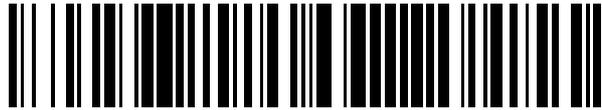


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 446 927**

51 Int. Cl.:

**C12Q 1/68**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.03.2007 E 07715848 (3)**

97 Fecha y número de publicación de la concesión europea: **18.12.2013 EP 1991698**

54 Título: **Detección rápida de SNP basada en secuencias utilizando ensayos de ligación**

30 Prioridad:

**01.03.2006 US 777514 P**

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

**10.03.2014**

73 Titular/es:

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

72 Inventor/es:

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

74 Agente/Representante:

**MORGADES MANONELLES, Juan Antonio**

**ES 2 446 927 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Detección rápida de SNP basada en secuencias utilizando ensayos de ligación.

## 5 CAMPO DE LA INVENCION

La presente invención se refiere al campo de la biología molecular y de la biotecnología. En particular, la presente invención se refiere al campo de la detección de ácidos nucleicos, más particularmente, al diseño y a la composición de (colecciones) de sondas que se pueden utilizar para la detección rápida de ácidos nucleicos. La presente invención se refiere asimismo a procedimientos para la detección de ácidos nucleicos utilizando las sondas y las composiciones. La presente invención proporciona además sondas que pueden hibridarse con una secuencia elegida como objetivo de interés, cebadores para la amplificación de sondas ligadas, la utilización de dichas sondas y cebadores en la identificación y/o detección de secuencias de nucleótidos que están relacionadas con una amplia variedad de características genéticas y genes y equipos de reactivos de cebadores y/o sondas aptos para utilizar en el procedimiento según la presente invención. La presente invención encuentra aplicabilidad en el campo de la detección rápida de secuencias de nucleótidos elegidas como objetivo, de origen sintético, vegetal, animal o humano o combinaciones de los mismos. La presente invención encuentra aplicación particular en el campo de la determinación rápida del genotipo.

## 20 ANTECEDENTES DE LA INVENCION

Cada vez existe más interés en la detección de secuencias específicas de ácidos nucleicos. Este interés no ha surgido únicamente a partir de la secuencia preliminar de nucleótidos descrita recientemente del genoma humano y muchos otros genomas y la presencia en el mismo, así como en los genomas de muchos otros organismos, de una cantidad abundante de polimorfismos de nucleótido único (SNP) y polimorfismos de pequeñas inserciones / deleciones (indel), sino asimismo de tecnologías de marcadores (tales como AFLP), SNPWave y el reconocimiento general de la importancia de la detección de secuencias específicas de ácido nucleico como indicación de, por ejemplo, enfermedades genéticas hereditarias. La detección de los distintos alelos del gen del cáncer de mama BRCA 1 para realizar una selección con respecto a la susceptibilidad hacia el cáncer de mama es únicamente uno de muchos ejemplos. El reconocimiento de que la presencia de sustituciones e indeles de nucleótidos únicos y en los genes proporciona una amplia variedad de información se ha atribuido asimismo a dicho aumento del interés. En la actualidad se reconoce generalmente que dichas sustituciones de nucleótido único constituyen una de las principales causas de un número significativo de enfermedades heredadas monogénica y multigenéticamente, por ejemplo, en seres humanos, o de algún otro modo se encuentran implicadas en el desarrollo de fenotipos complejos, tales como las características de rendimiento de especies vegetales y de ganado. Por lo tanto, las sustituciones de un nucleótido único en muchos casos están asimismo relacionadas con, o por lo menos indican, características importantes de seres humanos, y especies vegetales y animales.

El análisis de dichas sustituciones e indeles de nucleótido único proporcionarán una gran cantidad de información valiosa, que tendrá amplias implicaciones en medicina y agricultura en el sentido más amplio posible. En general se prevé, por ejemplo, que dichos desarrollos proporcionen una la medicación específica para cada paciente. Para analizar dichos polimorfismos genéticos, existe una necesidad cada vez mayor de procedimientos adecuados, fiables y rápidos que permitan la manipulación de grandes cantidades de muestras y grandes cantidades de (predominantemente) SNP de un modo rápido, sin comprometer significativamente la calidad de los datos obtenidos. Uno de los procedimientos principales utilizados en el análisis de los ácidos nucleicos de una secuencia conocida se basa en la hibridación de dos sondas con una secuencia elegida como objetivo y, cuando las sondas se hibridan de un modo adyacente a la secuencia elegida como objetivo, ligando las sondas.

Se ha descrito el principio del OLA (ensayo de ligación de oligonucleótidos), entre otros, en la patente US n.º 4.988.617 (Landegren *et al.*). Esta publicación describe un procedimiento para determinar la secuencia de ácido nucleico en una región de una secuencia de ácido nucleico conocida que presenta una mutación posible conocida. Para detectar la mutación, los oligonucleótidos se seleccionan para que se hibriden con los segmentos adyacentes de la secuencia a determinar de inmediato. Una de las sondas de oligonucleótidos seleccionadas presenta una región terminal en la que uno de los nucleótidos de la región terminal es complementario al nucleótido normal o al mutado de la posición correspondiente en la secuencia de ácido nucleico conocida. Se proporciona una ligasa que conecta covalentemente las dos sondas cuando se han emparejado correctamente y se encuentran inmediatamente adyacentes entre sí. La presencia, ausencia o cantidad de sondas enlazadas es una indicación de la presencia de la secuencia y/o mutación conocida.

Se han descrito otras variantes de técnicas basadas en el OLA, entre otros, en Nilsson *et al. Human mutation*, 2002, 19, 410-415; *Science* 1994, 265: 2085-2088; los documentos US n.º 5.876.924; WO 98/04745; WO 98/04746; US n.º 6.221.603; US n.º 5.521.065; US n.º 5.962.223; EP 185494B1; US n.º 6.027.889; US n.º 4.988.617; EP 246864B1; US n.º 6.156.178; EP 745140 B1; EP 964704 B1; WO 03/054511; US n.º 2003/0119004; US n.º 2003/190646; EP 1313880; US n.º 2003/0032016; EP 912761; EP 956359; US n.º 2003/108913; EP 1255871; EP 1194770; EP 1252334; WO96/15271; WO7/45559; US n.º 2003/0119004A1; US n.º 5.470.705.

Se han publicado avances particulares en las técnicas del OLA en Keygene, Wageningen, Países Bajos. En los documentos WO 2004/111271, WO2005/021794, WO2005/118847 y WO03/052142, se han descrito diversos procedimientos y diseños de la sonda que mejoran la fiabilidad de los ensayos de ligación de oligonucleótidos. Dichas solicitudes dan a conocer además la mejora significativa en los niveles múltiple que se pueden alcanzar. Sin embargo, todas las publicaciones anteriores adolecen del inconveniente de que se basan en procedimientos de detección electroforéticos o basados en matrices. Un inconveniente adicional es la amplia variación de la longitud de las sondas utilizadas, lo que puede provocar que la amplificación sea menos constante.

Es evidente que existe una necesidad continua de sondas de oligonucleótidos que combinen las ventajas y eviten los inconvenientes específicos de los diversos tipos de sonda de ligación y de los procedimientos de detección conocidos en la técnica. Existe asimismo una necesidad de continuar mejorando la tecnología proporcionando sondas que presenten ventajas adicionales. Constituye uno de los objetivos de la presente invención proporcionar dichas sondas. Constituye un objetivo adicional de la presente invención evitar los inconvenientes de las sondas conocidas comúnmente tal como se ha mencionado anteriormente en la presente memoria. Constituye un objetivo adicional de la presente invención proporcionar sondas que sean aptas para los procedimientos de detección rápida. Constituye asimismo un objetivo de la presente invención proporcionar un procedimiento para un procedimiento eficiente, fiable y/o rápido para la detección de secuencias de nucleótidos elegidas como objetivo, preferentemente realizando ensayos de ligación de oligonucleótidos.

Los presentes inventores se han propuesto eliminar o por lo menos disminuir los problemas existentes en la técnica anterior, tratando de mantener al mismo tiempo los muchos aspectos ventajosos de la misma, y continuar mejorando la tecnología. Otros problemas de la técnica anterior y las soluciones proporcionadas a los mismos por la presente invención se pondrán de manifiesto a lo largo de la descripción, las figuras y las diversas formas de realización descritas en la presente memoria.

## SUMARIO DE LA INVENCIÓN

Los presentes inventores han podido combinar tecnologías de secuenciación rápida novedosas con la versatilidad de los ensayos basados en la ligación de oligonucleótidos. En particular, la presente invención se refiere a un procedimiento de detección rápida de secuencias de nucleótidos elegidas como objetivo basándose en ensayos de ligación de oligonucleótidos, en el que las sondas utilizadas en los ensayos de ligación se modifican de tal modo que se puede utilizar un procedimiento de secuenciación rápida para revelar inequívocamente la presente ausencia de la cantidad de las una o más secuencias de nucleótidos elegidas como objetivo.

De este modo, los presentes inventores han descubierto que incorporando un identificador de oligonucleótido único en por lo menos una de las sondas que se utilizan en el ensayo OLA para la detección de cada secuencia elegida como objetivo en la muestra y la detección posterior de dicho identificador tras la ligación y las etapas de amplificación mediante procedimientos de secuenciación rápida proporciona una mejora muy eficiente y fiable de la tecnología existente. A diferencia de las sondas conocidas en la técnica, tanto lineales como circularizables, las sondas utilizadas en el procedimiento de la presente invención puede presentar la misma longitud o una muy similar. Dicha longitud uniforme es ventajosa cuando las sondas ligadas se amplifican ya que las eficiencias de amplificación para todas las sondas ligadas son similares, mientras que se ha observado que, con una longitud distinta de las sondas ligadas, la eficiencia de amplificación puede diferir ampliamente, comprometiendo de este modo la fiabilidad del ensayo como en su conjunto. La longitud uniforme facilita asimismo la detección ya que el identificador se encuentra en la misma posición para todas las sondas ligadas. Al mejorar los ensayos OLA de este modo, se realiza un paso significativo al proporcionar ensayos cada vez más uniformes que son fáciles de diseñar para una secuencia elegida como objetivo específica, pueden discriminar de un modo fiable entre las secuencias o muestras elegidas como objetivo y se pueden realizar de un modo rápido y muy multiplexado.

En ciertas formas de realización, se proporcionan procedimientos para la detección rápida de una o más secuencias de nucleótidos elegidas como objetivo. En ciertas formas de realización, el procedimiento proporciona la detección rápida de una o más secuencias de nucleótidos elegidas como objetivo que pueden obtenerse a partir de una o más muestras. En ciertas formas de realización, el procedimiento comprende proporcionar para cada secuencia de nucleótidos elegida como objetivo una primera sonda y una segunda sonda.

En ciertas formas de realización, la primera sonda comprende una sección específica elegida como objetivo en su extremo 3'. En ciertas formas de realización, la primera sonda comprende además una primera sección etiquetada. En ciertas formas de realización, la primera sección etiquetada no es complementaria a la secuencia de nucleótidos elegida como objetivo. En ciertas formas de realización, la primera sección etiquetada comprende además una primera secuencia de enlace con el cebador.

En ciertas formas de realización, la segunda sonda comprende una segunda sección específica de elegidas como objetivo en su extremo 5'. En ciertas formas de realización, la segunda sonda comprende una segunda sección etiquetada. En ciertas formas de realización, la segunda sección etiquetada no es complementaria a la secuencia de nucleótidos elegida como objetivo. En ciertas formas de realización, la segunda sección etiquetada comprende además una segunda secuencia de enlace con el cebador. En ciertas formas de realización, la primera o la segunda

sección etiquetada comprende una secuencia identificadora. En ciertas formas de realización, tanto la primera como la segunda sección etiquetada comprenden una secuencia identificadora. En ciertas formas de realización, la secuencia identificadora se encuentra entre la primera secuencia de enlace con el cebador y la primera secuencia específica elegida como objetivo. En ciertas formas de realización, el identificador se encuentra entre la segunda secuencia de enlace con el cebador y la segunda secuencia específica elegida como objetivo.

En ciertas formas de realización se permite que las sondas primera y segunda se hibriden con la secuencia elegida como objetivo de la muestra. Las secciones específicas de elegidas como objetivo correspondientes primera y segunda de las sondas se hibridan con las secciones preferentemente sustancialmente adyacentes de la secuencia elegida como objetivo, aunque en algunas formas de realización se puede encontrar presente un espacio de separación entre las dos secciones. En ciertas formas de realización, las sondas primera y segunda se encuentran ligadas, es decir, unidas entre sí. La ligación de las sondas primera y segunda proporciona sondas ligadas.

En ciertas formas de realización se amplifican las sondas ligadas para proporcionar amplicones. En ciertas formas de realización se utilizan uno o más cebadores en la amplificación. En ciertas formas de realización, se utiliza un primer y, en ciertas formas de realización, un segundo cebador para la amplificación.

En ciertas formas de realización, se puede utilizar un primer y, en ciertas formas de realización un segundo cebador que, independientemente, pueden contener un sitio de reconocimiento de enzimas de restricción para proporcionar amplicones. Los amplicones se digieren con las enzimas de restricción correspondientes para cuyos sitios de reconocimiento se encuentran presentes en el primer y, en ciertas formas de realización, el segundo cebador. Las secuencias para un tercer y, en ciertas formas de realización un cuarto cebador, se pueden ligar posteriormente con los amplicones digeridos para proporcionar una plantilla para la amplificación. En ciertas formas de realización, se utilizan el tercer y el segundo y, en ciertas formas de realización, el tercer y el cuarto cebador en la amplificación.

En ciertas formas de realización las sondas ligadas o, en ciertas formas de realización, los amplicones obtenidos a partir de la amplificación de las sondas ligadas, se someten a tecnologías de secuenciación rápida para determinar por lo menos una parte de la secuencia de nucleótidos de las sondas ligadas o amplicones. En ciertas formas de realización, la(s) parte(s) de la secuencia de nucleótidos que se determina(n) sometiendo las sondas ligadas o amplicones a tecnologías de secuenciación rápida comprenden por lo menos la(s) secuencia(s) identificadora(s).

En ciertas formas de realización, la presencia de la secuencia de nucleótidos elegida como objetivo de la muestra se identifica determinando la presencia o ausencia de la secuencia identificadora asociada a la secuencia de nucleótidos.

## BREVE DESCRIPCIÓN DE LOS DIBUJOS

La presente invención se ilustra mediante las figuras siguientes:

Figura 1: en la figura 1, se ilustran esquemáticamente distintos tipos de sonda (A, B, C, D, E, F), enfrentados a una secuencia de nucleótidos elegida como objetivo (T) de interés, que presenta un polimorfismo A/C. Se han representado diversos componentes de las sondas, utilizando representaciones idénticas en todas las figuras.

La figura 1A representa un tipo de sonda lineal, en el que una primera sonda (1) comprende una primera sección específica elegida como objetivo (TS1) y una primera sección etiquetada que comprende un primer identificador (ID1) y una primera secuencia de enlace con un cebador (PBS1), que se puede hibridar con un primer cebador (P1). Una segunda sonda (2) comprende una segunda sección específica elegida como objetivo (TS2) y una segunda sección etiquetada que comprende un primer identificador (ID2) y una segunda secuencia de enlace con un cebador (PBS2), que se puede hibridar con un segundo cebador (P2). En formas de realización para la detección específica de alelos, TS2 puede comprender, preferentemente en su extremo 3', un nucleótido específico de alelo (G o T), junto con un identificador diferente (ID2 o ID2'). La combinación locus alelo se puede determinar (obtener el genotipo) a continuación detectando la presencia o ausencia de la combinación de ID1 con ID2 o ID2'. De un modo similar, se puede obtener el genotipo de todas las variantes alélicas de un polimorfismo. Se pueden diseñar sondas específicas de alelos para el otro tipo de sonda descrito de un modo similar.

La figura 1B representa un tipo de sonda circularizable, comprendiendo la sonda circularizable una primera sección específica elegida como objetivo (TS1) y una segunda sección específica elegida como objetivo (TS2), dispuestas cada una de las mismas en el extremo (los extremos 3' y 5' correspondientes) de la sonda circularizable y una sección etiquetada que comprende un primer identificador (ID1), una primera secuencia de enlace con el cebador (PBS1) que puede hibridarse con un primer cebador (P5), una segunda secuencia de enlace con un cebador (PBS2) que puede hibridarse con un segundo cebador (P2) y un segundo identificador opcional (ID2). ID1 e ID2 pueden encontrarse adyacentes o no. En ciertas formas de realización, la utilización de dicho tipo de sonda se asocia a un tratamiento con exonucleasa para eliminar las sondas sin ligar que puedan provocar la genotipificación con falsos positivos. En algunos diseños, la combinación ID1 e ID2 puede representar combinaciones de locus alelos. La amplificación de las sondas circularizadas produce amplicones cortos que se pueden secuenciar y la determinación de la presencia de ID1 e/o ID2 proporciona la genotipificación positiva del polimorfismo pretendido.

La figura 1C representa un tipo de sonda circularizable alternativo. La sonda comprende los mismos componentes, pero el posicionamiento relativo y la orientación de ID1, ID2, PBS1 y PBS2 es tal que la amplificación se producirá únicamente cuando la sonda se circulariza. Ello evita cualquier eliminación de las sondas no ligadas.

5 La figura 1D representa un tipo de sonda lineal que es similar al tipo de sonda lineal de la figura 1A. Además, se han incorporado las secciones de sujeción C1 y C2 en la sección etiquetada, preferentemente en el extremo, pero se pueden encontrar asimismo entre ID y TS o entre PBS e ID. C1 y C2 se emparejan / hibridan entre sí, imitando de este modo el comportamiento de candado y, en particular, la cinética de hibridación mejorada en comparación con los tipos de sonda lineal convencionales (figura 1A), al mismo tiempo que se puede evitar la concatenación de los amplicones del tipo de sonda de las figuras 1B o C.

10 La figura 1E representa la conformación de la sonda del compuesto y la elongación de la misma. La primera sonda (1) comprende preferentemente una primera sección específica elegida como objetivo (TS1). La segunda sonda (2) comprende una segunda sección específica elegida como objetivo (TS2) y una segunda sección etiquetada que comprende un segundo identificador (ID2) y una segunda secuencia de enlace con el cebador (PBS2). Después de, o simultáneamente con, la etapa de hibridación / ligación, una sonda del compuesto (C) se hibrida con una parte de TS1 con una sección que puede hibridarse con (parte de) la primera sección específica elegida como objetivo (TSP1). (C) comprende además una sección que contiene una sección de enlace con el cebador (PBS1) que puede enlazarse a un cebador (P1). La elongación de la sonda del compuesto produce una sección complementaria a la segunda sección específica elegida como objetivo (TSP2), complementaria al segundo identificador (ID2) y complementaria a la segunda secuencia de enlace con el cebador (cPBS2) que puede hibridarse con el cebador CP2. La amplificación utilizando los cebadores P1 y CP2 produce amplicones que se pueden secuenciar.

15 La figura 1F representa la conformación de un conjunto de sondas asimétricas. La primera sonda (1) comprende una sección específica elegida como objetivo (TS1) y es resistente a la exonucleasa (estrella). La segunda sonda (2) comprende una segunda sección específica elegida como objetivo (TS2) y una sección etiquetada que comprende dos sitios de enlace con cebadores (PBS1 y PBS2, respectivamente) y un segundo identificador ID2 que se encuentra entre PBS1 y PBS2. Una ligación satisfactoria y la eliminación de las sondas sin ligar seguido de la amplificación proporcionó amplicones que se pueden secuenciar.

20 La figura 2 representa la estructura principal de las sondas ligadas / amplicones tras la ligación / amplificación para cada uno de los tipos de sondas de la figura 1A a 1F.

25 La figura 3 representa esquemáticamente la etapa de secuenciación rápida de la presente invención en la que las sondas ligadas / amplicones se une a una superficie (una perla en el caso de la PCR de emulsión seguida por la pirosecuenciación en el caso de la tecnología 454 o la superficie de las celdas de flujo). La superficie presenta una secuencia (CPBS) que puede emparejarse a una o ambas secuencias de enlace con cebadores PBS. Tras la hibridación de la sonda se ligó / amplicón con la superficie, la secuencia hibridada se puede amplificar para cargar la perla con secuencias amplificadas o generar grupos de secuencias amplificadas en la superficie de la celda de flujo. Posteriormente, se pueden determinar dichas secuencias utilizando las tecnologías de secuenciación rápida descritas. La secuencia se puede determinar de un modo unidireccional o bidireccional, añadiendo cebadores de la secuenciación, nucleótidos y enzimas.

30 La figura 4 a 15 representan un conjunto de sondas de ligación redundantes (es decir, para un alelo 'A' y 'G') que comprenden, además, o como sustitución, de los elementos representados en la figura 1 tales como las secciones elegidas como objetivo (TS1, TS2 etc.), un sitio de enlace con el cebador separado para la etapa de secuenciación (sPBS) y un sitio de enlace con el cebador de la PCR (inverso) (PBS2). El sitio de enlace con el cebador (OPBS) de la muestra específica (inversa) se degenera introduciendo una parte variable que actúa como identificador de la muestra (SIS). Cada muestra se puede amplificar selectivamente utilizando cebadores específicos de las muestras (SSP), representados aquí para las muestras 1, 2 hasta n. La utilización de múltiples cebadores inversos redundantes es más económica y se pueden utilizar en una pluralidad de ensayos. Las muestras combinadas se pueden amplificar utilizando un conjunto de cebadores redundantes correspondientemente. Se puede proporcionar una eficiencia equivalente en el enlace con el cebador y la amplificación de todos los alelos o muestras utilizando secuencias de anclaje GC. Las sondas pueden comprender, además, una C opcional dispuesta en el extremo 5' del sitio de enlace con el cebador, con una C correspondiente que se encuentra en el extremo 3' del cebador (inverso) (SSP). Las sondas pueden comprender, además, un identificador de alelos (AIS) para cada alelo a investigar. Los identificadores de alelos y muestras son de 5 pb, sin dos bases consecutivas idénticas (homopolímeros) y los identificadores difieren en por lo menos dos bases.

## 60 DESCRIPCIÓN DETALLADA DE LA INVENCIÓN

La presente invención en un primer aspecto se refiere a un procedimiento para la detección rápida de una o más secuencias de nucleótidos elegidas como objetivo en una o más muestras, comprendiendo el procedimiento las etapas de:

(a) proporcionar para cada secuencia de nucleótidos elegida como objetivo una primera sonda y una segunda sonda, comprendiendo la primera sonda una sección específica de como objetivo en su extremo 3' y una primera sección etiquetada que no es complementaria a la secuencia de nucleótidos elegida como objetivo y que comprende una primera secuencia de enlace con un cebador, comprendiendo la segunda sonda una segunda sección específica elegida como objetivo en su extremo 5' y una segunda sección etiquetada que no es complementaria a la secuencia de nucleótidos elegida como objetivo y que comprende una segunda secuencia de enlace con el cebador, comprendiendo la primera o la segunda sección etiquetada(s), o ambas, una secuencia identificadora que se encuentra entre la secuencia primera o segunda correspondiente de enlace con el cebador y la sección específica elegida como objetivo primera o segunda correspondiente,

(b) permitir que las sondas primera y segunda se hibriden con la secuencia elegida como objetivo,

(c) ligar las sondas primera y segunda, cuando las secciones específicas elegidas como objetivo correspondientes de las sondas se hibridan con secciones sustancialmente adyacentes en la secuencia elegida como objetivo para proporcionar sondas ligadas,

(d) opcionalmente, amplificar las sondas ligadas con un primer y un segundo cebador para proporcionar amplicones,

(e) someter las sondas ligadas o amplicones a la tecnología de secuenciación rápida para determinar por lo menos una parte de la secuencia de nucleótidos, comprendiendo por lo menos la secuencia identificadora, de las sondas ligadas o los amplicones,

(f) identificar la presencia de la secuencia de nucleótidos elegida como objetivo en la muestra determinando la presencia o ausencia de la secuencia identificadora en la secuencia de nucleótidos de la etapa (e), tal como se define en las reivindicaciones independientes adjuntas.

El procedimiento se inicia proporcionando una o más muestras (que se pueden combinar o agrupar) que pueden comprender la secuencia de interés. Se añade a dicha muestra el conjunto de sondas (para cada secuencia elegida como objetivo se pueden proporcionar distintos conjuntos de sondas) y se permite que las secciones específicas elegidas como objetivo de las sondas hibriden con la secuencia elegida como objetivo en unas condiciones aptas. Tras la hibridación, se liga cualquier sonda de hibridación adyacente de la secuencia elegida como objetivo para obtener sondas ligadas. Las sondas ligadas se pueden amplificar o, alternativamente, someter directamente a una secuenciación mediante procedimientos de secuenciación rápida basados en la secuenciación por síntesis. Con la etapa de secuenciación y la posterior identificación del identificador, se determina la presencia de la secuencia elegida como objetivo en la muestra y se completa la genotipificación.

Un aspecto de la presente invención se refiere al diseño ventajoso de las sondas utilizadas en la presente invención. Dichas sondas se describirán más detalladamente a continuación en la presente memoria. Otro aspecto ventajoso de la presente invención radica en la relación entre el estado actual de las tecnologías de secuenciación rápida como una plataforma de detección para ensayos de ligación de oligonucleótidos y la capacidad discriminadora de los ensayos basados en el OLA. Los ensayos OLA conocidos actualmente se han ideado únicamente para plataformas de detección que se basan en la separación de la longitud / movilidad (es decir, en el análisis electroforético), la hibridación (basadas en matrices) o en la determinación de la masa (espectrometría de masas / MALDI-TOF), al mismo tiempo que no se han desarrollado sondas aptas que se puedan utilizar en plataformas de detección por secuenciación rápida. Los presentes solicitantes han observado que, aparte de las innovaciones en diseño de sondas, asimismo los procedimientos para realizar ensayos OLA junto con la secuenciación rápida requieren modificaciones importantes tanto en lo que se refiere a las sondas como en los procedimientos.

### SECUENCIA DE NUCLEÓTIDOS ELEGIDA COMO OBJETIVO

En su definición más amplia, la secuencia elegida como objetivo puede ser cualquier secuencia de nucleótidos de interés. La secuencia elegida como objetivo puede ser cualquier secuencia de la que se pretenda su determinación / detección, por ejemplo, porque es indicativa, referida o representativa de una determinada enfermedad, constitución genética o trastorno. La secuencia elegida como objetivo es preferentemente una secuencia de nucleótidos que comprende, representa o está asociada a un polimorfismo. El término polimorfismo se refiere en la presente memoria a que se producen dos o más secuencias o alelos alternativos genéticamente determinados en una población. Un marcador o sitio polimórfico es el locus en el que se produce la divergencia de la secuencia. Los marcadores preferidos presentan por lo menos dos alelos, produciéndose cada uno con una frecuencia superior al 1%, y más preferentemente superior al 10% o al 20% de una población seleccionada. Un locus polimórfico puede ser tan pequeño como un par de bases. Los marcadores polimórficos comprenden polimorfismos de longitud de fragmentos de restricción, un número variable de repeticiones en tándem (VNTR), regiones hipervariables, minisatélites, repeticiones de dinucleótidos, repeticiones de trinucleótidos, repeticiones de tetranucleótidos, repeticiones de secuencias simples y elementos de inserción tales como Alu. La primera forma alélica identificada se designa arbitrariamente como forma de referencia y las otras formas alélicas se designan como alelos alternativos o variantes. Se hace referencia a veces a la forma alélica más frecuente en una población seleccionada como forma de referencia. Los organismos diploides (y tetraploides / hexaploides) pueden ser homocigotos o heterocigotos con

respecto a las formas alélicas. Un polimorfismo dialélico presenta dos formas. Un polimorfismo trialélico presenta tres formas. Se produce un polimorfismo de un nucleótido único en un sitio polimórfico ocupado por un nucleótido único, que es el sitio de variación entre las secuencias alélicas. El sitio está normalmente precedido y seguido por secuencias muy conservadas del alelo (por ejemplo, secuencias que varían en menos de 1/100 o 1/1000 miembros de las poblaciones). Un polimorfismo de nucleótido único normalmente aparece debido a la sustitución de un nucleótido por otro en el sitio polimórfico. Los polimorfismos de nucleótido único pueden aparecer asimismo a partir de una delección de un nucleótido o de una inserción de un nucleótido con respecto a un alelo de referencia. Otros polimorfismos comprenden (pequeñas) delecciones o inserciones de diversos nucleótidos, conocidas como indeles. En la presente memoria, a veces se hace referencia al proceso de analizar las variaciones genéticas particulares (polimorfismos) existentes en una muestra de ADN individual usando los procedimientos descritos actualmente como genotipificación o genotipificación SNP en el caso de polimorfismos de nucleótido único.

## ADN

En la muestra de ácido nucleico, los ácidos nucleicos que comprenden el elegido como objetivo pueden ser cualquier ácido nucleico de interés. Aunque los ácidos nucleicos de la muestra se encontrarán generalmente en forma de ADN, la información de la secuencia de nucleótidos contenida en la muestra puede proceder de cualquier fuente de ácidos nucleicos, entre ellas por ejemplo ARN, ARN poliA +, ADNc, ADN genómico, ADN de orgánulos tales como el ADN mitocondrial o el ADN de cloroplastos, ácidos nucleicos sintéticos, genotecas de ADN (tales como las genotecas BAC / clones de BAC agrupados), bancos de clones o cualquier selección o combinación de los mismos. El ADN de la muestra de ácido nucleico puede ser bicatenario, monocatenario y ADN bicatenario desnaturalizado en ADN monocatenario. La desnaturalización de secuencias bicatenarias produce dos fragmentos monocatenarios, pudiendo analizarse uno o ambos mediante sondas específicas para las cadenas correspondientes. Las muestras preferidas de ácido nucleico comprenden secuencias elegidas como objetivo de ADNc, ADN genómico, fragmentos de restricción, fragmentos de restricción ligados a adaptadores, fragmentos de restricción ligados a adaptadores amplificados, fragmentos AFLP<(R)> o fragmentos obtenidos en una preamplificación de una plantilla de AFLP.

## MUESTRAS

Se prefiere que una muestra contenga dos o más secuencias elegidas como objetivo distintas, es decir, dos o más referidas a la identidad en lugar de la cantidad de las secuencias elegidas como objetivo de la muestra. En particular, la muestra comprende por lo menos 100, preferentemente por lo menos 250, más preferentemente por lo menos 500, más en particular por lo menos 1000, preferentemente por lo menos 2500, más preferentemente por lo menos 5000 y más preferentemente aún por lo menos 10.000 secuencias elegidas como objetivo adicionales. En la práctica, el número de secuencias elegidas como objetivo en una muestra que se puede analizar se ve limitado, entre otros motivos, por el número de amplicones que se puede detectar. Los procedimientos de detección empleadas actualmente permiten un gran número relativo de secuencias elegidas como objetivo.

## SONDA

Las secciones de las sondas de oligonucleótidos que son complementarias a la secuencia elegida como objetivo se diseñan de tal modo que para cada secuencia elegida como objetivo de una muestra, se proporciona un par de una primera y una segunda sonda, por lo que cada sonda contiene una sección en su extremo terminal que es complementaria a una parte de la secuencia elegida como objetivo (una primera y una segunda parte de la secuencia elegida como objetivo, respectivamente) y las partes complementarias correspondientes de la secuencia elegida como objetivo se encuentran preferentemente sustancialmente adyacentes entre sí.

En ciertas formas de realización, se puede proporcionar unas sondas adicionales primera y/o segunda, correspondientes a distintos alelos de un locus. En ciertas formas de realización, el nucleótido específico del alelo se encuentra en la posición de tanto la primera como la segunda sonda en la que se va a producir la ligación, es decir, en el extremo de la sección específica elegida como objetivo (véase la figura 1A).

En ciertas formas de realización, en un par de sondas de oligonucleótidos, la primera sonda de oligonucleótidos presenta una sección en su extremo 5' (fosforilado) que es complementaria a una primera parte de una secuencia elegida como objetivo y la segunda sonda de oligonucleótidos presenta una sección en su extremo 3' (hidroxilo) que es complementaria a una segunda parte de la secuencia elegida como objetivo. Por lo tanto, cuando el par de sondas se hibrida con partes complementarias de una secuencia elegida como objetivo, el extremo 5' de la primera sonda de oligonucleótidos es sustancialmente adyacente al extremo 3' de la segunda sonda de oligonucleótidos de tal modo que los extremos correspondientes de las dos sondas se pueden ligar para formar un enlace fosfodiéster o un enlace covalente de algún otra modo apto. En ciertas formas de realización, en un par de sondas de oligonucleótidos, la primera sonda de oligonucleótidos presenta una sección en su extremo 3' que es complementaria a una primera parte de una secuencia elegida como objetivo y la segunda sonda de oligonucleótidos presenta una sección en su extremo 5' que es complementaria a una segunda parte de la secuencia elegida como objetivo. Por lo tanto, cuando el par de sondas se hibrida con partes complementarias de una secuencia elegida como objetivo, el extremo 3' de la primera sonda de oligonucleótidos es sustancialmente

adyacente al extremo 5' de la segunda sonda de oligonucleótidos de tal modo que los extremos correspondientes de las dos sondas se pueden ligar para formar un enlace fosfodiéster o un enlace covalente de algún otra modo apto.

En ciertas formas de realización, para cada secuencia elegida como objetivo para la que se debe determinar la presencia, ausencia o cantidad en una muestra, se ha diseñado un par específico de sondas de oligonucleótidos primera y segunda, cada sonda con una sección complementaria a la parte complementaria adyacente de cada secuencia elegida como objetivo, tal como se ha descrito anteriormente. Por lo tanto, en el procedimiento de la presente invención, para cada secuencia elegida como objetivo que se encuentra presente en una muestra, se puede obtener una sonda ligada o un amplicón correspondiente (específico) en la muestra amplificada. En ciertas formas de realización, se proporciona una pluralidad de sondas de oligonucleótidos primera y segunda complementarias a una pluralidad de secuencias elegidas como objetivo en una muestra. Un par de sondas de oligonucleótidos primera y segunda para una secuencia elegida como objetivo determinada en una muestra diferirá por lo menos en la secuencia de nucleótidos de los pares de sondas para otras secuencias elegidas como objetivo u otras muestras, y puede diferir en la longitud y/o la masa de los pares de sondas para otros objetivos (aunque, tal como se ha indicado anteriormente, ello es menos preferido). Más preferentemente, un par de sondas para un objetivo determinado producirá una sonda y/o amplicón ligado o conectado que difiere de la secuencia de las sondas conectadas correspondientes a otros objetivos de la muestra tal como se describirá a continuación.

## VARIACIONES EN LAS SONIDAS

### CANDADO (PADLOCK)

En ciertas formas de realización de la presente invención se pueden utilizar sondas circularizables o sondas candado. Las sondas primera y segunda se combinan en una sonda. La sonda circularizable es un oligonucleótido lineal que, cuando se hibrida con la secuencia elegida como objetivo, y cuando se liga, presenta una conformación circular que está topológicamente bloqueada con la secuencia elegida como objetivo. En ciertas formas de realización, un tratamiento de la muestra con exonucleasa tras la etapa de ligación y antes de la amplificación, preferentemente amplificación por PCR, permite eliminar cualquier sonda circular no ligada y evita la amplificación de cualquier sonda no ligada. Las propias sondas circularizables son conocidas en la técnica, por ejemplo a partir del documento EP745140 o de Van Eijk *et al.*, *Nucleic Acids Research*, 2004, 32, e47. Las sondas conocidas se amplifican normalmente utilizando amplificación por círculo rodador o la reacción en cadena de la polimerasa obteniéndose concatémoros. Además, los sitios de enlace con el cebador de las sondas circularizables conocidas se orientan de tal modo que se amplifica toda la sonda circularizada, comprendiendo cualquier sección específica elegida como objetivo. Para eludir los productos concatémoros durante la amplificación por PCR, se puede incorporar una modificación del bloqueo en la sonda de ligación entre los dos sitios de enlace con el cebador del tipo descrito en el documento WO03/052142.

En ciertas formas de realización, los sitios de enlace con el cebador de las presentes sondas circularizables se orientan de tal modo que preferentemente se amplifica únicamente la sección que comprende los sitios de enlace con el cebador y el identificador, y preferentemente las secciones específicas ligadas elegidas como objetivo no se amplifican. Preferentemente, junto con el tratamiento con exonucleasa para eliminar las sondas circularizables sin ligar, ello proporciona amplicones de una longitud relativamente corta en comparación con los amplicones convencionales obtenidos a partir de la amplificación convencional de sondas circularizadas. Ello evita la formación de grandes concatémoros y la amplificación adicional innecesaria de toda la sonda circularizada.

En ciertas formas de realización, el identificador se encuentra sustancialmente adyacente a una de las secuencias de enlace con el cebador y, preferentemente, entre el primer y el segundo sitio de enlace con el cebador, de tal modo que la amplificación de los amplicones comprende por lo menos uno de los dos sitios de enlace con el cebador y el identificador intermitente. La secuenciación rápida posterior de la amplificación proporcionará la secuencia del identificador y, por lo tanto, la presencia de la secuencia elegida como objetivo en la muestra.

### BLOQUEO (KEYLOCK)

En ciertas formas de realización, para cada secuencia elegida como objetivo determinada a detectar, se diseña preferentemente por lo menos un par de dos sondas de tal modo que cada sonda del par puede hibridarse con una parte de la secuencia elegida como objetivo y las sondas correspondientes del par más comprendiendo cada una sección que es complementaria a una sección correspondiente de la otra sonda del par de tal modo que ambas sondas pueden hibridarse entre sí. Las dos sondas del par se diseñan de tal modo que cuando se hibridan entre sí pueden asimismo hibridarse con una secuencia elegida como objetivo. Cuando se hibridan entre sí, las dos sondas imitan o actúan como sondas candado cuando se utilizan en un ensayo de ligación de oligonucleótidos para la detección de una secuencia de nucleótidos elegida como objetivo, mientras que en las etapas posteriores de amplificación y detección las sondas actúan como producto de ligación lineal. Dicho tipo de sonda se ha denominado "Keylock" y se da a conocer, entre otros, en el documento WO2004111271.

En ciertas formas de realización, la primera sonda de oligonucleótidos presenta una sección en su extremo 5' que es complementaria a una primera parte de una secuencia elegida como objetivo y la segunda sonda de

oligonucleótidos presenta una sección en su extremo 3' que es complementaria a una segunda parte de la secuencia elegida como objetivo. En ciertas formas de realización, la primera sonda de oligonucleótidos comprende además una sección de sujeción que puede hibridarse con una sección de sujeción complementaria que se encuentra en la segunda sonda de oligonucleótidos por lo que las secciones de sujeción son sustancialmente no complementarias a la secuencia elegida como objetivo. En ciertas formas de realización, la presente invención se refiere a un procedimiento para determinar la presencia, ausencia de una secuencia elegida como objetivo o las cantidades de por lo menos una secuencia elegida como objetivo en una muestra tal como se describe en la presente memoria, comprendiendo el procedimiento las etapas de proporcionar un par de sondas que comprenden secciones de sujeción como se describe en la presente memoria para cada secuencia elegida como objetivo a detectar. En ciertas formas de realización, las sondas de oligonucleótidos pueden hibridarse con la secuencia elegida como objetivo, proporcionando unos medios para unir las sondas de oligonucleótidos primera y segunda y permitiendo unir las sondas de oligonucleótidos primera y segunda para producir una sonda conectada correspondiente con una secuencia elegida como objetivo en la muestra.

La sección de sujeción se encuentra preferentemente en el extremo, o en la proximidad del mismo, de la sonda que es distal a la sección elegida como objetivo, es decir, cuando la sección elegida como objetivo se encuentra en el extremo 3', la sección de sujeción se encuentra más hacia el extremo 5' y viceversa. La sección de sujeción no se encuentra necesariamente más distal en el extremo terminal 5' o 3'; puede seguirse por otras secciones tal como se comenta en otro lugar de la presente memoria. Las secciones de sujeción se diseñan preferentemente de tal modo que no pueden hibridarse con las secciones elegidas como objetivo. Las secciones de sujeción de las sondas primera y segunda del par pueden hibridarse preferentemente entre sí. Las secciones de sujeción se diseñan preferentemente de tal modo que dos secciones de sujeción complementarias presentan una afinidad de enlace entre sí superior a la afinidad de enlace de la sección elegida como objetivo de la sonda por su parte complementaria en la secuencia de nucleótidos elegida como objetivo.

Ello significa en la práctica que las secciones de sujeción, cuando se hibridan entre sí, forman un dúplex más fuerte que el híbrido entre la sección elegida como objetivo y su complemento en la secuencia de nucleótidos elegida como objetivo y/o la hibridación de sujeciones complementarias se realiza a temperaturas superiores a la hibridación de la sección complementaria elegida como objetivo de las sondas con respecto al objetivo. Es decir, la sección de sujeción hibridada se desnaturaliza, en unas condiciones de otro modo comparables, a una temperatura superior o en unas condiciones más rigurosas que la temperatura de desnaturalización de las secciones complementarias elegidas como objetivo del par de sondas. Ello permite seleccionar las condiciones durante el procedimiento de la presente invención de tal modo que la sujeción hibridada o bloqueada permanece hibridada o cerrada por lo menos hasta que las sondas se conectan para producir una sonda conectada. La sujeción bloqueada se puede abrir desnaturalizando la sonda (conectada) a una temperatura o en unas circunstancias que permitan la desnaturalización de la sujeción bloqueada.

Un par de sondas que presentan sujeciones bloqueadas expresa una cinética de hibridación y un comportamiento similares o idénticos al de las sondas circularizables o candado. Las dos sondas de un par se pueden añadir por separado una vez se han hibridado las secciones de sujeción entre sí en la muestra o, alternativamente, se pueden bloquear las dos sondas antes de añadirse a la muestra.

En una forma de realización preferida, la sujeción presenta una temperatura de desnaturalización (o temperatura de fusión,  $T_m$ ) que supera la temperatura de desnaturalización de las secciones complementaria elegidas como objetivo del par de sondas en por lo menos 1 grado Celsius, preferentemente 5 grados Celsius, más preferentemente 10 grados Celsius superior a la  $T_m$  más baja de la sección T1 o T2. La temperatura de desnaturalización de una secuencia de oligonucleótidos se puede calcular / estimar a partir de la composición de nucleótidos utilizando la fórmula general para  $T_m = (4 \cdot G + C) + (2 \cdot A + T)$  o  $T_m = (4 \cdot G/C) + 2 \cdot A/T - 5$  C (Meinkoth *et al. Anal. Biochem.* (1984) 138: 267-284). Se pueden aplicar asimismo otras fórmulas ya que la esencia radica en la diferencia en la temperatura de desnaturalización entre las secciones ( $T_m[\text{sujeción}] - T_m[\text{objetivo}]$ ). Ello se puede alcanzar no únicamente variando la longitud de las secciones de sujeción, sino asimismo variando el contenido en GC de la sujeción, ya que el par de bases GC aumenta  $T_m$  en aproximadamente 2 grados Celsius en comparación con un par de bases AT.

Una sección de sujeción normal comprende entre 10 y 30, preferentemente entre 15 y 25 y más preferentemente entre 18 y 24 nucleótidos. Cuando el contenido en GC es inferior, este número de nucleótidos puede aumentar mientras se obtengan las características de hibridación pretendidas. Alternativamente, se pueden utilizar nucleótidos modificados que aumenten la hibridación entre las dos secciones de sujeción. Los ejemplos de los mismos son los nucleótidos que han mejorado las características de hibridación, tales como los ácidos nucleicos bloqueados tales como los descritos en los documentos WO 99/14226, WO 00/56748, WO 00/66604 y WO01/25478, ácidos nucleicos peptídicos o mediante otras moléculas que estabilizan o mejoran la hibridación del ADN, tales como aglutinantes menores de la escisión y otros, tales como los descritos en el documento EP 0 974 672.

Puede variar el contenido en GC de la sujeción, variando el contenido en GC de la sección de sujeción desde más del 50% hasta el 100%, preferentemente más del 60%, más preferentemente más del 70%, más preferentemente más del 80% y se encuentra preferentemente comprendido entre el 90% y el 100%. Por lo tanto, la mayoría de

secciones de sujeción comprenderán combinaciones A/T de un modo más incidental o estructural. Un grupo preferido de secciones de sujeción son secuencias ZIP enriquecidas con GC (Iannone *et al.* (2000), *Cytometry* 39: p. 131-140). Preferentemente, la sección de sujeción comprende por lo menos uno, preferentemente por lo menos 2, 3, 4 o 5 nucleótidos seleccionados del grupo que comprende G y C de, más que cada uno de T1 y T2.

En ciertas formas de realización, cuando se encuentran implicados grupos de pares de sondas, puede proporcionarse una sección de sujeción distinta para cada par de sondas del grupo. Se diseña la sección de sujeción de tal modo que una sujeción para un primer par de sondas y sujeciones para un segundo u otros pares de sondas se pueden distinguir entre sí y, preferentemente, no realizan hibridación por cruzamiento en las condiciones utilizadas en el ensayo de ligación. Cada par de sondas comprende una sujeción única, evitando de este modo la hibridación por cruzamiento entre las sujeciones de distintos pares de sondas de una muestra. Para ello, la sección de sujeción puede comprender nucleótidos adicionales o las secuencias de oligonucleótidos de la sección de sujeción pueden ser únicas en el grupo o en una combinación de las mismas, proporcionando de este modo una discriminación entre dos sujeciones en un grupo o para proporcionar un aumento del número de posibles secuencias de sujeción con enlaces no cruzados para utilizar en ensayos de ligación múltiple. Esta última forma de realización permite la detección simultánea de una pluralidad de secuencias elegidas como objetivo en una muestra. Esta forma de realización permite asimismo la detección de una o más secuencias elegidas como objetivo distintas en una pluralidad de muestras utilizando posteriormente la misma colección de pares de sondas. Esta forma de realización permite que el mismo grupo de pares se pueda utilizar una y otra vez para la detección de distintas secuencias elegidas como objetivo.

Preferentemente, cuando se utilizan distintas sujeciones en un grupo de pares de sondas, dichas sujeciones presentan una Tm que se encuentra dentro de un intervalo pequeño, preferentemente entre aproximadamente 60 y 90 grados Celsius, más preferentemente entre 65 y 88 grados Celsius, más preferentemente entre 70 y 85 grados Celsius. Se conoce que las características de hibridación de los ácidos nucleicos se ven influidas asimismo por las concentraciones salinas. Tal como se utiliza en la presente memoria, la comparación de las características de hibridación en general o las temperaturas de desnaturalización en particular de oligonucleótidos se consideran en unas concentraciones salinas comparables excepto cuando se indique lo contrario. Las sujeciones alternativas que se pueden utilizar en la presente invención son ácidos nucleicos que contienen enlaces fotodegradables. Tras la ligación, se elimina el enlace fotodegradable y se amplifica y/o detecta la sonda conectada.

En ciertas formas de realización, las sondas de la presente invención presentan dichas secciones de sujeción. En ciertas formas de realización, la sección de sujeción de las sondas primera y/o segunda o parte de la misma se proporciona resistente a la exonucleasa tal como se describe en otro lugar de la presente memoria para otras sondas. En ciertas formas de realización, la sección de sujeción o parte de la misma se puede utilizar para el aislamiento de cualquier primera sonda no ligada y/o sondas conectadas basándose en la hibridación de la sección de sujeción tal como se describe en otro lugar de la presente memoria para la retirada basada en la hibridación u otros procedimientos de aislamiento basados en la hibridación utilizando placas de vidrio, matrices, perlas, partículas paramagnéticas, etc. En ciertas formas de realización, la sección de sujeción se desnaturaliza antes del tratamiento con exonucleasa y/o el aislamiento de la primera sonda sin ligar o de las sondas conectadas.

En las sondas del tipo bloqueo (*Keylock*), el identificador se encuentra entre la sección de enlace con el cebador y la sección específica elegida como objetivo de la sonda. Ambas mitades de una sonda *Keylock* pueden presentar un identificador.

## SONDA DEL COMPUESTO

En ciertas formas de realización de la presente invención, se utiliza un conjunto de sondas que se describe en el documento WO2005021794. La secuencia elegida como objetivo se pone en contacto con unas sondas primera y segunda, comprendiendo la primera sonda una primera sección específica elegida como objetivo que es complementaria a la secuencia elegida como objetivo y en la que la primera sonda no comprende preferentemente una primera secuencia de enlace con el cebador en una primera sección etiquetada opcional. La segunda sonda comprende una segunda sección específica elegida como objetivo y una segunda sección etiquetada en la que la segunda sección etiquetada comprende una segunda secuencia de enlace con el cebador. La segunda sección etiquetada puede comprender un identificador entre la segunda secuencia de enlace con el cebador y la segunda sección específica elegida como objetivo. Tras la hibridación y la ligación de las dos sondas, o simultáneamente a las mismas, se proporciona una sonda del compuesto que comprende una sección que puede hibridarse con (parte de) la primera sección específica elegida como objetivo y comprende además una sección que contiene la sección de enlace con el cebador. La sonda del compuesto se hibrida con las sondas primera y segunda ligadas. La elongación de la sonda del compuesto en las sondas primera y segunda ligadas proporciona la sonda del compuesto alargado que posteriormente se puede amplificar utilizando los cebadores primero y segundo que se pueden enlazar con los sitios de enlace con el cebador primero y segundo. Los amplicones resultantes se pueden detectar utilizando tecnologías de secuenciación rápida tal como se describe en la presente memoria y se puede identificar la secuencia elegida como objetivo de la muestra mediante la presencia o ausencia del identificador.

## SONDAS ASIMÉTRICAS

En ciertas formas de realización, se utiliza un conjunto de sondas que se describe en el documento WO2005118847. En esta forma de realización, la secuencia elegida como objetivo se pone en contacto con unas sondas primera y segunda. La primera sonda comprende una primera sección específica elegida como objetivo y preferentemente comprende la misma, es decir, no contiene una primera sección etiquetada que puede comprender una primera secuencia de enlace con el cebador. La primera sección específica elegida como objetivo puede, preferentemente en el extremo distal del posible punto de ligación, comprender unos medios con los que la sonda es resistente a exonucleasas tales como la biotina o nucleótidos modificados tales como modificaciones de fosfomonotioato, fosforditioato y fosforoamidato, y PNA y LNA (véase asimismo el documento WO2005118847 para más detalles al respecto). La segunda sonda comprende una segunda sección específica elegida como objetivo y una segunda sección etiquetada. La segunda sección etiquetada comprende normalmente una primera y una segunda secuencia de enlace con el cebador y un identificador dispuesto de un modo intermitente. Una hibridación y una ligación satisfactorias producen sondas conectadas. Se realiza una etapa con exonucleasa para eliminar todas las sondas no ligadas, es decir, las segundas sondas. Únicamente se alcanzará la amplificación para sondas ligadas que sean resistentes a la exonucleasa por la presencia de la primera sonda. La amplificación de las sondas ligadas con un conjunto de un primer y un segundo cebador provocará la generación de amplicones cortos, que consisten principalmente en los dos cebadores y el identificador intermitente. Se pueden identificar dichos amplicones cortos (y el identificador de los mismos y, por consiguiente, la secuencia elegida como objetivo asociada) utilizando los procedimientos de secuenciación rápida de la presente invención.

## SECCIÓN ETIQUETADA

El término sección etiquetada se utiliza para indicar las partes de las sondas que no pueden hibridarse con las secuencias de nucleótidos elegidas como objetivo. Las secciones etiquetadas comprenden generalmente identificadores y sitios de enlace con el cebador y, en algunas ocasiones, tal como se indica en otro lugar de la presente memoria, secciones de sujeción.

## SECUENCIA DE ENLACE CON EL CEBADOR

Se pueden incorporar a las sondas secuencias de enlace con el cebador para facilitar la amplificación, ya sea lineal o exponencial. Los sitios de enlace con el cebador se encuentran preferentemente en partes de la sonda distintas a la sección específica elegida como objetivo, preferentemente en la sección etiquetada que es sustancialmente no complementaria a la secuencia elegida como objetivo. Los sitios de enlace con el cebador pueden unirse a cebadores para iniciar la elongación de la amplificación. Preferentemente dentro de un grupo de pares de sondas, los sitios de enlace con el cebador son universales, es decir, únicamente un grupo predeterminado de sitios de enlace con el cebador se incorpora a la sonda para permitir la elongación del cebador de amplificación múltiple o de un número limitado de cebadores, tales como cebadores que comprendan uno o más bases selectivas en su extremo 3', tal como se conocen a partir de AFLP (EP 0 534 858). Entre los grupos de pares de sondas, los sitios de enlace con el cebador pueden ser distintos. En ciertas formas de realización, la  $T_m$  de los cebadores que pueden unirse a los diferentes sitios de enlace con el cebador puede ser distinta entre los grupos de pares de sondas.

Las funciones del identificador y de los sitios de enlace con el cebador en una sonda se pueden combinar e interrelacionar en el sentido de que una parte específica de la sonda puede funcionar como (parte de) un primer sitio de enlace con el cebador para la elongación / amplificación (selectiva) del cebador y, al mismo tiempo o en otro, como (parte de) un identificador para proporcionar la diferencia pretendida y la diferencia basada en la plataforma tal como se describe en otra parte de la presente memoria.

## SECUENCIA IDENTIFICADORA

En ciertas formas de realización, la sonda de oligonucleótido de la presente invención comprende, además, un identificador o una secuencia identificadora. La secuencia identificadora es una secuencia de oligonucleótido de una secuencia variable. La longitud del identificador varía de 4 a 30. El identificador es una secuencia única. El carácter único se puede explicar como una secuencia con codificación ZIP del tipo descrito por Iannone *et al.* (2000), *Citometry* 39: p. 131-140. Con un identificador de 6 nucleótidos, se puede realizar un máximo de 4096 combinaciones únicas ( $= 4^6$ ). Preferentemente, el identificador contiene una secuencia de anclaje con un par de bases GC (o segmento corto definido rico en G/C) en el extremo 3' para garantizar la igualdad en la eficiencia de la afinidad de enlace y la amplificación. Además, el identificador no comprende dos bases consecutivas idénticas y todos los identificadores utilizados en un conjunto de identificadores difieren en por lo menos dos bases para garantizar una secuencia de reconocimiento inequívoca. Cuando se utiliza una pluralidad de muestras, se puede identificar cada muestra utilizando un conjunto específico de identificadores. El identificador se encuentra generalmente de tal modo que la amplificación de las sondas ligadas utilizando las secuencias de enlace con el cebador incorporará el identificador para que el amplicón resultante contenga la secuencia identificadora. Normalmente, ello significa que en la sonda ligada, el identificador se encuentra entre los sitios de enlace con el cebador. En formas de realización que utilizan dos o más identificadores, por ejemplo en combinaciones locus - alelo, los identificadores se encuentran asimismo entre los sitios de enlace con el cebador. En ciertas formas de

realización, se proporcionan dos identificadores, uno en cada sonda. Una de las sondas se puede considerar como sonda de locus, es decir, dirigida a un locus específico y comprende un identificador específico de locus. La otra sonda puede ser una sonda específica del alelo, es decir, puede comprender un nucleótido específico del alelo, preferentemente en su punto de ligación. La sonda específica del alelo puede comprender un identificador específico del alelo. De este modo se identifica la presencia o ausencia de una combinación específica de locus - alelo por la presencia / ausencia de los identificadores combinados. Cuando se analiza toda la variación alélica de un polimorfismo, únicamente se necesita una sonda de locus, combinado con 4 sondas específicas de alelos. En ciertas formas de realización, únicamente la sonda específica del alelo puede comprender un identificador que contenga una sección identificadora específica del locus y una sección identificadora específica del alelo, por ejemplo, en forma de identificador de locus de 5 pb, seguida por un identificador del alelo de 2 pb.

## HIBRIDACIÓN

En ciertas formas de realización, las sondas se ponen en contacto de hibridación con la secuencia elegida como objetivo de la muestra. Posteriormente, se permite que los pares de sondas de oligonucleótidos se emparejen con las partes complementarias, preferentemente adyacentes, de la secuencia elegida como objetivo de la muestra. Los procedimientos y las condiciones para el emparejamiento específico de las sondas de oligonucleótidos con secuencias complementarias elegidas como objetivo resultan muy conocidos en la técnica (véase por ejemplo, Sambrook and Russel (2001) "*Molecular Cloning: A Laboratory Manual*" (3ª edición), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press). Generalmente, tras mezclar las sondas de oligonucleótidos y secuencias elegidas como objetivo, los ácidos nucleicos se desnaturalizan por incubación (normalmente a una temperatura comprendida entre 94 grados Celsius y 96 grados Celsius) durante un corto período de tiempo (por ejemplo entre 30 segundos y 5 minutos) en una sal amortiguadora. A continuación, se deja enfriar la muestra que comprende las sondas desnaturalizadas y las secuencias elegidas como objetivo hasta una temperatura de hibridación óptima para el emparejamiento específico de las sondas y las secuencias elegidas como objetivo, que habitualmente es de aproximadamente 5 grados Celsius por debajo de la temperatura de fusión del híbrido entre la sección complementaria (sección elegida como objetivo) de la sonda y su secuencia complementaria (en la secuencia elegida como objetivo). Para evitar una hibridación inespecífica o ineficiente de una de las dos sondas de un par, o en una muestra con una pluralidad de secuencias elegidas como objetivo, se prefiere que, en una muestra, las secciones de las sondas que son complementarias a las secuencias elegidas como objetivo presenten unas temperaturas de fusión similares, preferentemente idénticas, entre las distintas secuencias elegidas como objetivo presentes en la muestra. Por lo tanto, las secciones complementarias de las sondas primera y segunda difieren preferentemente menos de 20, 15, 10, 5 o 2 grados Celsius en la temperatura de fusión. Ello se facilita utilizando secciones complementarias de las sondas primera y segunda con una longitud similar y/o un contenido similar en G/C, las secciones complementarias difieren preferentemente en menos de 20, 15, 10, 5 o 2 nucleótidos de longitud y sus contenidos en G/C difieren en menos del 30, 20, 15, 10 o 5%. El término complementario, tal como se utiliza en la presente memoria, significa que una primera secuencia de nucleótidos puede hibridarse específicamente con la segunda secuencia de nucleótidos en unas condiciones de rigurosidad normal. Una secuencia de nucleótidos que se considera complementaria a otra secuencia de nucleótidos puede comprender una cantidad inferior, es decir, preferentemente inferior al 20, 15, 10, 5 o 2 %, de los emparejamientos incorrectos. Alternativamente, puede ser necesario compensar los emparejamientos incorrectos, por ejemplo, incorporando los denominados nucleótidos universales, tal como por ejemplo se describe en el documento EP-A 974 672, o incorporando ciertos nucleótidos modificados que pueden compensar los emparejamientos incorrectos, por ejemplo, aumentando la temperatura de fusión o aumentando la especificidad tales como con los LNA. Puesto que el emparejamiento de sondas con secuencias elegidas como objetivo depende de la concentración, el emparejamiento se realiza preferentemente en un pequeño volumen, es decir, inferior a 25 microlitros, preferentemente inferior a 10 microlitros. En dichas condiciones de hibridación, el emparejamiento de sondas con secuencias elegidas como objetivo es generalmente rápido y no dura más de 5, 10 o 15 minutos, aunque se puede utilizar un mayor tiempo de emparejamiento superior siempre que se mantenga la temperatura de hibridación para evitar emparejamientos inespecíficos. Unos períodos de emparejamiento superiores son más importantes / requeridos en las aplicaciones cuantitativas que se basan en la ocupación completa del objetivo mediante sondas de ligación para permitir la supervisión o cantidades relativas de secuencias elegidas como objetivo entre las muestras.

En ciertas formas de realización de la presente invención, se han obtenido unos resultados excelentes con períodos prolongados de hibridación tales como la hibridación durante la noche o por hibridación repetida, por ejemplo, 10 ciclos de 1 hora. Los períodos prolongados de hibridación pueden resultar ventajosos en dichos ensayos ya que la diferencia en la señal debida a eficiencias distintas de hibridación está reducida y se considera aconsejable alcanzar la hibridación y la ligación completa de todas las sondas para las que está presente una secuencia elegida como objetivo. Se han obtenido unos excelentes resultados mediante una etapa de hibridación - ligación combinada utilizando una ligasa termoestable descrita en la presente memoria. En esta forma de realización, se realizó la hibridación - ligación permitiendo que las sondas se hibridasen durante 1 hora en presencia de una ligasa termoestable, seguido por una etapa de desnaturalización. La repetición de estas etapas por lo menos 2 veces proporcionó unos buenos resultados. La repetición de estas etapas 10 veces proporcionó resultados excelentes.

Para evitar la evaporación durante la desnaturalización y el emparejamiento, las paredes y las tapas de las cámaras de reacción (es decir, tubos o pocillos de microvaloración) se pueden calentar asimismo hasta por lo menos la

misma temperatura que la mezcla de reacción que se alcanza normalmente utilizando un equipo de amplificación de ADN comercial o proporcionando un recubrimiento de aceite mineral. En las sondas de oligonucleótido preferidas, la longitud de la sección complementaria a la elegida como objetivo presenta preferentemente por lo menos 15, 18 o 20 nucleótidos y preferentemente no más de 30, 40 o 50 nucleótidos, y las sondas presentan preferentemente una temperatura de fusión de la sección elegida como objetivo de por lo menos 50 grados Celsius, 55 grados Celsius o 60 grados Celsius.

## LIGACIÓN

Los extremos 5'-fosforilados y 3'-hidroxilados correspondientes de un par de sondas de oligonucleótidos primera y segunda o de la sonda circularizable cuyas secciones específicas elegidas como objetivo se emparejan sustancialmente adyacentes entre sí con las partes complementarias de una secuencia elegida como objetivo se conectan para formar un enlace covalente mediante cualquier medio apto conocido en la técnica. Los extremos de las sondas pueden conectarse enzimáticamente en un enlace fosfodiéster mediante una ligasa, preferentemente una ADN ligasa. Las ADN ligasas son enzimas capaces de catalizar la formación de un enlace fosfodiéster entre (los extremos de) dos cadenas de polinucleótidos unidas a sitios adyacentes en una cadena complementaria. Las ADN ligasas requieren habitualmente ATP (EC 6.5.1.1) o NAD (EC 6.5.1.2) como cofactor para sellar las mellas en el ADN bicatenario.

Las ADN ligasas aptas para utilizar en la presente invención son la ADN ligasa T4, la ADN ligasa de *E. coli* o, preferentemente, una ligasa termoestable como por ejemplo, la ligasa de *Thermus aquaticus* (Taq), la ADN ligasa de *Thermus thenophilus* o la ADN ligasa de *Pyrococcus*.

Alternativamente, la ligación química de los extremos de polinucleótidos modificados adecuadamente se puede utilizar para ligar dos sondas de oligonucleótidos emparejadas en sitios adyacentes de las partes complementarias de una secuencia elegida como objetivo. Los ejemplos de grupos reactivos en los extremos de polinucleótidos modificados comprenden, pero sin limitarse a los mismos, los grupos fosforotioato y tosiloato o yoduro, ésteres e hidrazida, RC(O)S, RCH<sub>2</sub>S y [α]-haloacilo, tiofosforilo y bromoacetamida, y S-pivaloiloiloximetil-4-tiotimidina.

Los agentes de ligación química comprenden, sin limitación, agentes activadores, condensadores y reductores, tales como carbodiimida, bromuro de cianógeno (BrCN), N-cianoimidazol, imidazol, 1-metilimidazol / carbodiimida / cistamina, ditiotretol (DTT) y la luz ultravioleta. La autoligación, es decir, la ligación espontánea sin un agente de ligación, está comprendida asimismo dentro del alcance de la presente invención. Los protocolos detallados de los procedimientos de ligación química y las descripciones de grupos reactivos aptos se pueden encontrar, entre otros, en Xu *et al.*, *Nucleic Acid Res.*, 27: 875-81 (1999); Gryaznov and Letsinger, *Nucleic Acid Res.* 21: 1403-08 (1993); Gryaznov *et al.*, *Nucleic Acid Res.* 22: 2366-69 (1994); Kanaya and Yanagawa, *Biochemistry* 25: 7423-30 (1986); Luebke and Dervan, *Nucleic Acids Res.* 20: 3005-09 (1992); Sievers and von Kiedrowski, *Nature* 369: 221-24 (1994); Liu and Taylor, *Nucleic Acids Res.* 26: 3300-04 (1999); Wang and Kool, *Nucleic Acids Res.* 22: 2326-33 (1994); Purmal *et al.*, *Nucleic Acids Res.* 20: 3713-19 (1992); Ashley and Kushlan, *Biochemistry* 30: 2927-33 (1991); Chu and Orgel, *Nucleic Acids Res.* 16: 3671-91 (1988); Sokolova *et al.*, *FEBS Letters* 232:153-55 (1988); Naylor and Gilham, *Biochemistry* 5:2722-28 (1966); y la patente US n.º 5.476.930.

Tanto la ligación química como la enzimática se producen de un modo mucho más eficiente en los complejos entre la secuencia elegida como objetivo y la sonda emparejados perfectamente en comparación con los complejos en los que una o ambas sondas presentan un emparejamiento incorrecto con la secuencia elegida como objetivo en sitio de la ligadura, o en la proximidad del mismo (Wu and Wallace, 1989, *Gene* 76: 245-254; Xu and Kool, *supra*). Para aumentar la especificidad de ligación, es decir, las eficiencias de ligación correspondientes de oligonucleótidos que coinciden perfectamente en comparación con oligonucleótidos con un emparejamiento incorrecto, la ligación se realiza preferentemente a temperaturas elevadas. Por lo tanto, en ciertas formas de realización de la presente invención, se utiliza una ADN ligasa que permanece activa entre 50 y 65 grados Celsius durante períodos prolongados, pero que se inactiva fácilmente a temperaturas más elevadas, por ejemplo, si se utiliza en la etapa de desnaturalización durante una PCR, habitualmente entre 90 y 100 grados Celsius. Una de dichas ADN ligasas es una ADN ligasa que requiere NAD de una bacteria grampositiva (cepa MRCH 065) tal como se conoce a partir del documento WO01/61033. Se conoce dicha ligasa como "ligasa 65" y se encuentra disponible comercialmente de MRC Holland, Ámsterdam. En ciertas formas de realización, se utiliza una ligasa Taq. En ciertas formas de realización, la ligasa se inactiva tras la ligación de las sondas primera y segunda. En ciertas formas de realización, la sonda conectada se desnaturaliza a partir de la secuencia elegida como objetivo.

## LIGACIÓN DE HUECOS

En una forma de realización alternativa, por ejemplo dirigida a la identificación de indeles, los extremos correspondientes de las secciones complementarias elegidas como objetivo de las sondas primera y segunda se pueden emparejar de tal modo que queda un espacio incompleto. En ciertas formas de realización, las partes primera y segunda de la secuencia de nucleótidos elegida como objetivo no se encuentran adyacentes. Es decir, las secciones específicas primera y segunda elegidas como objetivo de las sondas primera y segunda no se hibridan con las partes primera y segunda de la secuencia de nucleótidos elegida como objetivo que se encuentra adyacente.

Ello es fundamentalmente distinto de otras variedades de esta tecnología, por ejemplo, las descritas en particular en los documentos EP 185494, US n.º 5.521.065, US n.º 5.692.223 y WO 03054311. Dicho espacio incompleto se puede llenar con un (tercer) (oligo)nucleótido apto y ligarse. Dichos procedimientos se conocen en la técnica como "ligación de huecos" y se dan a conocer, entre otros, en los documentos WO 00/77260; US n.º 5.185.243; EP439182; EP320308; WO90/01069. Otra posibilidad para rellenar dichos huecos es mediante la extensión de un extremo de la sonda utilizando una polimerasa y una ligasa junto con nucleótidos individuales o múltiples, opcionalmente preseleccionados de entre A, T, C, o G, o dinucleótidos, trinucleótidos u otros oligonucleótidos pequeños. Sin embargo, en el caso de que la secuencia elegida como objetivo sea de ARN, otra posibilidad para llenar el hueco es por extensión de un extremo de la sonda utilizando una transcriptasa inversa y una ligasa junto con nucleótidos individuales o múltiples, opcionalmente preseleccionados entre A, T, C, o G, o dinucleótidos, trinucleótidos u otros oligonucleótidos pequeños. La ligación de huecos puede encontrar aplicación tanto en la detección de SNP / indeles individuales o SNP múltiples (haplotipificación) que se encuentran muy próximos.

## AMPLIFICACIÓN

En el procedimiento de la presente invención, las sondas conectadas se amplifican para producir una muestra amplificada que comprende sondas conectadas (detectables) amplificadas (amplicones) que son representaciones de la secuencia de nucleótidos elegida como objetivo mediante cualquier procedimiento de amplificación de ácido nucleico apto conocido en la técnica. Los procedimientos de amplificación de ácidos nucleicos utilizan generalmente uno o dos cebadores, dNTP, y una (ADN) polimerasa. Un procedimiento preferido para la amplificación es la PCR. La "PCR" o "reacción en cadena de la polimerasa" 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 la ADN polimerasa y un exceso de trifosfatos de desoxinucleótidos, los oligonucleótidos que se hibridan específicamente con la secuencia elegida como objetivo principal inicia la síntesis de nuevo ADN. Se prefiere que la polimerasa sea una ADN polimerasa que no exprese actividad de desplazamiento catenario o, por lo menos, no significativamente. Los ejemplos de las mismas son AmpliTaq y AmpliTaq Gold (proveedor Perkin Elmer) y AccuPrime (Invitrogen). Una ronda de resultados de la síntesis en nuevas cadenas de una longitud determinada que, como las cadenas originales, pueden hibridarse con los cebadores al desnaturalizarse y emparejarse. El segundo ciclo de desnaturalización, emparejamiento y síntesis produce dos productos - que juntos constituyen un producto discreto bicatenario, exactamente la longitud entre los extremos del cebador. Dicho producto discreto se acumula exponencialmente con cada ronda sucesiva de amplificación. Durante el transcurso de aproximadamente 20 a 30 ciclos, se puede alcanzar una amplificación de muchos millones de veces del fragmento discreto. Los protocolos de la PCR son muy conocidos en la técnica, y se describen en los libros de texto de laboratorio convencionales, por ejemplo, Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc. (1995). Las condiciones aptas para la aplicación de la PCR en el procedimiento de la presente invención se describen en los documentos EP-A 0 534 858 y Vos *et al.* (1995; *Nucleic Acids Res.* 23: 4407-4414), en los que una pluralidad de fragmentos de ADN de entre 70 y 700 nucleótidos y que comprenden secuencias de enlace con el cebador idénticas se amplifican con la misma eficiencia cerca utilizando un par de cebadores. En ciertas formas de realización, la polimerasa se inactiva tras la amplificación. Otros procedimientos de amplificación múltiplex y/o isotérmica que pueden aplicarse comprenden la amplificación por círculo rodador (RCA), la reacción en cadena de la ligasa (LCR), la replicación por secuencia autosostenida (3SR), amplificación del ARN en la que interviene la Q-B replicasa, o la amplificación por desplazamiento catenario (SDA). En algunos casos, ello puede requerir un diseño distinto de las sondas y los cebadores sin apartarse del espíritu de la presente invención.

## CEBADORES

La sonda conectada se amplifica utilizando un par de cebadores correspondientes a los sitios de enlace con el cebador. En ciertas formas de realización, el par de cebadores comprende únicamente un cebador y la amplificación es lineal en lugar de exponencial. En ciertas formas de realización, el par comprende un primer cebador que puede emparejarse con la primera sección de enlace con el cebador y puede iniciar la amplificación o la elongación. En ciertas formas de realización, el par comprende además un segundo cebador que puede emparejarse con la segunda sección de enlace con el cebador y puede iniciar la amplificación o la elongación. En ciertas formas de realización, el segundo cebador presenta la misma secuencia que el segundo sitio de enlace con el cebador en la sonda. En una forma de realización preferida, por lo menos uno de los cebadores, o el mismo par de cebadores, se utiliza en la amplificación de dos o más sondas conectadas distintas de una muestra, preferentemente para la amplificación de todas las sondas conectadas de una muestra. A veces se hace referencia a dicho cebador como cebador universal ya que dichos cebadores pueden iniciar la amplificación de todas las sondas conectadas que comprenden el sitio de enlace con el cebador universal correspondiente y, por consiguiente, de todas las sondas ligadas que comprenden el sitio de enlace con el cebador universal. Los distintos cebadores que se utilizan en la amplificación en la etapa (i) son preferentemente sustancialmente iguales en el emparejamiento y la eficiencia de la iniciación. Por lo tanto, los cebadores de una muestra difieren preferentemente en menos de 20, 15, 10, 5, o 2 grados Celsius en la temperatura de fusión. Ello se puede realizar tal como se describe en la presente memoria para las secciones específicas del objetivo de las sondas de oligonucleótidos. A diferencia de la secuencia de las secciones específicas elegidas como objetivo, la secuencia de los cebadores no viene dictada por la secuencia elegida como objetivo. Por lo tanto, las secuencias de los cebadores se pueden diseñar convenientemente uniendo la secuencia de tetrámeros de nucleótidos en los que cada tetrámero comprende una A, T, C y G o mediante otros

modos que garanticen que el contenido de G/C y la temperatura de fusión de los cebadores sean idénticos o muy similares. La longitud de los cebadores (y de los sitios de enlace con los cebadores correspondientes de la sección etiquetada de la segunda sonda) es preferentemente de por lo menos 12, 15 o 17 nucleótidos y preferentemente no superior a 25, 30, 40 nucleótidos.

En cierta forma de realización, por lo menos dos de las segundas sondas de oligonucleótidos que son complementarias a por lo menos dos secuencias elegidas como objetivo distintas de una muestra comprenden cada una una sección etiquetada que comprende una sección de enlace con el cebador que es complementaria a una secuencia única del cebador.

En ciertas formas de realización, para garantizar una eficiencia de iniciación similar en comparación con otros cebadores que presentan la misma secuencia de anclaje, el cebador puede comprender una secuencia de anclaje 3', preferentemente una secuencia de anclaje de 2 pb, preferentemente una secuencia de anclaje GC. Normalmente, la secuencia de enlace con el cebador correspondiente presentará asimismo el complemento de la misma.

Por lo tanto, preferentemente por lo menos uno de los cebadores primero y segundo de un par de cebadores se utiliza en la amplificación de las sondas conectadas correspondientes a por lo menos dos secuencias elegidas como objetivo distintas de una muestra, más preferentemente en la amplificación de las sondas conectadas correspondientes a todas las secuencias elegidas como objetivo de una muestra. Preferentemente se utiliza únicamente un único primer cebador y en algunas formas de realización se utilizan solamente un primer y un único segundo cebador en la amplificación de todas las sondas conectadas. La utilización de cebadores comunes en la amplificación de una pluralidad de fragmentos distintos resulta generalmente ventajosa para la eficiencia de la etapa de amplificación. Las sondas conectadas obtenidas de la ligación de las secciones de sondas emparejadas adyacentes se amplifican, utilizando un par de cebadores, que comprenden preferentemente un par de cebadores para cada una de las sondas conectadas de la muestra. El par de cebadores comprende cebadores que son complementarios a secuencias de enlace con el cebador que se encuentran presentes en la sonda conectada. Un par de cebadores comprende habitualmente un primer y por lo menos un segundo cebador, pero puede comprender únicamente un único cebador que realice la iniciación en ambas direcciones. Se han obtenido unos resultados excelentes utilizando cebadores que son conocidos en la técnica tales como los cebadores AFLP, tal como se describe, entre otros, en el documento EP534858 y en Vos *et al.*, *Nucleic Acid Research*, 1995, vol. 23, 4407-44014 y que se describirán más detalladamente posteriormente en la presente memoria.

### CEBADORES SELECTIVOS

En ciertas formas de realización, uno o más de los cebadores utilizados en la etapa de amplificación de la presente invención es un cebador selectivo. En la presente memoria se define cebador selectivo como un cebador que, además de su secuencia universal que es complementaria a un sitio de enlace con el cebador de la sonda, contiene una región que comprende los denominados "nucleótidos selectivos". La región que comprende los nucleótidos selectivos se encuentra en el extremo 3' del cebador universal. El principio de los nucleótidos selectivos se da a conocer, entre otros, en el documento EP534858 y en Vos *et al.*, *Nucleic Acid Research*, 1995, vol. 23,4407-44014 aunque en un contexto diferente, es decir, la obtención de la huella genética. Los nucleótidos selectivos son complementarios a los nucleótidos de las sondas (ligadas) que se encuentran adyacentes a la secuencia de enlace con el cebador. Los nucleótidos selectivos generalmente no forman parte de la región en las sondas (ligadas) que se representa como la secuencia de enlace con el cebador. Los cebadores que comprenden nucleótidos selectivos se denominan cebadores +N, representando N el número de nucleótidos selectivos presentes en el extremo 3' del cebador. N se selecciona preferentemente de entre A, C, T o G.

### AMPLICONES

El término "amplicón" tal como se utiliza en la presente memoria, se refiere al producto de la etapa de amplificación de las sondas conectadas. El término "amplicón", tal como se utiliza en la presente memoria, se refiere por lo tanto a una sonda conectada amplificada. Tras la etapa de ligación en la que las dos secciones específicas elegidas como objetivo se conectan mediante 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 produce unas representaciones lineales amplificadas de la sonda conectada. Preferentemente un amplicón es una representación monomérica de la sonda conectada amplificada. En ciertas formas de realización, el amplicón comprende y consiste preferentemente en los nucleótidos del primer y el segundo cebador opcional y el/los identificador(es) que se encuentra(n) en medio. Las diversas formas de realización de la presente invención proporcionarán más detalles en este sentido (figura 2).

### SECUENCIACIÓN RÁPIDA

La secuenciación o cribado rápido, a menudo abreviado como HTS, es un método de experimentación científica especialmente importante en los campos de la biología y la química. Combinando la robótica moderna y otros equipos de laboratorio especializados, el investigador puede cribar eficazmente grandes cantidades de muestras simultáneamente.

Se prefiere que la secuenciación se realice utilizando procedimientos de secuenciación rápida, 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 454 Life Sciences), por Seo et al. (2004) *Proc. Natl. Acad. Sci. USA* 101:5488-93, y las tecnologías de Helios, Solexa, US Genomics.

#### TECNOLOGÍA 454

En ciertas formas de realización, se prefiere más que la secuenciación se realice utilizando el aparato y/o el método que se dan 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). La técnica descrita permite la secuenciación de hasta 40 millones de bases en una sola serie y es 100 veces más rápida y más económica que la técnica de la competencia. La técnica de secuenciación comprende aproximadamente 5 etapas: 1) fragmentación del ADN y ligación de adaptadores específicos para crear una genoteca de ADN monocatenario (ADNss); 2) hibridación del ADNss con perlas, emulsificación de las perlas en microrreactores de agua en aceite y realización de la PCR en emulsión para amplificar las moléculas individuales de ADNss en las perlas; 3) selección / enriquecimiento de las perlas que contienen moléculas de ADNss amplificadas en la superficie de las mismas 4) deposición del ADN que presenta las perlas en una placa PicoTiter™; y 5) secuenciación simultánea en 100.000 pocillos mediante la generación de una señal luminosa de pirofosfato. El procedimiento se explicará más detalladamente a continuación.

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

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

En la primera etapa (a), los adaptadores de la secuenciación se ligan a los fragmentos en la genoteca de combinación. Dicho adaptador de secuenciación comprende por lo menos una región "clave" para emparejarse con una perla, una región del cebador de la secuenciación y una región del cebador de la PCR. De este modo, se obtienen los fragmentos adaptados. En una primera etapa, los fragmentos adaptados se emparejan con perlas, emparejándose cada perla con un único fragmento adaptado. Para la mezcla de fragmentos adaptados, se añaden suficientes perlas en exceso para garantizar la hibridación de un fragmento adaptado único por perla para la mayoría de las perlas (distribución de Poisson).

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

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

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

La propia pirosecuenciación resulta conocida en la técnica y se describe, entre otros, en [www.biotagebio.com](http://www.biotagebio.com); [www.pyrosequencing.com](http://www.pyrosequencing.com) / sección tecnología. La tecnología se aplica asimismo en, 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).

En la presente invención, las perlas presentan preferentemente secuencias de (enlace con los) cebadores y/o secciones de la sujeción o partes de las mismas que pueden enlazarse con los amplicones o las sondas ligadas, como puede ser el caso. En otras formas de realización, las sondas o los cebadores utilizados en la amplificación presentan secuencias que permiten el enlace de los amplicones o las sondas ligadas con las perlas para permitir la polimerización en emulsión posterior seguida por la secuenciación. Los amplicones secuenciados de las sondas ligadas proporcionarán la identidad del identificador y, de este modo, la presencia o ausencia de la secuencia elegida como objetivo en la muestra.

**TECNOLOGÍAS DE SOLEXA**

Uno de los procedimientos de secuenciación rápida se encuentra disponible en Solexa, Reino Unido (www.solexa.co.uk) y se describen, entre otros, en los documentos WO0006770, WO0027521, WO0058507, WO0123610, WO0157248, WO0157249, WO02061127, WO03016565, WO03048387, WO2004018497, WO2004018493, WO2004050915, WO2004076692, WO2005021786, WO2005047301, WO2005065814, WO2005068656, WO2005068089, WO2005078130. Sustancialmente, el procedimiento se inicia con fragmentos ligados a adaptadores de ADN genómico. El ADN ligado a adaptadores se une aleatoriamente a un tamiz denso de cebadores que se unen a una superficie sólida, normalmente en una celda de flujo. El otro extremo del fragmento ligado al adaptador se hibrida con un cebador complementario en la superficie. Los cebadores se extienden en presencia de nucleótidos y polimerasas en la denominada amplificación en puente en fase sólida para proporcionar fragmentos bicatenarios. La desnaturalización y la repetición de la amplificación en puente en fase sólida producen unos grupos densos de fragmentos amplificados distribuidos por la superficie. Se inicia la secuenciación añadiendo cuatro nucleótidos de terminación reversibles etiquetados de un modo distinto, cebadores y polimerasa a la celda de flujo. Tras la primera ronda de extensión del cebador, se detectan las etiquetas, se registra la identidad de las primeras bases incorporadas y se eliminan el 3' terminal bloqueado y el fluoróforo de la base incorporada. A continuación, se determina la identidad de la segunda base del mismo modo y, por lo tanto, continúa la secuenciación.

En la presente invención, las sondas ligadas o los amplicones, se unen con la superficie mediante la secuencia de enlace con el cebador, la secuencia del cebador o, en algunas formas de realización, la sección de sujeción o una combinación de los mismos. La secuencia se determina tal como se ha indicado, comprendiendo la secuencia del identificador y la secuencia elegida como objetivo asociada y se identifica su presencia o ausencia.

En formas de realización alternativas dirigidas a los procedimientos de secuenciación descritos en la presente memoria, las sondas o los cebadores utilizados en la amplificación pueden comprender secciones específicas (como alternativa al cebador descrito en la presente memoria o a las secuencias de enlace con el cebador) que se utilizan en la etapa de secuenciación posterior para unir las sondas ligadas / amplicones con la superficie. Estos se indican generalmente como la región clave.

A lo largo de la presente memoria, las figuras y las reivindicaciones adjuntas, las nociones "primero/a" y "segundo/a" se utilizan para distinguir entre las sondas utilizadas en el ensayo y sus componentes correspondientes. Los conceptos de "primero/a" y "segundo/a" no se utilizan en la presente memoria como sumatorios, es decir, no significan que únicamente puede existir un segundo componente, cuando existe asimismo un primer componente. Por motivos de coherencia y facilidad de referencia, dichos conceptos se utilizan asimismo cuando la propia forma de realización comprende dos sondas o dos componentes. Por ejemplo, una sonda circularizable, siendo únicamente una sonda, comprende todavía una primera y una segunda sección específica elegida como objetivo. Asimismo, en la figura 1A, tanto la primera como la segunda sonda pueden contener un identificador. En el caso de la primera sonda ello se representa como el primer identificador y en el caso de la segunda sonda ello se representa como el segundo identificador. En caso de que la segunda sonda comprenda un identificador y no la primera sonda, se puede hacer referencia a dicho identificador en la presente solicitud como segundo identificador sin implicar la existencia de un primer identificador.

**EJEMPLO**

Se aisló el ADN a partir de 2 progenitores y 88 descendientes utilizando procedimientos convencionales. Los progenitores (2x) y la descendencia (=4x) se encontraban en dúplex con distintos identificadores para probar la reproducibilidad. Las etiquetas utilizadas para distinguir muestras o alelos de SNP difieren entre sí en por lo menos 2 nucleótidos de cualquier otra etiqueta utilizada en los experimentos para la muestra o identificación de alelos, respectivamente. Se analiza la calidad en las distintas etapas utilizando geles de agarosa.

**Ejemplo 1:**

Para cada muestra de ADN se realizó una etapa de ligación utilizando conjuntos de sondas diseñados para detectar 30 locus SNP, comprendiendo cada uno 2 alelos. Las secuencias del cebador de amplificación se basan en las secuencias de hibridación que se encuentran en la superficie del sistema de secuenciación rápida Solexa. En particular, la secuencia P5 (PBS2) se sitúa en la parte 5' de la sonda de hibridación para TS2. La secuencia P7 (PBS1) se encuentra en el extremo 3' de la sonda que se hibrida con TS1.

Adyacente al extremo 3' de la secuencia P5, una secuencia que identifica una muestra degenerada (SIS) se encuentra comprendiendo una secuencia de NNNNN. Una secuencia de anclaje con el dinucleótido GC sigue el SIS. Es posible la identificación de alelos localizando una secuencia identificadora de 5 bases adyacente al extremo 3' de la sujeción GC. El complemento inverso del sitio de enlace con el cebador de la secuenciación (sPBS) sigue la secuencia identificadora del alelo (AIS).

La mezcla de ligación se amplifica utilizando las secuencias de cebador P5 y P7 comprendiendo el cebador P5 una secuencia de identificación de la muestra y una sujeción GC en su extremo 3'. Los productos son examinados según el rendimiento y la distribución de la longitud apropiada utilizando la detección de agarosa y se purifican posteriormente utilizando columnas Sephadex™. Se determinaron y se normalizaron las concentraciones y se crearon las mezclas. Se sometieron las mezclas a secuenciación masiva en paralelo basándose en la tecnología Solexa que comprende la amplificación en puente y la secuenciación seguida por el análisis de datos para determinar los genotipos de los progenitores y de la descendencia.

En un escenario alternativo se amplificó la mezcla de ligación mediante un método de 2 etapas.

Se realiza una primera amplificación utilizando las secuencias de los cebadores P5 y P7 sin secuencias de identificación de la muestra. Los productos son examinados según el rendimiento y la distribución de la longitud apropiada utilizando la detección de agarosa. La segunda PCR utiliza un cebador P7 junto con un cebador P5 que comprende un SIS y una sujeción GC en su extremo 3'. Los productos son examinados según el rendimiento y la distribución de la longitud apropiada utilizando la detección de agarosa. Se normalizaron las concentraciones y se mezclaron las muestras. Los productos se purificaron mediante columnas Qiagen.

En una segunda alternativa, la detección basada en la secuencia se realiza mediante la secuenciación de Sanger tras clonar los productos amplificados en un vector plasmídico bacteriano con la posterior transformación y amplificación por PCR de las colonias.

El análisis de los datos proporciona los genotipos.

**Ejemplo 2:**

La detección basada en la secuencia de productos de ligación amplificados se realizó mediante la secuenciación de Sanger.

La secuencia experimental implica una etapa de ligación, una primera amplificación utilizando cebadores sin SIS y sujeción GC, una segunda amplificación utilizando cebadores que comprenden SIS y sujeción GC, la clonación de los productos de la segunda amplificación en un plásmido bacteriano con la posterior transformación a un hospedador bacteriano, la amplificación por PCR de una colonias única y la determinación de la secuencia mediante la secuenciación de Sanger.

Se utilizaron y secuenciaron las estirpes progenitoras de maíz B73 y Mo17. Se secuenciaron más de 200 productos de ligación amplificados fueron para proporcionar el estudio demostrativo preliminar para la detección de SNP basada en la secuencia.

Se seleccionaron las estirpes progenitoras de maíz B73 y MO17. Se prepararon los productos de ligación utilizando conjuntos de sondas de ligación diseñados para detectar 30 locus de SNP. Se realizó la amplificación utilizando cebadores sin SIS y sujeción GC. Se prepararon los fragmentos para la clonación utilizando cebadores que comprendían SIS y sujeción GC. Se realizó la clonación utilizando un kit de clonación TOPO TA de Invitrogen.

Se prepararon los fragmentos de plantilla para la secuenciación de Sanger realizando una amplificación por PCR en colonias individuales.

Se identificaron los marcadores de ligación basados en la secuencia extrayendo etiquetas de secuencias de 27 pb observadas adyacentes al sitio de enlace con el cebador de la secuencia con frecuencias distintas en B73 y Mo17.

Los datos de los marcadores de ligación basados en las secuencias se compararon con las puntuaciones obtenidas por los marcadores SNPWave mediante una obtención convencional de la huella genética SNPWave utilizando la detección basada en longitud.

Se demostró la viabilidad de la detección de marcadores de ligación basada en la secuencia mediante la secuenciación de Sanger, con la que se puntuó un número inferior de marcadores de la ligación utilizando la detección basada en la secuencia que con las placas de gel convencionales. Sin embargo, la escala de secuenciación aumentó rápidamente en gran medida utilizando la tecnología de secuenciación masiva en paralelo de, por ejemplo, Solexa. Las comparaciones de datos de marcadores indicaron unos patrones de segregación idénticos entre la detección basada en la secuencia y la detección en placas de gel.

## REFERENCIAS CITADAS EN LA DESCRIPCIÓN

5 *La presente lista de referencias citadas por el solicitante se presenta únicamente para la comodidad del lector. No forma parte del documento de patente europea. Aunque la recopilación de las referencias se ha realizado muy cuidadosamente, no se pueden descartar errores u omisiones y la Oficina Europea de Patentes declina toda responsabilidad en este sentido.*

## Documentos de patente citados en la descripción

- US 4988617 A, Landegren [0004] [0005]
- US 5876924 A [0005]
- WO 9804745 A [0005]
- WO 9804746 A [0005]
- US 6221603 B [0005]
- US 5521065 A [0005] [0059]
- US 5962223 A [0005]
- EP 185494 B1 [0005]
- US 6027889 A [0005]
- EP 246864 B1 [0005]
- US 6156178 A [0005]
- EP 745140 B1 [0005]
- EP 964704 B1 [0005]
- WO 03054511 A [0005]
- US 20030119004 A [0005]
- US 2003190646 A [0005]
- EP 1313880 A [0005]
- US 20030032016 A [0005]
- EP 912761 A [0005]
- EP 956359 A [0005]
- US 2003108913 A [0005]
- EP 1255871 A [0005]
- EP 1194770 A [0005]
- EP 1252334 A [0005]
- WO 9615271 A [0005]
- WO 745559 A [0005]
- US 20030119004 A1 [0005]
- US 5470705 A [0005]
- WO 2004111271 A [0006] [0033]
- WO 2005021794 A [0006] [0045]
- WO 2005118847 A [0006] [0046]
- WO 03052142 A [0006] [0030]
- EP 745140 A [0030]
- WO 9914226 A [0039]
- WO 0056748 A [0039]
- WO 0066604 A [0039]
- WO 0125478 A [0039]
- EP 0974672 A [0039]
- EP 0534858 A [0048] [0060]
- EP 974672 A [0051]
- US 5476930 A [0057]
- WO 0161033 A [0058]
- EP 185494 A [0059]
- US 5692223 A [0059]
- WO 03054311 A [0059]
- WO 0077260 A [0059]
- US 5185243 A [0059]
- EP 439182 A [0059]
- EP 320308 A [0059]
- WO 9001069 A [0059]
- EP 534858 A [0064] [0065]
- WO 03004690 A [0068] [0069] [0075]
- WO 03054142 A [0068] [0069] [0075]
- WO 2004069849 A [0068] [0069] [0075]
- WO 2004070005 A [0068] [0069] [0075]
- WO 2004070007 A [0068] [0069] [0075]
- WO 2005003375 A [0068] [0069] [0075]
- WO 0006770 A [0077]
- WO 0027521 A [0077]
- WO 0058507 A [0077]
- WO 0123610 A [0077]
- WO 0157248 A [0077]
- WO 0157249 A [0077]
- WO 02061127 A [0077]
- WO 03016565 A [0077]
- WO 03048387 A [0077]
- WO 2004018497 A [0077]
- WO 2004018493 A [0077]
- WO 2004050915 A [0077]
- WO 2004076692 A [0077]
- WO 2005021786 A [0077]
- WO 2005047301 A [0077]
- WO 2005065814 A [0077]
- WO 2005068656 A [0077]
- WO 2005068089 A [0077]
- WO 2005078130 A [0077]

Documentos que no corresponden a patentes citados en la descripción

- **NILSSON et al.** *Human mutation*, 2002, vol. 19, 410-415 [0005]
- *Science*, 1994, vol. 265, 2085-2088 [0005]
- **VAN EIJK et al.** *Nucleic Acids Research*, vol. 32, e47 [0030]
- **MEINKOTH et al.** *Anal. Biochem.*, 1984, vol. 138, 267-284 [0038]
- **GRYAZNOV ; LETSINGER.** *Nucleic Acid Res.*, 1993, vol. 21, 1403-08 [0057]
- **GRYAZNOV et al.** *Nucleic Acid Res.*, 1994, vol. 22, 2366-69 [0057]
- **KANAYA ; YANAGAWA.** *Biochemistry*, 1986, vol. 25, 7423-30 [0057]
- **LUEBKE ; DERVAN.** *Nucleic Acids Res.*, 1992, vol. 20, 3005-09 [0057]
- **SIEVERS ; VON KIEDROWSKI.** *Nature*, 1994, vol. 369, 221-24 [0057]
- **LIU ; TAYLOR.** *Nucleic Acids Res.*, 1999, vol. 26, 3300-04 [0057]
- **WANG ; KOOL.** *Nucleic Acids Res.*, 1994, vol. 22, 2326-33 [0057]
- **PURMAL et al.** *Nucleic Acids Res.*, 1992, vol. 20, 3713-19 [0057]
- **ASHLEY ; KUSHLAN.** *Biochemistry*, 1991, vol. 30, 2927-33 [0057]
- **IANNONE et al.** *Cytometry*, 2000, vol. 39, 131-140 [0040] [0050]
- **SAMBROOK ; RUSSEL.** *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, 2001 [0051]
- **XU et al.** *Nucleic Acid Res.*, 1999, vol. 27, 875-81 [0057]
- **CHU ; ORGEL.** *Nucleic Acids Res.*, 1988, vol. 16, 3671-91 [0057]
- **SOKOLOVA et al.** *FEBS Letters*, 1988, vol. 232, 153-55 [0057]
- **NAYLOR ; GILHAM.** *Biochemistry*, 1966, vol. 5, 2722-28 [0057]
- **WU ; WALLACE.** *Gene*, 1989, vol. 76, 245-254 [0058]
- **AUSUBEL et al.** *Current Protocols in Molecular Biology.* John Wiley & Sons, Inc, 1995 [0060]
- **VOS et al.** *Nucleic Acids Res.*, 1995, vol. 23, 4407-4414 [0060]
- **VOS et al.** *Nucleic Acid Research*, 1995, vol. 23, 4407-44014 [0064] [0065]
- **SEO et al.** *Proc. Natl. Acad. Sci. USA*, 2004, vol. 101, 548893 [0068]

## REIVINDICACIONES

1. Procedimiento para la detección rápida de por lo menos 100 secuencias de nucleótidos elegidas como objetivo en una pluralidad de muestras, comprendiendo el procedimiento las etapas de:

(a) proporcionar para cada secuencia de nucleótidos elegida como objetivo en cada una de las muestras:

I. una primera sonda y una segunda sonda, comprendiendo la primera sonda una sección específica elegida como objetivo en su extremo 3' y una primera sección etiquetada que no es complementaria a la secuencia de nucleótidos elegida como objetivo y que comprende una primera secuencia de enlace con un cebador, comprendiendo la segunda sonda una segunda sección específica elegida como objetivo en su extremo 5' y una segunda sección etiquetada que no es complementaria a la secuencia de nucleótidos elegida como objetivo y que comprende una segunda secuencia de enlace con el cebador, comprendiendo la primera o la segunda sección etiquetada(s), o ambas, una secuencia identificadora que se encuentra entre la secuencia primera o segunda correspondiente de enlace con el cebador y la sección específica elegida como objetivo primera o segunda correspondiente,

o

II. una sonda circularizable, en la que las sondas primera y segunda tal como se definen en la alternativa I se combinan en una sonda;

o

III. una primera y una segunda sonda tal como se definen en la alternativa I, comprendiendo además la primera sonda una sección de sujeción que puede hibridarse con una sección de sujeción complementaria que se encuentra en la segunda sonda, por lo que las secciones de sujeción son sustancialmente no complementarias a la secuencia elegida como objetivo.

o

IV. una primera y una segunda sonda tal como se definen en la alternativa I, excepto en que la primera sonda no comprende una primera secuencia de enlace con el cebador y en que se proporciona una sonda del compuesto que comprende una sección que puede hibridarse con una parte de la primera sección específica elegida como objetivo la primera sonda y comprende además una sección de enlace con el cebador, y en el que el alargamiento de la sonda del compuesto proporciona una sonda del compuesto alargada amplificable,

(b) permitir que la primera y la segunda sección específica elegidas como objetivo de las sondas primera y segunda correspondientes o la sonda circularizable para hibridarse con la secuencia elegida como objetivo,

(c) ligar las sondas primera y segunda o la sonda circularizable cuando las secciones específicas elegidas como objetivo correspondientes de las sondas o la sonda circularizable se hibridan con secciones sustancialmente adyacentes en la secuencia elegida como objetivo para proporcionar sondas ligadas,

(d) opcionalmente, amplificar las sondas ligadas con un primer y un segundo cebador para proporcionar amplicones,

(e) someter las sondas ligadas o amplicones a la tecnología de secuenciación rápida para determinar por lo menos una parte de la secuencia de nucleótidos, comprendiendo por lo menos la(s) secuencia(s) identificadora(s), de las sondas ligadas o los amplicones,

(f) identificar la presencia de la secuencia de nucleótidos elegida como objetivo en la muestra determinando la presencia o ausencia de la secuencia identificadora en la secuencia de nucleótidos de la etapa (e).

2. Procedimiento según la reivindicación 1, en el que las sondas primera y segunda se emparejan con secciones no adyacentes en la secuencia elegida como objetivo y se proporcionan una polimerasa y nucleótidos de tal modo que se llena la distancia entre las sondas primera y segunda.

3. Procedimiento según cualquiera de las reivindicaciones 1 a 2, en el que la secuenciación se realiza mediante secuenciación rápida sobre un soporte sólido.

4. Método según la reivindicación 1, en el que la secuenciación rápida comprende las etapas de:

- emparejar las sondas ligadas o amplicones con perlas, emparejándose cada perla con una única sonda ligada o amplicón;

- emulsionar las perlas en microrreactores, comprendiendo cada microrreactor de agua en aceite una única perla;

- realizar la PCR en emulsión para amplificar sondas ligadas o amplicones en la superficie de las perlas;

- opcionalmente, seleccionar / enriquecer las perlas que comprenden fragmentos amplificados ligados adaptadores;

- cargar las perlas en pocillos, comprendiendo cada pocillo una única perla; y

- generar una señal de pirofosfato.

5. Método según la reivindicación 1, en el que la secuenciación rápida comprende las etapas de:

- emparejar las sondas ligadas o amplicones con perlas, emparejándose cada perla con una única sonda ligada o amplicón;

- emulsionar las perlas en microrreactores, comprendiendo cada microrreactor de agua en aceite una única perla;

- realizar la PCR en emulsión para amplificar sondas ligadas o amplicones en la superficie de las perlas;

- opcionalmente, seleccionar / enriquecer las perlas que comprenden fragmentos amplificados ligados adaptadores;

- cargar las perlas en pocillos, comprendiendo cada pocillo una única perla; y

- determinar la secuencia de nucleótidos de las sondas ligadas amplificadas o amplicones utilizando nucleótidos de terminación reversibles etiquetados.

6. Método según la reivindicación 1, en el que la secuenciación rápida comprende las etapas de:

- emparejar las sondas ligadas o amplicones con una superficie que comprende los cebadores primero y segundo o las secuencias de enlace con el cebador primera y segunda, respectivamente,
- realizar la amplificación en puente para proporcionar grupos de sondas amplificadas o amplicones,
- determinar la secuencia de nucleótidos de las sondas ligadas amplificadas o amplicones utilizando nucleótidos de terminación reversibles etiquetados.

7. Método según la reivindicación 1, en el que la secuenciación rápida comprende las etapas de:

- emparejar las sondas ligadas o amplicones con una superficie que comprende los cebadores primero y segundo o las secuencias de enlace con el cebador primera y segunda, respectivamente,
- realizar la amplificación en puente para proporcionar grupos de sondas amplificadas o amplicones,
- generar una señal de pirofosfato.

8. Procedimiento según la reivindicación 1, en el que el identificador es un identificador de muestras y/o un identificador combinado de locus / alelos.

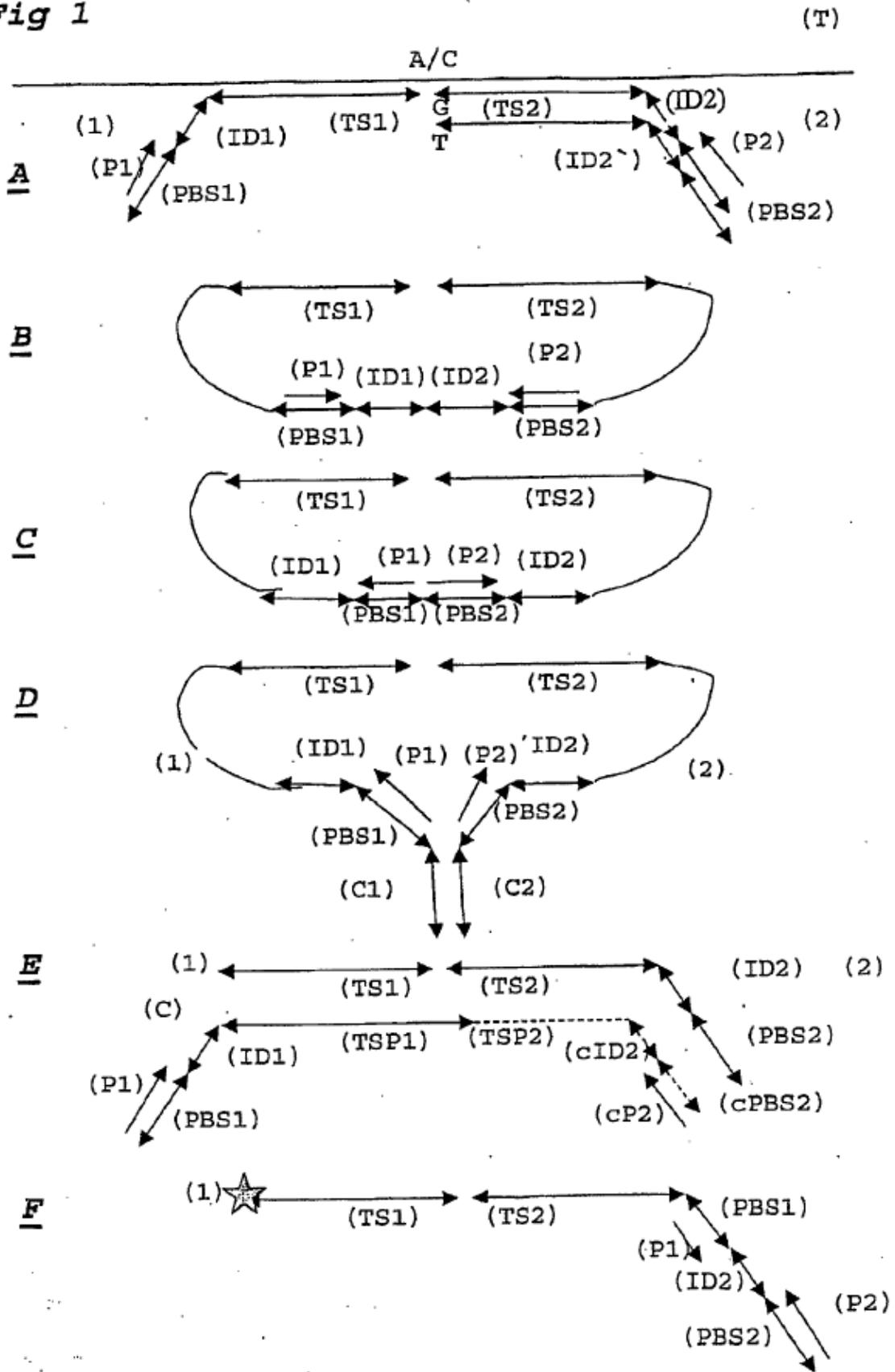
9. Procedimiento según la reivindicación 1, en el que las sondas ligadas se amplifican con un primer y un segundo cebador para proporcionar amplicones.

10. Procedimiento según la reivindicación 9, en el que el cebador comprende una secuencia de anclaje 3'.

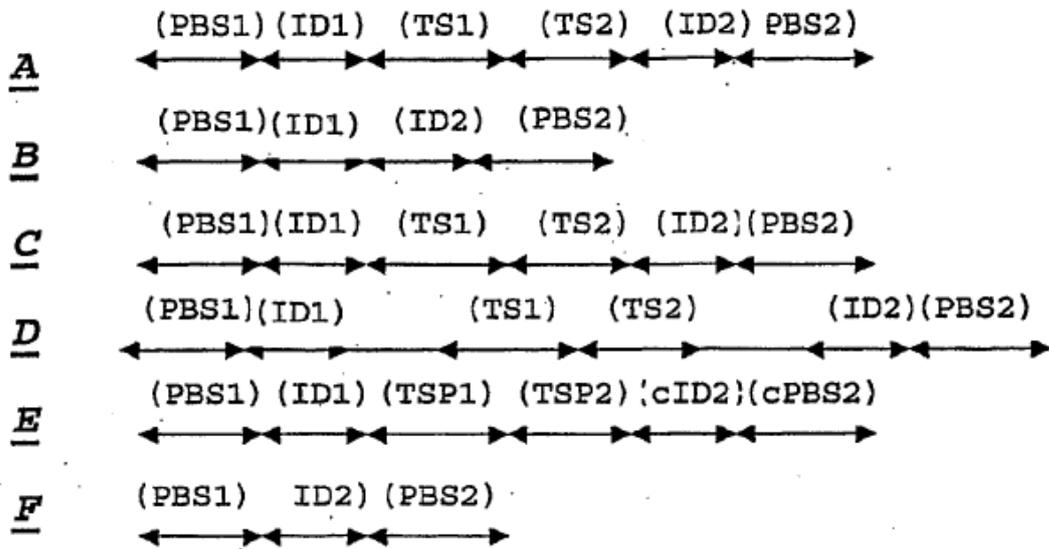
11. Procedimiento según la reivindicación 1, en el que para dos o más muestras o para dos o más combinaciones de locus / alelo, las secuencias identificadoras se utilizan para determinar el genotipo de la(s) muestra(s) para una o más secuencias y/o polimorfismos, tales como SNP y/o indeles.

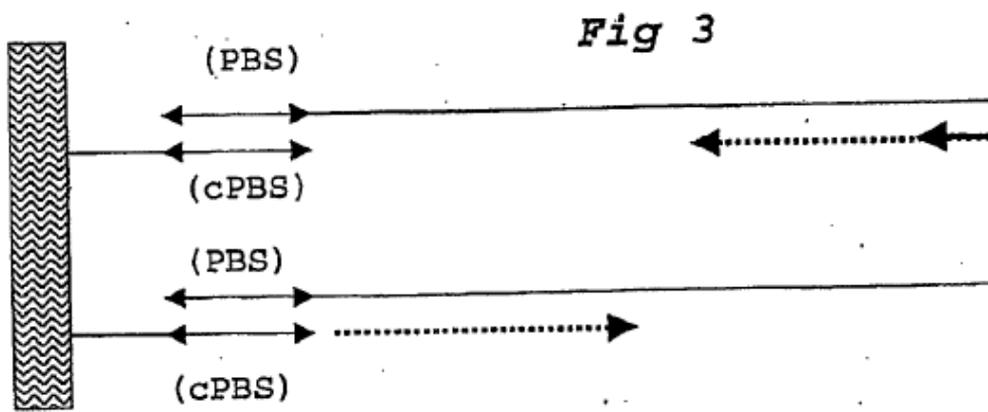
12. Uso de la secuenciación rápida en la determinación de la presencia o ausencia de por lo menos 100 secuencias elegidas como objetivo en una pluralidad de muestras utilizando un ensayo basado en la ligación, en el que para la pluralidad de secuencias elegidas como objetivo se proporcionan una primera sonda y una segunda sonda, comprendiendo la primera sonda una sección específica de como objetivo en su extremo 3' y una primera sección etiquetada que no es complementaria a la secuencia de nucleótidos elegida como objetivo y que comprende una primera secuencia de enlace con un cebador, comprendiendo la segunda sonda una segunda sección específica elegida como objetivo en su extremo 5' y una segunda sección etiquetada que no es complementaria a la secuencia de nucleótidos elegida como objetivo y que comprende una segunda secuencia de enlace con el cebador, comprendiendo la primera o la segunda sección etiquetada(s), o ambas, una secuencia identificadora que se encuentra entre la secuencia primera o segunda correspondiente de enlace con el cebador y la sección específica elegida como objetivo primera o segunda correspondiente, en el que el identificador no comprende 2 o más bases consecutivas idénticas y en el que todos los identificadores difieren entre sí en por lo menos dos bases, y en el que las sondas primera y segunda se ligan cuando las secciones específicas elegidas como objetivo correspondientes de las sondas se hibridan con secciones sustancialmente adyacentes de la secuencia elegida como objetivo para proporcionar sondas ligadas y las sondas ligadas se someten a secuenciación rápida para determinar por lo menos una parte de la secuencia de nucleótidos, comprendiendo por lo menos la(s) secuencia(s) identificador(as), de las sondas ligadas.

Fig 1

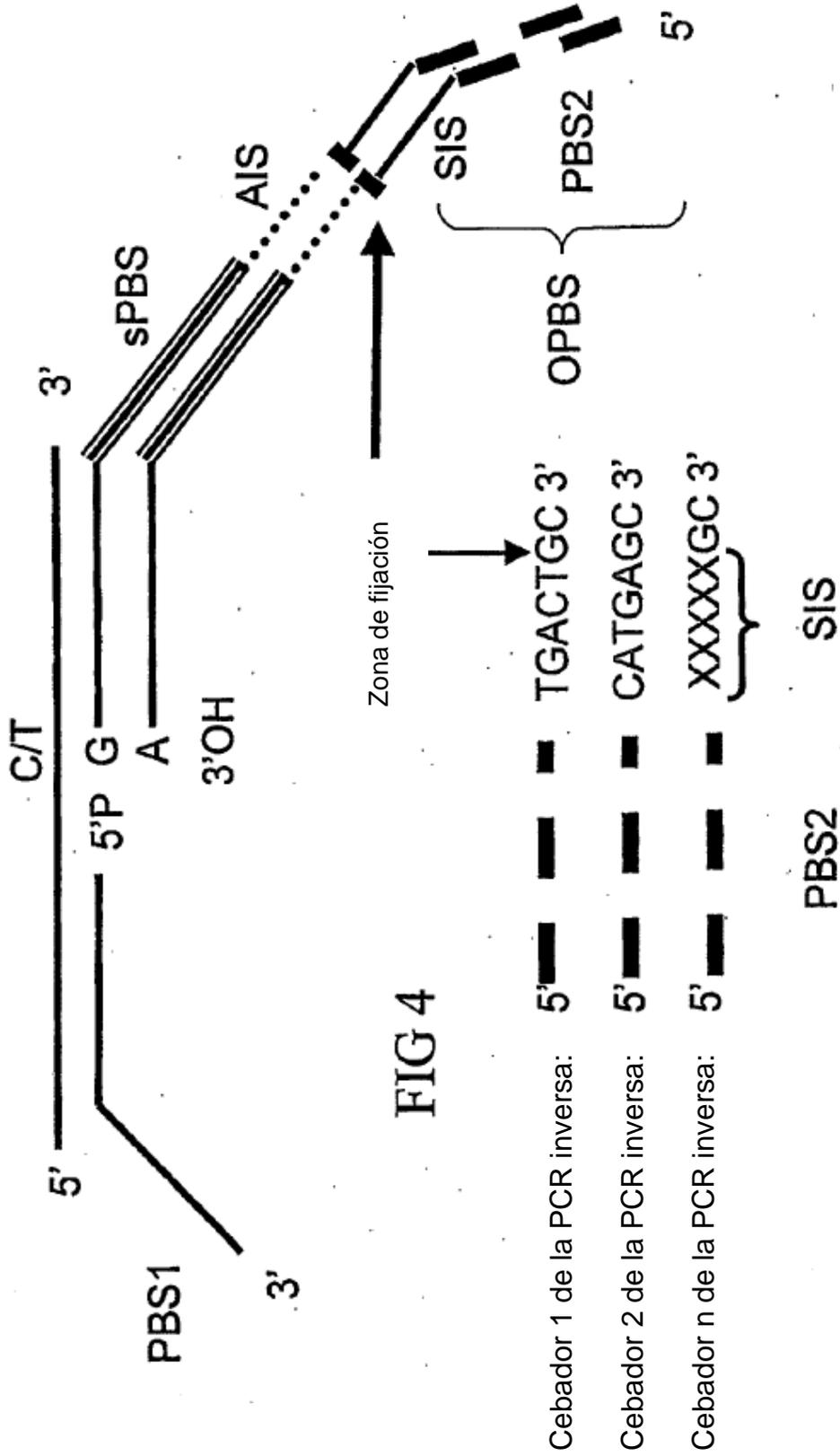


**Fig 2**





Secuenciación unidireccional o bidireccional por síntesis



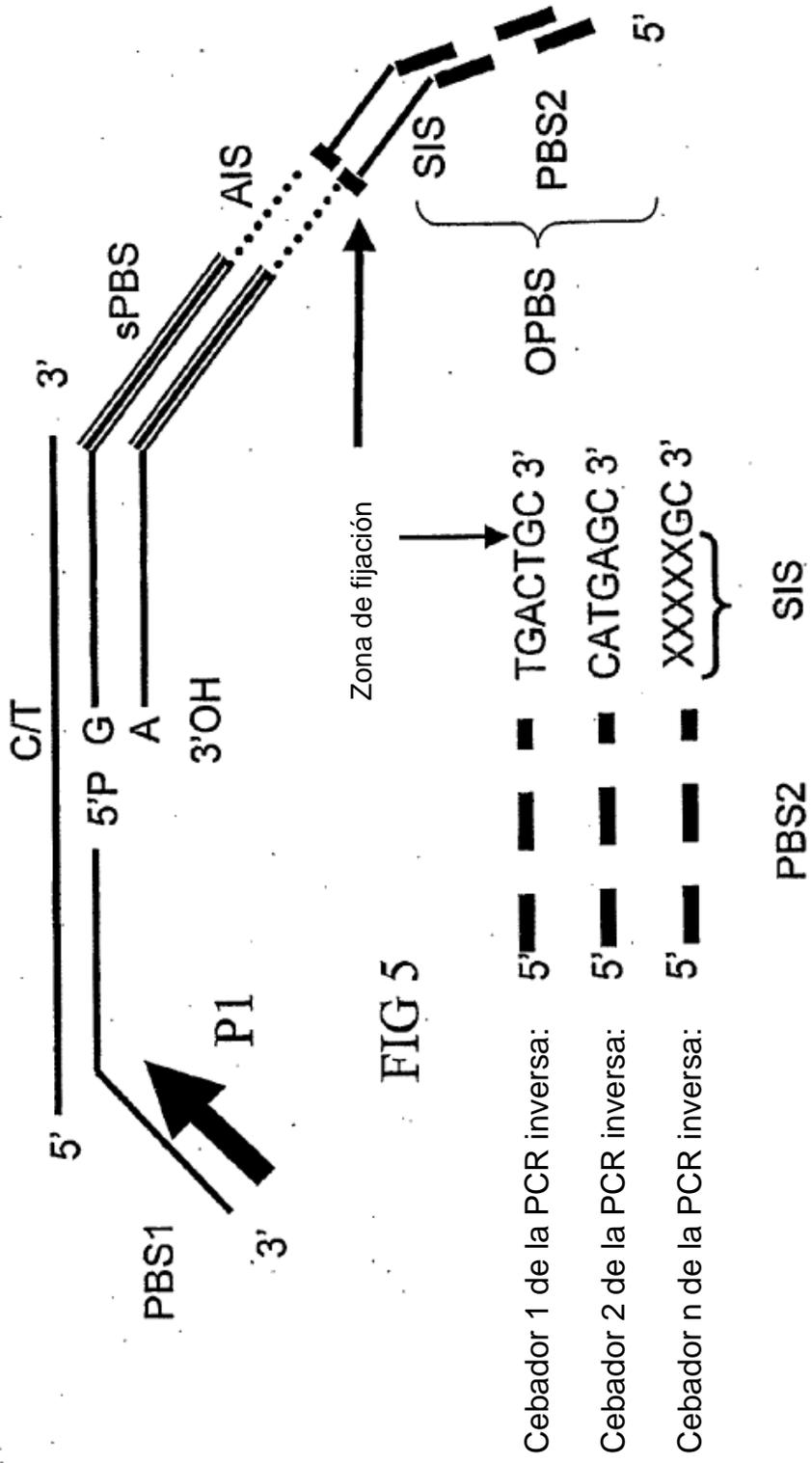


FIG 5

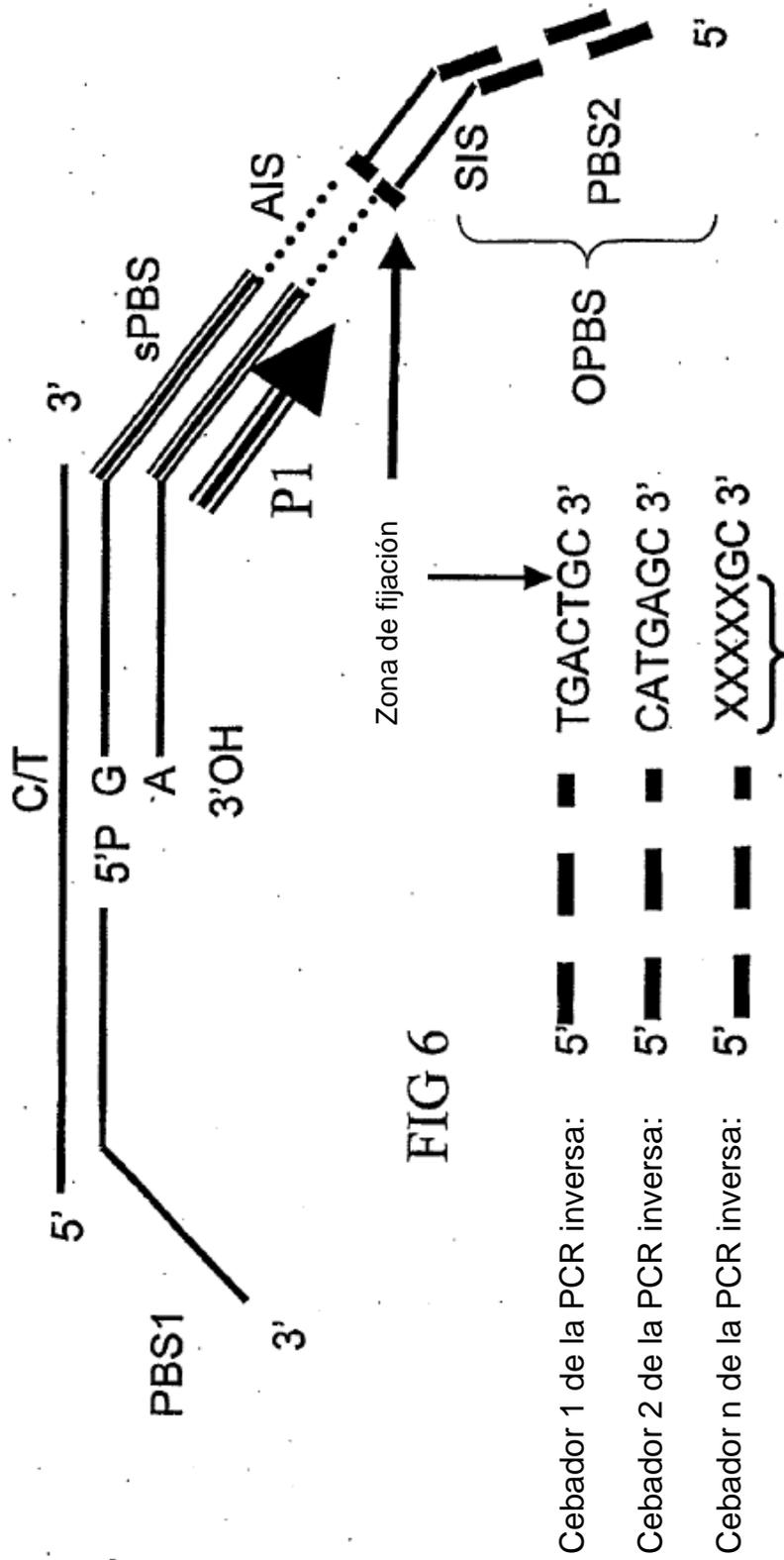


FIG 6

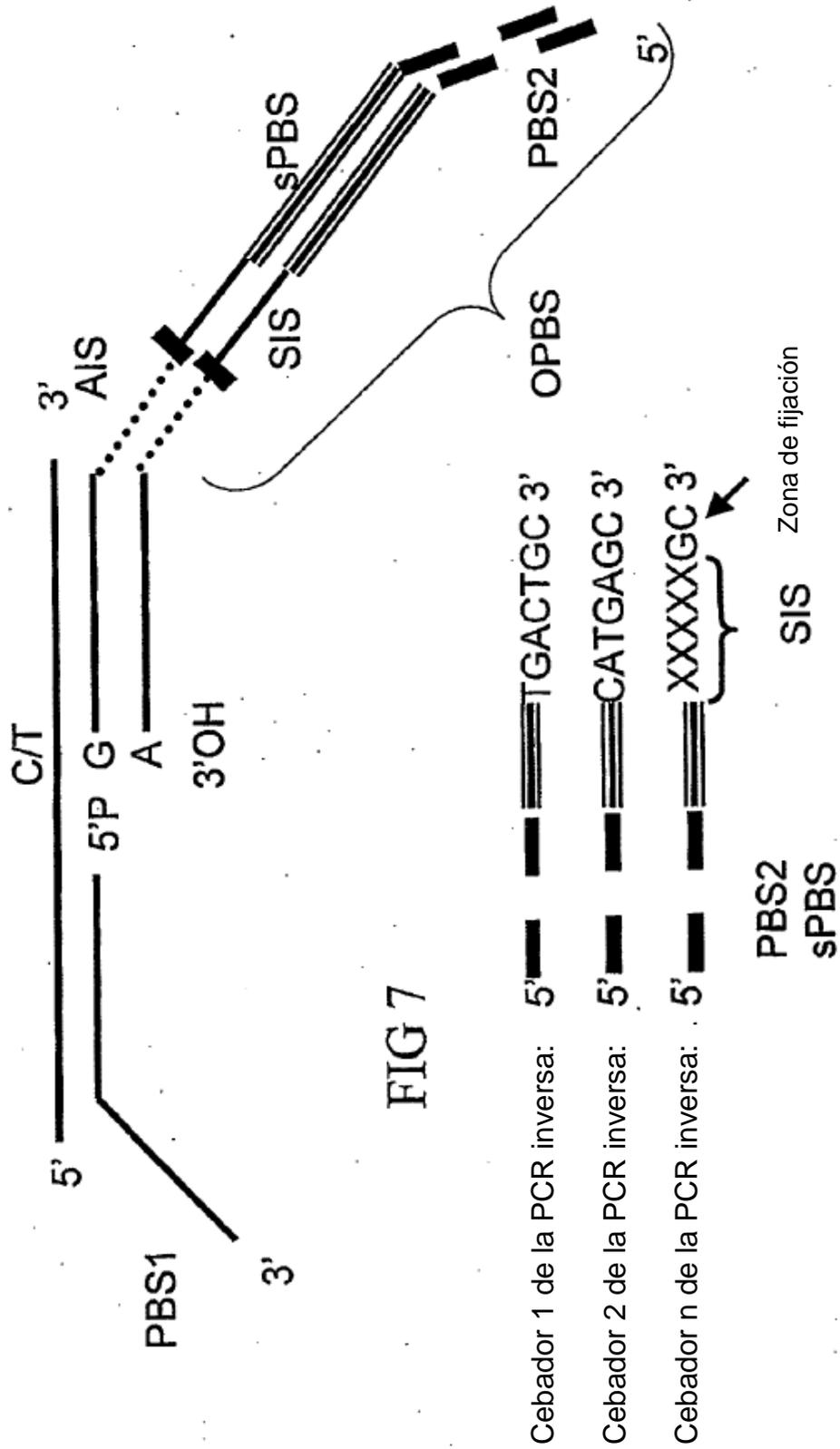


FIG 7

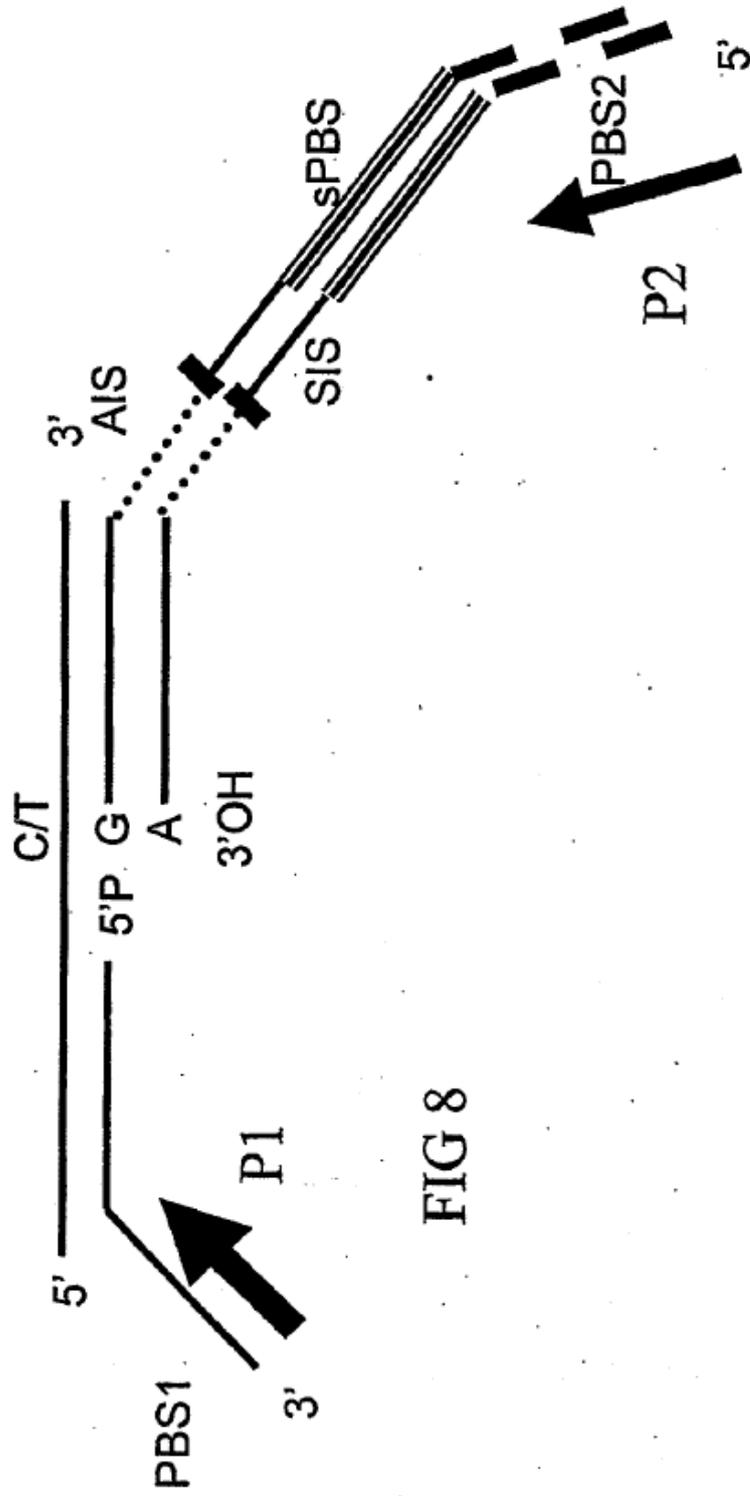


FIG 8

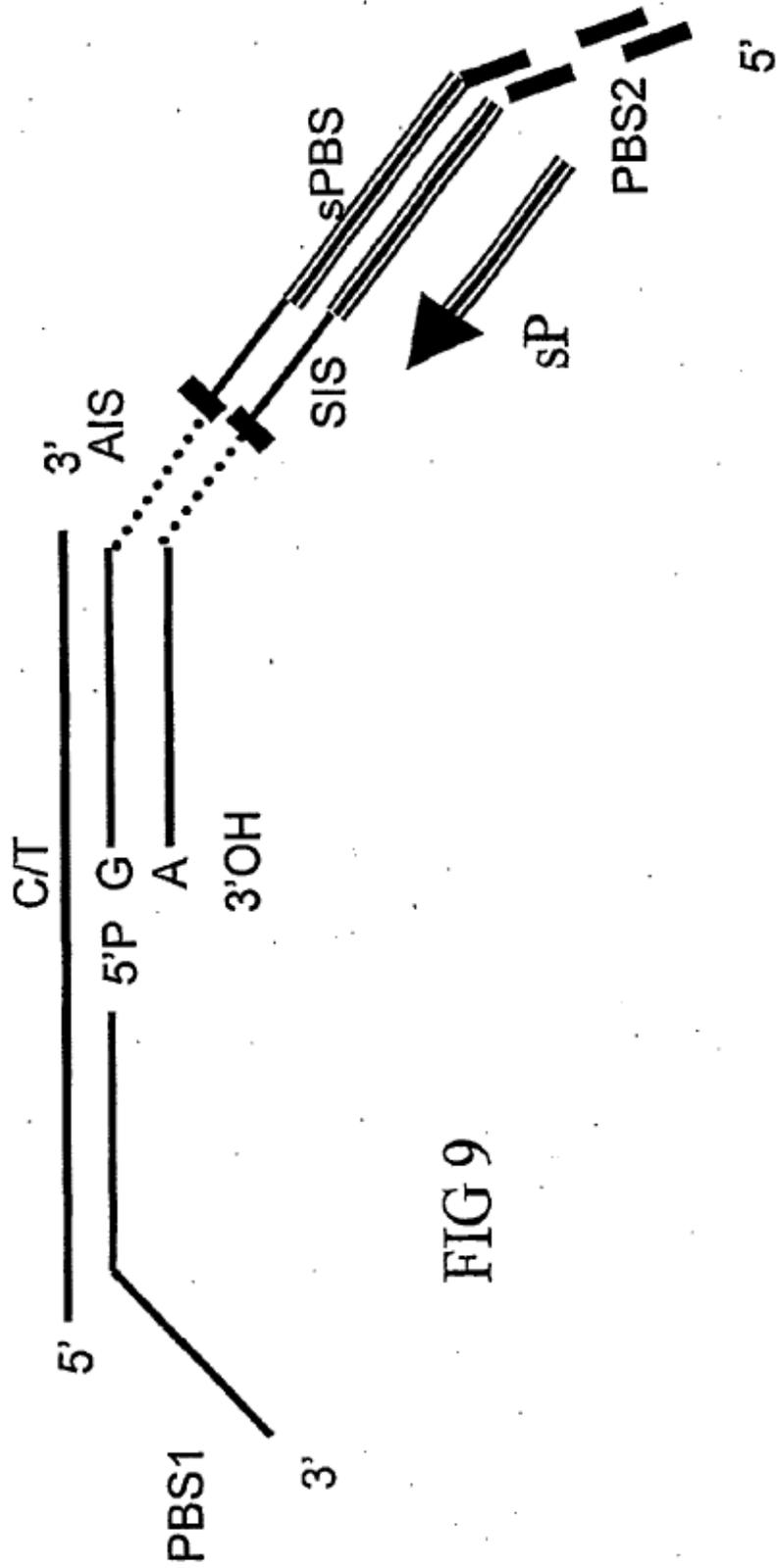


FIG 9

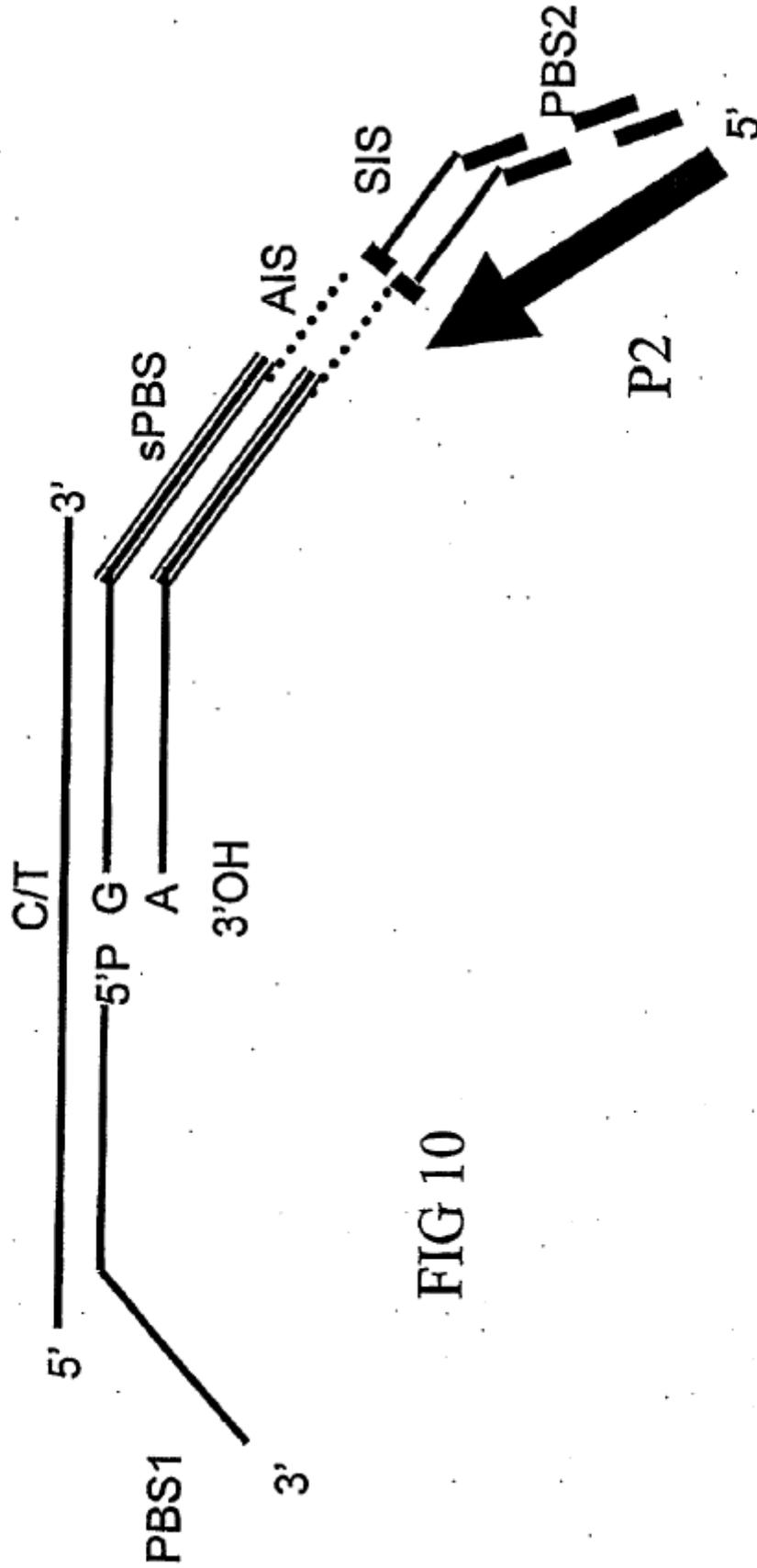


FIG 10

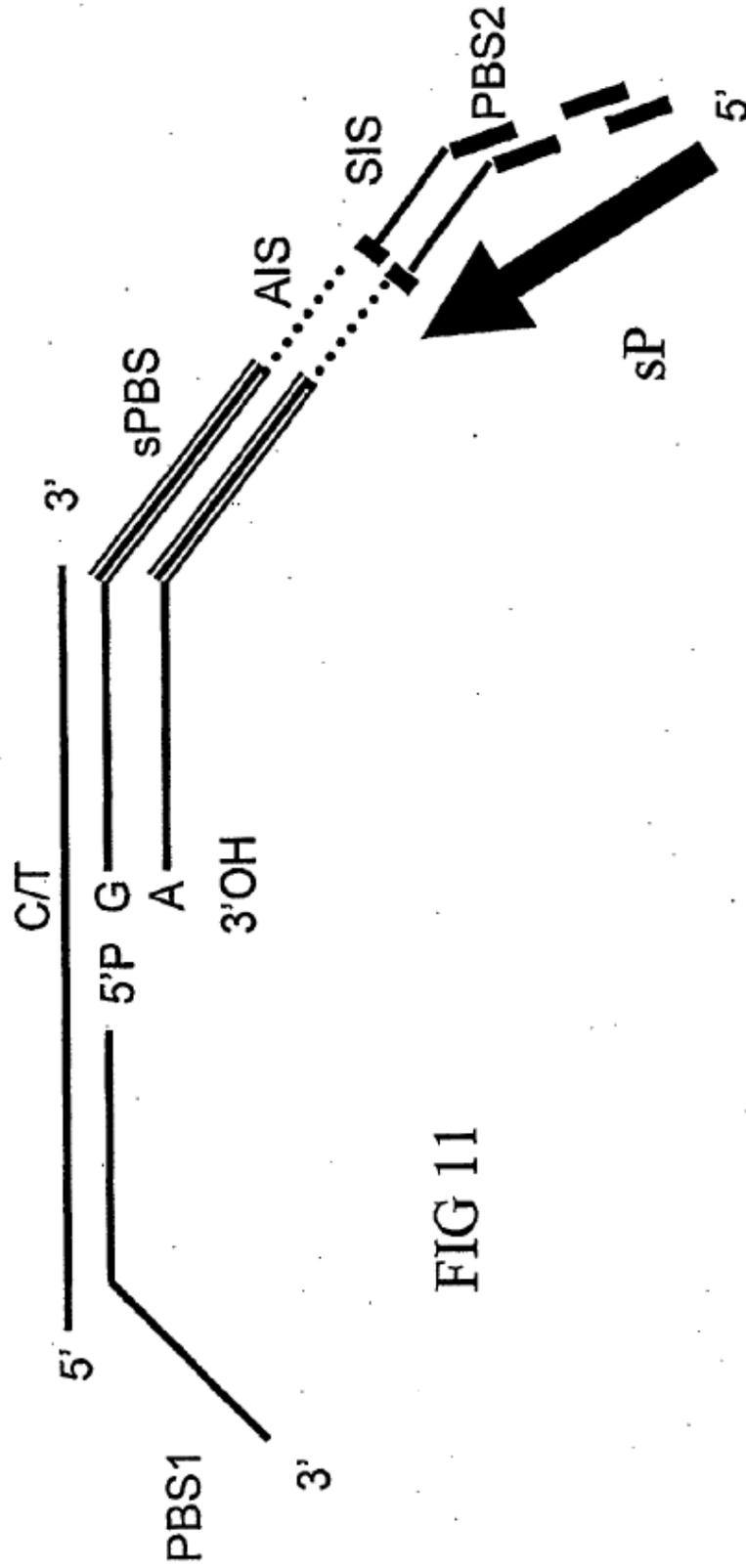


FIG 11

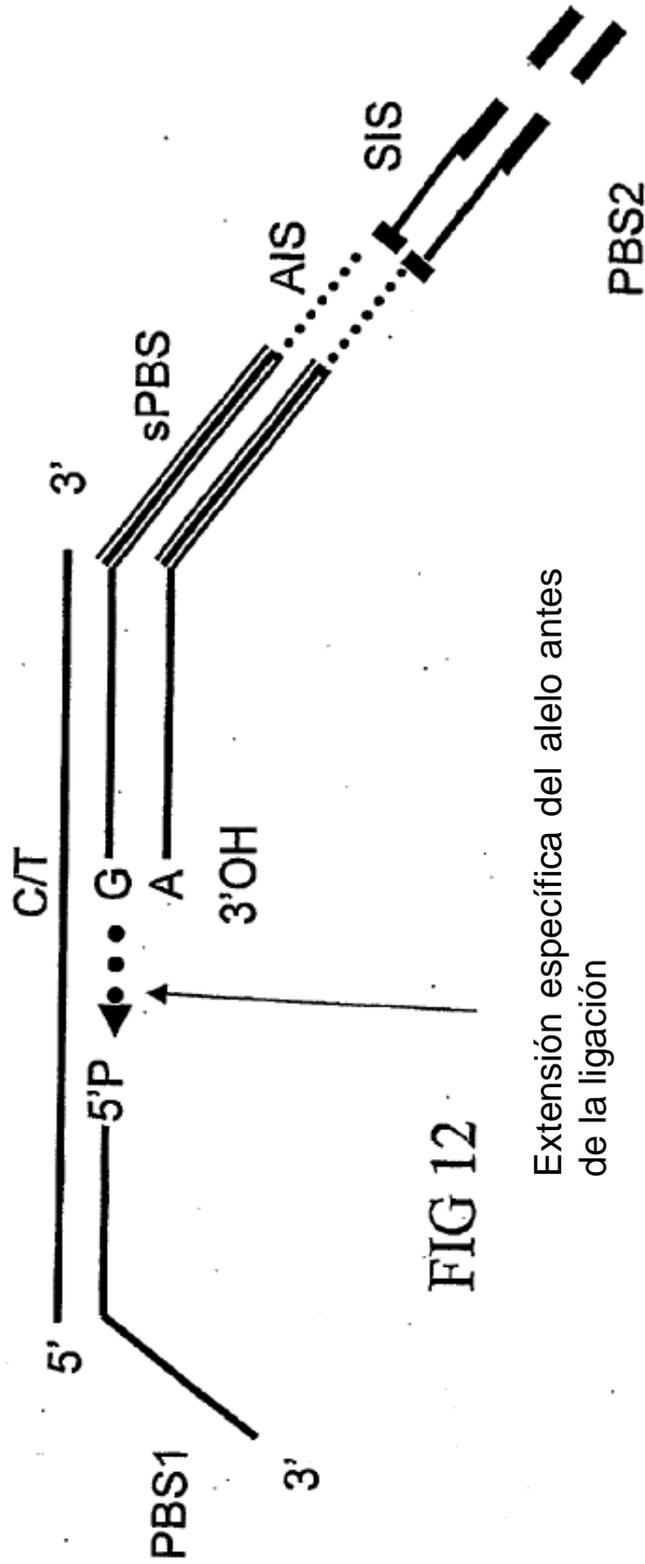


FIG 12

Extensión específica del alelo antes de la ligación

Fig 13

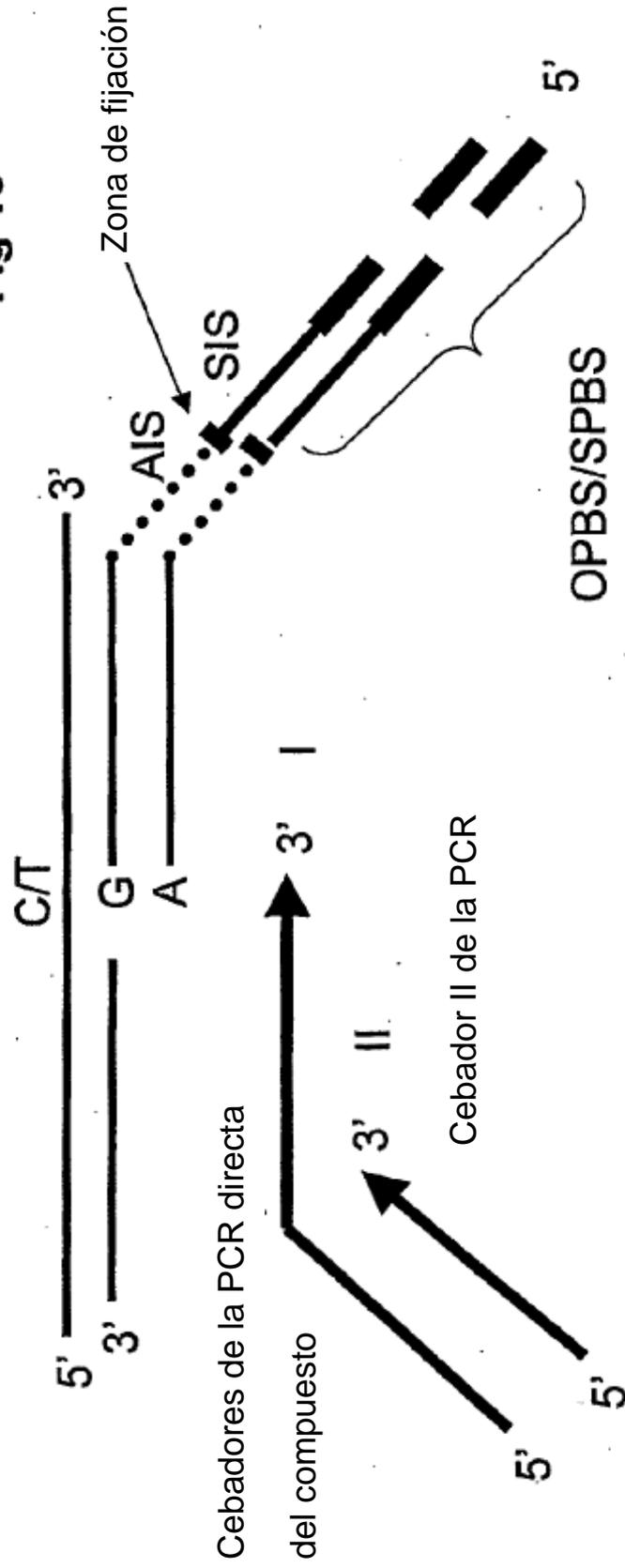


Fig 14

