos, incluyendo por ejemplo ARN, ARN poliA +, ADNc, ADN genómico, ADN organular, tal como ADN mitocondrial o de cloroplastos, ácidos nucleicos sintéticos, bibliotecas de ADN (tales como bibliotecas BAC/clones BAC agrupados), bancos clonales o cualquier selección o combinaciones de los mismos. El ADN de la muestra puede ser de doble cadena, de cadena sencilla y ADN de doble cadena desnaturalizado en ADN de una sola cadena. La desnaturalización de secuencias de doble cadena produce dos fragmentos de cadena sencilla, una o ambas pueden analizarse por sondas específicas para las cadenas respectivas. Las muestras de

ácido nucleico preferidas comprenden secuencias diana en ADNc, ADN genómico, fragmentos de restricción, fragmentos de restricción ligados a adaptadores, fragmentos de restricción ligados a adaptadores amplificados, fragmentos AFLP® o fragmentos obtenidos en una preamplificación de plantilla de AFLP.

5 Sonda

[0034] Las secciones de las sondas de oligonucleótidos que son complementarias a la secuencia diana se diseñan de tal manera que para cada secuencia diana en una muestra, se proporciona un par de una primera y una segunda sonda, mediante lo cual las sondas contienen cada una una sección en su extremo que es complementaria a una parte de la secuencia diana (una primera y una segunda parte de la secuencia diana, respectivamente) y las partes complementarias correspondientes de la secuencia diana se encuentran preferiblemente esencialmente adyacentes entre sí.

[0035] En ciertas realizaciones, se pueden proporcionar la primera y/o la segunda sondas adicionales, correspondientes a diferentes alelos de un locus. En ciertas realizaciones, el nucleótido específico de alelo se encuentra en la posición de la primera o la segunda sonda en la que se produce la ligación, es decir, en el extremo de la sección específica de diana.

[0036] En ciertas realizaciones, dentro de un par de sondas de oligonucleótidos, la primera sonda de oligonucleótidos tiene una sección en su extremo 5' (fosforilado) que es complementaria a una primera parte de una secuencia diana y la segunda sonda de oligonucleótidos tiene una sección en su extremo 3' (hidroxilo) que es complementaria a una segunda parte de la secuencia diana. De este modo, cuando el par de sondas se hibridan con partes complementarias de una secuencia diana, el extremo 5' de la primera sonda de oligonucleótidos es esencialmente adyacente al extremo 3' de la segunda sonda de oligonucleótidos de tal manera que los extremos respectivos de las dos sondas pueden ligarse para formar un enlace fosfodiéster o conectarse covalentemente en cualquier otra forma adecuada. En ciertas realizaciones, dentro de un par de sondas de oligonucleótidos, la primera sonda de oligonucleótidos tiene una sección en su extremo 3' que es complementaria a una primera parte de una secuencia diana y la segunda sonda de oligonucleótidos tiene una sección en su extremo 5' que es complementaria a una segunda parte de la secuencia diana. De este modo, cuando el par de sondas se hibrida con partes complementarias de una secuencia diana, el extremo 3' de la primera sonda de oligonucleótidos es esencialmente adyacente al extremo 5' de la segunda sonda de oligonucleótidos de tal manera que los extremos respectivos de las dos sondas se pueden ligar para formar un enlace fosfodiéster o conectarse de forma covalente en otra forma adecuada.

[0037] Para la detección específica de alelo, se prefiere que la sonda específica de alelo tenga su sección específica de diana en el extremo 3' de la sonda. Al contrario, la sonda específica de alelo en el extremo 5' de la sonda es menos preferida.

[0038] En ciertas realizaciones, para cada secuencia diana para la que se determina la presencia, ausencia o cantidad en una muestra, se diseña un par específico de primera y segunda sondas de oligonucleótidos, cada sonda con una sección complementaria a la parte complementaria adyacente de cada secuencia diana, tal como se describe anteriormente. Por lo tanto, en el procedimiento de la invención, para cada secuencia diana que está presente en una muestra, se puede obtener una sonda ligada o un amplicón (específico) correspondiente en la muestra amplificada. En ciertas realizaciones, se proporciona una multiplicidad de primera y segunda sondas de oligonucleótidos complementarias a una multiplicidad de secuencias diana en una muestra. Un par de primera y segunda sondas de oligonucleótidos para una secuencia diana determinada en una muestra diferirá al menos en la secuencia de nucleótidos de pares de sondas para otras secuencias diana u otras muestras, y pueden diferir en la longitud y/o la masa de pares de sondas para otras dianas (aunque, como se ha señalado anteriormente, esto es menos preferido). Más preferiblemente, un par de sondas para una diana determinada producirá una sonda ligada (a veces indicada como sonda conectada) y/o amplicón que difiere en la secuencia de sondas ligadas y/o amplicones correspondientes a otras dianas en la muestra.

[0039] Existe un número de variaciones posibles de la sonda dentro del alcance de la presente invención que se pueden utilizar como alternativas a la primera y segunda sonda (a veces indicadas como "sondas lineales") descritas en el presente documento. Los ejemplos se denominan sondas candado ("padlock") y sondas de "key lock". Estas variantes de sonda se pueden utilizar indistintamente, es decir, se pueden utilizar en un ensayo combinaciones de sondas lineales, sondas candado y Keylock.

60 Sondas candado

[0040] En ciertas realizaciones de la invención, se pueden utilizar sondas circularizables o sondas candado. Las primera y segunda sondas se combinan a continuación en una sonda. La sonda circularizable es un oligonucleótido lineal que, cuando se hibrida con la secuencia diana, y cuando se liga, tiene una conformación circular que está topológicamente bloqueada a la secuencia diana. En ciertas realizaciones, un tratamiento con exonucleasa de la muestra después de la etapa de ligación y antes de la amplificación, preferiblemente amplificación por PCR, sirve para eliminar cualquier sonda circular no ligada y para evitar la amplificación de cualquier sonda no ligada. Las

sondas circularizables son en sí mismas conocidas en la técnica, por ejemplo de EP745140 o de Van Eijk et al, Nucleic Acids Research, 2004, 32, E47. Las sondas candado conocidas se amplifican habitualmente mediante amplificación por círculo rodante o la reacción en cadena de la polimerasa dando lugar a concatámeros. Además, los sitios de unión al cebador en las sondas circularizables conocidas están orientados de tal manera que la sonda circularizada completa se amplifica incluyendo cualquier secciones específica de diana. Con el fin de evitar productos concatámeros durante la amplificación por PCR, se puede incorporar una modificación de bloqueo en la sonda de ligación circularizable entre los dos sitios de unión a cebador del tipo descrito en el documento WO03/052142. En ciertas realizaciones, los sitios de unión a cebador en las presentes sondas circularizables están orientados de tal manera que preferiblemente sólo la sección que comprende los sitios de unión a cebador y el identificador se amplifican y preferiblemente las secciones específicas de diana ligadas no se amplifican. Preferiblemente, en combinación con el tratamiento con exonucleasa para eliminar las sondas circularizables no ligadas, esto proporciona amplicones de longitud relativamente corta en comparación con amplicones convencionales obtenidos a partir de sondas circularizadas convencionalmente amplificantes. Esto evita la formación de grandes concatámeros y la posterior amplificación innecesaria de la sonda circularizada completa. En ciertas realizaciones, el identificador se encuentra esencialmente adyacente a una de las secuencias de unión a cebador, y preferiblemente entre el primer y el segundo sitio de unión a cebador, de manera que tras la amplificación de los amplicones comprende al menos uno de los dos sitios de unión a cebador y el identificador intermitente. Dicho al menos uno, preferiblemente dos sitios de reconocimiento de la endonucleasa de restricción están situados preferiblemente de manera que los sitios de reconocimiento abarcan el primer y/o el segundo identificador y la primera y segunda secciones específicas de diana de la sonda. La posterior secuenciación de alto rendimiento de los amplicones o sonda ligada restringida proporcionará la secuencia del identificador y/o (parte de) la secuencia de la sección o secciones diana y por lo tanto de la presencia de la secuencia diana en la muestra. En esta realización, la presencia de la secuencia de reconocimiento para la endonucleasa de restricción permite que los concatámeros se reduzcan a fragmentos secuenciables.

#### Sondas "Keylock"

[0041] En ciertas realizaciones, para cada secuencia diana determinada a detectar, se diseña preferiblemente al menos un par de dos sondas de manera que cada sonda en el par es capaz de hibridarse a una parte de la secuencia diana y las sondas respectivas en el par comprender cada una además una sección que es complementaria a una sección correspondiente de la otra sonda en el par de tal manera que ambas sondas son capaces de hibridarse entre sí. Las dos sondas en el par están diseñadas de tal manera que cuando se hibridan entre sí también son capaces cada una de hibridarse con una secuencia diana. Cuando se hibridan entre sí, las dos sondas mimetizan o actúan como sondas de candado cuando se utilizan en un ensayo de ligación de oligonucleótidos para la detección de una secuencia de nucleótidos diana, mientras que en las posteriores etapas de amplificación y detección las sondas actúan como un producto de ligación lineal. Este tipo de sonda ha sido llamada "Keylock" y se da a conocer, entre otras, en WO2004111271. En esta realización, la presencia de la secuencia de reconocimiento para la endonucleasa de restricción, situada entre las secciones de agarre y las secciones específicas de diana, permite la liberación de los keylocks de al menos sus secciones de agarre.

#### Sonda de compuesto

[0042] En ciertas realizaciones de la presente invención, se utiliza un conjunto de sondas que se describen en WO2005021794. La secuencia diana se pone en contacto con una primera y una segunda sonda, en el que la primera sonda contiene una primera sección específica de diana que es complementaria a la secuencia diana y en la que la primera sonda preferiblemente no contiene una primera secuencia de unión a cebador en una primera sección etiqueta opcional. La segunda sonda contiene una segunda sección específica de diana y una segunda sección etiqueta en la que la segunda sección etiqueta contiene una segunda secuencia de unión a cebador. La segunda sección etiqueta puede contener un identificador entre la segunda secuencia de unión a cebador y la segunda sección específica de diana. Después, o simultáneamente con, la hibridación y la ligación de las dos sondas, se proporciona una sonda de compuesto que contiene una sección que es capaz de hibridarse con (parte de) la primera sección específica de diana de la primera sonda y contiene además una sección que contiene una sección de unión a cebador. Tanto la primera sección etiqueta como la sección de sonda de compuesto que contiene un sitio de unión a cebador pueden contener adicionalmente un sitio de restricción. El sitio de restricción en la primera sección etiqueta se encuentra entre el sitio de unión a cebador y la sección específica de diana. El sitio de restricción en la sonda de compuesto se encuentra entre el sitio de unión a cebador y la sección que puede hibridarse a la primera sección diana. La sonda de compuesto se hibrida con la primera y segunda sonda ligadas. El alargamiento de la sonda de compuesto a lo largo de la primera y segunda sonda ligadas proporciona la sonda de compuesto alargada que posteriormente puede amplificarse utilizando el primer y segundo cebadores que pueden unirse al primer y segundo sitio de unión a cebador. Los amplicones resultantes se pueden restringir usando una o más endonucleasas de restricción y se detectan usando las tecnologías de secuenciación de alto rendimiento tal como se describe en el presente documento y la secuencia diana en la muestra se puede identificar por medio de la presencia o ausencia del identificador o identificadores y/o la información de locus/alelo.

#### Sección etiqueta

**[0043]** El término sección etiqueta se utiliza para indicar las partes de las sondas que no son capaces de hibridarse a las secuencias de nucleótidos diana. Las secciones etiqueta por lo general contienen identificadores y sitios de unión a cebador y en algunas ocasiones, tal como se describe en el presente documento en otra parte, secciones de agarre.

5

#### Secuencia de unión a cebador

**[0044]** Las secuencias de unión a cebador se pueden incorporar en las sondas para facilitar la amplificación, ya sea lineal o exponencial. Los sitios de unión a cebador se encuentran preferiblemente en otras partes de la sonda que en la sección específica de diana, preferiblemente en la sección etiqueta que es esencialmente no complementaria a la secuencia diana. Los sitios de unión a cebador son capaces de unir cebadores para iniciar la elongación o amplificación del cebador. Preferiblemente dentro de un grupo de pares de sondas (por ejemplo, tal como se utiliza dentro de una muestra), los sitios de unión a cebador son universales, es decir, sólo un grupo predeterminado de sitios de unión a cebador se incorporan en la sonda para permitir la elongación o amplificación de cebadores multiplex a partir de un número limitado de cebadores, tales como cebadores que comprenden una o más bases selectivas en su extremo 3', tal como se conocen de AFLP (EP 0 534 858). Entre los grupos de pares de sondas, los sitios de unión a cebador pueden ser diferentes (es decir, tienen una secuencia diferente). En ciertas realizaciones, la  $T_m$  de los cebadores capaces de unirse a los diferentes sitios de unión a cebador puede ser diferentes entre grupos de pares de sondas. Habitualmente, una secuencia de unión a cebador puede tener una longitud de entre 6 y 200 nucleótidos, con una preferencia en el área entre 8 y 50, más preferiblemente entre 10 y 25 nucleótidos.

10

15

20

#### Hibridación

**[0045]** Tal como se ha mencionado anteriormente en el presente documento, las sondas se ponen en contacto mediante hibridación con la secuencia diana en la muestra. Los pares de sondas de oligonucleótidos se les permite posteriormente hibridarse con las partes complementarias, preferiblemente adyacentes, de la secuencia diana en la muestra. Los procedimientos y condiciones para la hibridación específica de las sondas de oligonucleótidos a secuencias diana complementarias son bien conocidos en la técnica (véase por ejemplo, en Sambrook y Russel (2001) "Molecular Cloning: A Laboratory Manual (tercera edición), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press). Habitualmente, después de la mezcla de las sondas de oligonucleótidos y secuencias diana, los ácidos nucleicos se desnaturalizan mediante incubación (generalmente a entre 94 grados Celsius y 96 grados Celsius) durante un corto período de tiempo (por ejemplo, de 30 segundos a 5 minutos) en un tampón salino. La muestra que contiene las sondas desnaturalizadas y las secuencias diana se deja a continuación enfriar hasta una temperatura de hibridación óptima para la hibridación específica de las sondas y las secuencias diana, que por lo general es de aproximadamente 5 grados Celsius por debajo de la temperatura de fusión del híbrido entre la sección complementaria (sección diana) de la sonda y su secuencia complementaria (en la secuencia diana). Con el fin de evitar una hibridación no específica o ineficaz de una de las dos sondas de un par, o en una muestra con múltiples secuencias diana, se prefiere que, dentro de una muestra, las secciones de las sondas que son complementarias a las secuencias diana tengan una temperatura de fusión similar, preferiblemente idéntica, entre las diferentes secuencias diana presentes en la muestra. De este modo, las secciones complementarias de la primera y segunda sondas difieren preferiblemente en menos de 20, 15, 10, 5, o 2 grados Celsius en la temperatura de fusión. Esto se facilita mediante el uso de secciones complementarias de la primera y segunda sondas con una longitud similar y/o un contenido de G/C similar, las secciones complementarias difieren preferiblemente en menos de 20, 15, 10, 5, o 2 nucleótidos de longitud y su contenido de G/C difieren en menos de 30, 20, 15, 10, o 5%. Complementario tal como se utiliza en el presente documento significa que una primera secuencia de nucleótidos es capaz de hibridarse específicamente a una segunda secuencia de nucleótidos en condiciones normales de rigurosidad. Una secuencia de nucleótidos que se considera complementaria a otra secuencia de nucleótidos puede contener una cantidad menor, es decir, preferiblemente menos de 20, 15, 10, 5 o 2%, de los desajuste en el emparejamiento. Alternativamente, puede ser necesario compensar los desajustes en el emparejamiento, por ejemplo, mediante la incorporación de los llamados nucleótidos universales, tales como, por ejemplo, los descritos en el documento EP-A 974672 o mediante la incorporación de ciertos nucleótidos modificados que son capaces de compensar los desajustes en el emparejamiento, por ejemplo, mediante el aumento de la temperatura de fusión o el aumento de la especificidad, tal como LNA. Dado que la hibridación de las sondas a secuencias diana depende de la concentración, la hibridación se realiza preferiblemente en un pequeño volumen, es decir, menos de 25 microlitros, preferiblemente menos de 10 microlitros. Bajo estas condiciones de hibridación, la hibridación de sondas a secuencias diana habitualmente es rápida y no necesita proceder durante más de 5, 10 ó 15 minutos, aunque se pueden utilizar tiempos de hibridación más largos siempre que se mantenga la temperatura de hibridación para evitar una hibridación no específica. Los tiempos de hibridación más largos son más importantes/necesarios para aplicaciones cuantitativas que se basan en la ocupación completa de la diana por las sondas de ligación con el fin de permitir la monitorización de cantidades relativas de secuencias diana entre las muestras.

25

30

35

40

45

50

55

60

65

**[0046]** En ciertas realizaciones de la invención, se han obtenido excelentes resultados mediante tiempos de hibridación prolongados, tales como hibridación durante una noche o mediante hibridación repetida, tal como 10 ciclos de 1 hora. Los tiempos de hibridación prolongados pueden ser ventajosos en estos ensayos ya que se reduce la diferencia en la señal debido a diferentes eficiencias de hibridación y se considera deseable conseguir una

hibridación y la ligación completas de todas las sondas para las que una secuencia diana está presente. Se han obtenido excelentes resultados mediante una etapa combinada de ligación-hibridación usando una ligasa termoestable descrita en este documento. En esta realización, la hibridación-ligación se realizó permitiendo que las sondas se hibriden durante 1 hora en presencia de una ligasa termoestable, seguida por una etapa de desnaturalización. La repetición de estas etapas durante al menos 2 veces proporcionó buenos resultados. La repetición de estas etapas 10 veces proporcionó excelentes resultados. Para evitar la evaporación durante la desnaturalización y la hibridación, las paredes y tapas de las cámaras de reacción (es decir, tubos o pocillos de microtitulación) también se pueden calentar hasta al menos la misma temperatura que la mezcla de reacción que se logra habitualmente mediante el uso de un equipo de amplificación de ADN comercial o mediante la disposición de una capa de aceite mineral. En sondas de oligonucleótidos preferidas, la longitud de la sección complementaria de diana es preferiblemente de al menos 15, 18 o 20 nucleótidos y preferiblemente no más de 30, 40, o 50 nucleótidos y las sondas tienen preferiblemente una temperatura de fusión desde la sección diana de al menos 50 grados Celsius, 55 grados Celsius o 60 grados Celsius.

#### 15 Ligación

**[0047]** Los extremos 5'-fosforilado y 3'-hidroxilado respectivo de un par de primera y segunda sondas de oligonucleótidos o de la sonda circularizable de las cuales las secciones específicas de diana se hibridan esencialmente adyacentes entre sí a las partes complementarias en una secuencia diana se conectan para formar un enlace covalente mediante cualquier medio adecuado conocido en la técnica. Los extremos de las sondas pueden conectarse enzimáticamente en un enlace fosfodiéster por una ligasa, preferiblemente una ADN ligasa. Las ADN ligasas son enzimas capaces de catalizar la formación de un enlace fosfodiéster entre (los extremos de) dos cadenas polinucleotídicas unidas a sitios adyacentes en una cadena complementaria. Las ADN ligasas habitualmente requieren ATP (EC 6.5.1.1) o NAD (EC 6.5.1.2) como cofactor para sellar las mellas en el ADN de doble cadena. La ADN ligasa adecuada para usar en la presente invención son ADN ligasa de T4, ADN ligasa de E. coli o preferiblemente una ligasa termoestable como, por ejemplo ligasa de *Thermus aquaticus* (Taq), ADN ligasa de *Thermus thermophilus*, o ADN ligasa de *Pyrococcus*. Alternativamente, la ligación química de extremos de polinucleótidos modificados adecuadamente se puede utilizar para ligar dos sondas de oligonucleótidos hibridados en sitios adyacentes en las partes complementarias de una secuencia diana. Los ejemplos de grupos reactivos en los extremos de polinucleótidos modificados incluyen, pero no se limitan a, grupos de fosforotioato y tosiloato o yoduro, ésteres e hidrazida, RC (O)S, RCH<sub>2</sub>S y [alfa]-haloacilo, tiofosforilo y bromoacetamida y S-pivaloiloximetil-4-tiotimidina. Los agentes de ligación química incluyen, sin limitación, agentes activantes, de condensación, y reductores, tales como carbodiimida, bromuro de cianógeno (BrCN), N-cianoimidazol, imidazol, 1-metilimidazol/carbodiimida/cistamina, ditiotreitil (DTT) y luz ultravioleta. La autoligación, es decir, la ligación espontánea en ausencia de un agente de ligación, también está dentro del alcance de la invención. Los protocolos detallados para los procedimientos de ligación química y las descripciones de grupos reactivos apropiados se pueden encontrar, entre otros lugares, en Xu et al., *Nucleic Acid Res.*, 27: 875-81 (1999); Gryaznov y Letsinger, *Nucleic Acid Res.* 21: 1403-1408 (1993); Gryaznov et al., *Nucleic Acid Res.* 22: 2366-69 (1994); Kanaya y Yanagawa, *Biochemistry* 25: 7423-30 (1986); Luebke y Dervan, *Nucleic Acids Res.* 20: 3005-09 (1992); Sievers y von Kiedrowski, *Nature* 369: 221-24 (1994); Liu y Taylor, *Nucleic Acids Res.* 26: 3300-04 (1999); Wang y Kool, *Nucleic Acids Res.* 22: 2326-33 (1994); Purmal et al., *Nucleic Acids Res.* 20: 3713-19 (1992); Ashley y Kushlan, *Biochemistry* 30: 2927-33 (1991); Chu y Orgel, *Nucleic Acids Res.* 16: 3671-91 (1988); Sokolova et al., *FEBS Letters* 232: 153-55 (1988); Naylor y Gilham, *Biochemistry* 5: 2722-28 (1966); y la patente de Estados Unidos No. 5.476.930. La ligación tanto química como enzimática tiene lugar de manera mucho más eficiente en complejos de secuencia diana-sonda perfectamente emparejados en comparación con complejos en los que una o ambas de las sondas forman un desajuste en el emparejamiento con la secuencia diana en, o cerca del sitio de la ligación (Wu y Wallace, 1989, *Gene* 76: 245-254; Xu y Kool, supra). Con el fin de aumentar la especificidad de ligación, es decir, las eficiencias relativas de ligación de oligonucleótidos perfectamente emparejados en comparación con oligonucleótidos no emparejados, la ligación se realiza preferiblemente a temperaturas elevadas. De este modo, en ciertas realizaciones de la invención, se emplea una ADN ligasa que permanece activa a 50 - 65 grados Celsius durante tiempos prolongados, pero que se inactiva fácilmente a mayores temperaturas, por ejemplo, utilizada en la etapa de desnaturalización durante una PCR, habitualmente 90 - 100 grados Celsius. Una de estas ADN ligasas es una ADN ligasa que requiere NAD de una bacteria Gram-positiva (cepa MRCH 065) tal como se conoce del documento WO01/61033. Esta ligasa se conoce como "ligasa 65" y está disponible comercialmente en MRC Holanda, Amsterdam. En ciertas realizaciones, se usa una Taq ligasa. En ciertas realizaciones, la ligasa se inactiva después de la ligación de la primera y segunda sondas. En ciertas realizaciones, la sonda ligada se desnaturaliza de la secuencia diana.

**[0048]** En ciertas realizaciones de la presente invención, la hibridación y la ligación se llevan a cabo en una etapa combinada. Dicha etapa combinada de hibridación y ligación se puede realizar usando un perfil de temperatura cíclico y una ligasa termoestable.

#### Ligación de espacios

**[0049]** En una realización alternativa, por ejemplo dirigida a la identificación de indels, los extremos respectivos de las secciones complementarias dianas de las primera y segunda sondas pueden hibridarse de tal manera que se deja

un espacio. En ciertas realizaciones, las primera y segunda partes de la secuencia de nucleótidos diana no se encuentran adyacentes. En otras palabras, la primera y segunda secciones específicas de diana de la primera y segunda sondas no se hibridan a las primera y segunda partes de la secuencia de nucleótidos diana que se encuentran adyacentes. Esto es fundamentalmente diferente de otras variedades de esta tecnología, tales como las descritas, entre otras, en los documentos EP185494, US5521065, US5692223 y WO 03054311. Este espacio se puede rellenar con un (tercer) (oligo) nucleótido adecuado y ligarse. Dichos procedimientos son conocidos en la técnica como "ligación de espacios" (ensayos Illumina Golden Gate) y se describen, entre otros, en los documentos WO 00/77260; US5185243; EP439182; EP320308; W090/01069. Otra posibilidad para rellenar este espacio es mediante la extensión de un extremo de la sonda usando una polimerasa y una ligasa en combinación con nucleótidos individuales o múltiples, opcionalmente preseleccionados entre A, T, C, o G, o di-, tri- u otros oligonucleótidos pequeños. En caso de que la secuencia diana sea ARN, sin embargo, otra posibilidad para rellenar el espacio es mediante la extensión de un extremo de la sonda usando la transcriptasa inversa y una ligasa en combinación con nucleótidos individuales o múltiples, opcionalmente preseleccionados entre A, T, C o G, o di-, tri- u otros oligonucleótidos pequeños. La ligación de espacios puede encontrar aplicación en la detección de SNP individuales/indels o SNP múltiples (haplotipado) que se encuentran muy cerca. En esta realización, la etapa de secuenciación comprendería preferentemente la determinación de la secuencia del espacio.

#### Amplificación

**[0050]** En el procedimiento de la invención, las sondas ligadas pueden amplificarse para producir una muestra amplificada que comprende sondas ligadas amplificadas (amplicones) que son representaciones de la secuencia de nucleótidos diana mediante cualquier procedimiento de amplificación de ácidos nucleicos adecuado conocido en la técnica. Los procedimientos de amplificación de ácidos nucleicos emplean generalmente uno o dos cebadores, dNTP, y una (ADN) polimerasa. Un procedimiento preferido para la amplificación es la PCR. "PCR" o "Reacción en cadena de la polimerasa", tal como se utiliza aquí como ejemplo de un procedimiento de amplificación, es un procedimiento rápido para la amplificación enzimática in vitro de un segmento específico de ADN. El ADN a amplificar se desnaturaliza calentando la muestra. En presencia de ADN polimerasa y desoxinucleótido trifosfatos en exceso, los oligonucleótidos que se hibridan específicamente con la secuencia diana preparan una nueva síntesis de ADN. Se prefiere que la polimerasa sea una ADN polimerasa que no expresa actividad de desplazamiento de cadena o al menos no significativamente. Ejemplos de las mismas son Amplitaq y Amplitaq Gold (proveedor Perkin Elmer) y AccuPrime (Invitrogen). Una ronda de síntesis da lugar a nuevas cadenas de longitud determinada, que, como las cadenas parentales, pueden hibridarse con los cebadores sobre tras la desnaturalización e hibridación. El segundo ciclo de desnaturalización, hibridación y síntesis produce dos productos de cadena sencilla que juntos componen un producto de doble cadena discreto, exactamente la longitud entre los extremos del cebador. Este producto discreto se acumula exponencialmente con cada ronda sucesiva de amplificación. En el transcurso de alrededor de 20 a 30 ciclos, se puede conseguir una amplificación de muchos millones de veces del fragmento discreto. Los protocolos de PCR son bien conocidos en la técnica, y se describen en libros de texto estándar de laboratorio, por ejemplo, Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, Inc. (1995). Las condiciones adecuadas para la aplicación de PCR en el procedimiento de la invención se describen en el documento EP-A 0 534 858 y Vos et al. (1995; Nucleic Acids Res.23: 4407-4414), donde múltiples fragmentos de ADN entre 70 y 700 nucleótidos y que contienen secuencias de unión a cebador idénticas se amplifican con una eficacia casi igual usando un par de cebadores. En ciertas realizaciones, la polimerasa se inactiva después de la amplificación. Otros procedimientos de amplificación múltiplex y/o isotérmicos que se pueden aplicar incluyen, por ejemplo amplificación por círculo rodante (RCA), reacción en cadena con ligasa (LCR), replicación de secuencia auto-sostenida (3SR), amplificación de ARN mediada por Q-B-replicasa, o amplificación por desplazamiento de cadena (SDA). En algunos casos, esto puede requerir un diseño diferente de las sondas y cebadores sin apartarse de la esencia de la invención.

**[0051]** Dentro de la presente invención, la amplificación puede realizarse en varios puntos en el tiempo. La amplificación puede llevarse a cabo para la preparación de una biblioteca (aumento en el material de partida para la secuenciación), por ejemplo, después de la etapa de ligación y/o como parte de la etapa de secuenciación (es decir, PCR en emulsión (Roche, Ion Torrent) o amplificación por puente (Illumina).

#### Amplicón

**[0052]** El término "amplicón", tal como se utiliza aquí, se refiere al producto de la etapa de amplificación de las sondas ligadas. El término "amplificación", tal como se utiliza aquí, por tanto, se refiere a una sonda ligada amplificada. Después de la etapa de ligación en la que las dos secciones específicas de diana están conectadas por medio de una ligasa, la sonda conectada o ligada se combina con uno o más cebadores y una polimerasa y se amplifica para producir amplicones. La sonda ligada, los cebadores, la polimerasa y/u otros parámetros y variables son tales que la amplificación da lugar a representaciones amplificadas (lineales) de la sonda ligada. Preferiblemente un amplicón es una representación monomérica de la sonda conectada amplificada. En ciertas realizaciones, el amplicón comprende y, preferiblemente, consiste en nucleótidos del primer y segundo cebador opcional y el identificador o identificadores que están situados en el medio. En ciertas realizaciones, el amplicón puede contener nucleótidos de la sección específica de diana. Las diversas realizaciones de la presente invención proporcionarán más detalles a este respecto (Figura 2).

Endonucleasas de restricción

**[0053]** Enzima 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 de doble cadena, y escindirá las dos cadenas de la molécula de ADN en o cerca de cada sitio diana, dejando un extremo romo o un extremo escalonado (es decir, que contiene un saliente de uno o más nucleótidos). También existen endonucleasas de restricción que cortan ss-ADN (EndoTT, Exo I, Exo T) y son de uso en la presente invención cuando las sondas ligadas no se amplifican sino que directamente de cortan antes de la secuenciación.

**[0054]** Una endonucleasa de restricción Tipo-IIS es una endonucleasa que tiene una secuencia de reconocimiento que está distante del sitio de restricción. En otras palabras, las endonucleasas de restricción de tipo IIs escinden fuera de la secuencia de reconocimiento a un lado. Ejemplos de las mismas son NmeAIII (GCCGAG (21/19) y FokI, AlwI, Mme I. Existen enzimas Tipo IIs que cortan fuera de la secuencia de reconocimiento en ambos lados.

**[0055]** Los cortadores frecuentes y cortadores raros son indicaciones para enzimas de restricción que típicamente tienen secuencias de reconocimiento que varían en número de nucleótidos desde 4 (tal como MseI) a 6 (EcoRI) e incluso 8 (NotI). Las enzimas de restricción utilizadas pueden ser cortadores frecuentes y raros. El término "frecuente" a este respecto se suele utilizar en relación con el término "raro". Las endonucleasas de corte frecuente (también conocidas como cortadores frecuentes) son endonucleasas de restricción que tienen una secuencia de reconocimiento relativamente corta. Los cortadores frecuentes suelen tener 4 o 5 nucleótidos que reconocen y posteriormente cortan. Por lo tanto, un cortador frecuente corta en promedio una secuencia de ADN cada 256-1024 nucleótidos. Los cortadores raros son endonucleasas de restricción que tienen una secuencia de reconocimiento relativamente larga. Los cortadores raros suelen tener 6 o más nucleótidos que reconocen y posteriormente cortan. Por lo tanto, un cortador raro de 6 corta en promedio una secuencia de ADN cada 4096 nucleótidos, dando lugar a fragmentos más largos. Se observa de nuevo que la definición de frecuente y raro es una con respecto a la otra, lo que significa que cuando una enzima de restricción de 4 pb, tal como MseI, se utiliza en combinación con un cortador de 5, tal como Avall, se observa Avall como el cortador raro y MseI como el cortador frecuente.

**[0056]** Isoesquizómeros; los isoesquizómeros son pares de enzimas de restricción específicas para la misma secuencia de reconocimiento y cortan en la misma ubicación. Por ejemplo, Sph I (GCATG<sup>^</sup>C) y Bbu I (GCATG<sup>^</sup>C) son isoesquizómeros el uno del otro. La primera enzima para reconocer y cortar una secuencia determinada se conoce como el prototipo, todas las enzimas posteriores que reconocen y cortan esa secuencia son isoesquizómeros. Una enzima que reconoce la misma secuencia pero corta de otra manera es un neoesquizómero. Los isoesquizómeros son un tipo específico (subconjunto) de neoesquizómeros. Por ejemplo, Sma I (CCC<sup>^</sup>GGG) y Xma I (C<sup>^</sup>CCGGG) son neoesquizómeros (no isoesquizómeros) uno del otro.

**[0057]** Fragmento de restricción es el término usado para las moléculas de ADN producidas mediante la digestión de ADN con una endonucleasa de restricción. Cualquier genoma determinado (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 resultan de la escisión con endonucleasa de restricción se pueden utilizar adicionalmente en una variedad de técnicas, tales como la secuenciación. Los fragmentos de restricción pueden tener extremos romos o un saliente. El saliente se puede eliminar usando una técnica descrita como pulido. El término "secuencia interna" de un fragmento de restricción se utiliza normalmente para indicar que el origen de la parte del fragmento de restricción reside en el genoma de la muestra, es decir, no forma parte de un adaptador. La secuencia interna deriva directamente del genoma de la muestra, su secuencia es, por tanto, parte de la secuencia del genoma bajo investigación. El término secuencia interna se utiliza para distinguir sobre adaptadores, restos de secuencia de reconocimiento de enzimas de restricción etc.

Secuencia Identificadora

**[0058]** En ciertas realizaciones, la sonda de oligonucleótidos de la presente invención comprende, además, un identificador o una secuencia identificadora. La secuencia identificadora es una secuencia de oligonucleótidos de una secuencia variable. La longitud del identificador varía de 1 a 30, preferiblemente de 2 a 20, más preferiblemente de 3 a 10 y lo más preferido de 4 a 6 nucleótidos. El identificador es una secuencia única. Única tal como se utiliza aquí significa que un identificador o (combinación de) identificadores identifica de manera inequívoca una secuencia diana específica en una muestra o una pluralidad de muestras como diferentes de cualquier otra secuencia diana, alelo, locus en la muestra o pluralidad de muestras. El carácter único se puede explicar como una secuencia codificada con un CP del tipo descrito por Iannone et al. (2000), Cytometry 39: pp. 131-140. Con un identificador de 6 nucleótidos, se pueden realizar un máximo de 4096 combinaciones únicas (= 4 exp 6). En ciertas realizaciones, el identificador contiene una secuencia de anclaje de 2 bases GC (u otra corta definida rica en G/C) en el extremo 3' para asegurar la misma afinidad de unión y eficiencia de amplificación. Además se prefiere que el identificador no contenga dos bases consecutivas idénticas y se prefiere, además, que todos los identificadores utilizados en un conjunto de identificadores difieran en al menos dos bases con el fin de garantizar un reconocimiento inequívoco de secuencia. Cuando se utilizan múltiples muestras se prefiere que cada muestra pueda identificarse utilizando un conjunto específico de identificadores. El identificador se encuentra generalmente, de manera que la amplificación o

restricción de las sondas ligadas utilizando las secuencias de unión a cebador y/o endonucleasas de restricción incorporarán el identificador a fin de que el amplicón resultante o sonda ligada restringida contenga la secuencia de identificador.

5 **[0059]** Normalmente, esto significa que en la sonda ligada, el identificador está situado cerca de la sección diana y  
entre el primer sitio de unión a cebador y la posición del segundo sitio de unión a cebador opcional (véase la figura  
1B). En realizaciones que utilizan dos o más identificadores, por ejemplo, combinaciones locus-alelo, los  
10 identificadores también se encuentran entre los sitios de unión a cebador. En ciertas realizaciones, se pueden  
proporcionar dos identificadores, uno en cada sonda. Una de las sondas puede verse como una sonda de locus, es  
decir, dirigida a un locus específico y contiene un identificador específico de locus. La otra sonda puede ser una  
sonda específica de alelo, es decir, contiene un nucleótido específico de alelo, preferiblemente en su punto de  
15 ligación. La sonda específica de alelo puede contener un identificador específico de alelo. De esta manera, la  
presencia o ausencia de una combinación específica de locus-alelo se identifica por la presencia/ausencia de los  
identificadores combinados. Cuando se prueba para toda la variación alélica de un polimorfismo, sólo se necesita  
una sonda de locus, en combinación con 4 sondas específicas de alelo. En ciertas realizaciones, solo la sonda  
20 específica de alelo puede comprender un identificador que comprende una sección de identificador específica de  
locus y una sección de identificador específica de alelo, por ejemplo en forma de un identificador de locus de 5 pb,  
seguido de un identificador de alelo de 2 pb. O un identificador de muestra de 5 pb seguido de un identificador de  
alelo de 2 pb (en una sonda o en ambas sondas), o y parte de la sección diana para identificar el locus.

**[0060]** También son posibles identificadores basados en la muestra, solos, o en combinación con identificadores de  
locus y/o alelo. Los identificadores de muestra pueden proporcionarse de antemano en la sonda, pero también se  
pueden ligar a las sondas o amplicones restringidos. Los identificadores basados en la muestra también pueden  
25 estar presentes en los adaptadores que se ligan a las sondas o amplicones restringidos.

**[0061]** En base a esta guía están ahora disponibles para el experto en la materia una variedad de posibilidades. Por  
lo tanto, cuando se analizan múltiples muestras en un proceso de secuenciación, uno de los identificadores puede  
usarse para la identificación de la muestra en la pluralidad de muestras. Para fines de identificación, también es  
30 posible utilizar una combinación de la secuencia de la sección específica de diana (que identifica el locus y/o alelo),  
un (o parte de un) identificador que identifica el alelo y/o locus y otro (parte del) identificador que identifica la  
muestra.

**[0062]** En ciertas realizaciones, los identificadores se pueden utilizar para la identificación de la muestra y el alelo, y  
el locus se puede identificar por al menos parte de la secuencia de la sección específica de diana.

**[0063]** En resumen, se pueden introducir identificadores (ID), de forma independiente, en la sección etiqueta de la  
sonda, en un adaptador ligado a las sondas ligadas restringidas o amplicones, introducidos a través de un cebador  
durante una etapa de amplificación y/o en las propias secciones específicas de diana (información de locus (L)/alelo  
(AI)). Los identificadores se pueden situar, de forma independiente, en la información de locus/alelo, entre el sitio de  
40 restricción (RE) y la secuencia de alelo/locus (secciones específicas de diana), entre el adaptador que está ligado a  
las sondas ligadas restringidas o amplicones restringidos. Tanto la introducción como la posición se pueden disponer  
de forma independiente en una o ambas sondas.

**[0064]** En la figura 3B se proporciona una representación esquemática de algunas de las diversas posiciones  
45 individuales de los identificadores en las sondas (ligadas).

**[0065]** En una realización particular preferida de la invención, las sondas comprenden una sección diana y una  
secuencia de reconocimiento y, opcionalmente, una secuencia de unión a cebador. Después de la ligación, las  
sondas ligadas son restringidas o amplificadas, seguido de restricción/digestión para producir sondas ligadas  
50 restringidas (RLP) o amplicones restringidos (RA). A los RLP/RA resultantes, se ligan uno o dos adaptadores que  
contienen uno o más identificadores. Los RLP/RA ligados a adaptador resultantes son secuenciados. La  
combinación alelo/locus de la secuencia diana se identifica por la información de secuencia de la sección diana. La  
muestra se identifica basándose en el identificador o identificadores en el adaptador o adaptadores ligados. Esta es  
una manera eficiente de análisis de múltiples secuencias diana en una muestra, que combina los resultados de una  
55 pluralidad de muestras y se analizan las muestras combinadas. Se ilustra en la figura 1C.

#### Cebadores

**[0066]** La sonda ligada se amplifica usando un par de cebadores correspondientes a los sitios de unión a cebador.  
60 En ciertas realizaciones, el par de cebadores contiene un solo cebador y la amplificación es lineal en lugar de  
exponencial. En ciertas realizaciones, el par comprende un primer cebador que es capaz de hibridarse a la primera  
sección de unión a cebador y es capaz de iniciar la amplificación o elongación. En ciertas realizaciones, el par  
comprende, además, un segundo cebador que es capaz de hibridarse a la segunda sección de unión a cebador y es  
capaz de iniciar la amplificación o elongación. En ciertas realizaciones, el segundo cebador tiene la misma  
65 secuencia que el segundo sitio de unión a cebador en la sonda, es decir, es un cebador inverso. En una realización  
preferida, al menos uno de los cebadores o el mismo par de cebadores se usa para la amplificación de dos o más

sondas ligadas diferentes en una muestra, preferiblemente para la amplificación de todas las sondas ligadas en una muestra. Dicho cebador se conoce a veces como cebador universal ya que estos cebadores son capaces de cebar la amplificación de todas las sondas conectadas que contienen el sitio de unión a cebador universal correspondiente y por consiguiente de todas las sondas ligadas que contienen el sitio de unión a cebador universal. Los diferentes cebadores que se utilizan en la amplificación en la etapa (i) son preferiblemente esencialmente iguales en la eficiencia de hibridación y cebado. Por lo tanto, los cebadores en una muestra difieren preferiblemente menos de 20, 15, 10, 5, o 2 grados Celsius en la temperatura de fusión. Esto se puede lograr tal como se describe en el presente documento en otra parte para las secciones específicas de diana de las sondas de oligonucleótidos. A diferencia de la secuencia de las secciones específicas de diana, la secuencia de los cebadores no está dictada por la secuencia diana. Por lo tanto, las secuencias de cebadores se pueden diseñar de manera convenientemente mediante el ensamblaje de la secuencia de tetrámeros de nucleótidos, en los que cada tetrámero contiene una A, T, C y G o por otros modos que aseguren que el contenido de G/C y la temperatura de fusión de los cebadores son idénticos o muy similares. La longitud de los cebadores (y los sitios de unión a cebador correspondientes en la sección etiqueta de la segunda sonda) es preferiblemente de al menos 12, 15 ó 17 nucleótidos y preferiblemente no más de 25, 30, 40 nucleótidos. En una cierta realización, al menos dos de las segundas sondas de oligonucleótidos que son complementarias a al menos dos secuencias diana diferentes en una muestra comprenden cada una una sección etiqueta que comprende una sección de unión a cebador que es complementaria a una única secuencia de cebador.

[0067] En ciertas realizaciones, para garantizar una eficiencia de cebado similar en comparación con otros cebadores que albergan la misma secuencia de anclaje, el cebador puede comprender una secuencia de anclaje 3', preferiblemente una secuencia de anclaje de 2 pb, preferiblemente una secuencia de anclaje GC. Habitualmente, la secuencia de unión a cebador correspondiente también albergará el complemento de la misma.

[0068] Por lo tanto, preferiblemente al menos uno de los primer y segundo cebadores en un par de cebadores se usa para la amplificación de sondas ligadas correspondientes a al menos dos secuencias diana diferentes en una muestra, más preferiblemente para la amplificación de sondas conectadas correspondientes a todas las secuencias diana en una muestra. Preferiblemente sólo se utiliza un primer cebador único y en algunas realizaciones solamente se utiliza un único primer y un único segundo cebador para la amplificación de todas las sondas (ligadas) conectadas. La utilización de cebadores comunes para la amplificación de múltiples fragmentos diferentes normalmente es ventajoso para la eficacia de la etapa de amplificación. Las sondas ligadas obtenidas a partir de la ligación de las secciones de sonda hibridadas se amplifican, utilizando un par de cebadores, que consiste preferiblemente en un par de cebadores para cada una de las sondas ligadas en la muestra. El par de cebadores comprende cebadores que son complementarios a secuencias de unión a cebador que están presentes en la sonda ligada. Un par de cebadores habitualmente comprende un primer y al menos un segundo cebador, pero puede consistir en solamente un único cebador que ceba en ambas direcciones. Se han obtenido excelentes resultados usando cebadores que son conocidos en la técnica como cebadores AFLP, tal como se describe, entre otros, en EP534858 y en Vos et al., Nucleic Acid Research, 1995, vol. 23, 4407-4414 y se discute en más detalle en este documento a continuación.

#### 40 Secuenciación de alto rendimiento

[0069] La secuenciación o cribado de alto rendimiento, a menudo abreviada como HTS, es un procedimiento para la experimentación científica especialmente relevante para los campos de la biología y la química. A veces se refiere también como Next Generation Sequencing y se describe ampliamente en Janitz Ed. Next Generation Genome sequencing, Wiley VCH, 2008. A través de una combinación de la robótica moderna y otros equipos de laboratorio especializados, se permite a un investigador el cribado de manera efectiva de grandes cantidades de muestras de forma simultánea.

[0070] Se prefiere que la secuenciación se realice utilizando procedimientos de secuenciación de alto rendimiento, tales como los procedimientos descritos en los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007 y WO 2005/003375 (todos a nombre de de 454 Life Sciences, ahora Roche Diagnostics), por Seo et al. (2004) Proc. Natl. Acad. Sci. USA 101: 5488-93, y tecnologías de Helicos, Solexa, US Genomics, etcétera.

#### 55 [0071] Secuenciación de alto rendimiento basada en tecnologías de Roche GS FLX

[0072] En ciertas realizaciones, se prefiere que la secuenciación se lleve a cabo utilizando el aparato y/o procedimiento dado a conocer en los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007 y WO 2005/003375 (todos a nombre de 454 Life Sciences, ahora Roche Diagnostics). La tecnología descrita permite la secuenciación de 40 millones de bases en un solo proceso y es 100 veces más rápido y más barato que tecnología competidora. La tecnología de secuenciación consiste aproximadamente de 5 etapas:

- 1) fragmentación de ADN y ligación de adaptadores específicos para crear una biblioteca de ADN de cadena sencilla (ADNss);
- 2) hibridación de ADNss a microesferas, emulsificación de las microesferas en microrreactores de agua en aceite y realización de la PCR en emulsión para amplificar las moléculas de ADNss individuales sobre microesferas;

- 3) selección de enriquecimiento para las microesferas que contienen moléculas de ADNss amplificadas en su superficie;
- 4) deposición de microesferas que transportan ADN en una placa PicoTiter (TM); y
- 5) secuenciación simultánea en más de 1.000.000 de pocillos mediante la generación de una señal luminosa de pirofosfato. El procedimiento se explicará en más detalle a continuación.

**[0073]** En una realización preferida, la secuenciación comprende las etapas de:

- (a) hibridación de fragmentos adaptados a microesferas, estando cada microesfera hibridada con un solo fragmento adaptado;
- 10 (b) emulsionar las microesferas en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una única microesfera;
- (c) cargar las microesferas en pocillos, comprendiendo cada pocillo una única microesfera; y generar 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 puede incluir un identificador adicional y más secuencias para
- 15 hibridarse a una microesfera, una región de cebadores de secuenciación y una región de cebadores de PCR. De este modo, se obtienen fragmentos adaptados. En una primera etapa, los fragmentos adaptados se hibridan a microesferas, hibridándose cada microesfera con un único fragmento adaptado. Al conjunto de fragmentos adaptados, se añaden microesferas en exceso para garantizar la hibridación de un único fragmento adaptado por microesfera para la mayoría de las microesferas (distribución de Poisson). En una siguiente etapa, las microesferas se emulsionan en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una sola microesfera. Los reactivos de PCR están presentes en los microrreactores de agua en aceite lo que permite que se lleve a cabo una reacción de PCR dentro de los microrreactores. Posteriormente, los microrreactores se rompen, y las microesferas que comprenden ADN (microesferas positivas en ADN) se enriquecen. En una siguiente etapa, las microesferas se cargan en pocillos, comprendiendo cada pocillo una única microesfera. Los pocillos son
- 20 preferiblemente parte de una placa PicoTiter (TM) que permite la secuenciación simultánea de una gran cantidad de fragmentos. Después de la adición de microesferas que transportan enzimas, se determina la secuencia de los fragmentos usando pirosecuenciación. En etapas sucesivas, la placa PicoTiter (TM) y las microesferas, así como las microesferas con enzimas en las mismas, se someten a diferentes desoxirribonucleótidos en presencia de reactivos de secuenciación convencionales, y tras la incorporación de un desoxirribonucleótido se genera una señal de luz
- 25 que se registra. La incorporación del nucleótido correcto generará una señal de pirosecuenciación que se puede detectar.

**[0074]** La pirosecuenciación en sí es conocida en la técnica y se describe, entre otros, en [www.biotagebio.com](http://www.biotagebio.com), [www.pyrosequencing.com/section technology](http://www.pyrosequencing.com/section%20technology). La tecnología se aplica adicionalmente, por ejemplo, en los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007 y WO 2005/003375 (todos a nombre de 454 Life Sciences, ahora Roche Diagnostics). En la presente invención, las microesferas están equipadas preferiblemente con secuencias para cebador (de unión) y/o secciones de agarre o partes de las mismas que son capaces de unirse a los amplicones o las sondas ligadas, según sea el caso. En otras realizaciones, las sondas o los cebadores usados en la amplificación en emulsión están equipadas con secuencias que permiten la unión de los amplicones o las sondas ligadas a las microesferas con el fin de permitir la posterior amplificación en emulsión seguido por secuenciación. Los amplicones o sondas ligadas secuenciadas revelarán la identidad del identificador o identificadores y, opcionalmente parte de la secuencia diana y, de este modo, de la presencia o ausencia de la secuencia diana en la muestra.

#### 45 Secuenciación de alto rendimiento basada en tecnologías Illumina Genoma Analyzer/HiSeq/Miseq

**[0075]** Uno de los procedimientos para la secuenciación de alto rendimiento se describe, entre otros, en WO0006770, WO0027521, WO0058507, WO0123610, WO0157248, WO0157249, WO02061127, WO03016565, WO03048387, WO2004018497, WO2004018493, WO2004050915, WO2004076692, WO2005021786, WO2005047301, WO2005065814, WO2005068656, WO2005068089, WO2005078130. En esencia, el procedimiento comienza con fragmentos ligados a adaptadores de ADN. El ADN a utilizar en la tecnología de secuenciación descrita en la presente es sondas ligadas restringidas (RLP) o amplicones restringidos (RA). El ADN ligado a adaptador se hibrida al azar a un campo denso de cebadores que se unen a una superficie sólida, habitualmente en una celda de flujo. Después de la elongación, el extremo de los fragmentos recién formados se hibrida a un cebador que está unido al soporte sólido en las proximidades del fragmento. Este cebador se extiende en presencia de nucleótidos y polimerasas para proporcionar fragmentos ds. Los cebadores se extienden en presencia de nucleótidos y polimerasas en la denominada amplificación por puente en fase sólida para proporcionar fragmentos de cadena doble. La desnaturalización y la repetición de la amplificación por puente en fase sólida dan lugar a densos grupos de fragmentos amplificados distribuidos sobre la superficie. La secuenciación se inicia mediante la adición de cuatro nucleótidos terminadores reversibles marcados de manera diferente, cebadores y polimerasa a la celda de flujo. Después de la primera ronda de la extensión del cebador, 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. A continuación, se determina la identidad de la segunda base de la misma manera y, de este modo, continúa la secuenciación. En la presente invención, las sondas ligadas o los amplicones se unen a la superficie a través de la secuencia de unión a cebador, la secuencia del cebador o en algunas realizaciones, la sección de agarre o una combinación de las mismas. La secuencia se determina tal como se ha indicado, incluyendo

la secuencia identificadora y la secuencia diana asociada y se identifica su presencia o ausencia.

#### Secuenciación de alto rendimiento basada en tecnologías de Ion Torrent

5 **[0076]** Uno de los procedimientos para la secuenciación de alto rendimiento se describe, entre otros, en US2010137143, WO2010008480, US2010282617, WO2009158006, WO2010016937, WO2010047804, US2010197507, US2010304982, WO2010138182, WO2010138186, WO2010138187, WO2010138188. El procedimiento se basa en la fragmentación de ADN de muestra, la ligación de adaptadores, generación de cadenas de ADNss, captura de las cadenas en microsferas seguido por PCR en emulsión y la posterior hibridación de un oligonucleótido para cebar la síntesis de ADN. En esencia, se basa en una matriz para la medición del protón liberado que se produce cuando dos dNTPs están acoplados el uno al otro. Cada vez que se agrega un nucleótido, se libera un protón. La medición de la liberación del protón es una medida de la incorporación exitosa de los nucleótidos en el oligonucleótido.

15 **[0077]** La detección de un nucleótido específico en una cadena de ADN en crecimiento se produce en el interior de un pocillo fabricado de un chip semiconductor específico. El chip secuenciación captura mediciones de voltaje de la liberación directa de iones hidrógeno después de la polimerización del ADN. El número total de mediciones independientes, o lecturas de secuencia, es función del número de sensores y los pocillos fabricados que un chip contiene.

#### Adaptadores

20 **[0078]** En algunas de las realizaciones de la presente invención, uno o más adaptadores se ligan a uno o ambos extremos de los amplicones restringidas o sondas ligadas restringidas.

25 **[0079]** Los adaptadores, tal como se usan aquí, son moléculas cortas de ADN de doble cadena con un número limitado de pares de bases, por ejemplo, de aproximadamente 10 a aproximadamente 50 pares de bases de longitud, que están diseñados de tal manera que se pueden ligar a los extremos de fragmentos de restricción. Los adaptadores se componen generalmente de dos oligonucleótidos sintéticos que tienen secuencias de nucleótidos que son parcialmente complementarias entre sí. Cuando se mezclan los dos oligonucleótidos sintéticos en solución bajo las condiciones apropiadas, se hibridarán entre sí formando una estructura de doble cadena. Después de la hibridación, un extremo de la molécula adaptadora está diseñado de tal manera que es 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 puede ligar, pero esta necesidad no es el caso (adaptadores dobles ligados).

35 **[0080]** Preferiblemente, los adaptadores se ligan antes de la etapa de secuenciación. A las sondas ligadas restringidas o a los amplicones restringidos se ligan adaptadores que se pueden utilizar en una etapa de amplificación posterior que forma parte de la etapa de secuenciación (PCR en emulsión o amplificación por puente), por ejemplo para hibridarse a un portador (tal como una microesfera) utilizado en la tecnología de secuenciación y para proporcionar funcionalidades adicionales que pueden ser útiles durante la etapa de secuenciación, tales como sitios de unión a cebador para facilitar una etapa de amplificación como parte del protocolo de secuenciación. Dichos adaptadores se refieren generalmente como "adaptadores de secuenciación" y su diseño y funcionalidad se ejemplificarán aquí a continuación. Ejemplos de dichos adaptadores de secuenciación son conocidos en la técnica como adaptadores P5 y P7 y se utilizan en la tecnología Illumina. Otras tecnologías emplean adaptadores conceptualmente similares.

40 **[0081]** De este modo, en ciertas realizaciones, un adaptador (doble cadena) se liga al extremo del fragmento de restricción proporcionado por la endonucleasa de restricción. El adaptador está construido de tal manera que es ligable al extremo del fragmento de restricción. Si el extremo del fragmento de restricción es romo, ya sea por el pulido o por la endonucleasa, el adaptador también tiene preferiblemente extremos romos. El adaptador puede construirse o diseñarse de tal manera que sólo una cadena es ligable al extremo del fragmento de restricción, mientras que la otra cadena del adaptador puede diseñarse de manera que no se liga, por ejemplo a través del uso de un nucleótido no fosforilado. Si el extremo del fragmento de restricción está escalonado, se prefiere utilizar un adaptador escalonado que contenga preferiblemente al menos un extremo ligable. Un extremo ligable en este contexto es un extremo que es al menos complementario a los restos del sitio de restricción de la endonucleasa de restricción. Si se utiliza una enzima de restricción en las sondas ligadas o en las sondas ligadas amplificadas, se puede utilizar un adaptador. También es posible utilizar múltiples adaptadores cuando se utiliza una enzima de restricción en las sondas ligadas o en las sondas ligadas amplificadas. El uso de múltiples adaptadores puede proporcionar una funcionalidad adicional, por ejemplo en la separación (selección) de parte de las sondas ligadas o en las sondas ligadas amplificadas como parte de una reducción de la complejidad. En ciertas realizaciones, cuando se usan dos (una primera y una segunda) endonucleasas de restricción, dos (un primer y un segundo) adaptadores se pueden ligar a los respectivos extremos de los fragmentos de restricción.

50 **[0082]** En ciertas realizaciones, el adaptador puede ser un adaptador en forma de Y (a veces llamado un "adaptador de horquilla"). Un adaptador en forma de Y puede tener un extremo escalonado o un extremo romo. En general, un adaptador en forma de Y se produce a partir de dos fragmentos de ADN de cadena sencilla. Los dos fragmentos de

ADN de cadena sencilla contienen cada uno una sección en un extremo de la cadena que es complementaria a la otra, de tal manera que las secciones son capaces de hibridar las dos cadenas entre sí. Los fragmentos de ADN de cadena sencilla contienen cada uno una sección adicional que no es complementaria a la otra y que no se hibrida. El extremo complementario permite al adaptador en forma de Y ligarse a los extremos de los amplicones restringidos o sondas ligadas restringidas. El extremo complementario puede ser de cualquier longitud adecuada y puede ser de 1-50 nucleótidos de longitud. El uso de un adaptador en forma de Y permite la introducción de dos cadenas diferentes de ADN usando sólo un tipo de adaptador. Una representación esquemática del mismo se proporciona en la figura 5A.

5  
10 **[0083]** Los adaptadores pueden contener, además, identificadores y el adaptador en forma de Y puede contener identificadores diferentes en los dos brazos del adaptador en forma de Y.

15 **[0084]** En cierta realización, el adaptador en forma de Y puede diseñarse de tal manera que un adaptador en forma de Y es capaz de ligarse a ambas cadenas de un fragmento de restricción, mientras que al mismo tiempo se evita la autoligación del adaptador en forma de Y. En esta realización, el saliente del fragmento de restricción está parcialmente relleno antes de la etapa de ligación, permitiendo así que el adaptador en forma de Y se ligue sólo al fragmento y no a otros adaptadores en forma de Y. Una representación esquemática del mismo se proporciona en la figura 5B.

20 **[0085]** Como alternativa, los adaptadores también pueden ser del tipo conocido como "adaptadores de horquilla" que son capaces de hibridarse y ligarse a ADN de cadena sencilla, proporcionando ADN parcialmente de doble cadena.

25 **[0086]** En otra realización relacionada con el uso de sondas ligadas de cadena sencilla en la etapa en la que las sondas ligadas son restringidas antes de la etapa de secuenciación, se contempla el uso de nucleasas de corte de cadena sencilla. Como alternativa, se puede proporcionar un oligonucleótido que proporcionará una cadena ds local que posteriormente se puede cortar usando la endonucleasa de restricción. Los adaptadores también pueden ser del tipo conocido como "adaptadores de horquilla" que son capaces de hibridarse y ligarse al ADN de cadena sencilla, proporcionando ADN parcialmente de doble cadena que, posteriormente, se puede cortar con una endonucleasa. Una representación esquemática de las dos últimas variantes se proporciona en la figura 5C.

30 **[0087]** Las tecnologías de secuenciación descritas actualmente contienen algunas variaciones en sus protocolos de secuenciación. El uso de estas variaciones de secuenciación puede tener influencia en el diseño de las diferentes sondas y cebadores usados en la presente invención, la forma en que se obtienen los datos de secuencia y la calidad, fiabilidad y la cantidad de datos generados.

35 Secuenciación de lectura única unidireccional

40 **[0088]** Con una secuenciación de lectura única, la sonda ligada restringida o el amplicón restringido se ligan a uno o dos adaptadores (adaptadores de secuenciación) y se secuencian en una dirección utilizando un cebador. La secuencia de nucleótidos que se somete finalmente a secuenciación se indica comúnmente en esta descripción como el "fragmento de secuenciación".

45 **[0089]** Esta realización se representa esquemáticamente en la Figura 4A. En esta realización, el fragmento ligado al adaptador de secuenciación se secuencia empezando a partir del cebador (cebador de secuenciación, SEQ PR) en el fragmento, secuenciando de ese modo al menos parte de la sección etiqueta que quedaba después de la restricción, al menos parte de la secuencia diana (es decir, secciones específicas de diana). Cualquier identificador situado 3' del cebador de secuenciación se secuenciará a lo largo como lo hará (parte de) la secuencia diana. Este identificador puede estar presente en el adaptador de secuenciación y/o en la sección etiqueta que quedaba después de la restricción. La secuencia diana puede identificarse a continuación por el identificador o identificadores o la secuencia diana o una combinación de los mismos.

50 Secuenciación de doble etiquetado de lectura única unidireccional

55 **[0090]** Con doble etiquetado de lectura única unidireccional, (doble cebado de lectura única unidireccional) el amplicón restringido o sonda ligada restringida se ligan a uno o dos adaptadores de secuenciación y se secuencian en una dirección, pero con dos cebadores (SEQ PR1, SEQ PR2). Esta realización se representa esquemáticamente en la Figura 4B. En esta realización, idéntica a la secuenciación de lectura única, el fragmento de secuenciación se secuencia empezando a partir del cebador de secuenciación SEQ PR1 en el fragmento, secuenciando de ese modo al menos parte de la sección etiqueta que permanecía después de la restricción, al menos parte de la secuencia diana (es decir, secciones específicas de diana). En esta realización, la lectura de secuencia resultante de esta etapa se indica como "lectura larga". El segundo cebador (SEQ PR2) se dirige a una segunda parte de la sonda ligada restringida o el amplicón restringido ligados a un adaptador o adaptadores de secuenciación y amplifica una segunda parte de los mismos, habitualmente indicado como "lectura corta".

65 **[0091]** Cuando se realiza el reagrupamiento de los fragmentos de secuenciación para la secuenciación, el segundo cebador de secuenciación también puede dar lugar a una lectura larga (véase la figura 4C). En el reagrupamiento,

los fragmentos de secuenciación se someten a un protocolo en el que se hibridan al portador utilizando el otro adaptador de secuenciación (que da lugar a un puente del fragmento de secuenciación en el portador y la posterior deshibridación del primer adaptador de secuenciación para el portador en el que se hibridaron y unieron). El resultado es que la orientación del fragmento frente al portador se desplaza (reagrupado) y la secuenciación se puede realizar de nuevo. Un ejemplo de dicho reagrupamiento y la secuenciación desde ambos lados (secuenciación de extremos emparejados) se describe en Bentley et. Nature 2008, 456, 53-59. Habiendo realizado un reagrupamiento se puede, por tanto, llegar a dos lecturas largas (Long 1, Long 2).

#### Secuenciación de doble etiquetado bidireccional

**[0092]** Con la secuenciación de doble etiquetado bidireccional (secuenciación de doble de cebado bidireccional), representado en la Figura 4D, el fragmento de secuenciación se secuencia usando secuenciación de extremos emparejados en el que el fragmento se secuencia desde ambos lados.

**[0093]** En esta realización, un tercer identificador puede estar presente en el fragmento de secuenciación que puede ser dirigido mediante el uso de un cebador que tiene una orientación inversa que puede resultar en una segunda lectura larga (Long 2). El tercer identificador se puede dirigir (secuenciar) mediante el segundo cebador o un tercer cebador que se dirige específicamente para identificar el tercer identificador (ejemplificado en la figura por una superposición diferente, pero éste no tiene por qué ser el caso). En una realización adicional, el cebador con la orientación inversa no se utiliza para identificar el identificador, de manera que el tercer identificador puede omitirse y la etapa de secuenciación en la dirección inversa sirve para proporcionar datos de la secuencia del fragmento de secuenciación.

#### Secuenciación de extremos emparejados

**[0094]** Tal como se utiliza aquí, "secuenciación de extremos emparejados" en un procedimiento que se basa en secuenciación de alto rendimiento, en particular basado en las plataformas que actualmente son comercializadas por Illumina y Roche. Illumina ha dado a conocer un módulo de hardware (el módulo PE) que se puede instalar en el secuenciador existente como una actualización, que permite la secuenciación de ambos extremos de la plantilla, generando de ese modo lecturas de extremos emparejados. En particular, es preferente utilizar la secuenciación de extremos emparejados, en particular, utilizando la tecnología de Roche o Illumina, en los procedimientos de acuerdo con la presente invención. Ejemplos de secuenciación de extremos emparejados se describen por ejemplo en US20060292611 y en publicaciones de Roche (secuenciación 4540).

#### Secuenciación de pares acoplados ("mate pair")

**[0095]** La secuenciación de pares acoplados ("mate pair") es una variante de la secuenciación de extremos emparejados en el que los extremos se acoplan. Los fragmentos de ADN a secuenciar, tales como productos de ligación tratados con endonucleasa de restricción (o amplicones a partir de los mismos) se circularizan, se fragmentan y los fragmentos que contienen los extremos del ADN original se secuencian posteriormente obteniendo de este modo información de secuencia a partir de ambos extremos del ADN original en una etapa de secuenciación. La secuenciación de pares acoplados se puede aplicar a cualquiera de los fragmentos de secuenciación que se describen en este documento. Véase también [www.illumina.com](http://www.illumina.com) para ejemplos de secuenciación de pares acoplados.

**[0096]** La información detallada sobre el concepto de secuenciación de pares acoplados se proporciona en la figura 6. A modo de ejemplo se usa uno de los productos de ligación, pero el principio de secuenciación de pares acoplados se aplica a cualquiera de los fragmentos de secuenciación de la presente invención, independientemente de los elementos presentes en los fragmentos de secuenciación. Para ilustrar esto, se generaliza un fragmento de secuenciación en una línea sólida que representa un fragmento de secuenciación de ADN. El fragmento se circulariza y se fragmenta (a través de cizalla (corte) o restricción). A los extremos del fragmento, se ligan adaptadores de secuenciación y la cadena de ADN resultante se somete a secuenciación, preferiblemente desde ambos extremos (emparejados).

**[0097]** A lo largo de esta memoria, figuras y reivindicaciones adjuntas las nociones "primera" y "segunda" se utilizan para distinguir entre elementos, tales como las sondas, adaptadores, cebadores, etc. utilizados en el ensayo y sus respectivos componentes. Las nociones "primeros" y "segunda" no se utilizan en el presente documento como sumas, es decir, no es para que sólo pueda haber un segundo componente cuando también hay un primer componente. Por razones de coherencia y facilidad de referencia estas nociones también se utilizan cuando la propia realización no está constituida de dos sondas o dos componentes. Por ejemplo, una sonda circularizable, siendo sólo una sonda, todavía contiene una primera y segunda sección específica de diana. Asimismo, en la figura 1, por ejemplo, ya sea una de la primera y segunda sonda puede contener un identificador. En el caso de la primera sonda éste se representa como el primer identificador y en el caso de la segunda sonda éste se representa como el segundo identificador. En caso de que la segunda sonda contiene un identificador y la primera sonda no, este identificador puede contemplarse en esta solicitud como el segundo identificador sin implicar la existencia de un primer identificador.

Ejemplo de detección de polimorfismo de un solo nucleótido (SNP) en Melón

1. Aislamiento de ADN

5 **[0098]** Se aisló el ADN genómico dos descendientes F2 de una población segregante de melón Charantais de material de hojas usando un procedimiento CTAB modificado descrito por Stuart y Via (Stuart, C.N., Jr y Via, L.E. (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques. Vol. 14, 748-750). Las muestras de ADN se diluyeron hasta una concentración de 100 ng/μl en TE (Tris-HCl 10 mM pH 8,0, EDTA 1 mM) y se almacenaron a -20°C.

2. Amplificación de ADN

15 **[0099]** Con el fin de tener suficiente ADN disponible para múltiples pruebas, el ADN aislado se amplificó utilizando el kit de amplificación de ADN Illustra GenomiPhi v2 (GE Healthcare) de acuerdo con las especificaciones del fabricante.

3. Selección de los SNP de melón

20 **[0100]** Los SNP de melón se seleccionaron de una colección que previamente se había incorporado en ensayos BeadXpress (Illumina) y se utilizaron para determinar el genotipo de varias de las muestras antes mencionadas. Se seleccionaron un total de 23 SNP que contenían una amplia variedad de SNP, véase la tabla 1. Los SNP sirven como ejemplos y no limitan el concepto general de la tecnología de la invención.

25 TABLA 1: SNP seleccionados de melón

SNP	SNP	Alelos	SNP	SNP	Alelos
1	SBG0004r	A/G	13	SBG0023	T/C
2	SBG0005	T/C	14	SBG0025	A/G
3	SBG0008	A/G	15	SBG0026	T/C
4	SBG0009	C/G	16	SBG0027	T/C
5	SBG0010	C/T	17	SBG0028	T/C
6	SBG0013	A/C	18	SBG0030	A/T
7	SBG0014	A/C	19	SBG0033	T/C
8	SBG0015	A/T	20	SBG0036	T/C
9	SBG0016	T/G	21	SBG0037	T/C
10	SBG0018	T/C	22	SBG0039	A/G
11	SBG0021	C/G	23	SBG0040	A/T
12	SBG0022	A/G			

4. Diseño de sondas de oligonucleótidos para la reacción de ligación de oligonucleótidos

30 **[0101]** Se diseñaron sondas de oligonucleótidos (orientación 5'-3') utilizando procedimientos comunes basados en la secuencia conocida de los loci y se seleccionaron para discriminar los alelos de SNP para cada uno de los 23 loci descritos en la Tabla 1. Se incluyeron las regiones de unión a cebador de PCR. Todas las sondas fueron fosforiladas en el extremo 5'. Para cada SNP, se diseñaron dos sondas de alelo que contenían el alelo específico y una sonda inversa. La fosforilación de las sondas inversa es funcional, mientras que la fosforilación de las sondas específicas de alelo es meramente un resultado de la reducción de costes.

35 **[0102]** Todos los oligonucleótidos fueron adquiridos de Metabion, Martinsried, Alemania. La concentración de los oligonucleótidos se ajustó a 1 μM.

40 **[0103]** Se preparó una mezcla de 4x sonda combinando 1 μl de cada sonda de alelo (= 46x) y 2 μl de cada sonda inversa (= 23x). La mezcla de 4x sonda se diluyó 4 veces con agua MilliQ para obtener una mezcla de 1x sonda.

5. Diseño de los cebadores de amplificación de PCR

**[0104]** Las secuencias de los cebadores utilizados para la amplificación por PCR de los productos de ligación de oligonucleótidos eran complementarios a la regiones de unión a cebador de PCR incorporadas en las sondas de ligación descritas en "4. Diseño de sondas de oligonucleótidos para la reacción de ligación de oligonucleótidos". Las secuencias de los cebadores de PCR se derivan de las secuencias de adaptadores utilizadas en el proceso AFLP tal como se describe por Zabeau y Vos, 1993: Selective restriction fragment amplification; a general method for DNA fingerprinting. EP 0534858-A1, B1; patente US 6045994) y Vos et al (Vos, P., Högers. R., Bleeker. M., Reijans. M., van de Lee. T., Hornes, M., Frijters. A., Pot. J., Peleman J., Kuiper M. et al (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acids Res, 21, 4407-4414). Específicamente, el extremo 3' de las secuencias de los cebadores se modificó para albergar una parte del sitio de reconocimiento de la enzima de restricción para EcoRI (sonda específica de alelo) o HindIII (sonda inversa (= específica de locus)).

#### 6. Tampones y reactivos

**[0105]** La concentración de los tampones fue: tampón de ligación de oligonucleótidos multiplex (10x): Tris-HCl 200 mM pH 7,6, KAc 250 mM, MgAc 100 mM, NAD 10 mM, ditiotreitól 100 mM, Triton-X100 al 1%. Tampón de PCR (10x): Tris-HCl 100 mM pH 8,3, KCl 500 mM, MgCl<sub>2</sub> 15 mM, gelatina al 0,01% (p/v). Tampón de ligación de restricción (5x): 5x tampón de ADNasa (Affymetrix), ditiotreitól 25 mM, BSA 250 µg/ml. Tampón de elución Minelute incluyendo Tween: Tris 10 mM pH 8,5, Tween X-100 al 0,1%.

#### 7. Reacción de ligación de oligonucleótidos multiplex y amplificación

**[0106]** Las reacciones de ligación se llevaron a cabo por duplicado para cada una de las dos muestras de ADN aislado de la siguiente manera: se combinaron 100 a 200 ng de ADN genómico (amplificado) con 1 µl de tampón de ligación de oligonucleótidos multiplex, 4 unidades de ADN ligasa Taq (New England Biolabs), 0,4 µl de mezcla de 1x Sonda y agua MilliQ hasta un total de 10 µl. La mezcla de reacción se incubó durante 2 minutos a 94°C, seguido por 4 horas a 60°C. Las reacciones se mantuvieron a 4°C hasta su uso posterior. Las reacciones de ligación se diluyeron 4 veces con 1x tampón de ligación de oligonucleótidos multiplex. La amplificación de los productos de ligación se llevó a cabo de la siguiente manera: 10 µl de reacción de ligación diluida 4 veces, 30 ng de cada cebador (E00LF y H00LR), 0,2 µl de mezcla 20 mM de cada dNTP, 2 µl de 10x tampón de PCR, 0,4 unidades de AmpliTaq-Gold (Applied Biosystems) y agua MilliQ hasta un total de 20 µl. Las reacciones de amplificación se realizaron por duplicado. El perfil de termociclado se realizó en un PE9700 (Perkin Elmer Corp.) con un bloque de oro o plata usando las siguientes condiciones: Etapa 1: Incubación Pre PCR: 12 minutos a 94°C. Etapa 2: Desnaturalización: 30 segundos a 94°C; hibridación; 30 segundos a 65°C en el primer ciclo. En cada ciclo siguiente esta temperatura de hibridación se redujo en un 0,7°C. Extensión: Extensión de 1 minuto a 72°C.

**[0107]** El número total de ciclos es 13 Etapa 3: desnaturalización: 30 segundos a 94°C. Hibridación; 30 segundos a 56°C. Extensión: extensión de 1 minuto a 72°C. Número total de ciclos es de 23.

**[0108]** Etapa 4: Extensión: 7 minutos a las 72°C. Las reacciones se mantuvieron a 4°C hasta su uso posterior. Los productos de amplificación (20 µl) se purificaron usando el kit MinElute (Qiagen) y se eluyeron en 10 µl de agua MilliQ.

#### 8. Reacción de ligación de restricción

**[0109]** Los productos de amplificación de la etapa 7 (7 µl) fueron digeridos con las enzimas de restricción EcoRI (20 unidades) e HindIII (20 unidades) en un volumen total de 40 µl que contenía 1x tampón de ligación de restricción e incubación a 37°C durante 2 horas.

**[0110]** Se ligaron adaptadores a los productos de digestión a través de la adición de 5 pmol de adaptador genérico de HindIII, 5 pmol de adaptador de EcoRI que contenía ID de muestra, 0,1 µl de ATP 100 mM, 2 µl de 5x tampón de ligación de restricción, 1 unidad de ADN ligasa T4 en un volumen total de 10 µl e incubación a 37°C durante 3 horas.

**[0111]** La composición de la secuencia del adaptador de EcoRI que contenía un ID de la muestra era tal que la cadena superior contenía un ID de muestra situado en 3' (5 nt) y la cadena inferior contenía un ID de muestra situado en 5' (5 nt) que estaba en complemento inverso de la correspondiente cadena superior. Los ID de muestra diferían entre la muestra 1 y muestra 2.

**[0112]** El extremo 3' del adaptador de cadena inferior fue modificado con un grupo amino.

**[0113]** Los adaptadores se prepararon mezclando cantidades iguales (pmol) de cada oligo (superior e inferior) en un tubo eppendorf. La concentración final de los adaptadores era de 50 µM.

**[0114]** La composición de la secuencia del adaptador de HindIII se diseñó y sintetizó de manera análoga al adaptador de EcoRI, con y sin un identificador.

#### 9. Amplificación de los productos de ligación de restricción

**[0115]** Los productos de ligación de restricción se diluyeron 10 veces en agua MilliQ. Se mezclaron cinco  $\mu$ l de producto de ligación de restricción diluida 10 veces con 5 ng de cebador directo, 30 ng de cebador inverso, 0,2 20 mM de cada dNTP, 2  $\mu$ l de 10x tampón PCR, 0,08  $\mu$ l de ADN polimerasa AmpliTaq 5 U/ $\mu$ l (Applied Biosystems) y 12,02  $\mu$ l de agua MilliQ. La mezcla de reacción se colocó en un termociclador (PE9700, bloque de oro o plata) y se aplicó el siguiente perfil: preincubación: 2 minutos 72°C, Ciclado 50 veces: 30 segundos 94°C, 2 minutos 58°C, 2 minutos 72°C. De cada reacción de amplificación se agruparon 5  $\mu$ l de los cuales 130  $\mu$ l se purificaron usando una columna de MinElute (Qiagen). El producto purificado se eluyó en 30  $\mu$ l de tampón de elución con la adición de Tween.

#### 10. Secuenciación de los productos de amplificación

**[0116]** La secuenciación de los productos amplificados de la etapa 9 se realizó en el Genome Analyzer II (Illumina) que es una plataforma de secuenciación por síntesis y utiliza la tecnología Clonal Single Molecule Array (CSMATM) que utiliza diferentes protocolos de secuenciación incluyendo secuenciación de lectura única, secuenciación de doble etiquetado de lectura única unidireccional, secuenciación de doble etiquetado bidireccional, secuenciación de extremos emparejados y secuenciación de pares acoplados.

#### 11. Procesado de datos y determinación del genotipo

**[0117]** Las lecturas de secuencia obtenidas se cribaron para la presencia de los identificadores. Permanecieron un total de 1.644.183 lecturas. Todas las etiquetas de ID se detectaron con un número promedio de lecturas por muestra de 411.046. El número de lecturas por muestra varió de 308.105 (Muestra 1) a 603.889 (Muestra 3). Se realizó un control de calidad adicional en las lecturas que contenían las etiquetas de ID de la muestra. Esto incluía la presencia del sitio de reconocimiento de EcoRI, la ausencia de lecturas que contenían homopolímeros (definido como tramos contiguos del mismo nucleótido sobre más de 20 posiciones), la ausencia de lecturas con emparejamiento positivo frente a la base de datos del cloroplasto de NCBI, la ausencia de lecturas que contienen 'N' en la secuencia y la ausencia de lecturas con puntuación baja de calidad (QS promedio <15) en los primeros 50 nucleótidos de la lectura. El número de lecturas eliminadas por muestra variaba de 12.445 (Muestra 4) a 28.447 (Muestra 1), con un promedio de 18.769. El número de lecturas eliminadas por muestra representaba un pequeño porcentaje del número total de lecturas (promedio de 4,6%). Como consecuencia, el porcentaje promedio de lecturas que pasaban los filtros de control de calidad fue alta (95,4%). Las lecturas que pasaban el control de calidad se utilizaron como datos de entrada para determinar los genotipos de las muestras para cada uno de los SNP. Este proceso implicó la alineación de las lecturas a las secuencias de referencia de los loci utilizando el software BWA, el procesamiento de los datos de salida con SAMtools (incluyendo la clasificación, fusión e indexación), la determinación de las apariciones de los alelos en las muestras y la determinación de los genotipos para cada muestra basada en las relaciones de las apariciones de los alelos. SAMtools está disponible a través de <http://samtools.sourceforge.net>. Cuando una posición degenerada en las secuencias de referencia estaba presente, se utilizó una clasificación alfabética para la base para reemplazar la posición ambigua.

**[0118]** Se detectaron SNP en todas las 23 diana y se determinaron los genotipos utilizando todos los tipos de secuenciación.

#### 12. Validación del genotipo

**[0119]** La comparación de los genotipos generados a partir de los duplicados (muestras 1 y 2, muestras 3 y 4) y de los diferentes protocolos de secuenciación mostró que el 100% de los genotipos identificados eran idénticos entre los duplicados.

**[0120]** Los genotipos determinados en la etapa 11 se compararon con los genotipos disponibles que se generaron utilizando la tecnología BeadXpress (Illumina). Los resultados de la comparación mostraron que:

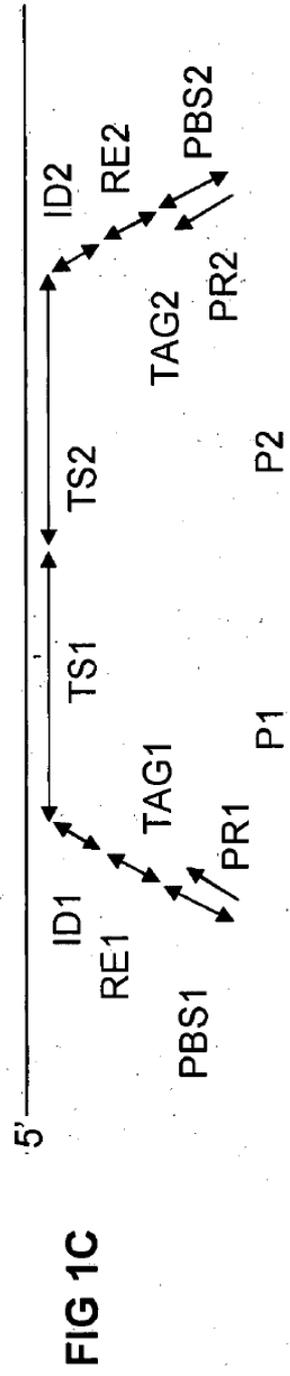
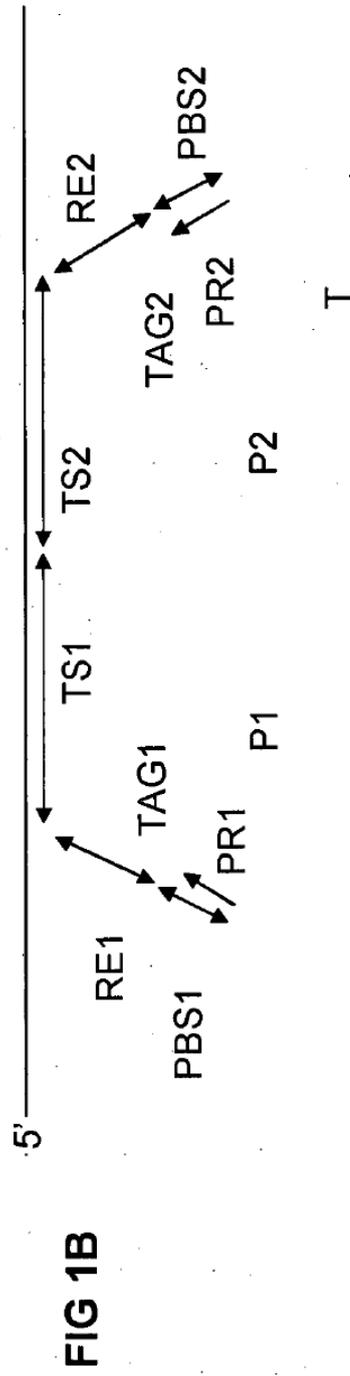
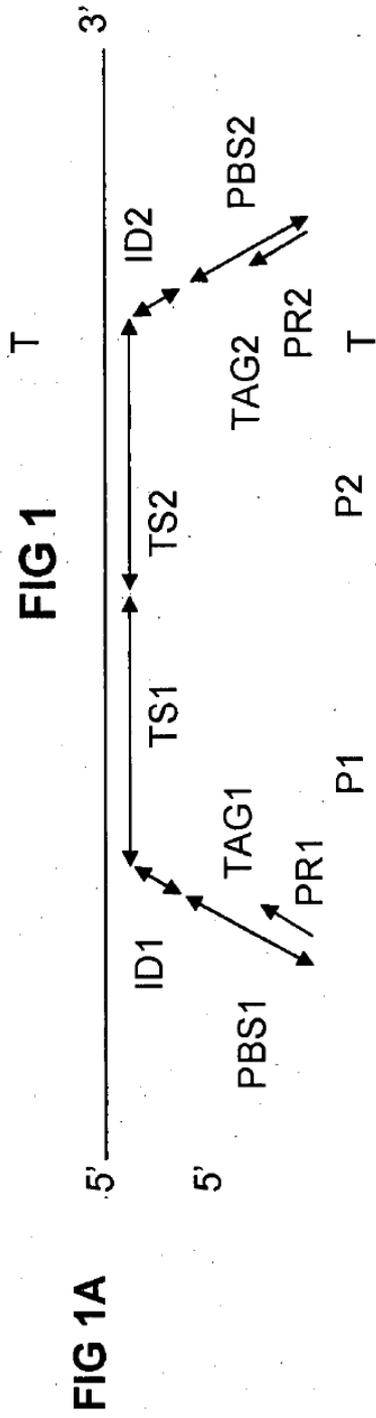
- 21 de los loci de SNP mostraron 100% de correlación entre la estrategia actual y los datos de BeadXpress.
- 1 locus de SNP (SBG0014) no se puntuó en el conjunto de datos BeadXpress, es decir, puntuaciones U (= desconocido), mientras que en este experimento la estrategia utilizada generó un genotipo claro.
- 1 locus de SNP (SBG0039) mostró consistentemente una discrepancia homocigota (estrategia actual) frente a heterocigótica (BeadXpress).

## REIVINDICACIONES

1. Procedimiento para la determinación de una secuencia de nucleótidos diana en una muestra que comprende las etapas de:
- 5 (a) proporcionar para cada secuencia de nucleótidos diana (T) una primera sonda (P1) y una segunda sonda (P2), en el que la primera sonda comprende una primera sección específica de diana (TS1) y una primera sección etiqueta (TAG1) que es no complementaria a la secuencia de nucleótidos diana y que comprende opcionalmente una primera secuencia de unión a cebador (PBS1), en el que la primera sección etiqueta comprende una primera secuencia de reconocimiento (RE1) para una primera endonucleasa de restricción;
- 10 en el que la segunda sonda comprende una segunda sección específica de diana (TS2) y una segunda sección etiqueta (TAG2) que es no complementaria a la secuencia de nucleótidos diana y que comprende una segunda secuencia de unión a cebador (PBS2) opcional, en el que la segunda sección etiqueta comprende una segunda secuencia de reconocimiento (RE2) para una segunda endonucleasa de restricción;
- 15 (b) permitir que la primera y segunda sección específica de diana de la respectiva primera y segunda sonda se hibriden con la secuencia diana;
- (c) ligar la primera y segunda sonda cuando las secciones específicas de diana respectivas de las sondas se hibridan con secciones esencialmente adyacentes en la secuencia diana para proporcionar sondas ligadas (LP);
- (d) opcionalmente amplificar las sondas ligadas con un primer cebador opcional y/o un segundo cebador opcional para proporcionar amplicones (A);
- 20 (e) restringir las sondas ligadas o amplicones con la primera y/o segunda endonucleasa de restricción para proporcionar sondas ligadas restringidas (RLP) o amplicones restringidos (RA), ligar un primer y/o un segundo adaptador que contiene un identificador de base adaptador (AD ID1, AD ID2) a las sondas ligadas restringidas (RLP) o amplicones restringidos (RA);
- 25 (f) someter las sondas ligadas restringidas (RLP) ligadas a adaptador o amplicones restringidos (RA) ligados a adaptador a tecnología de secuenciación de alto rendimiento para determinar al menos parte de la secuencia de nucleótidos de las sondas ligadas restringidas o amplicones restringidos
- (g) identificar la presencia, ausencia o cantidad de la secuencia de nucleótidos diana en la muestra.
2. Procedimiento, según la reivindicación 1, en el que una primera secuencia de identificador basada en una sonda (ID1) se encuentra en la primera sección etiqueta y/o en el que una segunda secuencia de identificador basada en una sonda (ID2) se encuentra en la segunda sección etiqueta.
3. Procedimiento, según las reivindicaciones 1-2, en el que el primer identificador basado en una sonda se encuentra entre la primera secuencia de reconocimiento para una endonucleasa de restricción y la primera sección específica de diana y/o en el que el segundo identificador basado en una sonda se encuentra entre la segunda secuencia de reconocimiento para una endonucleasa de restricción y la segunda sección específica de diana.
- 35 4. Procedimiento, según las reivindicaciones 1-3, en el que la hibridación, ligación y el relleno de huecos opcional se realizan en una etapa combinada.
- 40 5. Procedimiento, según las reivindicaciones 1-4, en el que la secuencia de reconocimiento para la endonucleasa de restricción para la primera sección etiqueta tiene una secuencia de nucleótidos diferente en comparación con el sitio de reconocimiento para la endonucleasa de restricción para la segunda sección etiqueta.
- 45 6. Procedimiento, según las reivindicaciones 1-5, en el que la secuenciación de alto rendimiento comprende secuenciación mediante síntesis.
7. Procedimiento, según las reivindicaciones 1-5, en el que la secuenciación de alto rendimiento comprende amplificación por puente o amplificación en emulsión.
- 50 8. Procedimiento, según las reivindicaciones 1-5, en el que la secuenciación de alto rendimiento comprende secuenciación de lectura única unidireccional, secuenciación de doble cebado de lectura única unidireccional, secuenciación bidireccional (extremos emparejados) o secuenciación de pares acoplados ("mate pair").
- 55 9. Procedimiento, según las reivindicaciones 1-8, en el que se determina una pluralidad de secuencias diana en una muestra o en el que se determina una secuencia diana en una pluralidad de muestras o en el que se determina una pluralidad de secuencias diana en una pluralidad de muestras.
- 60 10. Procedimiento, según las reivindicaciones 1-9, en el que las sondas son sondas circularizables, sondas "keylock" y/o sondas de compuesto.
11. Procedimiento para el genotipado de una muestra biológica para la presencia, ausencia o cantidad de una secuencia diana en la muestra utilizando un ensayo de ligación de nucleótidos que comprende al menos dos sondas, en el que al menos una de las sondas comprende además una sección diana una secuencia de reconocimiento para una endonucleasa de restricción, en el que el procedimiento comprende además una etapa de ligación para proporcionar sondas ligadas, en el que después de la ligación las sondas ligadas se restringen o

amplifican seguido por la restricción para proporcionar sondas ligadas restringidas (RLP) o amplicones restringidos (RA), en el que a los RLP/RA resultantes se ligan uno o dos adaptadores que contienen uno o más identificadores, y en el que se secuencian los RLP/RA ligados a adaptador resultantes.

- 5 12. Procedimiento, según la reivindicación 11, en el que el procedimiento comprende además una etapa de amplificación para proporcionar amplicones.
- 10 13. Procedimiento, según la reivindicación 12, en el que el procedimiento comprende además una etapa en la que los amplicones o sondas ligadas se ponen en contacto con una endonucleasa de restricción para proporcionar amplicones restringidos y/o sondas ligadas restringidas.
14. Procedimiento, según las reivindicaciones 11-13, en el que el genotipado es genotipado codominante que comprende al menos dos sondas específicas de alelos.
- 15 15. Procedimiento, según las reivindicaciones 11-14, en el que se determina al menos parte de la secuencia de parte de los amplicones restringidos o sondas ligadas restringidas.



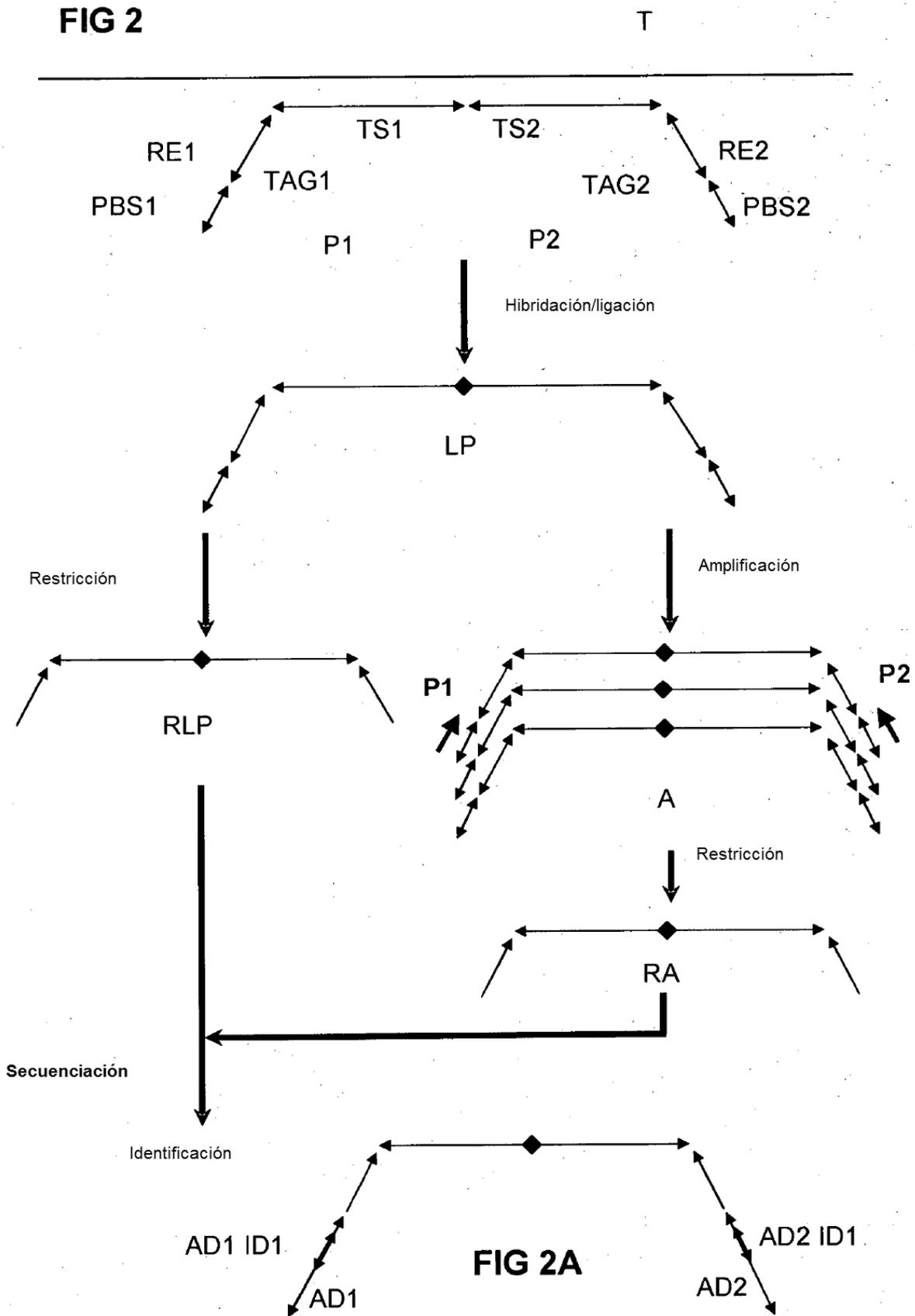


FIG 3



FIG 3A

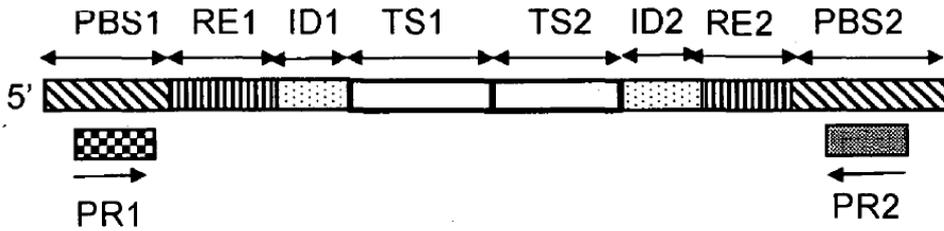


FIG 3B

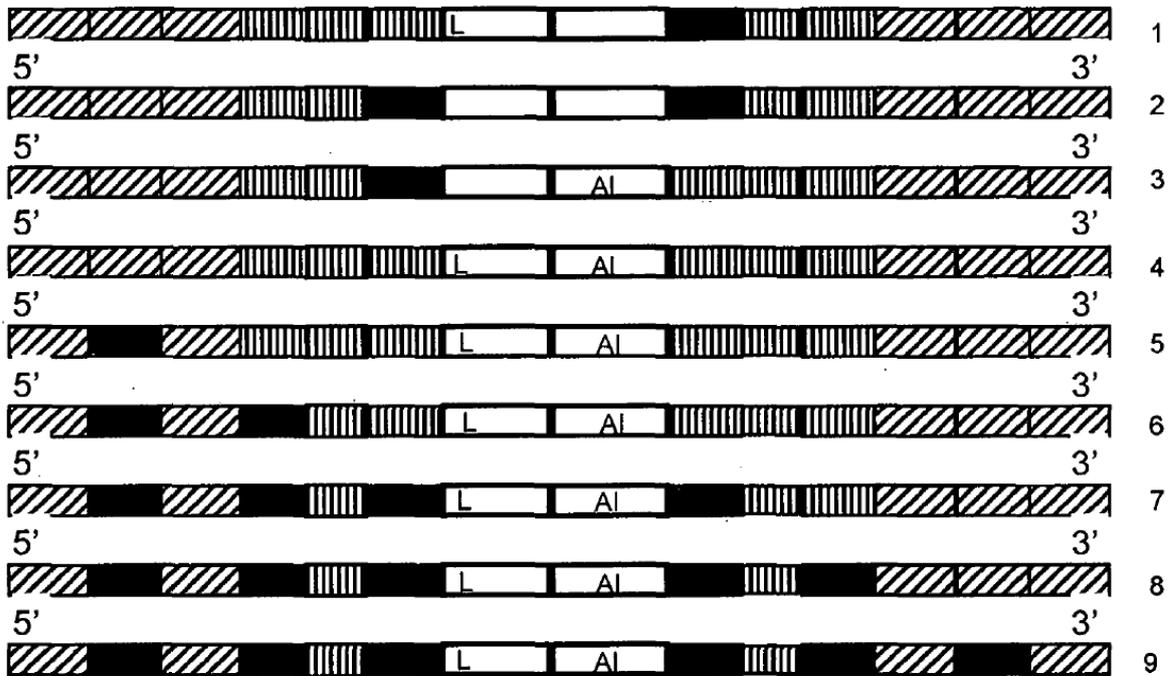


FIG 4

FIG 4A

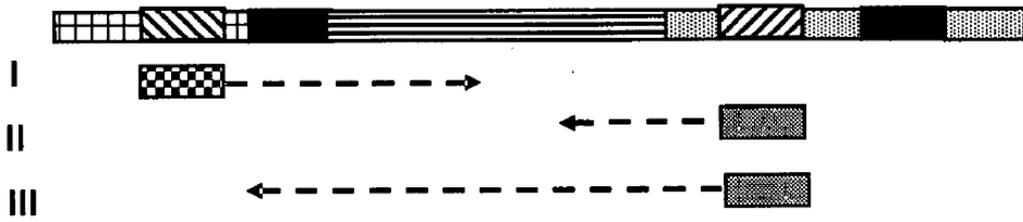


FIG 4B

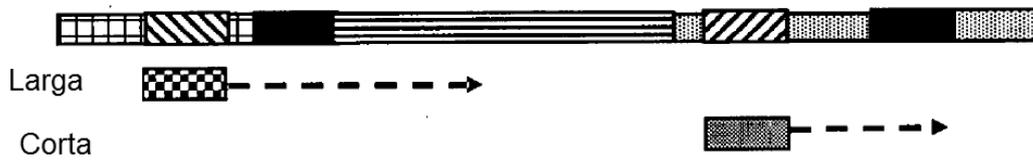


FIG 4C

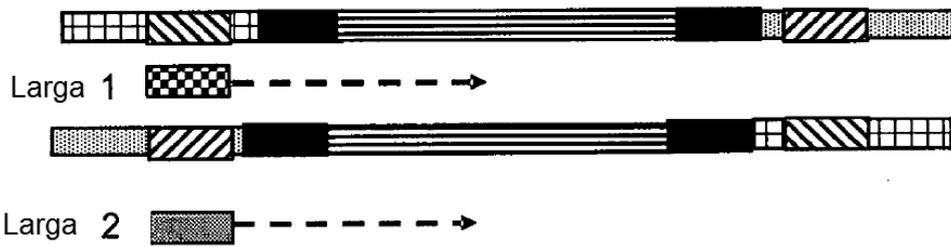


FIG 4D

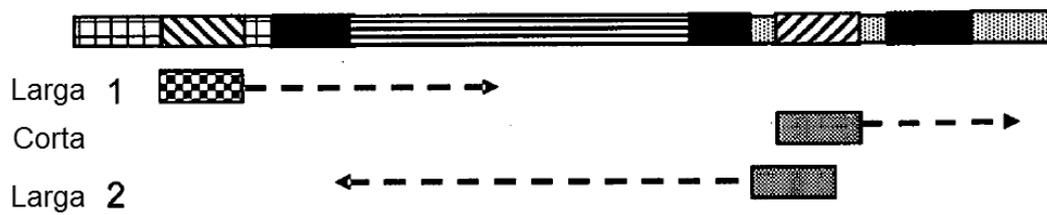


FIG 5A

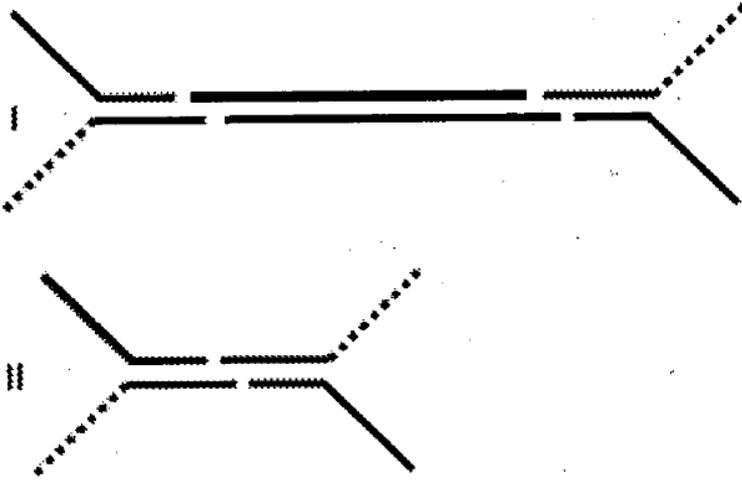


FIG 5

FIG 5B

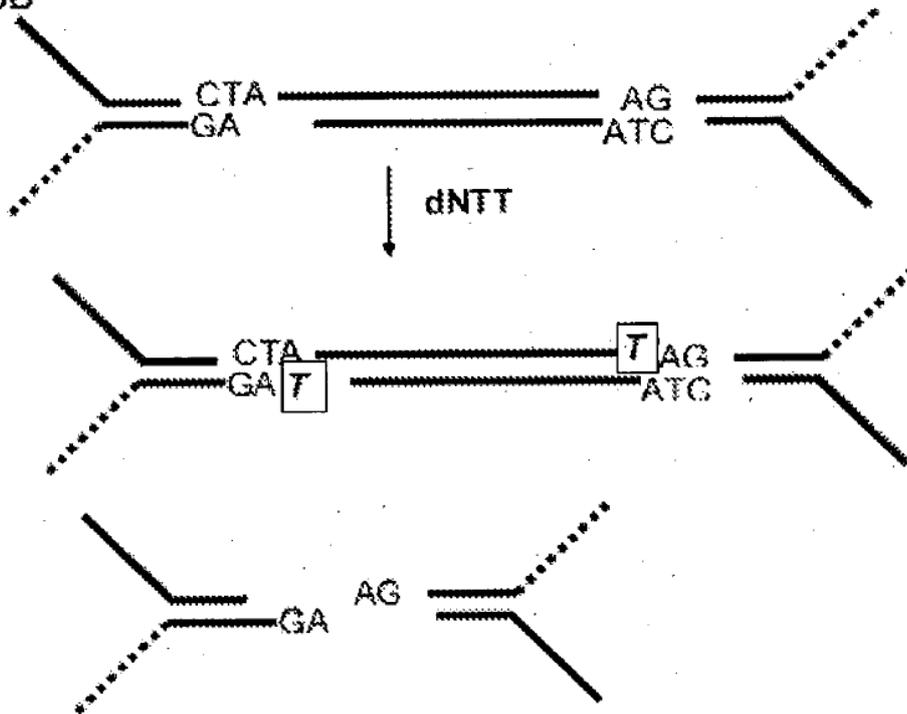


FIG 5C

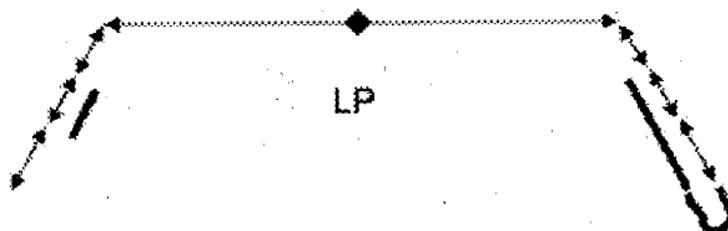


FIG 6

