

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 741 099**

51 Int. Cl.:

C12Q 1/6806 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **26.02.2013** **E 17175537 (4)**

97 Fecha y número de publicación de la concesión europea: **19.06.2019** **EP 3287531**

54 Título: **Método de fijación de una secuencia de recuento para una muestra de ácido nucleico**

30 Prioridad:

28.02.2012 US 201261604360 P

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

10.02.2020

73 Titular/es:

AGILENT TECHNOLOGIES, INC. (100.0%)
5301 Stevens Creek Blvd.
Santa Clara, CA 95051, US

72 Inventor/es:

OSBORNE, ROBERT;
CASBON, JAMES;
CLAAS, ANDREAS;
MIKAWA, GI y
MUSGROVE-BROWN, ESTHER

74 Agente/Representante:

MILTENYI , Peter

ES 2 741 099 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

Método de fijación de una secuencia de recuento para una muestra de ácido nucleico

5 **Antecedentes**

En ciertos métodos, una pluralidad de etiquetas oligonucleotídicas se unen a los polinucleótidos de la muestra de ácido nucleico inicial para proporcionar una muestra marcada en la que esencialmente cada molécula polinucleotídica diana en la muestra se une a una etiqueta oligonucleotídica diferente, es decir, una etiqueta oligonucleotídica que tiene una secuencia que es diferente a otras etiquetas oligonucleotídicas. Después de que los polinucleótidos diana se hayan amplificado, el número de polinucleótidos diana en la muestra de ácido nucleico inicial puede estimarse contando el número de etiquetas diferentes que están asociadas con los polinucleótidos diana amplificados. Se puede pensar en la pluralidad de etiquetas como una secuencia de recuento puesto que proporcionan una forma en que se pueden contar las moléculas polinucleotídicas individuales, incluso después de que se amplifiquen.

Sumario

La invención es como se define en las reivindicaciones. En ciertas realizaciones, el método de la invención comprende: (a) hibridar una población de cebadores directos de la fórmula 5'-A-Y-Z con una población de moléculas de ácido nucleico molde, donde (i) la región A de los cebadores directos proporciona un sitio de unión a cebador cuando se copia y es la misma en cada cebador de la población de cebadores directos; (ii) la región Y de los cebadores directos varía entre los diferentes cebadores de la población de cebadores y proporciona una secuencia de recuento; y (iii) la región Z de los cebadores directos es complementaria a un primer sitio en un polinucleótido diana en una población de moléculas de ácido nucleico molde y es la misma en cada cebador de la población de cebadores. En la población de cebadores directos, la región Y puede definirse por una ejecución de bases degeneradas (EBD) que comprende al menos una base nucleotídica seleccionada entre: R, Y, S, W, K, M, B, D, H, V y N.

En este método, la etapa de hibridación se realiza en presencia de un segundo cebador que comprende la secuencia de la región A y un cebador inverso que se hibrida con un sitio que está corriente abajo de los cebadores directos. El cebador inverso contiene una cola 5' (es decir, una secuencia que no se hibrida con el ácido nucleico diana y proporciona un sitio de unión a cebador para un tercer cebador). La hibridación se realiza en presencia de un tercer cebador que tiene la secuencia de la cola 5' del cebador inverso. Después de que se haya realizado la etapa de hibridación inicial, el método prosigue por medio de un ciclo a través de dos rondas de extensión del cebador. Las dos rondas de extensión del cebador comprenden: (b) extender los cebadores directos que se hibridan con el polinucleótido diana utilizando el polinucleótido diana como un molde para producir una población de primeros productos de extensión que comprenden un sitio de unión para el cebador inverso; (c) hibridar el cebador inverso con el sitio de unión para el cebador inverso en la primera población de productos de extensión producidos por la etapa (b); (d) extender el cebador inverso para producir una población de segundos productos de extensión que comprende, en sus extremos 3', el complemento de la región A y el complemento de la región Y, donde el complemento de la región Y es diferente en cada molécula en la población de segundos productos de extensión.

Después de que se hayan completado las primeras dos rondas de extensión del cebador con el fin de producir una población de segundos productos de extensión, el método comprende (e) desactivar selectivamente, después de la etapa (d) cualquier cebador directo que no esté extendido, donde en esta etapa "desactivar selectivamente" significa evitar selectivamente que los cebadores directos, pero no el segundo cebador que comprende la secuencia de la región A, se extiendan en la siguiente reacción de PCR (descrita más adelante). Dependiendo de cómo se implemente esta etapa, los cebadores directos que se han hibridado con su diana pero no se han extendido, así como los cebadores directos que no se hibridan con su diana pueden ser desactivados selectivamente. Además, en ciertos casos, la etapa de desactivación selectiva también puede alterar físicamente los cebadores directos extendidos (es decir, los cebadores directos que se han extendido para convertirse en los primeros productos de extensión) en el sentido de que pueden ser escindidos o secuestrados además de los cebadores directos no extendidos. Tal alteración de los cebadores directos extendidos no debería afectar el resultado general del método ya que el segundo producto de extensión (que contiene una copia de la secuencia de recuento y no contiene el cebador directo extendido) se amplifica en la etapa final de PCR del método. Los ejemplos descritos a continuación proporcionan varios ejemplos de cómo los cebadores directos, pero no el segundo cebador que comprende la secuencia de la región A, pueden desactivarse selectivamente. Los cebadores directos pueden desactivarse selectivamente en una variedad de formas diferentes, que incluyen: (i) escindir o degradar selectivamente los cebadores directos no extendidos por tratamiento térmico, químicamente, enzimáticamente o por fotoescisión. En algunas de estas realizaciones, los cebadores directos pueden contener, en ciertos casos, uno o más nucleótidos o enlaces que no se encuentran en el ADN natural que hace que los cebadores directos sean susceptibles a una condición de escisión específica. Por ejemplo, los cebadores directos pueden contener uno o más ribonucleótidos, o uno o más enlaces que los hacen escindibles cuando están expuestos a condiciones apropiadas (p. ej., calor, luz, una actividad enzimática particular, etc.). En otra realización, los cebadores directos pueden inactivarse selectivamente (ii) por inducción de un cambio conformacional en los cebadores directos no extendidos (p. ej., al

utilizar una temperatura que favorece la formación de una horquilla 3' en el extremo 3' de los cebadores directos no extendidos). En otras realizaciones, los cebadores directos pueden inactivarse selectivamente por (iii) realización de la última etapa de amplificación por PCR utilizando condiciones en las que los cebadores directos no funcionan (p. ej., utilizando una temperatura de hibridación óptima que es superior a la T_a de la región Z). En ciertos casos, la temperatura de hibridación de la PCR puede ser de al menos 5 °C (p. ej., al menos 10 °C, al menos 15 °C o 20 °C más alta) que la T_a de la región Z de los cebadores directos. Los cebadores directos pueden inactivarse selectivamente utilizando una variedad de otros métodos, incluso separando espacialmente los cebadores directos no extendidos de la población de segundos productos de extensión. Esto se puede llevar a cabo utilizando, p. ej., cebadores directos que se conjugan con perlas magnéticas, utilizando una celda de flujo (en cuyo caso los cebadores directos se pueden extraer de la solución mediante hibridación o utilizando estreptavidina si los cebadores directos están biotinilados), mediante centrifugación, o preparando la reacción inicial de extensión con cebador utilizando cebadores que se están inmovilizando en un soporte sólido, y luego separando los productos de extensión con cebador de los cebadores directos antes de realizar las siguientes etapas del método. En algunas realizaciones que dependen de la actividad de una exonucleasa, p. ej., una exonucleasa 3' a 5', el segundo cebador y, opcionalmente, el cebador inverso o el tercer cebador pueden ser resistentes a las nucleasas por la exonucleasa, es decir, protegidos de la degradación de la exonucleasa, por un enlace no escindible, p. ej., un enlace fosforotioato en un extremo.

Una vez que los cebadores directos no extendidos han sido inactivados selectivamente, el método comprende: (f) amplificar la población de segundos productos de extensión mediante la reacción en cadena de la polimerasa (PCR) utilizando: el segundo cebador que comprende la región A y el tercer cebador para producir una población de productos de PCR en la que los productos clonalmente relacionados están marcados con la misma secuencia de recuento y los productos que no están clonalmente relacionados se marcan con una secuencia diferente entre sí. En otras palabras, la población de productos de PCR producidos por este método contiene múltiples subpoblaciones de productos clonalmente relacionados, donde todas las moléculas de cualquier subpoblación única de productos de PCR se amplificaron a partir de una única molécula en la muestra inicial.

Todas las etapas principales del método (es decir, las etapas (a) a (f), como se ha descrito anteriormente) se realizan en un recipiente cerrado y no se añaden reactivos adicionales al recipiente desde el exterior del recipiente durante el método.

El cebador inverso es de la fórmula 5'-B-W, donde la región B, cuando se copia, proporciona un sitio de unión para el tercer cebador y W es complementario a un sitio corriente abajo en el polinucleótido diana. La etapa de amplificación (f) se realiza utilizando el cebador que comprende la secuencia de la región A y el tercer cebador para producir una población de producto de PCR que tiene una cadena superior de la siguiente fórmula: 5'-A-Y-Z-polinucleótido diana-W'-B'.

En ciertas realizaciones, la etapa de desactivación selectiva puede implicar la desactivación del cebador inverso además de los cebadores directos. En estas realizaciones del método, el cebador inverso contiene una cola 5', la hibridación se realiza en presencia de un tercer cebador que tiene la secuencia de la cola 5', la etapa de desactivación selectiva (e) comprende además la desactivación selectiva del cebador inverso; y la etapa de amplificación comprende la PCR utilizando el segundo cebador y el tercer cebador.

En ciertas realizaciones de los métodos descritos anteriormente, la desactivación selectiva comprende exponer los cebadores directos no extendidos a una condición que escinde los cebadores directos no extendidos. En ciertas realizaciones, la escisión del cebador directo produce un oligonucleótido más corto que bien tiene un extremo 3' extensible o bien es demasiado corto para hibridarse en las condiciones de PCR utilizadas en la etapa (f). En algunas de estas realizaciones, los cebadores directos no extendidos comprenden uno o más nucleótidos sensibles al calor, p. ej., ribonucleótidos o análogos similares (es decir, nucleótidos que tienen enlaces que se escinden a una temperatura más alta), y la desactivación comprende exponer los cebadores directos no extendidos a una temperatura de al menos 90 °C durante un periodo de tiempo suficiente para degradar los cebadores directos no extendidos. En otra realización, los cebadores directos no extendidos son selectivamente susceptibles a la escisión de la nucleasa, y la condición comprende la activación de una nucleasa. En esta realización, la exonucleasa puede introducirse en el producto de la etapa (d) a una temperatura inferior a 45 °C utilizando un gel termorreversible, p. ej., un gel termorreversible que es líquido a una temperatura inferior a 40 °C y sólido a una temperatura superior a 50 °C. Alternativamente, una actividad de nucleasa se puede activar incubando el producto de la etapa (d) a una temperatura que se optimiza para la actividad de la nucleasa, donde la nucleasa ya está en la composición.

En otras realizaciones, la desactivación selectiva puede comprender alterar la temperatura del producto de la etapa (d). En una realización, la T_a de unión de los primeros cebadores al polinucleótido diana es al menos 5 grados más baja que la T_a de unión del segundo cebador al complemento de la secuencia A. En esta realización, la etapa de hibridación (a) puede ser realizada a una temperatura que es al menos 5 grados más baja que la etapa de hibridación de la PCR de (f). Un cambio en la temperatura también puede inducir selectivamente un cambio conformacional en los cebadores directos no extendidos. Por ejemplo, si los cebadores directos contienen una horquilla 3', entonces la temperatura de la reacción se puede disminuir para producir una horquilla en el extremo 3' de los cebadores directos, inactivándolos de este modo.

En otras realizaciones, la desactivación selectiva puede comprender separar los cebadores directos no extendidos de la población de segundos productos de extensión. Esto se puede hacer de varias maneras diferentes, p. ej., utilizando un imán, centrifugación o moviendo la reacción en un dispositivo microfluídico para que los primeros cebadores no extendidos puedan eliminarse por afinidad a un agente de captura, p. ej., estreptavidina, si los primeros cebadores no extendidos están biotinilados.

En algunas realizaciones, tanto el cebador directo A como el cebador inverso tienen una secuencia de recuento. En estas realizaciones, el cebador directo puede ser de la fórmula A-Ya-Z y el cebador inverso puede ser de la fórmula B-Yb-W, donde Ya e Yb pueden ser la misma población de secuencias de recuento o una población de secuencias de recuento diferente. En este método, la etapa de hibridación se realiza en presencia de un segundo cebador que comprende la secuencia de la región A y un tercer cebador inverso que se hibrida con B. Después de que se haya realizado la etapa de hibridación inicial, el método prosigue por un ciclado a través de dos rondas de la extensión con cebador. Las dos rondas de extensión con cebador comprenden: (b) extender los cebadores directos que se hibridan con el polinucleótido diana utilizando el polinucleótido diana como un molde para producir una población de primeros productos de extensión que comprenden un sitio de unión para el cebador inverso; (c) hibridar los cebadores inversos con el sitio de unión para el cebador inverso en la primera población de productos de extensión producidos por la etapa (b); (d) extender los cebadores inversos para producir una población de segundos productos de extensión que comprende, en sus extremos 5', la región B y la región Ya, y en sus extremos 3', el complemento de la región A y el complemento de la región Yb. La combinación de las secuencias de conteo de las regiones 5' y 3' es, con una alta probabilidad estadística, diferente en cada molécula en la población de los segundos productos de extensión. Después de que se hayan completado las primeras dos rondas de extensión con cebador para producir una población de segundos productos de extensión, el método comprende (e) desactivar selectivamente, después de la etapa (d) cualquier cebador directo y cualquier cebador inverso que no se extiendan. Una vez que los cebadores inversos no extendidos y directos no extendidos hayan sido inactivados selectivamente, el método comprende: (f) amplificar la población de segundos productos de extensión por reacción en cadena de la polimerasa (PCR) utilizando: el segundo cebador que comprende la región A y el tercer cebador que comprende la región B para producir una población de productos de PCR en la que los productos clonalmente relacionados están marcados con las mismas secuencias de recuento y los productos que no están clonalmente relacionados están marcados con una secuencia diferente entre sí.

Los métodos resumidos anteriormente se describen con gran detalle a continuación.

También se divulgan composiciones. En ciertas realizaciones, la composición puede comprender una población de cebadores directos de la fórmula 5'-A-Y-Z-3', donde: (i) la región A proporciona un sitio de unión a cebador cuando se copia y es la misma en cada uno de los cebadores de la población de cebadores directos (ii) la región Y varía entre los diferentes cebadores de la población de cebadores y proporciona una secuencia de recuento; y (iii) la región Z es complementaria a un primer sitio en un polinucleótido diana en una población de moléculas de ácido nucleico molde y es la misma en cada uno de los cebadores de la población de cebadores; donde dichos cebadores directos contienen un resto escindible que, cuando se escinde, produce un oligonucleótido más corto. Los detalles sobre la composición se describen con mayor detalle a continuación.

También se divulga un kit para realizar el método resumido anteriormente. En ciertas realizaciones, el kit comprende la composición descrita anteriormente.

Breve descripción de los dibujos

La invención se comprenderá mejor a partir de la siguiente descripción detallada cuando se lee junto con los dibujos adjuntos. Se enfatiza que, según la práctica común, las diversas características de los dibujos no están a escala. De hecho, las dimensiones de las diversas características se amplían o reducen arbitrariamente para mayor claridad. En los dibujos se incluyen las siguientes figuras.

La Fig. 1 ilustra esquemáticamente una población de cebadores directos.

La Fig. 2 ilustra esquemáticamente un polinucleótido diana y algunos de los cebadores que se pueden utilizar en el método objeto.

La Fig. 3 ilustra esquemáticamente una realización del método.

La Fig. 4 ilustra esquemáticamente cómo se puede utilizar el método para generar poblaciones clonalmente relacionadas de productos de PCR.

Definiciones

A menos que se defina otra cosa, todos los términos técnicos y científicos utilizados en la presente memoria tienen el mismo significado que entiende comúnmente un experto en la materia a la que pertenece esta invención. Aun así, ciertos elementos se definen en aras de la claridad y facilidad de referencia.

Los términos y símbolos de la química del ácido nucleico, bioquímica, genética y biología molecular que se utilizan en la presente memoria siguen los de los tratados y textos convencionales del campo, p. ej. Kornberg y Baker, DNA Replication, segunda edición (W.H. Freeman, Nueva York, 1992); Lehninger, Biochemistry, segunda edición (Worth Publishers, Nueva York, 1975); Strachan y Read, Human Molecular Genetics, segunda edición (Wiley-Liss, Nueva York, 1999); Eckstein, editor, Oligonucleotides and Analogs: A Practical Approach (Oxford University Press, Nueva York, 1991); Gait, editor, Oligonucleotide Synthesis: A Practical Approach (IRL Press, Oxford, 1984); y similares.

"Amplicón" significa el producto de una reacción de amplificación de polinucleótidos. Es decir, es una población de polinucleótidos, generalmente bicatenarios, que se replican a partir de una o más secuencias de partida. Las una o más secuencias de partida pueden ser una o más copias de la misma secuencia, o pueden ser una mezcla de diferentes secuencias. Los amplicones pueden producirse por una variedad de reacciones de amplificación cuyos productos son múltiples replicados de uno o más ácidos nucleicos diana. En general, las reacciones de amplificación que producen amplicones son "estimuladas por molde", por que el apareamiento de bases de los reactivos, ya sean nucleótidos u oligonucleótidos, tienen en un polinucleótido molde los complementos que se requieren para la creación de productos de reacción. En un aspecto, las reacciones estimuladas por molde son extensiones con cebador con una polimerasa de ácido nucleico o ligamientos de oligonucleótidos con una ligasa de ácido nucleico. Dichas reacciones incluyen, entre otros, reacciones en cadena de la polimerasa (PCR), reacciones lineales de la polimerasa, amplificación basada en secuencias de ácido nucleico (ABSANS), amplificaciones de círculo rodante y similares, que se divulgan en las siguientes referencias: *Mullis et al*, patentes de EE.UU. n.º 4.683.195; 4.965.188; 4.683.202; 4.800.159 (PCR); Gelfand *et al*, patente de EE. UU. n.º 5.210.015 (PCR en tiempo real con sondas "TAQMAN™"); Wittwer *et al*, patente de EE. UU. n.º 6.174.670; Kacian *et al*, patente de EE.UU. n.º 5.399.491 ("ABSAN"); Lizardi, patente de EE.UU. n.º 5.854.033; Aono *et al*, publicación de patente japonesa JP 4-262799 (amplificación de círculo rodante); y similares. En un aspecto, los amplicones de la invención se producen mediante PCR. Una reacción de amplificación puede ser una amplificación en "tiempo real" si se dispone de una química de detección que permita medir un producto de reacción a medida que progresa la reacción de amplificación, p. ej. "PCR en tiempo real" que se describe a continuación, o "ABSAN en tiempo real", como se describe en Leone *et al*, Nucleic Acids Research, 26: 2150-2155 (1998), y referencias similares. Como se utiliza en la presente memoria, el término "amplificar" significa realizar una reacción de amplificación. Una "mezcla de reacción" significa una solución que contiene todos los reactivos necesarios para realizar una reacción, que puede incluir, entre otros, agentes de tamponamiento para mantener el pH en un nivel seleccionado durante una reacción, sales, cofactores, depuradores y similares.

El término "analizar" incluye cualquier forma de medición, e incluye determinar si un elemento está presente o no. Los términos "determinar", "medir", "analizar", "evaluar" y "examinar" se utilizan indistintamente e incluyen las determinaciones cuantitativas y cualitativas. El análisis puede ser relativo o absoluto. "Analizar la presencia de" incluye determinar la cantidad de algo presente y/o determinar si algo está presente o ausente. Como se utiliza en la presente memoria, los términos "determinar", "medir" y "analizar" y "examinar" se utilizan indistintamente e incluyen determinaciones tanto cuantitativas como cualitativas.

Los polinucleótidos que están "marcados asimétricamente" tienen dominios de adaptador a izquierda y derecha que no son idénticos. Este proceso se denomina genéricamente como unión de adaptadores que marcan asimétrica o asimétricamente un polinucleótido, p. ej., un fragmento polinucleotídico. La producción de polinucleótidos con extremos terminales de adaptador asimétrico se puede lograr de cualquier manera conveniente. Los adaptadores asimétricos a modo de ejemplo se describen en: las patentes de EE.UU. 5.712.126 y 6.372.434; publicaciones de patente de EE. UU. 2007/0128624 y 2007/0172839; y la publicación PCT WO/2009/032167. En ciertas realizaciones, los adaptadores asimétricos empleados son los descritos en la solicitud de patente de EE.UU. n.º 12/432.080, presentada el 29 de abril de 2009.

Como ejemplo, un usuario de la presente invención puede utilizar un adaptador asimétrico para marcar polinucleótidos. Un "adaptador asimétrico" es aquel que, cuando se liga a ambos extremos de un fragmento de ácido nucleico bicatenario, dará lugar a la producción de productos de extensión o amplificación de cebadores que tienen secuencias no idénticas que flanquean el inserto genómico de interés. El ligamiento suele ir seguido por etapas de procesamiento subsiguientes para generar las secuencias de adaptador de terminales no idénticos. Por ejemplo, la replicación de un fragmento(s) unido(s) a un adaptador asimétrico da como resultado productos polinucleotídicos en los que hay al menos una diferencia de secuencia de ácido nucleico, o modificación de nucleótido/nucleósido, entre las secuencias terminales del adaptador. La unión de adaptadores de forma asimétrica a los polinucleótidos (p. ej., fragmentos polinucleotídicos) da como resultado polinucleótidos que tienen una o más secuencias de adaptador en un extremo (p. ej., una o más regiones o dominios, p. ej., un sitio de cebador) que no están presentes o tienen una secuencia de ácido nucleico diferente en comparación con la secuencia del adaptador en el otro extremo. Se observa que un adaptador que se denomina "adaptador asimétrico" no es necesariamente estructuralmente asimétrico por sí mismo, ni el mero acto de unir un adaptador asimétrico a un fragmento polinucleotídico le vuelve inmediatamente asimétrico. Más bien, un polinucleótido unido a un adaptador asimétrico, que tiene un adaptador asimétrico idéntico en cada extremo, produce productos de replicación (o polinucleótidos monocatenarios aislados) que son asimétricos con respecto a las secuencias del adaptador de los extremos opuestos (p. ej., después de al menos una ronda de amplificación/extensión con cebador).

- "Complementario" o "esencialmente complementario" se refiere a la hibridación o apareamiento de bases o a la formación de un dúplex entre nucleótidos o ácidos nucleicos, tales como, por ejemplo, entre las dos cadenas de una molécula de ADN bicatenario o entre un cebador de oligonucleótidos y un sitio de cebador en un ácido nucleico monocatenario. Los nucleótidos complementarios son, generalmente, A y T (o A y U), o C y G. Dos moléculas de ARN o ADN monocatenario son aquellas que son esencialmente complementarias cuando los nucleótidos de una cadena, alineados óptimamente y comparados con inserciones o deleciones de nucleótidos apropiados, se aparean con al menos aproximadamente el 80 % de los nucleótidos de la otra cadena, por lo general con al menos aproximadamente 90 % a 95 %, y más preferentemente con aproximadamente 98 a 100 %. Alternativamente, existe complementariedad sustancial cuando una cadena de ARN o ADN se hibrida en condiciones de hibridación selectivas a su complemento. Normalmente, se producirá la hibridación selectiva cuando exista al menos aproximadamente el 65 % de complementariedad en un tramo de al menos 14 a 25 nucleótidos, p. ej., al menos aproximadamente 75 %, o al menos aproximadamente 90 % de complementariedad. Véase, M. Kanehisa *Nucleic Acids Res.* 12: 203 (1984).
- "Dúplex" significa que al menos dos oligonucleótidos y/o polinucleótidos que son total o parcialmente complementarios se someten a un apareamiento de bases de tipo Watson-Crick entre todos o la mayor parte de sus nucleótidos, de manera que se forma un complejo estable. Los términos "apareamiento" e "hibridación" se utilizan indistintamente para referirse a la formación de un dúplex estable. "Perfectamente apareados" en referencia a un dúplex significa que las cadenas poli- u oligonucleotídicas que constituyen el dúplex forman entre sí una estructura bicatenaria, de manera que cada uno de los nucleótidos en cada cadena experimenta el apareamiento de bases de Watson-Crick con un nucleótido en la otra cadena. Un dúplex estable puede incluir un apareamiento de bases de Watson-Crick y/o un apareamiento de bases que no sea de Watson-Crick entre las cadenas del dúplex (donde el apareamiento de bases significa la formación de enlaces de hidrógeno). En ciertas realizaciones, un par de bases que no sea de Watson-Crick incluye un análogo de nucleósido, tal como desoxiinosina, 2,6-diaminopurina, PNAs, LNAs y similares. En ciertas realizaciones, un par de bases que no sea de Watson-Crick incluye una "base oscilante", tal como desoxiinosina, 8-oxo-dA, 8-oxo-dG y similares, donde por "base oscilante" se entiende una base de ácido nucleico que puede formar un par de bases con una primera base de nucleótidos en una cadena complementaria de ácido nucleico, pero que, cuando se emplea como una cadena molde para la síntesis de ácidos nucleicos, conduce a la incorporación de una segunda base de nucleótidos diferente en la cadena de síntesis (las bases oscilantes se describen con más detalle más adelante). Un "apareamiento erróneo" en un dúplex entre dos oligonucleótidos o polinucleótidos significa que un par de nucleótidos en el dúplex no puede someterse a la unión de Watson-Crick.
- "Locus genético", "locus" o "locus de interés" en referencia a un genoma o polinucleótido diana, significa una sub-región o segmento contiguo del genoma o del polinucleótido diana. Como se utiliza en la presente memoria, el locus genético, locus o locus de interés puede referirse a la posición de un nucleótido, un gen o una parte de un gen en un genoma, incluido el ADN mitocondrial u otro ADN no cromosómico (p. ej., plásmido bacteriano), o puede referirse a cualquier parte contigua de la secuencia genómica, esté o no dentro de un gen o esté asociado a él. Un locus genético, locus o locus de interés puede tener desde un solo nucleótido hasta un segmento de unos pocos cientos o unos pocos miles de nucleótidos de longitud o más. En general, un locus de interés tendrá una secuencia de referencia asociada al mismo (véase la descripción de "secuencia de referencia" a continuación).
- "Kit" se refiere a cualquier sistema de administración para administrar materiales o reactivos para llevar a cabo un método de la invención. En el contexto de los ensayos de reacción, tales sistemas de administración incluyen sistemas que permiten el almacenamiento, transporte o administración de reactivos de reacción (p. ej., sondas, enzimas, etc. en los contenedores apropiados) y/o materiales de soporte (p. ej., tampones, instrucciones escritas para realizar el ensayo, etc.) de un lugar a otro. Por ejemplo, los kits incluyen uno o más recintos (p. ej., cajas) que contienen los reactivos de reacción y/o materiales de soporte pertinentes. Dichos contenidos se pueden administrar al receptor propuesto juntos o por separado. Por ejemplo, un primer contenedor puede contener una enzima para su uso en un ensayo, mientras que un segundo contenedor contiene sondas.
- "Ligamiento" significa formar un enlace o unión covalente entre los extremos terminales de dos o más ácidos nucleicos, p. ej., oligonucleótidos y/o polinucleótidos, en una reacción estimulada por molde. La naturaleza del enlace o unión puede variar ampliamente y el ligamiento se puede llevar a cabo enzimática o químicamente. Como se utiliza en la presente memoria, los ligamientos se llevan a cabo usualmente de forma enzimática para formar un enlace fosfodiéster entre el carbono 5' de un nucleótido terminal de un oligonucleótido con el carbono 3' de otro oligonucleótido. Una variedad de reacciones de ligamiento estimuladas por molde se describe en las siguientes referencias: Whiteley *et al*, patente de EE.UU. n.º 4.883.750; Letsinger *et al*, patente de EE.UU. n.º 5.476.930; Fung *et al*, patente de EE.UU. n.º 5.593.826; Kool, patente de EE.UU. n.º 5.426.180; Landegren *et al*, patente de EE.UU. n.º 5.871.921; Xu y Kool, *Nucleic Acids Research*, 27: 875-881 (1999); Higgins *et al*, *Methods in Enzymology*, 68: 50-71 (1979); Engler *et al*, *The Enzymes*, 15: 3-29 (1982); y Namsaraev, publicación de patente de EE.UU. 2004/0110213.
- "Identificador multiplex" (IDM), como se utiliza en la presente memoria, se refiere a una etiqueta o combinación de etiquetas asociadas con un polinucleótido cuya identidad (p. ej., la etiqueta de una secuencia de ADN) se puede utilizar para diferenciar polinucleótidos en una muestra. En ciertas realizaciones, se utiliza el IDM sobre un

polinucleótido para identificar la fuente a partir de la cual se deriva el polinucleótido. Por ejemplo, una muestra de ácido nucleico puede ser una agrupación de polinucleótidos derivados de diferentes fuentes, (p. ej., polinucleótidos derivados de diferentes individuos, diferentes tejidos o células, o polinucleótidos aislados en diferentes puntos temporales), donde los polinucleótidos de cada fuente diferente se marcan con un único IDM. Como tal, un IDM proporciona una correlación entre un polinucleótido y su fuente. En ciertas realizaciones, los IDM se emplean para marcar de forma única cada polinucleótido individual en una muestra. La identificación del número de los IDM únicos en una muestra puede proporcionar una lectura de cuántos polinucleótidos individuales están presentes en la muestra o de cuántos polinucleótidos originales se ha derivado una muestra de polinucleótido manipulado (véase, p. ej., la patente de EE.UU. 7.537.897, expedida el 26 de mayo de 2009) o ambos, si el IDM comprende dos regiones (que pueden ser pero no tienen que ser contiguas), una región que identifica la muestra particular de origen y una que identifica una molécula particular en una muestra particular. Los IDM pueden variar en longitud de 2 a 100 bases de nucleótidos o más, y pueden incluir múltiples subunidades, donde cada IDM diferente tiene una identidad y/o un orden de subunidades definidos. Las etiquetas de ácido nucleico a modo de ejemplo que encuentran utilidad como IDM se describen en la patente de EE.UU. n.º 7.544.473, expedida el 6 de junio 2009, y titulada "Nucleic Acid Analysis Using Sequence Tokens", así como la patente de EE. UU. n.º 7.393.665, expedida el 1 de julio 2008, y titulada "Methods and Compositions for Tagging and Identifying Polynucleotides". En ciertas realizaciones, un conjunto de IDM empleados para marcar una pluralidad de muestras no necesita tener ninguna propiedad particular común (p. ej., Tm, longitud, composición de bases, etc.), ya que los métodos descritos en la presente memoria se pueden adaptar a una amplia variedad de conjuntos de IDM únicos. Se destaca aquí que los IDM solamente necesitan ser únicos dentro de un experimento dado. Por ende, el mismo IDM se puede utilizar para marcar una muestra diferente que está siendo procesada en un experimento diferente. Además, en ciertos experimentos, un usuario puede utilizar el mismo IDM para marcar un subconjunto de muestras diferentes dentro del mismo experimento. Por ejemplo, todas las muestras derivadas de individuos que tienen un fenotipo específico pueden ser marcadas con el mismo IDM, p. ej., todas las muestras derivadas de sujetos de control (o de tipo silvestre) se pueden marcar con un primer IDM mientras que los sujetos que tienen una enfermedad se pueden marcar con un segundo IDM (diferente del primer IDM). Como otro ejemplo, puede ser deseable marcar diferentes muestras derivadas de la misma fuente con diferentes IDM (p. ej., muestras derivadas a lo largo del tiempo o derivadas de diferentes sitios dentro de un tejido). Además, los IDM se pueden generar en una variedad de diferentes maneras, p. ej., mediante un enfoque combinatorio de etiquetado en el que un IDM se une mediante ligamiento y un segundo IDM se une por extensión con cebador. De este modo, los IDM se pueden diseñar e implementar en una variedad de diferentes maneras para hacer un seguimiento de los fragmentos polinucleotídicos durante el procesamiento y análisis, y por ende no se pretende ninguna limitación a este respecto. Un IDM puede contener una secuencia identificadora de muestras así como una secuencia de recuento, como se describe más adelante.

"Nucleósido" como se utiliza en la presente memoria incluye los nucleósidos naturales, incluyendo las formas 2'-desoxi y 2'-hidroxilo, p. ej., como se describe en Kornberg y Baker, DNA Replication, 2ª Ed. (Freeman, San Francisco, 1992). "Análogos" en referencia a los nucleósidos incluye nucleósidos sintéticos que tienen restos de bases modificadas y/o restos de azúcar modificado, p. ej., descritos por Scheit, Nucleotide Analogs (John Wiley, Nueva York, 1980); Uhlman y Peyman, Chemical Reviews, 90: 543-584 (1990), o similares, con la condición de que sean capaces de hibridación específica. Dichos análogos incluyen nucleósidos sintéticos diseñados para mejorar las propiedades de unión, reducir la complejidad, aumentar la especificidad, y similares. Polinucleótidos que comprenden análogos con mejores propiedades de hibridación o de resistencia a la nucleasa están descritos en Uhlman y Peyman (citado anteriormente); Crooke *et al.*, Exp. Opin. Ther. Patents, 6: 855-870 (1996); Mesmaeker *et al.*, Current Opinion in Structural Biology, 5: 343-355 (1995); y similares. Los tipos a modo de ejemplo de polinucleótidos que son capaces de mejorar la estabilidad del dúplex incluyen N3'→P5' fosforamidatos de oligonucleótidos (denominados en la presente memoria "amidatos"), ácidos nucleicos peptídicos (denominados en la presente memoria "ANP"), oligo-2'-O-alkilribonucleótidos, polinucleótidos que contienen propinilpirimidinas C-5, ácidos nucleicos bloqueados ("ANB"), y compuestos similares. Tales oligonucleótidos o bien están disponibles comercialmente o bien se pueden sintetizar utilizando métodos descritos en la literatura. Haciendo referencia a cualquiera de los nucleótidos definidos por el código IUPAC (p. ej., R, Y, S, W, K, M, B, D, H, V y N incluye análogos de los mismos que tienen las mismas características de apareamiento de bases.

"Reacción en cadena de la polimerasa" o "PCR", significa una reacción para la amplificación *in vitro* de secuencias específicas de ADN mediante la extensión simultánea con cebador de cadenas complementarias de ADN. En otras palabras, la PCR es una reacción para fabricar múltiples copias o replicados de un ácido nucleico diana flanqueado por sitios de cebadores, comprendiendo dicha reacción una o más repeticiones de las siguientes etapas: (i) desnaturalizar el ácido nucleico diana, (ii) hibridar los cebadores a los sitios de cebador, y (iii) extender los cebadores por medio de una polimerasa de ácido nucleico en presencia de trifosfatos de nucleósidos. Normalmente, la reacción se hace en ciclos a través de diferentes temperaturas optimizadas para cada etapa en un instrumento termociclador. Las temperaturas particulares, las duraciones de cada etapa, y las velocidades de cambio entre etapas dependen de muchos factores bien conocidos por los expertos en la materia, p. ej., ejemplificados por las referencias: McPherson *et al.*, editors, PCR: A practical Approach and PCR2: A practical Approach (IRL Press, Oxford, 1991 y 1995, respectivamente). Por ejemplo, en una PCR convencional que utiliza la ADN polimerasa Taq, se puede desnaturalizar un ácido nucleico diana bicatenario a una temperatura >90 °C, los cebadores se pueden hibridar a una temperatura en el intervalo de 50-75 °C, y los cebadores se extienden a una temperatura en el intervalo de 72-78 °C. El término "PCR" engloba formas derivadas de la reacción, que incluyen, pero no se limitan a,

RT-PCR, PCR en tiempo real, PCR anidada, PCR cuantitativa, PCR multiplexada, y similares. Los volúmenes de reacción varían desde unas pocas decenas de picolitros, p. ej., 100 pl, a unos centenares de μl , p. ej., 200 μl . "PCR de transcripción inversa", o "RT-PCR", significa una PCR que va precedida por una reacción de transcripción inversa que convierte un ARN diana en un ADN complementario monocatenario, que se amplifica entonces, p. ej., Tecott *et al.*, patente de EE. UU. n.º 5.168.038, cuya patente se incorpora en la presente por referencia. "PCR en tiempo real" significa una PCR para la cual la cantidad del producto de reacción, es decir, el amplicón, se controla conforme avanza la reacción. Hay muchas formas de PCR en tiempo real que difieren principalmente en la química de detección utilizada para el control del producto de reacción, p. ej., Gelfand *et al.*, patente de EE.UU. n.º 5.210.015 ("TAQMAN™"); Wittwer *et al.*, patentes de EE.UU. n.º 6.174.670 y 6.569.627 (tintes intercalantes); Tyagi *et al.*, patente de EE.UU. n.º 5.925.517 (balizas moleculares). La química de detección para la PCR en tiempo real se revisa en Mackay *et al.*, *Nucleic Acids Research*, 30: 1292-1305 (2002), que también se incorpora en la presente por referencia. "PCR anidada" significa una PCR de dos fases donde el amplicón de una primera PCR se convierte en la muestra para una segunda PCR que utiliza un nuevo conjunto de cebadores, uno de los cuales al menos se une en una ubicación interior del primer amplicón. Como se utiliza en la presente memoria, "cebadores iniciales" en referencia a una reacción de amplificación anidada significan los cebadores utilizados para generar un primer amplicón, y "cebadores secundarios" significan los uno o más cebadores utilizados para generar un segundo amplicón, o amplicón anidado. "PCR multiplexada" significa una PCR donde múltiples secuencias diana (o una sola secuencia diana y una o más secuencias de referencia) se llevan a cabo simultáneamente en la misma mezcla de reacción, p. ej., Bernard *et al.*, *Anal. Biochem.*, 273: 221-228 (1999) (PCR de dos colores en tiempo real). Habitualmente, se emplean distintos conjuntos de cebadores para cada secuencia que se amplifica.

Como se utiliza en la presente memoria, la expresión "reactivos de PCR" se refiere a todos los reactivos que son requeridos para realizar una reacción en cadena de la polimerasa (PCR) en un molde. Como es conocido en la técnica, los reactivos de PCR incluyen esencialmente un primer cebador, un segundo cebador, una polimerasa termostable, y nucleótidos. En función de la polimerasa utilizada, también pueden estar presentes iones (p. ej., Mg^{2+}). Los reactivos de PCR pueden contener opcionalmente un molde a partir del cual se puede amplificar una secuencia diana.

"Polinucleótido" u "oligonucleótido" se utilizan de modo intercambiable y cada uno significa un polímero lineal de monómeros nucleotídicos. Los monómeros que constituyen los polinucleótidos y oligonucleótidos son capaces de unirse específicamente a un polinucleótido natural por medio de un patrón regular de interacciones monómero a monómero, tales como el apareamiento de bases de tipo Watson-Crick, apilamiento de bases, apareamiento de bases de tipos Hoogsteen o Hoogsteen inverso, apareamiento de bases oscilantes, o similares. Como se describe con detalle a continuación, por "base oscilante" se entiende una base de ácido nucleico que puede formar un par de bases con una primera base de nucleótido en una cadena complementaria de ácido nucleico, pero que, cuando se emplea como una cadena molde para la síntesis de ácido nucleico, da lugar a la incorporación de una segunda base de nucleótidos diferente a la cadena de síntesis. Dichos monómeros y sus uniones internucleosídicas pueden ser de origen natural o pueden ser análogos de los mismos, p. ej., análogos de origen natural o no natural. Los análogos de origen no natural pueden incluir ácidos nucleicos peptídicos (ANPs, p. ej., como se describe en la patente de EE. UU. n.º 5.539.082), ácidos nucleicos bloqueados (ANB, p. ej., como se describe en la patente de EE.UU. n.º 6.670.461), uniones internucleosídicas de fosforotioato, bases que contienen grupos de unión que permiten la fijación de etiquetas, tales como fluoróforos, o haptenos y similares. Siempre que el uso de un oligonucleótido o polinucleótido requiera un procesamiento enzimático, tal como extensión por una polimerasa, ligamiento por una ligasa, o similares, un experto en la materia debe entender que los oligonucleótidos o polinucleótidos en estos casos no deben contener ciertos análogos de uniones internucleosídicas, restos de azúcar, o bases en todas o algunas posiciones. Los polinucleótidos normalmente varían de tamaño desde algunas unidades monoméricas, p. ej., 5-40, cuando se denominan habitualmente "oligonucleótidos", hasta varios miles de unidades monoméricas. Siempre que un polinucleótido u oligonucleótido esté representado por una secuencia de letras (mayúsculas o minúsculas), tal como "ATGCCTG", se entenderá que los nucleótidos están en orden 5'→ 3', de izquierda a derecha y que "A" indica desoxiadenosina, "C" indica desoxicidina, "G" indica desoxiguanosina, y "T" indica timidina, "I" indica desoxiinosina, "U" indica uridina, a menos que se indique otra cosa o se deduzca del contexto. A menos que se indique otra cosa, los convenios de la terminología y de la numeración de átomos seguirán los divulgados en Strachan and Read, *Human Molecular Genetics 2* (Wiley-Liss, Nueva York, 1999). Habitualmente, los polinucleótidos comprenden los cuatro nucleósidos naturales (p. ej., desoxiadenosina, desoxicidina, desoxiguanosina, desoxitimidina para el ADN o sus homólogos de ribosa para el ARN) unidos por enlaces de fosfodiéster; sin embargo, también pueden comprender análogos de nucleótidos no naturales, p. ej., que incluyen bases modificadas, azúcares, o uniones internucleosídicas. Es evidente para los expertos en la materia que, cuando una enzima tiene requisitos específicos de sustrato de oligonucleótidos o polinucleótidos para su actividad, p. ej., ADN monocatenario, dúplex de ARN/ADN, o similares, entonces la selección de la composición apropiada para los sustratos de oligonucleótidos o polinucleótidos está dentro de los conocimientos de un experto en la materia, especialmente con la guía de tratados, tales como Sambrook *et al.*, *Molecular Cloning*, Segunda edición (Cold Spring Harbor Laboratory, Nueva York, 1989), y referencias similares.

"Cebador" significa un oligonucleótido, ya sea natural o sintético, que es capaz, tras formar un dúplex con un molde de polinucleótido, de actuar como un punto de iniciación de la síntesis de ácido nucleico y de extenderse desde su extremo 3' a lo largo del molde de modo que se forma un dúplex extendido. La secuencia de nucleótidos añadida

durante el proceso de extensión se determina por la secuencia del polinucleótido molde. Habitualmente, los cebadores se extienden por una ADN polimerasa. Los cebadores tienen generalmente una longitud compatible con su uso en la síntesis de productos de extensión con cebador, y normalmente caen en el intervalo de entre 8 a 100 nucleótidos de longitud, tal como 10 a 75, 15 a 60, 15 a 40, 18 a 30, 20 a 40, 21 a 50, 22 a 45, 25 a 40, y así sucesivamente, más normalmente en el intervalo de entre 18-40, 20-35, 21-30 nucleótidos de largo, y cualquier longitud entre los intervalos indicados. Los cebadores típicos pueden estar en el intervalo de entre 10-50 nucleótidos de longitud, tal como 15-45, 18-40, 20-30, 21-25 y así sucesivamente, y cualquier longitud entre los intervalos indicados. En algunas realizaciones, los cebadores habitualmente no tienen más de aproximadamente 10, 12, 15, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65 o 70 nucleótidos de longitud.

Los cebadores son normalmente monocatenarios para una máxima eficiencia en la amplificación, pero pueden ser alternativamente bicatenarios. Si es bicatenario, el cebador se trata habitualmente primero para separar sus cadenas antes de ser utilizado para preparar productos de extensión. Esta etapa de desnaturalización se ve normalmente afectada por el calor, pero alternativamente se puede llevar a cabo utilizando un álcali, seguido de neutralización. De este modo, un "cebador" es complementario a un molde, y se compleja mediante enlace de hidrógeno o hibridación con el molde para dar un complejo cebador/molde para la iniciación de la síntesis por una polimerasa, que se extiende por la adición de bases unidas covalentemente vinculadas en su extremo 3' complementario al molde en el proceso de síntesis de ADN.

Un "par de cebadores", como se utiliza en la presente memoria, se refiere a los primer y segundo cebadores que tienen la secuencia de ácido nucleico adecuada para la amplificación a base de ácido nucleico de un ácido nucleico diana. Dichos pares de cebadores incluyen generalmente un primer cebador que tiene una secuencia que es la misma o similar a la de una primera parte de un ácido nucleico diana, y un segundo cebador que tiene una secuencia que es complementaria a una segunda parte de un ácido nucleico diana para proporcionar la amplificación del ácido nucleico diana o de un fragmento del mismo. La referencia a "primer" y "segundo" cebadores en la presente memoria es arbitraria, a no ser que se indique específicamente lo contrario. Por ejemplo, el primer cebador puede ser diseñado como un "cebador directo" (que inicia la síntesis de ácido nucleico a partir de un extremo 5' del ácido nucleico diana) o como un "cebador inverso" (que inicia la síntesis de ácido nucleico a partir de un extremo 5' del producto de extensión producido a partir de la síntesis iniciada a partir del cebador directo). Asimismo, el segundo cebador puede ser diseñado como un cebador directo o como un cebador inverso.

"Sitio de cebador" (p. ej., un sitio de cebador de secuenciación, y sitio de cebador de amplificación, etc.), como se utiliza en la presente memoria, se refiere a un dominio en un polinucleótido que incluye la secuencia de un cebador (p. ej., un cebador de secuenciación) y/o la secuencia complementaria de un cebador. Cuando está presente en una forma monocatenaria (p. ej., en un polinucleótido monocatenario), un sitio de cebador puede ser bien la secuencia idéntica de un cebador o bien la secuencia complementaria de un cebador. Cuando está presente en forma bicatenaria, un sitio de cebador contiene la secuencia de un cebador hibridado con la secuencia complementaria del cebador. Por ende, un sitio de cebador es una región de un polinucleótido que es bien idéntica o complementaria con la secuencia de un cebador (cuando se encuentra en forma monocatenaria) o una región bicatenaria formada entre una secuencia de cebador y su complemento. Los sitios de cebador pueden estar presentes en un adaptador fijado a un polinucleótido. La orientación específica de un sitio de cebador puede deducirse por los expertos en la materia a partir de las características estructurales del polinucleótido y/o contexto pertinentes en los cuales se utiliza.

"Específico" o "especificidad", en referencia a la unión de una molécula con otra molécula, tal como una secuencia diana marcada para una sonda, significa el reconocimiento, contacto y formación de un complejo estable entre las dos moléculas, junto con esencialmente menos reconocimiento, contacto o formación de complejos de esa molécula con otras moléculas. En un aspecto, "específica" en referencia a la unión de una primera molécula con una segunda molécula significa que en la medida en que la primera molécula reconoce y forma un complejo con otra molécula en una reacción o muestra, forma el mayor número de los complejos con la segunda molécula. En algunos casos, este mayor número es al menos el cincuenta por ciento. Generalmente, las moléculas implicadas en un evento de unión específica tienen áreas en sus superficies o en cavidades que dan lugar al reconocimiento específico entre las moléculas que se unen entre sí. Los ejemplos de unión específica incluyen interacciones anticuerpo-antígeno, interacciones enzima-sustrato, formación de dúplex o tríplex entre polinucleótidos y/u oligonucleótidos, interacciones de biotina-avidina o biotina-estreptavidina, interacciones receptor-ligando, y similares. Como se utiliza en la presente memoria, "contacto", en referencia a la especificidad o unión específica significa que dos moléculas están lo suficientemente próximas para que interacciones químicas no covalentes débiles, tales como las fuerzas de Van der Waals, enlaces de hidrógeno, interacciones en apilamiento de bases, interacciones iónicas e hidrófobas, y similares, dominen en la interacción de las moléculas.

Como se utiliza en la presente memoria, el término " T_f " se utiliza en la presente memoria para referirse a la "temperatura de fusión" de una secuencia. La temperatura de fusión es la temperatura (p. ej., medida en °C) a la que una población de moléculas de ácido nucleico bicatenario llega a ser semidisociada de forma monocatenaria. Diversas ecuaciones para calcular la T_f de ácidos nucleicos se conocen en la técnica (véase, p. ej., Anderson y Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985). Otras referencias (p. ej., Allawi, H.T. y SantaLucia, J., Jr., Biochemistry 36, 10581-94 (1997)) incluyen métodos alternativos de cálculo que tienen en cuenta, para el cálculo de T_f las características estructurales y ambientales, así como las características de las

secuencias.

El término "T_a" se utiliza en la presente memoria para referirse a la "temperatura óptima de hibridación" de una secuencia. La temperatura óptima de hibridación es la temperatura (p. ej., medida en °C) que da como resultado un rendimiento óptimo de productos de PCR con una producción mínima de productos falsos. La temperatura óptima de hibridación de una secuencia puede determinarse experimentalmente o teóricamente. En una realización, se puede calcular una aproximación de la temperatura óptima de hibridación utilizando el método descrito en Rychlik *et al.* (Rychlik *et al.* Nucleic Acids Res 1990 18: 6409-12) utilizando la fórmula $T_a \text{ Opt} = 0,3 \times (T_f \text{ del cebador}) + 0,7 \times (T_f \text{ del producto}) - 14,9$; donde T_f del cebador es la temperatura de fusión del par del cebador molde menos estable y T_f del producto de la temperatura de fusión del producto. Véase también Innis, ed., PCR protocols: A guide to methods and applications. 1990 Academic Press San Diego, CA.

"Muestra" significa una cantidad de material procedente de una fuente biológica, ambiental, médica, o de pacientes en la que se busca la detección, medida, o marcado de los ácidos nucleicos diana. Por una parte, se pretende incluir un espécimen o cultivo (p. ej., cultivos microbiológicos). Por otra parte, se pretende incluir tanto las muestras biológicas como ambientales. Una muestra puede incluir un espécimen de origen sintético. Las muestras biológicas pueden ser de animales, incluyendo seres humanos, fluidos, sólidos (p. ej., heces) o tejidos, así como alimentos líquidos y sólidos y productos e ingredientes alimentarios tales como productos lácteos, verduras, carne y subproductos cárnicos y residuos. Las muestras biológicas pueden incluir materiales tomados de un paciente, que incluyen, entre otros, cultivos, sangre, saliva, líquido cefalorraquídeo, fluido pleural, leche, linfa, esputo, semen, aspirados con aguja, y similares. Las muestras biológicas se pueden obtener de todas las diversas familias de animales domésticos, así como de animales silvestres o salvajes, que incluyen, entre otros, animales tales como ungulados, osos, peces, roedores, etc. Las muestras ambientales incluyen material ambiental tal como materia superficial, suelo, agua y muestras industriales, así como las muestras obtenidas de los instrumentos, aparatos, equipo, utensilios, artículos desechables y no desechables de procesamiento de alimentos y productos lácteos. Estos ejemplos no se deben interpretarse como limitantes de los tipos de muestra aplicables a la presente invención.

Los términos "corriente arriba" y "corriente abajo" en la descripción de la orientación y/o polimerización de la molécula de ácido nucleico se utilizan en la presente memoria como son entendidos por los expertos en la materia. Como tal, "corriente abajo" significa generalmente proceder en la dirección 5' a 3', es decir, la dirección en la que una polimerasa de nucleótido extiende normalmente una secuencia, y "corriente arriba" significa generalmente lo contrario. Por ejemplo, un primer cebador que se hibrida "corriente arriba" de un segundo cebador sobre la misma molécula de ácido nucleico diana está ubicado en el lado 5' del segundo cebador (y por lo tanto la polimerización del ácido nucleico a partir del primer cebador se dirige hacia el segundo cebador). En otras palabras, el término "corriente abajo", en el contexto de un primer elemento de secuencia que está corriente debajo de un segundo elemento de secuencia, se refiere a un primer elemento de secuencia que es 3' relativo al segundo elemento. En casos particulares (y como será evidente por el contexto) algunos ácidos nucleicos se ilustran en la orientación 3' a 5' para facilitar la comprensión, un primer elemento de secuencia que está "corriente abajo" de un segundo elemento de secuencia puede colocarse en una figura a la izquierda del segundo elemento de secuencia.

El término "no extensible", en el contexto de un oligonucleótido que no es extensible en su extremo 3' cuando está hibridado con un ácido nucleico diana, se refiere a un oligonucleótido que no puede extenderse por una polimerasa dependiente de polimerasa molde, ya sea por que el extremo 3' del oligonucleótido está bloqueado en el extremo 3' (p. ej., por un nucleótido didesoxi o cualquiera de una multitud de nucleótidos que no son sustratos para la polimerasa) o por que el extremo 3' del oligonucleótido tiene un apareamiento erróneo con la diana, es decir, por que uno o más nucleótidos en el extremo 3' del oligonucleótido no son complementarios a los nucleótidos colocados correspondientemente en la secuencia diana).

Los términos "pluralidad", "población" y "colección" se utilizan indistintamente para referirse a algo que contiene al menos 2 miembros. En ciertos casos, una pluralidad, población o colección puede tener al menos 10, al menos 100, al menos 1.000, al menos 10.000, al menos 100.000, al menos 10⁶, al menos 10⁷, al menos 10⁸ o al menos 10⁹ o más miembros.

La expresión "que contiene un adaptador", en el contexto de un ácido nucleico que contiene un adaptador, se refiere a un ácido nucleico que se ha ligado a un adaptador, o a un ácido nucleico que se ha preparado mediante PCR utilizando un cebador de PCR que contiene una secuencia sin molde en su extremo 5'.

El término "heterólogo" en el contexto de un ácido nucleico que contiene una secuencia fuente y una secuencia que es heteróloga al ácido nucleico fuente, se refiere a una secuencia que no está asociada con la secuencia fuente en una célula huésped de tipo silvestre.

El término "PCR clonal" es una técnica de PCR en la que cada reacción se realiza en una única molécula molde, y las reacciones de PCR se mantienen separadas espacialmente una de otra. La PCR puente y PCR en emulsión, comúnmente utilizadas en aplicaciones de secuenciación de próxima generación, son ejemplos de plataformas de PCR clonal.

- La expresión "secuencia de recuento" es una secuencia única de nucleótidos que se adjunta a un polinucleótido diana donde, para cada polinucleótido diana, la secuencia de nucleótidos de la secuencia de recuento es diferente a esencialmente todas las demás secuencias de recuento añadidas a otros polinucleótidos diana. Como se describirá con mayor detalle más adelante, las secuencias de recuento permiten determinar el número de moléculas de polinucleótido diana iniciales que se han analizado, es decir, "contar" el número de moléculas de polinucleótido diana iniciales que se han analizado. En el contexto de una secuencia de recuento, "esencialmente todo" significa al menos 95 %, p. ej., al menos 96 %, al menos 97 %, al menos 98 %, al menos 99 % o al menos 99,5 % de los polinucleótidos diana que se asociarán con diferentes secuencias de recuento.
- La expresión "secuencia de identificador de muestra" es una secuencia de nucleótidos que se anexa a un polinucleótido diana, donde la secuencia identifica la fuente del polinucleótido diana (es decir, la muestra a partir de la cual se deriva el polinucleótido diana). En uso, cada muestra se marca con una secuencia de identificador de muestra diferente (p. ej., se anexa una secuencia a cada muestra, donde las diferentes muestras se anexan a secuencias diferentes), y las muestras marcadas se agrupan. Después de secuenciar la muestra agrupada, la secuencia del identificador de muestra se puede utilizar para identificar la fuente de las secuencias. Se puede añadir una secuencia de identificador de muestras al extremo 5' de un polinucleótido o al extremo 3' de un polinucleótido. En ciertos casos, parte de la secuencia del identificador de muestra puede estar en el extremo 5' de un polinucleótido y el resto de la secuencia del identificador de muestra puede estar en el extremo 3' del polinucleótido. En conjunto, las secuencias 3' y 5' identifican la muestra.
- En una célula, el ADN suele existir en una forma bicatenaria, y como tal, tiene dos cadenas complementarias de ácido nucleico a las que se hace referencia en la presente memoria como las cadenas "superior" e "inferior". En ciertos casos, las cadenas complementarias de una región cromosómica pueden denominarse cadenas "positivas" y "negativas", las cadenas "primera" y "segunda", las cadenas "codificantes" y "no codificantes", las cadenas de "Watson" y "Crick" o las cadenas "sentido" y "antisentido". La asignación de una cadena como una cadena "superior" o "inferior" es arbitraria y no implica ninguna orientación, función o estructura particular. Las secuencias de nucleótidos de la primera hebra de varias regiones cromosómicas de mamíferos a modo de ejemplo (p. ej., BAC, ensamblajes, cromosomas, etc.) son conocidos, y pueden encontrarse en la base de datos Genbank del NCBI, por ejemplo.
- El término "con cola", en el contexto de un cebador con cola o un cebador que tiene una cola 5', se refiere a un cebador que tiene una región (p. ej., una región de al menos 12-50 nucleótidos) en su extremo 5' que no se hibrida con la misma diana que el extremo 3' del cebador.
- Las expresiones "desactivación selectiva" e "impedido selectivamente", en el contexto de un cebador, se utilizan indistintamente para describir la alteración física de la composición (donde "físicamente" tiene por objeto englobar la adición de reactivos, la eliminación física o el secuestro de cebadores, alteración de la posición de los reactivos y otros cambios en la composición, p. ej., a la temperatura de la composición, etc.) en la que existe un cebador no extendido de manera que el cebador no extendido, pero no otros cebadores, se excluya de ser extendido en la siguiente etapa de un método de extensión del cebador.
- El término "no extendido" en el contexto de un cebador no extendido, se refiere a un cebador que no se ha extendido.
- La expresión "relacionado clónicamente" se refiere a un conjunto de moléculas que se han amplificado a partir de una única molécula inicial. Los productos típicos de PCR contienen múltiples subpoblaciones de productos que están relacionados clonalmente en el sentido de que todas las moléculas de cualquier subpoblación individual se amplificaron a partir de una única molécula en una muestra inicial.
- El término "variable", en el contexto de dos o más secuencias de ácidos nucleicos que son variables, se refiere a dos o más ácidos nucleicos que tienen diferentes secuencias de nucleótidos entre sí. En otras palabras, si los polinucleótidos de una población tienen una secuencia variable, entonces la secuencia de nucleótidos de las moléculas de polinucleótidos de la población varía de una molécula a otra. El término "variable" no debe leerse para requerir que cada molécula en una población tenga una secuencia diferente a las otras moléculas en una población.
- La siguiente descripción explica las fórmulas utilizadas en la presente divulgación. Ciertos polinucleótidos descritos en la presente memoria pueden ser referidos mediante una fórmula (p. ej., "A-Y-Z", "B-W"). Tales fórmulas siguen la convención establecida en el sentido de que describen un polinucleótido que está orientado en la dirección 5' a 3'. Los componentes de la fórmula, p. ej., "A", "Y" y "Z" se refieren a secuencias definibles por separado de nucleótidos dentro de un polinucleótido, donde las secuencias están unidas covalentemente de manera tal que un polinucleótido descrito por una fórmula es una sola molécula. Los componentes de la fórmula pueden estar inmediatamente adyacentes entre sí o separados unos de otros en la única molécula. En ciertos casos, otros elementos de secuencia, p. ej., otros sitios de unión a cebadores, códigos de barras moleculares, promotores, etc. pueden proporcionarse mediante secuencias que están entre los componentes de una fórmula. Además, cada uno de los diversos componentes de una fórmula puede tener funciones además de las descritas en la presente memoria. Siguiendo la convención, el complemento de una secuencia que se muestra en una fórmula se indicará con un

número primo (') tal que el complemento de la secuencia "A" será "A'". Es más, a menos que se indique lo contrario o esté implícito en el contexto, un polinucleótido definido por una fórmula puede tener una secuencia adicional en su extremo 3', en su extremo 5' o en ambos extremos 3'y 5'.

5 Otras definiciones de términos pueden aparecer a lo largo de la memoria descriptiva.

Se observa, además, que las reivindicaciones se pueden redactar para excluir cualquier elemento opcional. Por lo tanto, esta declaración tiene por objeto servir como base antecedente para el uso de una terminología tan exclusiva como "únicamente", "solamente" y similares en relación con la lectura de los elementos reivindicados, o el uso de una limitación "negativa".

Descripción detallada de la invención

15 Cuando se proporciona un intervalo de valores, se entiende que cada valor intermedio, hasta la décima de la unidad del límite inferior a menos que el contexto dicte claramente otra cosa, entre los límites superior e inferior de ese intervalo está también específicamente divulgado. Cada intervalo más pequeño entre cualquier valor establecido o valor intermedio en un intervalo establecido y cualquier otro valor establecido o valor intermedio en dicho intervalo establecido está incluido dentro de la invención. Los límites superior e inferior de estos intervalos más pequeños pueden estar independientemente incluidos o excluidos del intervalo, y cada intervalo en el que cualquiera, ninguno

20 o ambos límites están incluidos en los intervalos más pequeños está también incluido dentro de la invención, sujeto a cualquier límite específicamente excluido en el intervalo establecido. Cuando el intervalo establecido incluye uno o ambos límites, los intervalos que excluyen cualquiera o ambos de dichos límites incluidos, están incluidos también en la invención.

25 A menos que se defina otra cosa, todos los términos técnicos y científicos utilizados en la presente memoria tienen el mismo significado que es entendido comúnmente por los expertos en la materia a la que pertenece esta invención. Aunque cualquier método y material similar o equivalente a los descritos en la presente memoria pueden utilizarse en la práctica o ensayo de la presente invención, se describen ahora algunos métodos y materiales potenciales y preferidos. Todas las publicaciones mencionadas en la presente memoria se incorporan en la presente

30 memoria por referencia para divulgar y describir los métodos y/o los materiales en relación a los cuales se citan las publicaciones. Se entiende que la presente divulgación reemplaza cualquier divulgación de una publicación incorporada en la medida en que haya una contradicción.

35 Se ha de observar que, como se utiliza en la presente memoria y en las reivindicaciones adjuntas, las formas singulares "un", "uno", "una" y "el", "la" incluyen los referentes plurales a menos que el contexto dicte claramente otra cosa. Así, por ejemplo, la referencia a "un ácido nucleico" incluye una pluralidad de dichos ácidos nucleicos y la referencia al "el compuesto" incluye la referencia a uno o más compuestos y equivalentes de los mismos conocidos por los expertos en la materia, y así sucesivamente.

40 La práctica de la presente invención puede emplear, a menos que se indique otra cosa, técnicas y descripciones convencionales de la química orgánica, tecnología de polímeros, biología molecular (incluyendo técnicas recombinantes), biología celular, bioquímica e inmunología, que están dentro de la experiencia de la técnica. Tales técnicas convencionales incluyen la síntesis de matrices de polímeros, hibridación, ligamiento, y detección de la hibridación utilizando una etiqueta. Ilustraciones específicas de técnicas adecuadas se pueden haber tomado como

45 referencia para el siguiente ejemplo de la presente memoria. Sin embargo, también se pueden utilizar por supuesto, otros procedimientos convencionales equivalentes. Tales técnicas y descripciones convencionales se pueden encontrar en manuales de laboratorio estándar, tales como Genome Analysis: A Laboratory Manual Series (Vol. I-IV.), Using Antibodies: A Laboratory Manual, Cells: A Laboratory Manual, PCR Primer: A Laboratory Manual, and Molecular Cloning: A Laboratory Manual (todos de Cold Spring Harbor Laboratory Press), Stryer, L. (1995) Biochemistry (4ª Ed.) Freeman Nueva York, Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, Londres, Nelson y Cox (2000), Lehninger, A., Principles of Biochemistry 3ª Ed., W. H. Freeman Pub., Nueva York, N.Y. y Berg *et al.* (2002) Biochemistry, 5ª Ed., W.H. Freeman Pub., Nueva York, N. Y.

55 El método hace uso de una población de cebadores directos de la fórmula 5'-A-Y-Z donde: (i) la región A proporciona un sitio de unión a cebador cuando se copia y es la misma en todos los cebadores de la población de cebadores directos; (ii) la región Y varía entre los diferentes cebadores de la población de cebadores y proporciona una secuencia de recuento; y (iii) la región Z es complementaria a un primer sitio en un polinucleótido diana en una población de moléculas de ácido nucleico molde y es la misma en todos los cebadores de la población de cebadores. Este cebador se hibrida específicamente y ceba el ácido nucleico del complemento de la región Z en un polinucleótido diana. Tal población de cebadores directos se ilustra esquemáticamente en la Fig. 1. La población de

60 los cebadores directos 2 tiene, en este ejemplo, cuatro miembros **4a**, **4b**, **4c** y **4d**, donde cada miembro comprende: (i) la región A, que es la misma en todos los cebadores de la población, (ii) la región Y, que es diferente en cada uno de los cebadores, y (iii) la región Z, que es la misma en todos los cebadores de la población. Los cebadores directos ilustrados en la Fig. 1 tienen regiones Y₁, Y₂, Y₃ e Y₄, cada una de las cuales tiene una secuencia diferente. En la práctica, la región Y puede estar representada por al menos 10, al menos 100, al menos 1.000, al menos 10.000 o al menos 100.000 o más secuencias diferentes dependiendo de la complejidad deseada. En otras palabras, puede

65

haber al menos 10, al menos 100, al menos 1.000, al menos 10.000, al menos 100.000, al menos 1.000.000 o al menos 10.000.000 o más cebadores directos diferentes en la población de cebadores directos utilizados en el método, donde los cebadores directos difieren en la región Y. Una población de cebadores directos puede contener duplicados de cualquier secuencia. En algunas realizaciones y como se describirá con mayor detalle a continuación, la región Y puede estar representada por una región de bases degeneradas (RBD) que comprende uno o más (p. ej., al menos 2, al menos 3, al menos 4, al menos 5, o 5 a 30 o más) nucleótidos seleccionados entre R, Y, S, W, K, M, B, D, H, V, N (según lo definido por el código IUPAC). Además, uno o más de los nucleótidos en los cebadores directos, p. ej., un nucleótido en la región Y, pueden comprender una base sintética en lugar de una base de origen natural. Estos nucleótidos degenerados pueden ser contiguos o no contiguos.

También se emplea en el método: (i) un segundo cebador que comprende la secuencia de la región A (es decir, una secuencia que es la misma que la secuencia de la región A de los cebadores directos) y (ii) un cebador inverso de secuencia W que se hibrida con el polinucleótido diana. El cebador inverso tiene una cola 5' de la secuencia B y, el método emplea además un tercer cebador que tiene la secuencia B (es decir, una secuencia que es la misma que la secuencia de la región B del cebador inverso con cola). La Fig. 2 ilustra el posicionamiento relativo de un cebador directo **10** (que tiene la fórmula 5'-A-Y-Z, como se ha explicado anteriormente) en relación con el polinucleótido diana **12** y el cebador inverso **16**, que comprende una región 3' que se hibrida con el polinucleótido diana en un sitio que está corriente abajo de la secuencia Z. El cebador inverso **16** contiene una cola de la secuencia B, que se muestra con una línea de puntos en la Fig. 2. La región B proporciona un sitio de unión a cebador para el tercer cebador **18** (también se muestra en una línea de puntos) cuando se copia. W es complementario a un sitio en el polinucleótido diana. La Fig. 2 también muestra el segundo cebador **14**, que tiene la secuencia A. El segundo cebador **14** es capaz de cebar la síntesis de ácidos nucleicos cuando se hibrida con el complemento de A del cebador directo. Las longitudes de las distintas regiones con cada uno de los cebadores pueden variar mucho. En algunas realizaciones, las regiones pueden variar independientemente de 10 a 50 nucleótidos, p. ej., de 12 a 30 o más nucleótidos de longitud. Como sería evidente (pero no se muestra en la Fig. 2), el segundo cebador **14** y/o el tercer cebador **18**, pueden contener una cola 5' que introduce un elemento de secuencia adicional, p. ej., secuencia de identificador de muestra, otra secuencia de recuento, secuencia de otro cebador, o un promotor de bacteriófagos, en la población resultante de productos de PCR.

La Fig. 3 muestra algunas de las características generales de una realización del método. Con referencia a la Fig. 4, algunas realizaciones del método comprenden: (a) hibridar una población de cebadores directos **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, con una población de moléculas de ácido nucleico molde **12** en presencia de un el segundo cebador **14** que comprende la secuencia de la región A y un cebador inverso **16** que comprende una secuencia 3' W que se hibrida con el polinucleótido diana **12**. Para facilitar la explicación, solo se muestra la cadena inferior del polinucleótido diana **12** en la Fig. 3. La hibridación se realiza en presencia de un tercer cebador **18** que tiene la secuencia de la cola 5'. El tercer cebador **18** se indica como una línea de puntos.

En la realización mostrada, la hibridación da como resultado complejos **20** que contienen el polinucleótido diana **12** y una subpoblación de cebadores directos **10a** que se hibridan con el polinucleótido diana. Otra subpoblación de cebadores directos **10b** no se hibrida con el polinucleótido diana. Después de que los complejos **20** se produzcan por hibridación, el método comprende realizar dos rondas de extensión con cebador. Estas etapas de extensión con cebador incluyen: (b) extender cebadores directos **10a** que se hibridan con el polinucleótido diana **12** utilizando el polinucleótido diana **12** como molde para producir una población de primeros productos de extensión **22**. Debido a que cada molécula de extensión producida durante esta etapa se ceba utilizando un cebador directo que tiene una secuencia diferente en la región Y, cada uno de los primeros productos de extensión con cebador **22** tiene una secuencia diferente en su extremo 5'. Como se ilustra, la población de los primeros productos de extensión **22** contiene un sitio de unión **24** para el cebador inverso. Como se ilustra en la Fig. 3, la siguiente etapa incluye (c) hibridar el cebador inverso **16** con su sitio de unión en la primera población de productos de extensión **22**, dando como resultado complejos **26** adicionales. Después de que el cebador inverso **16** haya hibridado con su sitio de unión en la primera población de productos de extensión **22**, el método comprende (d) extender el cebador inverso **16** para producir una población de segundos productos de extensión **28** que comprende el complemento de la secuencia A (es decir, A') y el complemento de la secuencia Y (es decir, Y'), así como el complemento de Z (es decir, Z'). Debido a que cada molécula del primer producto de extensión **22** se cebó utilizando un cebador que tiene una secuencia diferente en la región Y, cada molécula en la población de los segundos productos de extensión tiene una secuencia diferente en su extremo 3', lo que permite que el número de moléculas en esa población sea contado después de haberse amplificado.

Después de que se haya producido la población de los segundos productos de extensión **28**, el método comprende (e) desactivar selectivamente la subpoblación de cebadores directos no extendidos **10b**. En esta etapa, la subpoblación de cebadores directos **10** que no hibridaron con un polinucleótido diana y/o no se extendieron se les impide selectivamente participar en la siguiente etapa del método. El segundo cebador **14** no se inactiva en esta etapa y se utiliza en la siguiente etapa de amplificación.

Esta inactivación evita que los cebadores directos no extendidos **10b** participen en rondas subsiguientes de extensión con cebador, que sobrescribirían eficazmente las secuencias de recuento que ya se han colocado en el extremo 3' de la población de los segundos productos de extensión. Como se describirá con mayor detalle a

continuación, dependiendo de cómo se implemente el método, los cebadores directos no extendidos **10b** pueden desactivarse mediante una variedad de métodos diferentes.

5 Como etapa final en esta realización, el método puede comprender: (f) amplificar la población de segundos productos de extensión mediante PCR utilizando el segundo cebador **14** y el tercer cebador **18** para producir una población de productos de PCR **30**. Las líneas de puntos indican la secuencia añadida.

10 Dependiendo de cómo se implemente el método, solo los cebadores directos **10** pueden desactivarse selectivamente en la etapa. En otras realizaciones, tanto los cebadores directos **10** como los cebadores inversos **16** pueden desactivarse selectivamente.

15 Como se ha señalado anteriormente, el cebador inverso **16** contiene una cola 5' que no se hibrida con el polinucleótido diana. El cebador inverso tiene la fórmula 5'-B-W, donde la región B proporciona un sitio de unión a un cebador cuando se copia y W es complementario a un sitio en el polinucleótido diana. La etapa de amplificación se realiza utilizando el segundo cebador **14** que comprende la secuencia de la región A y un tercer cebador **18** que comprende la secuencia de la región B (que se hibrida y se extiende desde el complemento del cebador inverso) para producir una población de producto de PCR que tiene una cadena superior de las siguientes regiones: 5'-A-Y-Z-polinucleótido diana-W'-B'. Esta población de productos de PCR se ilustra esquemáticamente en la Fig. 3. Como sería evidente, esta población de productos de PCR se puede amplificar y/o secuenciar utilizando cebadores universales que se hibridan con las secuencias terminales de los productos de PCR (p. ej., el cebador universal se hibrida con A' y B', por ejemplo).

25 La población resultante de productos de PCR **30** es aquella en la que los productos de PCR clonalmente relacionados están marcados con la misma secuencia de recuento, y los productos de PCR que no están clonalmente relacionados están marcados con una secuencia diferente entre sí. Este concepto se ilustra en la Fig. 4. Como se ilustra en la Fig. 4, se utiliza una población de cebadores directos **2** que comprende cebadores directos **4a**, **4b**, **4c** y **4d**, donde cada uno de los cebadores tiene una secuencia diferente en la región Y (secuencias Y₁; Y₂, Y₃ e Y₄, como se muestra). Los cebadores se hibridan con los polinucleótidos diana y se someten a dos rondas de extensión con cebador en presencia de un cebador inverso y un cebador de la secuencia A, como se ha analizado anteriormente. La población resultante de los segundos productos de extensión **28** comprende cuatro productos de extensión **28a**, **28b**, **28c** y **28d**, donde cada uno de los productos de extensión está marcado con una secuencia diferente (es decir, Y₁; Y₂, Y₃ e Y₄) que se deriva de los cebadores directos extendidos. Después de desactivar los cebadores directos no extendidos y la PCR utilizando el cebador inverso y el cebador de la secuencia A, la población resultante de los productos de PCR **30** tiene múltiples subpoblaciones de productos que están clonalmente relacionados en cada una de las diferentes subpoblaciones que se amplifica desde un solo segundo producto de extensión. En el ejemplo mostrado, la población de productos de PCR **30** tiene cuatro subpoblaciones clonalmente relacionadas de productos de PCR **30a**, **30b**, **30c** y **30d**, cuando las moléculas de la primera subpoblación **30a** se marcan con la secuencia Y₁; las moléculas de la segunda subpoblación **30b** se marcan con la secuencia Y₂, las moléculas de la tercera subpoblación **30c** se marcan con la secuencia Y₃, y las moléculas de la cuarta subpoblación **30d** se marcan con la secuencia Y₄. Como sería evidente, aunque puede haber varios miles o millones o más o moléculas en cualquiera de las subpoblaciones de productos de PCR clonalmente relacionadas y el número de moléculas en esas subpoblaciones clonalmente relacionadas puede variar mucho, el número de primeros productos de extensión producidos en la primera etapa del método se puede estimar contando el número de secuencias de recuento que están representadas en la población de productos de PCR. En este caso, debido a que hay cuatro secuencias de recuento representadas en la población de productos de PCR (es decir, secuencias Y₁; Y₂, Y₃ e Y₄), se produjeron cuatro primeros productos de extensión en la primera etapa del método. Este número es útil ya que, en ciertas realizaciones, la población de productos de PCR elaborados utilizando este método puede secuenciarse para producir una pluralidad de secuencias. El número de diferentes secuencias de recuento que están asociadas con las secuencias de un polinucleótido diana se puede contar, y este número se puede utilizar para estimar el número de moléculas de ácido nucleico molde inicial que se han secuenciado. En ciertos casos que dependen del diseño experimental, un polinucleótido diana único en la muestra inicial puede estar representado por dos secuencias de recuento diferentes (es decir, un polinucleótido diana único dará lugar a dos poblaciones clonalmente relacionadas de productos de PCR que se derivan de ese polinucleótido). En estas realizaciones, el número estimado de polinucleótidos diana iniciales en la muestra se puede ajustar para tener en cuenta esta variable.

55 En ciertos casos, p. ej., si el cebador directo contiene un enlazador escindible o si está vinculado a una partícula magnética, entonces los productos de extensión que contienen ese producto pueden ser escindidos o eliminados durante la etapa de desactivación. En estas realizaciones, el cebador inverso utilizado no debe contener el enlazador escindible ni estar unido a una partícula magnética, de modo que el segundo producto de extensión (que contiene el cebador inverso) pueda amplificarse por PCR.

65 Además, en cualquiera de las realizaciones descritas anteriormente, el cebador directo y/u opcionalmente, el cebador inverso, pueden contener una secuencia de identificador de muestra que marca cada una de las moléculas amplificadas con una secuencia que indica su fuente. Cuando está presente en los cebadores directos, el identificador de la muestra puede estar en cualquier posición entre la región A de los cebadores directos y la región Z de los cebadores directos. Por ejemplo, la secuencia del identificador de muestra puede estar entre las regiones A

e Y de los cebadores directos, entre los cebadores directos de las regiones Y o Z, o en ciertas realizaciones, el identificador de la muestra puede estar en la secuencia Y, p. ej., en las posiciones que están intercaladas entre los nucleótidos que codifican la secuencia de recuento, o integradas en la secuencia de recuento (en cuyo caso, cada muestra puede marcarse con un conjunto de secuencias de recuento que es único para la muestra).

5 Alternativamente, el identificador de muestra podría introducirse en 5' de A. Por ejemplo, si utiliza la plataforma Illumina, el cebador que incluye 5'-FP2-IDM-A-3' podría utilizarse para introducir un identificador de muestra. En este caso, FP2 es un cebador de celdas de flujo e IDM es el identificador de la muestra.

10 Como se ha indicado anteriormente, el método comprende, después de la segunda etapa de extensión con cebador, desactivar selectivamente la subpoblación de cebadores directos no extendidos **10b**. En algunas realizaciones, los cebadores directos no extendidos se pueden desactivar agregando reactivos adicionales al recipiente que contiene la población de los segundos productos de extensión (es decir, abriendo el recipiente y mezclando una solución que contiene un reactivo adicional con la solución que contiene el segundo producto de extensión). En otras realizaciones, la desactivación se puede realizar sin la adición de ningún reactivo al recipiente que contiene la población de segundos productos de extensión. En estas realizaciones, todas las etapas del método desde la etapa de hibridación a la etapa de PCR final pueden ser (es decir, las etapas (a) a (f)) pueden estar en un recipiente cerrado y no se añaden reactivos adicionales durante el desempeño del método.

20 Como se ha indicado anteriormente, el término "desactivar selectivamente" comprende alterar físicamente la composición en la que existe la población de segundos productos de extensión con cebador de modo que los cebadores directos no extendidos **10b** y opcionalmente, el cebador inverso **16**, se excluyen de la extensión o se extienden por el extremo 3' libre de un producto de extensión, en la siguiente reacción de PCR. Como se expondrá con más detalle a continuación, los cebadores directos no extendidos pueden desactivarse selectivamente en una variedad de formas diferentes que incluyen degradar los cebadores directos no extendidos, induciendo un cambio conformacional en los cebadores directos no extendidos, termociclar utilizando condiciones en las cuales los cebadores directos no funcionan, y separar los cebadores directos no extendidos de la población de segundos productos de extensión. Muchas de estas etapas pueden llevarse a cabo sin añadir reactivos adicionales al recipiente que alberga la población de los segundos productos de extensión.

30 En ciertas realizaciones, los cebadores directos pueden contener nucleótidos sensibles al calor, p. ej., ribonucleótidos o un análogo sensible al calor de los mismos, y los cebadores directos se escinden calentando los cebadores. En estas realizaciones, los cebadores pueden incubarse con un ion metálico, p. ej., un ion de la serie de lantánidos o un ion metálico divalente tal como Mn^{2+} Mg^{2+} o Zn^{2+} (que puede estar en una concentración de, p. ej., 5 mM a 200 mM) a una temperatura elevada (p. ej., en el rango de 50 °C a 95 °C) durante un periodo de tiempo, p. ej., de 1 minuto a 1 hora, como se describe, p. ej., en Brown *et al* (J. Am. Chem. Soc. 2002 124: 7950-7962). El ion metálico divalente ya puede estar presente en la reacción y no es necesario agregarlo. En una realización, dicho cebador puede fragmentarse mediante incubación con 10 mM de sulfato de zinc ($ZnSO_4$) o cloruro de zinc ($ZnCl_2$) en 25 mM de Tris-HCl (pH 7,4) a 60 °C durante 30 minutos, como se describe por Liu, *supra*. En otro caso, el ARN se puede incubar con $ZnCl_2$ 10 mM en Tris-HCl 10 mM pH 7 durante 15 minutos a 70 °C.

40 Como sería evidente, en ciertas realizaciones, las secuencias A y B añadidas por los cebadores directo e inverso, respectivamente, pueden ser compatibles con el uso en una plataforma de secuenciación de próxima generación, p. ej., el método del terminador reversible de Illumina, el método de pirosecuenciación de Roche (454), secuenciación por ligamiento de Life Technologies (la plataforma SOLiD) o la plataforma Ion Torrent de Life Technologies. Ejemplos de tales métodos se describen en las siguientes referencias: Margulies *et al* (Nature 2005 437: 376-80); Ronaghi *et al* (Analytical Biochemistry 1996 242: 84-9); Shendure (Science 2005 309: 1728); Imelfort *et al* (Brief Bioinform. 2009 10:609-18); Fox *et al* (Métodos Mol Biol. 2009;553:79-108); Appleby *et al* (Methods Mol Biol. 2009; 513: 19-39) y Morozova (Genomics. 2008 92: 255-64). En ciertos casos, las secuencias A y B pueden proporcionar dos conjuntos de sitios de unión a cebador, uno para amplificar los productos y el otro para secuenciar el producto resultante. En otros casos, los sitios de cebadores de secuenciación pueden añadirse amplificando los productos de PCR finales con cebadores con cola, donde las colas de esos cebadores proporcionan sitios de unión a cebador.

55 En otras realizaciones, los productos pueden secuenciarse utilizando secuenciación de nanoporos (p. ej., como se describe en Soni *et al* Clin Chem 53: 1996-2001 2007, o según lo descrito por Oxford Nanopore Technologies). La secuenciación de nanoporos es una tecnología de secuenciación de una sola molécula mediante la cual una única molécula de ADN se secuenciaría directamente a medida que pasa a través de un nanoporo. Un nanoporo es un agujero pequeño, del orden de 1 nanómetro de diámetro. La inmersión de un nanoporo en un fluido conductor y la aplicación de un potencial (voltaje) a través del mismo producen una ligera corriente eléctrica debido a la conducción de iones a través del nanoporo. La cantidad de corriente que fluye es sensible al tamaño y forma del nanoporo. A medida que una molécula de ADN pasa a través de un nanoporo, cada nucleótido en la molécula de ADN obstruye el nanoporo en un grado diferente, cambiando la magnitud de la corriente a través del nanoporo en diferentes grados. Por lo tanto, este cambio en la corriente a medida que la molécula de ADN pasa a través del nanoporo representa una lectura de la secuencia de ADN. La tecnología de secuenciación de nanoporos se divulga en la patente de EE.UU. n.º 5.795.782, 6.015.714, 6.627.067, 7.238.485 y 7.258.838 y las publicaciones de solicitud de patente de EE. UU. US2006003171 y US20090029477.

65

El método descrito anteriormente se puede utilizar para analizar un genoma o ADNc creado a partir de ARNm de cualquier entidad que contenga ácido nucleico, p. ej., cualquier organismo, fago o virus, etc. En ciertos casos, el método puede utilizarse para analizar un genoma de cualquier organismo, p. ej., plantas, animales (p. ej., reptiles, mamíferos tales como humanos y ratones, insectos, gusanos, peces, etc.), muestras de tejidos, bacterias, hongos (p. ej., levadura), fagos, virus, tejidos de cadáveres, muestras arqueológicas/antiguas, etc. En ciertas realizaciones, el ADN inicial utilizado en el método puede derivarse de un mamífero, donde en ciertas realizaciones el mamífero es un ser humano.

En ciertas realizaciones, el ADN inicial que se está analizando puede derivarse de una sola fuente (p. ej., un solo organismo, virus, tejido, célula, sujeto, etc.), mientras que en otras realizaciones, la muestra de ácido nucleico puede ser una agrupación de ácidos extraídos de una pluralidad de fuentes (p. ej., una agrupación de ácidos nucleicos de una pluralidad de organismos, tejidos, células, sujetos, etc.), donde "pluralidad" significa dos o más. Como tal, en ciertas realizaciones, una muestra de ácido nucleico puede contener ácidos nucleicos de 2 o más fuentes, 3 o más fuentes, 5 o más fuentes, 10 o más fuentes, 50 o más fuentes, 100 o más fuentes, 500 o más fuentes, 1000 o más fuentes, 5000 o más fuentes, hasta e incluir aproximadamente 10.000 o más fuentes. Se puede añadir una secuencia de identificador de muestra a cada una de las fuentes y puede permitir que las secuencias de diferentes fuentes se distingan después de que se analicen. Además, la reacción puede ser multiplexada, de manera que una pluralidad de polinucleótidos diana diferentes (p. ej., 10 a 1000) se dirijan a una sola reacción.

En ciertas realizaciones, los fragmentos de ácido nucleico que deben agruparse con fragmentos de ácido nucleico derivados de una pluralidad de fuentes (p. ej., una pluralidad de organismos, tejidos, células, sujetos, etc.), donde "pluralidad" significa dos o más. En tales realizaciones, los ácidos nucleicos derivados de cada fuente incluyen una secuencia de identificador de muestra tal que puede determinarse la fuente a partir de la cual se derivó cada fragmento de ácido nucleico marcado. En tales realizaciones, cada fuente de muestra de ácido nucleico se correlaciona con una secuencia de identificador de muestra, donde por secuencia de identificador de muestra se entiende que cada secuencia de identificador de muestra diferente empleada puede diferenciarse de cualquier otra secuencia de identificador de muestra empleada en virtud de al menos una característica, p. ej., la secuencia de ácido nucleico de la secuencia de identificador de muestra. Se puede utilizar cualquier tipo de secuencia de identificador de muestra, incluyendo, entre otros, las descritas en la solicitud de patente de EE. UU. en trámite junto con la presente número de serie 11/656.746, presentada el 22 de enero de 2007, y titulada "Nucleic Acid Analysis Using Sequence Tokens", así como la patente de EE. UU. n.º 7.393.665, publicada el 1 de julio de 2008, y titulada "Methods and Compositions for Tagging and Identifying Polynucleotides". En ciertas realizaciones, un conjunto de secuencias de identificador de muestras empleadas para marcar una pluralidad de muestras no necesita tener ninguna propiedad común particular (p. ej., T_f , longitud, composición de base, etc.), como los métodos de marcado asimétricos (y muchos métodos de lectura de etiquetas), incluyendo, entre otros, la secuenciación de la etiqueta o la medición de la longitud de la etiqueta) pueden acomodar una amplia variedad de conjuntos únicos de identificadores de muestras.

Las secuencias de recuento utilizadas en el método, que se derivan de una región de bases degeneradas (RBD) que se unen a las moléculas polinucleótidas de partida y posteriormente se secuencian (p. ej., después de que se realicen ciertas etapas del procedimiento, p. ej., amplificación y/o enriquecimiento, p. ej., PCR). Como se detalla a continuación, la evaluación del número (y en algunos casos, la combinación) de diferentes secuencias RBD presentes en un análisis de secuenciación permite el establecimiento del número (o número mínimo) de diferentes polinucleótidos de partida que se han secuenciado para un polinucleótido particular (o región de interés; RDI).

La secuenciación de ADN normalmente incluye una etapa de unión de un adaptador a los polinucleótidos en una muestra a secuenciar, donde el adaptador contiene un sitio de cebador de secuenciación (p. ej., mediante ligamiento). Como se utiliza en la presente memoria, un "sitio de cebador de secuenciación" es una región de un polinucleótido que es idéntica o complementaria a la secuencia de un cebador de secuenciación (cuando está en una forma monocatenaria) o una región bicatenaria formada entre una secuencia de cebador de secuenciación y su complemento. Los expertos en la materia pueden inferir la orientación específica de un sitio de cebador de secuenciación a partir de las características estructurales del polinucleótido específico que contiene el sitio de cebador de secuenciación.

Además del sitio de cebador de secuenciación, también se une una región de bases degeneradas (RBD) a los polinucleótidos para proporcionar las secuencias de recuento. Una RBD es una región que puede tener una composición o secuencia de bases variables (que se puede considerar como "aleatoria") en comparación con otros polinucleótidos marcados en la muestra. El número de secuencias de recuento diferentes en una población de polinucleótidos en una muestra dependerá del número de bases en la RBD utilizada, así como del número potencial de bases diferentes que pueden estar presentes en cada posición. Una RBD puede incluir 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más bases degeneradas, incluyendo 15 o más, 20 o más, etc. En ciertas realizaciones, la RBD tiene una longitud de 3 a 10 bases. Además, cada posición en una RBD puede tener una composición de base diferente. Por ejemplo, una RBD de 4 bases puede tener cualquiera de las siguientes composiciones: NNNN; NRSN; SWSW; BDHV (véase la Tabla 1 a continuación para el código de nucleótidos de la IUPAC).

65

Código IUPAC de nucleótidos	Base
A	Adenina
C	Citosina
G	Guanina
T (o U)	Timina (o uracilo)
R	A o G
Y	C o T
S	G o C
W	A o T
K	G o T
M	A o C
B	C o G o T
D	A o G o T
H	A o C o T
V	A o C o G
N	Cualquier base

En una realización, los códigos degenerados están diseñados para minimizar las interacciones no específicas entre los códigos y la porción específica diana de los cebadores. Se puede utilizar cualquier método adecuado para realizar esto, incluida la maximización de la distancia de edición de Levenshtein entre códigos degenerados y secuencias diana (para obtener una definición de la distancia de edición de Levenshtein, véase "Levinshtein *et al.*, Binary codes capable of correcting deletions, insertions, and reversals". Sovietic Physics Doklady 10: 707-10).

Aquí se señala que una secuencia de recuento puede estar representada por una única secuencia contigua (es decir, que tiene todas las bases de nucleótidos adyacentes entre sí) o puede estar presente en diferentes ubicaciones en un polinucleótido (es decir, las bases de la secuencia de recuento están separadas por secuencias no RBD, también llamadas RBD divididas). Por ejemplo, una secuencia de recuento puede tener una o más bases en un primer adaptador en una primera ubicación en un polinucleótido y una o más bases en un segundo adaptador en una segunda ubicación en el mismo polinucleótido (p. ej., en los extremos 5' y 3'). No se pretende ninguna limitación en este sentido.

Las RBD pueden diseñarse para facilitar la detección de errores que ocurren en las RBD durante los procedimientos de amplificación que se llevan a cabo antes del análisis de secuencia y/o los errores que ocurren en la propia reacción de secuenciación. En tales realizaciones, las secuencias RBD empleadas se diseñan de tal manera que un error en una secuencia RBD no necesariamente conduce a la generación de otra posible secuencia RBD (lo que da como resultado la identificación incorrecta de replicones derivados del mismo molde como provenientes de un molde diferente debido a una mutación RBD).

La descripción de secuencias de identificación de errores (o corrección de errores) a modo de ejemplo se puede encontrar en toda la literatura (p. ej., se describen en las publicaciones de solicitud de patente de EE. UU. US2010/0323348, titulada "Method and compositions for using error-detecting and/o error-correcting barcodes in nucleic acid amplification process", y US2009/0105959 titulada "System and method for identification of individual samples from a multiplex mixture"). Códigos de corrección de errores pueden ser necesarios para cuantificar números absolutos de moléculas, p. ej., en el caso de un título viral. Muchos informes en la literatura utilizan códigos desarrollados originalmente para la corrección de errores de sistemas binarios (códigos de Hamming, códigos de Reed Solomon, etc.) o los aplican a sistemas cuaternarios (p. ej., códigos cuaternarios de Hamming; véase Generalized DNA barcode design based on Hamming codes, Bystrykh 2012 PLoS One. 2012 7: e36852).

La población de cebadores directos utilizados en el método se puede hacer como un oligonucleótido único que contiene posiciones degeneradas (es decir, posiciones que contienen más de un tipo de nucleótido). Alternativamente, la población de cebadores directos se puede hacer fabricando una matriz de los oligonucleótidos utilizando métodos de síntesis *in situ*, y escindiendo oligonucleótidos del sustrato. Ejemplos de tales métodos se describen, p. ej., en Cleary *et al.* (Nature Methods 2004 1: 241-248) y LeProust *et al.* (Nucleic Acids Research 2010 38: 2522-2540).

En ciertos casos, se utilizan cebadores directos que comprenden uno o más enlaces escindibles. En algunas de estas realizaciones, el enlace utilizado se puede caracterizar por que: (1) no debe perjudicar significativamente la hibridación; (2) debe ser un molde para la síntesis de ADN, es decir, no bloquear la inserción de nucleótidos opuesta al enlace; (3) la escisión debe ser una reacción eficiente; y (4) que la escisión no debe dejar un punto de partida de la polimerización.

Los enlaces escindibles adecuados que pueden emplearse incluyen, entre otros, lo siguiente: sitios escindibles por bases tales como ésteres, particularmente succinatos (escindibles por, por ejemplo, amoníaco o trimetilamina), sales de amonio cuaternario (escindibles por, por ejemplo, diisopropilamina) y uretanos (escindibles por hidróxido de sodio

acuoso); sitios escindibles por ácidos tales como derivados de alcohol bencílico (escindible utilizando ácido trifluoroacético), aglicona de teicoplanina (escindible por ácido trifluoroacético seguido de base), acetales y tioacetales (también escindibles por ácido trifluoroacético), tioéteres (escindibles, por ejemplo, por HF o cresol) y sulfonilos (escindibles por ácido trifluorometano sulfónico, ácido trifluoroacético, tioanisol, o similares); sitios

5 escindibles por nucleófilos tales como ftalamida (escindible por hidracinas sustituidas), ésteres (escindibles por, por ejemplo, tricloruro de aluminio); y amida de Weinreb (escindible por hidruro de litio y aluminio); y otros tipos de sitios químicamente escindibles, incluyendo fosforotioato (escindible por plata o iones mercurícos) y diisopropildialcoxisililo (escindible por iones fluoruro). Otros enlaces escindibles resultarán evidentes para los expertos en la materia o se describen en la literatura y textos pertinentes (p. ej., Brown (1997) Contemporary Organic Synthesis 4 (3); 216-237).

10 En realizaciones particulares, se puede emplear un enlazador fotoescindible ("FE") (p. ej., un enlazador escindible por uv). Los enlazadores fotoescindibles adecuados para su uso pueden incluir enlazadores basados en orto-nitrobencilo, enlazadores fenacilo, enlazadores alcoxibenzoina, enlazadores complejos de cromo-areno, enlazadores NpSSPact y enlazadores pivaloilglicol, como se describe en Guillier *et al* (Chem Rev. 14 de junio de 2000; 100 (6): 2091-158). En realizaciones particulares, es deseable escindir en un enlace que libera un oligonucleótido escindido que no tiene un hidroxilo 3' y, por lo tanto, no puede extenderse. Los grupos de enlace a modo de ejemplo que pueden emplearse en los métodos sujeto se pueden describir en Guillier *et al*, *supra* y Olejnik *et al* (Methods in Enzymology 1998 291: 135-154), y se describen adicionalmente en U.S.P.N. 6.027.890; Olejnik *et al* (Proc. Natl Acad Sci, 92: 7590-94); Ogata *et al*. (Anal. Chem. 2002 74:4702-4708); Bai *et al*. (Nucl. Acids Res. 2004 32:535-541); Zhao *et al* (Anal. Chem. 2002 74: 4259-4268); y Sanford *et al* (Chem Mater. 1998 10: 1510-20), y pueden adquirirse en Ambergen (Boston, MA; NHS-PC-LC-Biotin), Link Technologies (Bellshill, Escocia), Fisher Scientific (Pittsburgh, PA; PIERCE EZ-LINK™ NHS-PC-LC-Biotin) y Calbiochem-Novabiochem Corp. (La Jolla, CA).

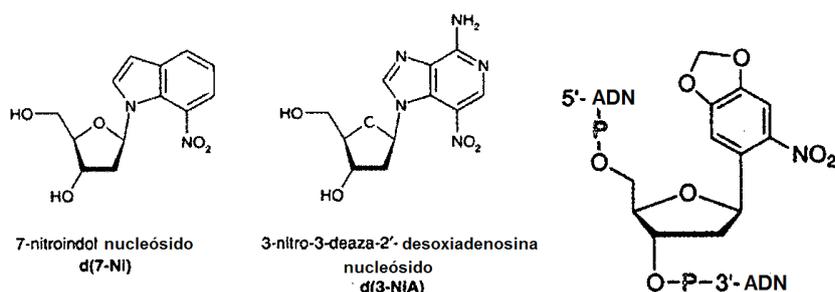
25 Las posibles modificaciones de FE incluyen timidina 5'C-o-nitrofenilatada (Dussy *et al* 2002 PIDM 17590954), 3-nitro-3-deaza-2'-desoxiadenosina (Berthet *et al* 2009 19586934), 7-nitroindol 2-desoxirribonucleósido (Crey-Desbiolles *et al* 2005 PIDM 15767278) y nitropiperonil-2'-desoxirribósido (Pirring *et al* 2001 PIDM 11300902). Si la fotoescisión no es satisfactoria, existen varias estrategias alternativas que no se basan en productos personalizados. La primera implica el uso de una ADN polimerasa capaz de derivación de lesiones o una combinación de una ADN polimerasa de derivación de lesiones y una ADN polimerasa termoestable. Las ADN polimerasas de derivación de lesión a modo de ejemplo incluyen la ADN polimerasa de Sulfolobus (disponible en New England Biolabs Inc), que puede derivar lesiones molde, como las de sitios abásicos (McDonald *et al* 2006 PIDM 16488882) y espaciadores fotoescindibles o ADN polimerasa KOD-1, que pueden incorporan las bases de las modificaciones de la cadena principal del ADN, como los ácidos nucleicos bloqueados (Efficiente enzymatic synthesis of LNA-modified DNA duplexes using KOD DNA polymerase, Veedu RN, Vester B, Wengel J. Org Biomol Chem. 7 de abril de 2009; 7(7): 1404-9 PIDM: 19300826). Después de dos ciclos iniciales, el espaciador fotoescindible se escinde y la reacción de PCR continúa. La segunda estrategia implica el uso de modificaciones de azobenceno que adoptan una forma *cis* cuando se fotoactivan, lo que reduce la temperatura de fusión del oligonucleótido (Asanuma *et al* 2007 PIDM 17401355). La tercera estrategia implica el uso de un oligonucleótido con una secuencia auto-complementaria y grupos de bloqueo que bloquean la auto-hibridación; lo que significa que el cebador está "activado" para la PCR. Después de la fotoescisión, los grupos de bloqueo se eliminaron y el oligonucleótido se autohibrida; desconectando el cebador para la PCR (Dieters 2010/0099159).

En algunas realizaciones, los enlazadores escindibles pueden producirse utilizando:

45 (1) 3-nitro-3-deaza-2'-desoxiadenosina [d(3-NiA)], que tiene propiedades de hibridación cercanas a dATP. dATP incorporó el opuesto de d(3-NiA), aunque la incorporación es ligeramente menos eficiente que dTTP opuesto. La escisión da como resultado un fosfato 3'.

50 (2) 7-nitroindol [(7-Ni)]. Las propiedades de hibridación son similares a las del 5-nitroindol. dATP incorporó el opuesto de d (7-Ni), aunque la incorporación es menos eficiente que las bases naturales opuestas. La escisión da como resultado un fosfato 3'.

(3) Nitropiperonil 2'-desoxirribósido (véase, p. ej., Pirring *et al* J. Org. Chem., 2001, 66 (6), págs. 2067-2071).



Las secuencias de recuento se pueden hacer utilizando RBD que contienen nucleótidos degenerados. Sin embargo, los errores durante la síntesis de oligonucleótidos, la amplificación por PCR y la secuenciación pueden convertir un recuento en otro y proporcionar un recuento falso. Hay dos métodos disponibles para corregir estos errores: el primero es sintetizar múltiples oligonucleótidos con recuentos de corrección de errores y el segundo es utilizar nucleótidos degenerados existentes pero desarrollar modelos de corrección de errores. Los recuentos de corrección de errores tienen una distancia de edición que permite la detección de un número determinado de errores. Para identificar errores, los nucleótidos degenerados pueden intercalarse con secuencias invariantes para proporcionar una manera la identificación de la ubicación aproximada de los errores en los nucleótidos degenerados. Los datos de simulación sugieren que los recuentos se pueden clasificar en recuentos de buena fe o mutantes según (i) el número de moléculas introducidas en la reacción, (ii) el número total de recuentos utilizados en el experimento y (iii) el número de lecturas de secuenciación del experimento. La clasificación se puede hacer utilizando k- que significa agrupación (k = 2 agrupaciones) u otros métodos de agrupación que utilizan la cantidad de lecturas de secuenciación asociadas con cada recuento. Si los recuentos o las lecturas de secuenciación limitan la distribución de lecturas por recuento, ya no serán bimodales y la agrupación generalmente subestima el número de recuentos.

Como tal, en ciertas realizaciones, el recuento puede proporcionar corrección de errores. En ciertos casos, después de que se secuencien los recuentos, las secuencias de los recuentos pueden ser sometidas a agrupación para identificar que los recuentos que contienen una mutación pueden identificarse y corregirse.

También se divulgan composiciones de reactivos y productos intermedios de reacción que son coherentes con el método descrito anteriormente. Por ejemplo, en ciertas realizaciones, la composición puede comprender una población de cebadores directos de la fórmula 5'-A-Y-Z-3', donde: (i) la región A proporciona un sitio de unión a cebador cuando se copia y es la misma en todos los cebadores de la población de cebadores directos; (ii) la región Y varía entre los diferentes cebadores de la población de cebadores y proporciona una secuencia de recuento; y (iii) la región Z es complementaria a un primer sitio en un polinucleótido diana en una población de moléculas de ácido nucleico molde y es la misma en todos los cebadores de la población de cebadores; donde los cebadores directos contienen un resto escindible que, cuando se escinde, produce un oligonucleótido más corto. Además, los elementos de la composición se pueden describir en el contexto del método actual.

Kits

Esta divulgación también divulga un kit para poner en práctica el método propuesto, como se ha descrito anteriormente. Un kit propuesto puede contener al menos: una composición de cebadores directos, como se ha descrito anteriormente, así como uno o más cebadores que pueden utilizarse en el método. Además, el kit también puede comprender reactivos para realizar la extensión del cebador y la PCR (p. ej., polimerasa, nucleótidos y tampón, etc.) y otras enzimas y/o reactivos para realizar el método, p. ej., una nucleasa, etc. Los diversos componentes del kit pueden estar presentes en contenedores separados o ciertos componentes compatibles pueden combinarse previamente en un solo contenedor, según se desee.

Además de los componentes mencionados antes, los presentes kits incluyen normalmente además instrucciones para utilizar los componentes del kit para poner en práctica los métodos propuestos, p. ej., para proporcionar instrucciones para análisis de muestras. Las instrucciones para la práctica de los métodos propuestos generalmente se registran en un medio de registro adecuado. Por ejemplo, las instrucciones se pueden imprimir sobre un sustrato, tal como papel o plástico, etc. Como tales, las instrucciones pueden estar presentes en los kits como un prospecto, en el etiquetado del contenedor del kit o de sus componentes (es decir, asociado con el empaquetado o sub-empaquetado), etc. En otras realizaciones, las instrucciones están presentes como un archivo electrónico de almacenamiento de datos presente en un medio de almacenamiento legible por un ordenador adecuado, p. ej., CD-ROM, disquete, etc. Todavía en otras realizaciones, las instrucciones reales no están presentes en el kit, pero se proporcionan medios para obtener las instrucciones de una fuente remota, p. ej., a través de Internet. Un ejemplo de esta realización es un kit que incluye una dirección web en la que se pueden ver las instrucciones y/o de la que se pueden descargar las instrucciones. Al igual que con las instrucciones, este medio para obtener las instrucciones se registra en un sustrato adecuado.

Utilidad

Como se ha resumido anteriormente, las secuencias de recuento se pueden añadir a los polinucleótidos que se someten a un análisis de secuencia para establecer el número de moléculas polinucleotídicas individuales que se originan en la misma región genómica de la misma muestra original que se han secuenciado. La inclusión de una secuencia de recuento en polinucleótidos sometidos a análisis de secuenciación encuentra uso en una variedad de análisis genéticos, incluido el aumento de la confianza en la nomenclatura de alelos al proporcionar un mecanismo para determinar un valor estadístico para una nomenclatura de alelos, un valor que no puede derivarse solo del número de lectura (véase, p. ej., el documento US20120071331. En este contexto, los recuentos se pueden utilizar, por ejemplo, para proporcionar una medida estadística de la confianza en las nomenclaturas alélicas, aumentando así la confianza en la toma de tales determinaciones de alelos (p. ej., cuando se denominan alelos homocigotos). Las RBD también permiten la identificación de posibles errores de secuenciación o amplificación que tienen un impacto negativo en el análisis genético si no se detectan. El método descrito anteriormente se puede emplear para,

p. ej., estimar el título de un patógeno como un virus, para contar el número de secuencias de ADNc que hay en una muestra, proporcionando así un perfil de expresión génica, para analizar un genoma, p. ej., para determinar el número de copias de las regiones de un genoma o aumentar la confianza de una nomenclatura alélica.

- 5 En ciertas realizaciones, este método puede comprender a) fijar un conjunto de oligonucleótidos que comprende una región de bases degeneradas (RBD) que comprende al menos una base nucleotídica seleccionada entre: R, Y, S, W, K, M, B, D, H, V, N a las moléculas de ácido nucleico de una muestra genómica utilizando los métodos descritos, donde la región diana polimórfica para producir una población de polinucleótidos unidos a adaptador, donde cada molécula de ácido nucleico que comprende la región diana polimórfica en la pluralidad de moléculas de ácido nucleico está unida a una secuencia RBD de un oligonucleótido del conjunto de oligonucleótidos; b) amplificar los polinucleótidos unidos a adaptador para producir polinucleótidos unidos al adaptador amplificados; c) secuenciar una pluralidad de los polinucleótidos unidos a adaptador amplificados que contienen la región diana polimórfica para producir una pluralidad de secuencias, donde la secuencia proporciona, para cada uno de los polinucleótidos unidos a adaptador amplificados que contienen la región diana polimórfica: (i) la secuencia de nucleótidos de al menos una parte de la región diana polimórfica y (ii) la secuencia RBD del oligonucleótido al que está unida la región diana polimórfica; d) calcular la probabilidad de que el alelo esté presente en la muestra genómica utilizando el número de secuencias diferentes contadas en la etapa c); y e) realizar una denominación alélica en función de la probabilidad calculada en la etapa d), donde una mayor probabilidad aumenta la confianza de la denominación alélica. En ciertos casos, el método puede comprender repetir la etapa de recuento c) en alelos adicionales de la región diana polimórfica, f) calcular independientemente la probabilidad de que cada uno de los alelos esté presente en la muestra genómica utilizando el número de secuencias diferentes contadas para cada uno de los alelos y g) denominar un genotipo para la muestra genómica basándose en la probabilidad de que los alelos estén presentes en la muestra genómica.

25 Ejemplos

Ejemplo 1

30 Varias ADN polimerasas termoestables, incluyendo las polimerasas Q5™, PHUSION™, Q5 HOT START™, PHUSION HOT START FLEX™, ONE TAQ™, LONGAMP™ *Taq*, VENT™ y DEEP VENT™ tienen una actividad de exonucleasa 3' a 5' que degrada cebadores libres. Con referencia a la Fig. 2, este ejemplo utiliza una población del cebador directo **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, un cebador inverso **16** que tiene la fórmula 5'-B-W, donde B es una cola 5', un segundo cebador **14** que comprende la secuencia de la región A y un tercer cebador **18** que tiene la secuencia de la cola 5'. En este ejemplo, el cebador directo **10** y el cebador inverso **16** son sensibles a la nucleasa, mientras que el segundo cebador **14** y el tercer cebador **18** están protegidos de la degradación de la nucleasa por un resto en el extremo 3'. Los restos a modo de ejemplo incluyen fosforotioatos (PTO) o cadenas principales de ácidos nucleicos bloqueados. En algunos casos, se incluye más de una modificación en el extremo 3'. Por ejemplo, los PTO especificados durante la síntesis de oligonucleótidos están usualmente presentes en una proporción aproximadamente igual de dos estereoisómeros de los cuales solo uno es resistente a la degradación de las nucleasas.

Estos cebadores se combinan con el molde y las dos primeras rondas de extensión con cebador se realizan en condiciones que favorecen la extensión con cebador en lugar de la actividad exonucleasa de la polimerasa. Una vez que se completan los primeros dos ciclos para producir la población de los segundos productos de extensión, la reacción da una incubación prolongada, p. ej., 30 minutos a 2 horas, p. ej., 65 °C a 75 °C sin ciclado. Durante esta incubación prolongada, la polimerasa aún está activa, pero como no hay ciclado, no se produce una amplificación adicional. El extremo 3' de los productos de extensión bicatenarios de los primeros 2x ciclos está en un flujo entre que es degradado por la actividad de exonucleasa y rellenado de nuevo por la actividad de polimerasa de la polimerasa. En cambio, los cebadores directos utilizados en los primeros 2x ciclos (que no están protegidos) son un sustrato para la actividad de exonucleasa pero no tienen una extensión de equilibrio. La incubación extendida debe eliminar constantemente al menos el extremo 3' de los cebadores no protegidos. La degradación no necesita estar completa, pero debería ser suficiente para eliminar las secuencias de recuento y, por lo tanto, evitar la sobreescritura de secuencias de recuento en los ciclos de PCR posteriores. Los cebadores protegidos están protegidos de la actividad exonucleasa de la polimerasa y se mantienen en niveles que permiten su correcto funcionamiento en reacciones posteriores. Después de la incubación prolongada, la PCR continúa con los cebadores genéricos.

Este método se puede llevar a cabo según el siguiente protocolo:

- 60 (1) mezclar cebadores específicos diana (no protegidos con PTO), cebadores genéricos (protegidos con PTO), con dNTP, molde, enzima, tampón, etc.;
- (2) arranque en caliente de la ADN polimerasa (p. ej., Phusion);
- (3) 2x ciclos de PCR;
- (4) incubación extendida; y
- (5) continuar la PCR para los ciclos restantes
- 65 Tenga en cuenta que no es necesario abrir el tubo aquí durante la reacción.

Ejemplo 2

Varios geles termorreversibles son sólidos a temperaturas más altas, pero líquidos a temperaturas más bajas. Por ejemplo, véase (i) Shim *et al.*, J. Biomed Mater Res. Agosto de 2002; 61(2):188-96; (ii) The Users Guide for Thermogelling PolyVivo PLGA-PEG-PLGA (AK12/AK24), disponible gratuitamente en Polysciotech en "www". seguido de "akinainc.com" seguido de "/pdf/PolyVivo%20AK12-AK24%20thermogelation.pdf"; (iii) Yu *et al.*, Journal of Polymer Science Part A: Polymer Chemistry, volumen 45, edición 6, páginas 1122-1133, 15 de marzo de 2007; y/o (iv) Fadnavis *et al.*, Biotechnol Prog. 1999 de enero; 15(1):98-104.

10 Por consiguiente, en este ejemplo, se mezcla una enzima exonucleasa que escinde ADNmc (p. ej., exonucleasa I) con la solución de gel (estado no sólido) y se pipetea en un tubo. El aumento de temperatura (p. ej., del refrigerador donde se almacena el gel) al tubo hace que el gel se solidifique. Esto atrapa la exonucleasa en un estado sólido.

15 Con referencia a la Fig. 2, este ejemplo utiliza una población del cebador directo **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, un cebador inverso **16** que tiene la fórmula 5'-B-W, donde B es una cola 5', un segundo cebador **14** que comprende la secuencia de la región A y un tercer cebador **18** que tiene la secuencia de la cola 5'. En este ejemplo, el cebador directo **10** y el cebador inverso **16** son sensibles a la nucleasa, mientras que el segundo cebador **14** y el tercer cebador **18** están protegidos de la degradación de la nucleasa por un enlace fosforotioato (PTO) en el extremo 3'.

20 Los cebadores **10**, **14**, **16** y **18** se combinan con el molde en el tubo que contiene la mezcla de gel solidificado/exonucleasa y las dos primeras rondas de extensión con cebador se realizan en condiciones que favorecen el mantenimiento del gel en estado sólido (es decir, una temperatura elevada, p. ej., a una temperatura de al menos 50 °C). Después de que se completen los primeros dos ciclos para producir la población de los productos de segunda extensión, la reacción se enfría a una temperatura adecuada para la actividad enzimática de la exonucleasa, p. ej., a 37 °C (y los contenidos del tubo se mezclan, p. ej., mediante agitación en vórtex) y apropiada para transformar el gel en un estado líquido, liberando así la exonucleasa para degradar los cebadores desprotegidos. Después de permitir que la exonucleasa degrade durante un tiempo suficiente los cebadores no protegidos, la temperatura se incrementa y se mantiene durante el tiempo suficiente para desnaturalizar la exonucleasa (p. ej., 65 °C, 75 °C, 80 °C, 85 °C, 90 °C, 95 °C, 98 °C, etc., dependiendo de la nucleasa). Sin embargo, una etapa separada de desnaturalización por calor puede no ser necesaria ya que la exonucleasa no podrá degradar los cebadores protegidos y la presencia de la exonucleasa no debería, por lo tanto, interferir con las etapas de reacción restantes. La reacción de PCR se continúa utilizando los cebadores protegidos. El aumento de temperatura necesario para continuar la reacción de PCR, después de la etapa de incubación a 37 °C, puede hacer que algunos geles se vuelvan a solidificar, p. ej., dependiendo del tipo de gel exacto utilizado. Sin embargo, la resolidificación no afectaría perjudicialmente a la reacción de PCR.

Este método se puede llevar a cabo según el siguiente protocolo:

- 40 (1) mezclar los cebadores específicos diana (no protegidos con PTO), los cebadores genéricos (protegidos con PTO), con dNTP, molde, enzima, tampón, etc., y añadir al tubo en el que ya se ha colocado el gel solidificado que contiene exonucleasa;
- (2) 2x ciclos de PCR (el gel permanece sólido);
- (3) temperatura más baja para licuar el gel y liberar la exonucleasa (incubación extendida);
- 45 (4) opcional: elevar la temperatura para la desnaturalización en caliente y destruir la actividad de la exonucleasa (incubación extendida); y
- (5) continuar la PCR para los ciclos restantes
- Tenga en cuenta que no es necesario abrir el tubo aquí durante la reacción.

50 Ejemplo 3

Un cebador puede modificarse con una base/cadena principal fotoescindible. La exposición a la luz de una longitud de onda específica es todo lo que se requiere para escindir el cebador. Este método requiere una fuente de luz para la fotoescisión. Las máquinas de PCR pueden diseñarse para incluir una fuente de luz de este tipo para que los tubos no tengan que retirarse de la máquina para la etapa de exposición a la luz. El diseño de tales máquinas puede ser similar al utilizado para la PCR en tiempo real, donde la luz se utiliza para la cuantificación de amplicones en lugar de para la fotoescisión.

60 Con referencia a la Fig. 2, este ejemplo utiliza una población del cebador directo **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, un cebador inverso **16** que tiene la fórmula 5'-B-W, donde B es una cola 5', y un segundo cebador **14** que comprende la secuencia de la región A. En este ejemplo, el cebador directo **10** se modifica en una o más (p. ej., dos o más, tres o más, cuatro o más, cinco o más, etc.) posiciones de nucleótidos con una base/cadena principal fotoescindible y, por lo tanto, se eliminan eficazmente de la reacción (es decir, se escinden) al exponerse a la longitud de onda de luz apropiada. El segundo cebador **14** y el tercer cebador inverso **16** no se modifican con una base/cadena principal fotoescindible y, por lo tanto, están protegidos de la escisión dependiente de la luz. En este ejemplo, el enlazador fotoescindible puede ser (3-(4,4'-Dimetoxitritil)-1-(2-nitrofenil)-propan-1-il-[(2-cianoetil)-(N,N-

diisopropil]-fosforamidita), que se puede atravesar utilizando ADN polimerasa IV de *Sulfolobus* (3A10 o 5D4); véase, p. ej., el documento EP 18012113, que se incorpora por referencia para la divulgación de esas enzimas.

Estos cebadores **10**, **14** y **16** se combinan con el molde y las dos primeras rondas de extensión con cebador se llevan a cabo en condiciones que favorecen la extensión con cebador en ausencia de la longitud de onda de la luz fotoescindible. Una vez que se completan los primeros dos ciclos para producir la población de segundos productos de extensión, la reacción se expone a la longitud de onda de la luz fotoescindible durante un periodo de tiempo apropiado, p. ej., aproximadamente 30 segundos o más (p. ej., 45 segundos o más, 1 minuto o más, 1 minuto, 30 segundos o más, 2 minutos o más, 3 minutos o más, 4 minutos o más, 5 minutos o más, 7,5 minutos o más, 10 minutos o más, 15 minutos o más, 20 minutos o más, etc.), para escindir el cebador directo **10**. La reacción de PCR continúa utilizando los cebadores **14** y **16** que no se pueden fotoescindir.

Este método se puede llevar a cabo según el siguiente protocolo:

- 15 (1) mezclar los cebadores específicos diana (fotoescindibles), los cebadores genéricos (no fotoescindibles), con dNTPs, molde, enzima, tampón, etc.;
- (2) 2x ciclos de PCR (ausencia de la longitud de onda de luz fotoescindibles);
- (3) exponer la reacción a la fotoescisión de la longitud de onda de la luz; y
- (4) continuar la PCR para los ciclos restantes
- 20 Tenga en cuenta que no es necesario abrir el tubo aquí durante la reacción.

Ejemplo 4

Un cebador puede modificarse de varias maneras diferentes, de modo que el cebador (y los productos de extensión que contienen el mismo) se puedan separar de otros componentes en la muestra. Los cebadores que comprenden la secuencia de recuento se pueden eliminar de la mezcla de reacción después de los primeros dos ciclos de extensión con cebador. Como un ejemplo no limitativo, un cebador puede modificarse con uno o más de varios restos diferentes (p. ej., biotina, fluoresceína, dinitrofenol (2,4-dinitrofenol), digoxigenina (DIG), etc.) que permiten que el cebador estén unidos específicamente por una proteína (p. ej., un cebador modificado con biotina puede estar específicamente unido por estreptavidina (o uno de los muchos derivados de estreptavidina) mientras que la fluoresceína, dinitrofenol y DIG pueden unirse específicamente por anticuerpos específicos). En una realización, el resto modificador de interés está unido al extremo 5' del oligonucleótido que comprende la secuencia de recuento. La proteína, a su vez, se puede unir a una perla paramagnética, p. ej., antes de ensamblar la mezcla de reacción de PCR. Alternativamente, un cebador puede unirse directamente a una perla paramagnética o el cebador puede incluir un resto paramagnético (p. ej., Fe₃O₄).

Con referencia a la Fig. 2, este ejemplo fusiona una población del cebador directo **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, un cebador inverso **16** que tiene la fórmula 5'-B-W, donde B es una cola 5', y un segundo cebador **14** que comprende la secuencia de la región A. En este ejemplo, el cebador directo **10** se magnetiza como se ha descrito anteriormente, mientras que el cebador inverso **16** y el segundo cebador **14** no están magnetizados.

Estos cebadores **10**, **14** y **16** se combinan con el molde y las dos primeras rondas de extensión con cebador se realizan en condiciones que favorecen la extensión del cebador en ausencia de un imán. Una vez que se completan los primeros dos ciclos para producir la población de los productos de segunda extensión, la temperatura de la mezcla de reacción se eleva a la temperatura de desnaturalización para disociar el ADN bicatenario en cadenas sencillas. Luego, la reacción se expone a un campo magnético producido por una fuente localizada (p. ej., un imán colocado sobre el tubo) durante una cantidad apropiada de, p. ej., aproximadamente 30 segundos o más (p. ej., 45 segundos o más, 1 minuto o más, 1 minuto, 30 segundos o más, 2 minutos o más, 3 minutos o más, 4 minutos o más, 5 minutos o más, 7,5 minutos o más, 10 minutos o más, 15 minutos o más, 20 minutos o más, etc.), para separar el cebador directo magnetizado **10** y cualquier producto de extensión con cebador **10** de otros componentes de reacción. En este ejemplo, cerca de la fuente magnética (p. ej., en la parte superior del tubo) hay un material inmiscible, o en gran medida inmiscible (p. ej., aceite mineral) que no se mezcla libremente con los componentes de reacción acuosos. El movimiento de los cebadores paramagnéticos y los productos de extensión en el material viscoso elimina efectivamente los cebadores de la reacción. La reacción de PCR se continúa utilizando los cebadores no magnetizados **14** y **18**. El cebador residual **16** no debe interferir con las reacciones de PCR posteriores. Si esto es motivo de preocupación, el diseño de los cebadores y sus concentraciones relativas se pueden ajustar para que el cebador **16** no interfiera con las reacciones de PCR posteriores.

Este método se puede llevar a cabo según el siguiente protocolo:

- 60 (1) mezclar cebadores específicos diana (magnetizados), cebadores genéricos (no magnetizados), con dNTP, molde, enzima, tampón, capa superior de aceite mineral, etc.;
- (2) 2x ciclos de PCR (ausencia de campo magnético, es decir, imán);
- (3) exponer la reacción a la fuente magnética (p. ej., parte superior del tubo); y
- 65 (4) continuar la PCR para los ciclos restantes.

Tenga en cuenta que no es necesario abrir el tubo aquí durante la reacción.

El concepto general de este método se puede adaptar a una celda de flujo, donde la reacción se puede mover a una región en la celda de flujo que contiene un agente de unión (p. ej., estreptavidina, etc.) que se une a los cebadores que se van a eliminar. Una vez que se han eliminado los cebadores, la reacción se puede mover a otra parte de la celda de flujo para continuar con el resto de las etapas.

Ejemplo 5

Se puede diseñar un cebador tal que los nucleótidos en el extremo 5' o el par de bases del extremo 3' con nucleótidos del mismo cebador formen una horquilla (es decir, un dúplex interno). Los cebadores de horquilla se pueden diseñar de modo que las horquillas se formen solo a temperaturas inferiores a una temperatura umbral particular, p. ej., por debajo de 60 °C o por debajo de 50 °C. Por consiguiente, los cebadores de horquilla no forman un dúplex interno a temperaturas en el umbral de temperatura o por encima del mismo y pueden funcionar como cebadores convencionales cuando la temperatura de hibridación de una reacción de PCR es igual o superior al umbral. Sin embargo, los cebadores de horquilla no son funcionales como cebadores de PCR cuando la temperatura de hibridación está por debajo del umbral ya que la horquilla interna impide que el cebador se una al ácido nucleico diana y/o impide que el cebador proporcione un hidroxilo 3' libre para la extensión.

Con referencia a la Fig. 2, este ejemplo utiliza una población del cebador directo **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, un cebador inverso **16** que tiene la fórmula 5'-B-W, donde B es una cola 5', un segundo cebador **14** que comprende la secuencia de la región A y un tercer cebador **18** que tiene la secuencia de la cola 5'. En este ejemplo, el cebador directo **10** y el cebador inverso **16** son cebadores de horquilla como se ha descrito anteriormente, mientras que el segundo cebador **14** y el tercer cebador **18** no son cebadores de horquilla. Además, los cebadores de horquilla están diseñados de modo que la temperatura de umbral sea significativamente superior (p. ej., 5 °C o más, 6 °C o más, 7 °C o más, 8 °C o más, 9 °C o más, 10 °C o más, 11 °C o más, 12 °C o más, etc.) la temperatura de hibridación utilizada para el segundo cebador **14** y el tercer cebador **18**.

Estos cebadores **10**, **14**, **16** y **18** se combinan con el molde y las dos primeras rondas de extensión con cebador se realizan en condiciones que favorecen la hibridación de los cebadores de horquilla (cebador directo **10** y cebador inverso **16**) con el ácido nucleico diana (es decir, temperatura de hibridación en o por encima del umbral de temperatura). Bajo tales condiciones, las horquillas no se forman. Después de que se completan los primeros dos ciclos para producir la población de los productos de segunda extensión, la reacción de PCR continúa con la temperatura de hibridación para los ciclos restantes por debajo de la temperatura umbral para los cebadores de horquilla. Por consiguiente, los cebadores de horquilla (cebador directo **10** y cebador inverso **16**) formarán dúplex internos (es decir, horquillas) y se eliminarán de manera efectiva de la reacción de PCR. Debido a que el segundo cebador **14** y el tercer cebador **18** no son cebadores de horquilla, los ciclos de PCR restantes utilizarán el segundo cebador **14** y el tercer cebador **18**.

Este método se puede llevar a cabo según el siguiente protocolo:

(1) mezclar cebadores específicos diana (cebadores de horquilla), cebadores genéricos (cebadores sin horquilla), con dNTP, molde, enzima, tampón, etc.;

(2) 2x ciclos de PCR (temperatura de hibridación en o por encima del umbral para los cebadores de horquilla); y

(3) continuar la PCR para los ciclos restantes (temperatura de hibridación por debajo del umbral para los cebadores de horquilla).

Tenga en cuenta que no es necesario abrir el tubo aquí durante la reacción.

Ejemplo 6

La temperatura de fusión (T_f) de un cebador se determina por la longitud del cebador, el grado de complementariedad entre el cebador y el ácido nucleico diana y el contenido de la secuencia (p. ej., % de pares de bases GC frente a % de pares de bases AT). Por consiguiente, los cebadores pueden diseñarse para hibridarse por debajo de un umbral particular, pero no por encima del umbral. Por ejemplo, se puede diseñar un cebador corto (p. ej., 8 nucleótidos, 9 nucleótidos, 10 nucleótidos, 11 nucleótidos, 12 nucleótidos, 13 nucleótidos, 14 nucleótidos, 15 nucleótidos, 16 nucleótidos, o 17 nucleótidos) que se hibridará con el ADN diana a bajas temperaturas de hibridación (p. ej., aproximadamente 35-40 °C), pero no a temperaturas superiores a la baja temperatura de hibridación. Por el contrario, se pueden diseñar cebadores más largos que se hibridarán con el ADN diana tanto a baja temperatura de hibridación como a altas temperaturas de hibridación.

Con referencia a la Fig. 2, este ejemplo utiliza una población del cebador directo **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, un cebador inverso **16** que tiene la fórmula 5'-B-W, donde B es una cola 5', un segundo cebador **14** que comprende la secuencia de la región A y un tercer cebador **18** que tiene la secuencia de la cola 5'. Además, este ejemplo utiliza un cebador adicional que comprende una parte de la secuencia de la región Y y una parte de la secuencia de la región Z. El cebador adicional está diseñado para tener una T_f baja de modo que el

cebador se hibridará con el ácido nucleico diana durante los primeros dos ciclos a baja temperatura de hibridación (p. ej., 35-40 °C), pero no se hibridará en ciclos posteriores que tengan temperaturas de hibridación más altas (p. ej., por encima de 45 °C, por encima de 55 °C o por encima de 55 °C).

5 Además, en este ejemplo, el cebador directo **10** está bloqueado de tal manera que no puede hibridarse con el ácido nucleico diana o proporcionar un hidroxilo 3' libre durante los primeros dos ciclos a baja temperatura de hibridación. Por ejemplo, el cebador directo **10** puede tener una secuencia de horquilla en el extremo 5' o 3' que forma un dúplex (horquilla) a la temperatura de hibridación baja, pero no forma una horquilla a las temperaturas de hibridación más altas subsiguientes. Adicionalmente, el cebador directo **10** está diseñado con una T_f tal que el cebador directo **10** se
10 hibridará con el ácido nucleico diana en una temperatura de hibridación intermedia (p. ej., aproximadamente 50-60 °C), pero no superior. El segundo cebador **14** y un tercer cebador **18** están diseñados con una T_f alta, tal que se hibridarán con el ácido nucleico diana a una alta temperatura de hibridación (p. ej., aproximadamente 65-70 °C).

15 Estos cebadores **10**, **14**, **16** y **18** y el cebador adicional se combinan con el molde y las dos primeras rondas de extensión con cebador se llevan a cabo en condiciones (baja temperatura de hibridación) que favorecen la hibridación del cebador adicional con el ácido nucleico diana. Durante estos dos primeros ciclos, el cebador de horquilla (cebador directo **10**) no se hibrida con el ácido nucleico diana (es decir, la temperatura de hibridación es demasiado baja y se forma la horquilla). El segundo cebador **14** y un tercer cebador **18** no se hibridan con el ácido nucleico diana durante los dos primeros ciclos ya que las regiones A y B aún no se han incorporado.

20 Una vez que se completan los primeros dos ciclos para producir la población de segundos productos de extensión, la reacción de PCR continúa durante un ciclo adicional con una temperatura de hibridación intermedia (p. ej., aproximadamente 50-60 °C) para permitir la hibridación y la extensión del cebador directo **10**. Esta etapa incorpora las regiones completas A, Y y Z en los productos de extensión. Los ciclos de PCR restantes se realizan a la
25 temperatura de hibridación más alta (p. ej., aproximadamente 65-70 °C) de modo que solo se utilizan el segundo cebador **14** y el tercer cebador **18**.

Este método se puede llevar a cabo según el siguiente protocolo:

30 (1) mezclar los cebadores específicos diana (cebador directo **10**, cebador adicional y cebador inverso **16**), cebadores genéricos (segundo cebador **14** y un tercer cebador **18**), con dNTP, molde, enzima, tampón, etc.;
(2) 2x ciclos de PCR (la temperatura de hibridación es baja);
(3) 1x ciclo de PCR (la temperatura de hibridación es intermedia); y
(4) continuar la PCR para los ciclos restantes (la temperatura de hibridación es alta).

35

Tenga en cuenta que no es necesario abrir el tubo aquí durante la reacción.

Ejemplo 7

40 La estabilidad térmica y los valores óptimos de temperatura de varias enzimas (p. ej., nucleasa, uracil-ADN glicosilasa, etc.) se pueden aprovechar para eliminar de manera efectiva un conjunto de cebadores de una mezcla de reacción de PCR mientras se deja intacto un segundo conjunto de cebadores. Un ejemplo de dicha enzima termorresistente es la uracil-ADN glicosilasa (UDG), que cataliza de manera eficiente la liberación de uracilo libre del ADN que contiene uracilo. Por consiguiente, se puede diseñar un cebador para incorporar una base de uracil ADN
45 en una o más posiciones nucleotídicas (p. ej., dos o más, tres o más, cuatro o más, cinco o más, etc.) y dicho cebador sería un sustrato para escisión por UDG. Además, la UDG funciona de manera óptima a una temperatura cercana a 37 °C y exhibe actividades reducidas a temperaturas de intervalo de PCR (p. ej., por encima de 50 °C), pero es relativamente resistente a la desnaturalización con calor.

50 Con referencia a la Fig. 2, este ejemplo utiliza una población del cebador directo **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, un cebador inverso **16** que tiene la fórmula 5'-B-W, donde B es una cola 5', un segundo cebador **14** que comprende la secuencia de la región A y un tercer cebador **18** que tiene la secuencia de la cola 5'. En este ejemplo, cuando se utiliza una UDG, el cebador directo **10** se modifica en una o más (p. ej., dos o más, tres o más, cuatro o más, cinco o más, etc.) posiciones de nucleótidos con un desoxirribonucleótido que contiene uracilo.

55

Alternativamente, se puede utilizar una exonucleasa termorresistente como la E. coli RecJ. Al igual que UDG, RecJ tiene una temperatura óptima de aproximadamente 37 °C, exhibe una actividad reducida a temperaturas de intervalo de PCR (p. ej., por encima de 50 °C) y es relativamente resistente a la desnaturalización con calor. Como tal, cuando se utiliza una nucleasa tal como RecJ, el cebador directo **10** es sensible a la nucleasa, mientras que el segundo
60 cebador **14** y el tercer cebador **18** están protegidos de la degradación por nucleasa por un enlace de fosforotioato (PTO) en el extremo 3'.

Los cebadores se combinan con el molde y las dos primeras rondas de extensión con cebador se realizan en condiciones que favorecen la extensión con cebador en lugar de la actividad UDG (o exonucleasa). Una vez que se completan los primeros dos ciclos para producir la población de los segundos productos de extensión, la reacción da
65 una incubación prolongada, p. ej., a aproximadamente 37 °C sin ciclado, para permitir que la UDG (o la nucleasa)

escinda el cebador directo **10** y cebador inverso **16**. Los breves tiempos que pasan a altas temperaturas (p. ej., 90-98 °C) durante los primeros dos ciclos no son suficientes para inactivar la UDG o la exonucleasa ya que las enzimas se eligen por su capacidad para soportar duraciones cortas a tales temperaturas. El segundo cebador **14** y el tercer cebador **18** están protegidos de la escisión ya que no tienen ningún desoxirribonucleótido que contenga uracilo incorporado (p. ej., cuando se utiliza UDG) o están protegidos de la degradación por nucleasa por un enlace PTO en el extremo 3' (p. ej., cuando se utiliza una exonucleasa). A continuación, la reacción da opcionalmente una incubación prolongada, p. ej., a aproximadamente 90-98 °C sin ciclado para inactivar con calor la UDG o la exonucleasa. La reacción de PCR se continúa utilizando el segundo cebador **14** y el tercer cebador **18**.

10 Este método se puede llevar a cabo según el siguiente protocolo:

(1) mezclar cebadores específicos diana (ya sea (i) no están protegidos con PTO o (ii) tienen uno o más desoxirribonucleótidos que contienen uracilo), cebadores genéricos (ya sea (i) que están protegidos con PTO o (ii) no tienen ningún desoxirribonucleótido que contenga uracilo), con dNTP, molde, ADN polimerasa, tampón, enzima termorresistente (p. ej., UDG, exonucleasa, etc.);

(2) 2x ciclos de PCR;

(3) temperatura más baja para permitir que la enzima termorresistente se active (incubación extendida);

(4) opcional: elevar la temperatura para desnaturalizar en caliente la enzima termorresistente (incubación extendida);

y

(5) continuar la PCR para los ciclos restantes.

Tenga en cuenta que no es necesario abrir el tubo aquí durante la reacción.

Ejemplo 8

Los cebadores pueden diseñarse para incluir nucleótidos de ARN en oposición a los nucleótidos de ADN en cualquier posición de nucleótidos, de modo que el cebador puede tener tan pocos como un nucleótido de ARN, donde los nucleótidos restantes son todos nucleótidos de ADN; o como mucho hasta 100 % de nucleótidos de ARN. Los cebadores que tienen nucleótidos de ARN (es decir, los cebadores de ARN), incluso si tienen un 100 % de nucleótidos de ARN, pueden servir como cebadores para la PCR (Shibata *et al.*, Genome Res. Noviembre de 1995; 5(4):400-3). Sin embargo, debido a que el cebador de ARN se incorpora en la primera cadena durante la primera ronda de extensión, las bases de ARN sirven como moldes durante la segunda ronda de extensión. Por consiguiente, la polimerasa utilizada en tal método debe ser capaz de transcripción inversa (polimerización de ADN a partir de un molde de ARN). Un ejemplo de una enzima adecuada es la ADN polimerasa *rTth* (Shibata *et al.*, Genome Res. Noviembre de 1995; 5(4): 400-3). La ADN polimerasa *rTth* tiene actividades de ADN polimerasa dependientes de ARN y ADN y se puede utilizar en condiciones de tampón modificado.

Con referencia a la Fig. 2, este ejemplo utiliza una población del cebador directo **10** de fórmula 5'-A-Y-Z, como se ha descrito anteriormente, un cebador inverso **16** que tiene la fórmula 5'-B-W, donde B es una cola 5', un segundo cebador **14** que comprende la secuencia de la región A y un tercer cebador **18** que tiene la secuencia de la cola 5'. En este ejemplo, el cebador directo **10** tiene, en uno o más (p. ej., dos o más, tres o más, cuatro o más, cinco o más, 10 o más, 15 o más, 20 o más, o todas) posiciones de nucleótidos, Un nucleótido de ARN en lugar de un nucleótido de ADN. El segundo cebador **14** y el tercer cebador **18** no tienen un nucleótido de ARN en ninguna posición de nucleótido.

En este ejemplo, existen al menos dos métodos alternativos para aprovechar los cebadores de ARN (el cebador directo **10** y el cebador inverso **16**):

Método 1 El ARN monocatenario (ARNmc) es menos estable químicamente que el ADNmc a altas temperaturas y, por lo tanto, el ARNmc es sometido a descomposición a altas temperaturas. Por lo tanto, los cebadores son cada vez menos estables a temperaturas más altas a medida que aumenta la fracción de nucleótidos que son ARN en lugar de ADN. Los cebadores descritos anteriormente se combinan con el molde y las dos primeras rondas de extensión con cebador se realizan en condiciones que favorecen la extensión con cebador. Una vez que se completan los primeros dos ciclos para producir la población de los productos de segunda extensión, la reacción da una incubación prolongada, p. ej., entre 80-98 °C sin ciclado, para permitir la degradación de los cebadores de ARN. Los cebadores que no tienen nucleótidos de ARN (segundo cebador **14** y tercer cebador **18**) son estables y no se degradan durante la incubación. La reacción de PCR se continúa utilizando el segundo cebador **14** y el tercer cebador **18**.

Este método se puede llevar a cabo según el siguiente protocolo:

(1) mezclar cebadores específicos de diana (cebadores de ARN, que tienen uno o más nucleótidos de ARN), cebadores genéricos (que no tienen ningún nucleótido de ARN), con dNTP, molde, ADN polimerasa, tampón, etc.;

(2) 2x ciclos de PCR;

(3) elevar la temperatura para degradar los cebadores de ARN (incubación extendida); y

(4) continuar la PCR para los ciclos restantes.

Tenga en cuenta que no es necesario abrir el tubo aquí durante la reacción.

Método 2 El ARN puede ser escindido por enzimas específicas del ARN (es decir, enzimas que escinden el ARN, p. ej., la RNasa H escinde el ARN hibridado, mientras que la RNasa A o la RNasa I escinde el ARN monocatenario).

- 5 Las versiones termoestables de estas enzimas están disponibles. Por lo tanto, los cebadores descritos anteriormente se combinan con el molde y las dos primeras rondas de extensión con cebador se realizan en condiciones que favorecen la extensión del cebador. Mientras que la enzima de escisión de ARN está presente durante los dos primeros ciclos, los cebadores de ARN están presentes en abundancia suficiente para permitir dos rondas de PCR. Después de que se completen los primeros dos ciclos para producir la población de los productos
- 10 de segunda extensión, la reacción da una incubación prolongada a la temperatura óptima para que la enzima que escinde el ARN para degradar cualquier cebador de ARN restante. Los cebadores que no tienen nucleótidos de ARN (segundo cebador **14** y tercer cebador **18**) son estables y no se degradan por la enzima de escisión del ARN. La reacción de PCR se continúa utilizando el segundo cebador **14** y el tercer cebador **18**.
- 15 Este método se puede llevar a cabo según el siguiente protocolo:
- (1) mezclar cebadores específicos diana (cebadores de ARN, que tienen uno o más nucleótidos de ARN), cebadores genéricos (que no tienen ningún nucleótido de ARN), con dNTP, molde, ADN polimerasa, enzima de escisión de ARN, tampón, etc.;
- 20 (2) 2x ciclos de PCR;
- (3) cambiar la temperatura a la temperatura óptima para la degradación de los cebadores de ARN por la enzima que escinde el ARN (incubación extendida); y
- (4) continuar la PCR para los ciclos restantes.
- 25 Tengas en cuenta que no es necesario abrir el tubo aquí durante la reacción.

REIVINDICACIONES

1. Un método de procesamiento de una muestra de ácido nucleico, que comprende:

5 (a) hibridar una población de cebadores directos (10) de la fórmula 5'-A-Y-Z con una población de moléculas de ácido nucleico molde (12), donde:

(i) la región A proporciona un sitio de unión a cebador cuando se copia por una polimerasa y es la misma en todos los cebadores de la población de cebadores directos;

10 (ii) la región Y varía entre los diferentes cebadores de la población de cebadores y proporciona una secuencia de recuento; y

15 (iii) la región Z es complementaria a un primer sitio en un polinucleótido diana en una población de moléculas de ácido nucleico molde y es la misma en todos los cebadores de dicha población de cebadores; y

dicha hibridación se realiza en presencia de

i) un segundo cebador (14) que comprende la secuencia de la región A;

20 ii) un tercer cebador (18) que tiene la secuencia B; y

iii) un cebador inverso (16) de fórmula 5'-B-W, donde la región B es una cola 5' de la secuencia B y la secuencia W es complementaria a un sitio en dicho polinucleótido diana;

25 donde el segundo y tercer cebadores están protegidos de la degradación por una exonucleasa 3' a 5';

(b) extender i) los cebadores directos que se hibridan con dicho polinucleótido diana en las etapas (a) e ii) el cebador inverso, utilizando una polimerasa termoestable que tiene una actividad de exonucleasa 3' a 5' que degrada los cebadores libres, utilizando dicho polinucleótido diana como molde para producir una población de primeros productos de extensión (22) que comprenden un sitio de unión (24) para dicho cebador inverso;

30 (c) hibridar dicho cebador inverso con el sitio de unión para dicho cebador inverso en la población de primeros productos de extensión producidos por la etapa (b);

35 (d) extender dicho cebador inverso para producir una población de segundos productos de extensión (28) que comprende el complemento de la región A y el complemento de la región Y, donde el complemento de la región Y es diferente en cada molécula en dicha población de segundos productos de extensión ;

40 (e) desactivar selectivamente, después de la etapa (d) cualquier cebador directo que no esté extendido, y

(f) amplificar dicha población de segundos productos de extensión mediante PCR utilizando: i) dicho segundo cebador que comprende la región A y ii) dicho tercer cebador que tiene la secuencia B, para producir una población de productos de PCR (30) en la que los productos clonalmente relacionados se marcan con la misma secuencia de recuento y los productos que no están relacionados clonalmente se marcan con una secuencia diferente entre sí;

donde dichas etapas (a) a (f) se realizan en un recipiente cerrado y no se añaden reactivos adicionales al recipiente durante dicho método.

50 2. El método de la reivindicación 1, donde la secuencia Y comprende una región de bases degeneradas (RBD) que comprende al menos una base de nucleótido seleccionada entre: R, Y, S, W, K, M, B, D, H, V, N, y sus versiones modificadas.

55 3. El método de la reivindicación 1 o 2, donde el cebador directo contiene uno o más nucleótidos o enlaces no encontrados en el ADN natural que hacen que el cebador directo sea susceptible a una condición de escisión específica.

4. El método de la reivindicación 3, donde el cebador directo contiene uno o más ribonucleótidos.

60 5. El método de la reivindicación 1 o 2, donde el cebador directo o el cebador inverso, o ambos, contiene una secuencia de identificador de muestra.

65 6. El método de una cualquiera de las reivindicaciones 1-3, donde el método comprende además la secuenciación en al menos parte de la población de dichos productos de PCR (30) para determinar las secuencias de al menos parte de una pluralidad de polinucleótidos diana (12) y sus secuencias de recuento asociadas.

7. El método de la reivindicación 4, que comprende contar el número de moléculas diana (12) secuenciadas utilizando dichas secuencias de recuento, donde dicho recuento comprende identificar secuencias de recuento que contienen una mutación por agrupación.
- 5 8. El método de la reivindicación 4, comprende evaluar el número de diferentes secuencias de recuento para establecer el número de polinucleótidos de partida diferentes que se han secuenciado para una región particular de interés.
- 10 9. El método de una cualquiera de las reivindicaciones precedentes, donde tanto el cebador directo como el cebador inverso tienen una secuencia de recuento.
10. El método de la reivindicación 7, donde la secuencia de recuento en el cebador directo es diferente de la secuencia de recuento en el cebador inverso.
- 15 11. El método de una cualquiera de las reivindicaciones precedentes, donde hay al menos 100 cebadores directos diferentes (10) en dicha población de cebadores directos, donde los cebadores directos difieren en la región Y.

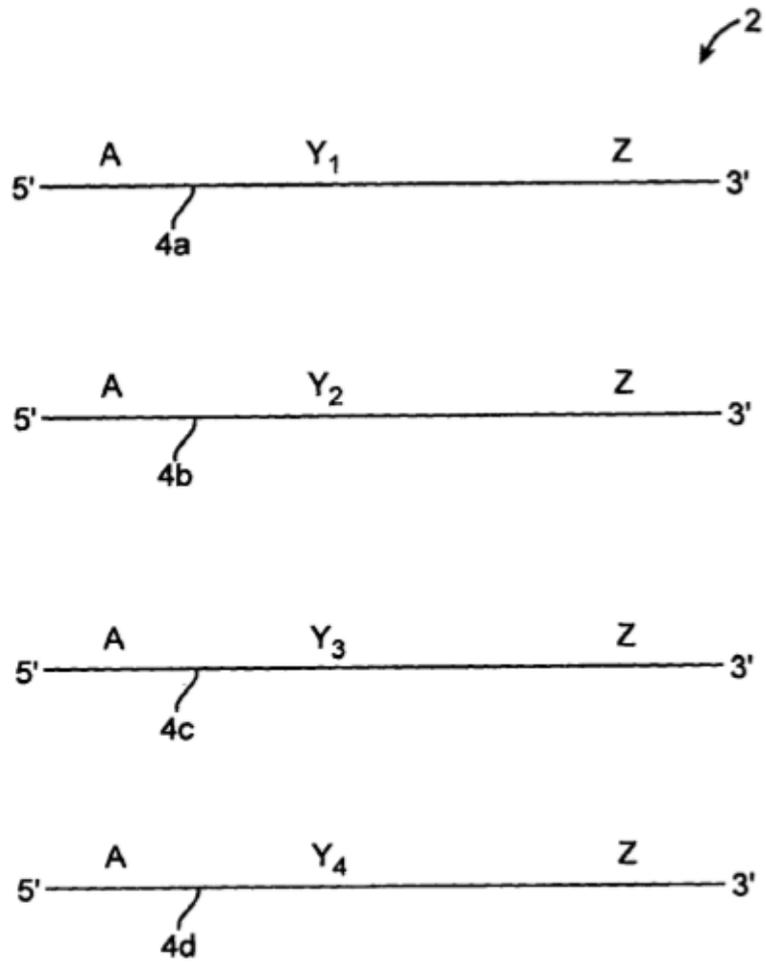


FIG. 1

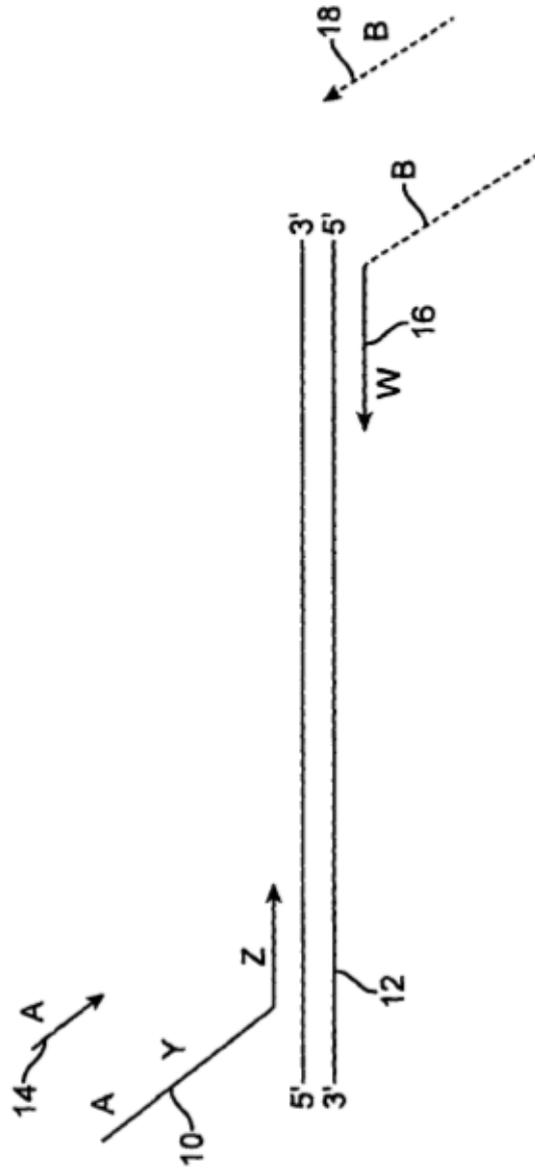


FIG. 2

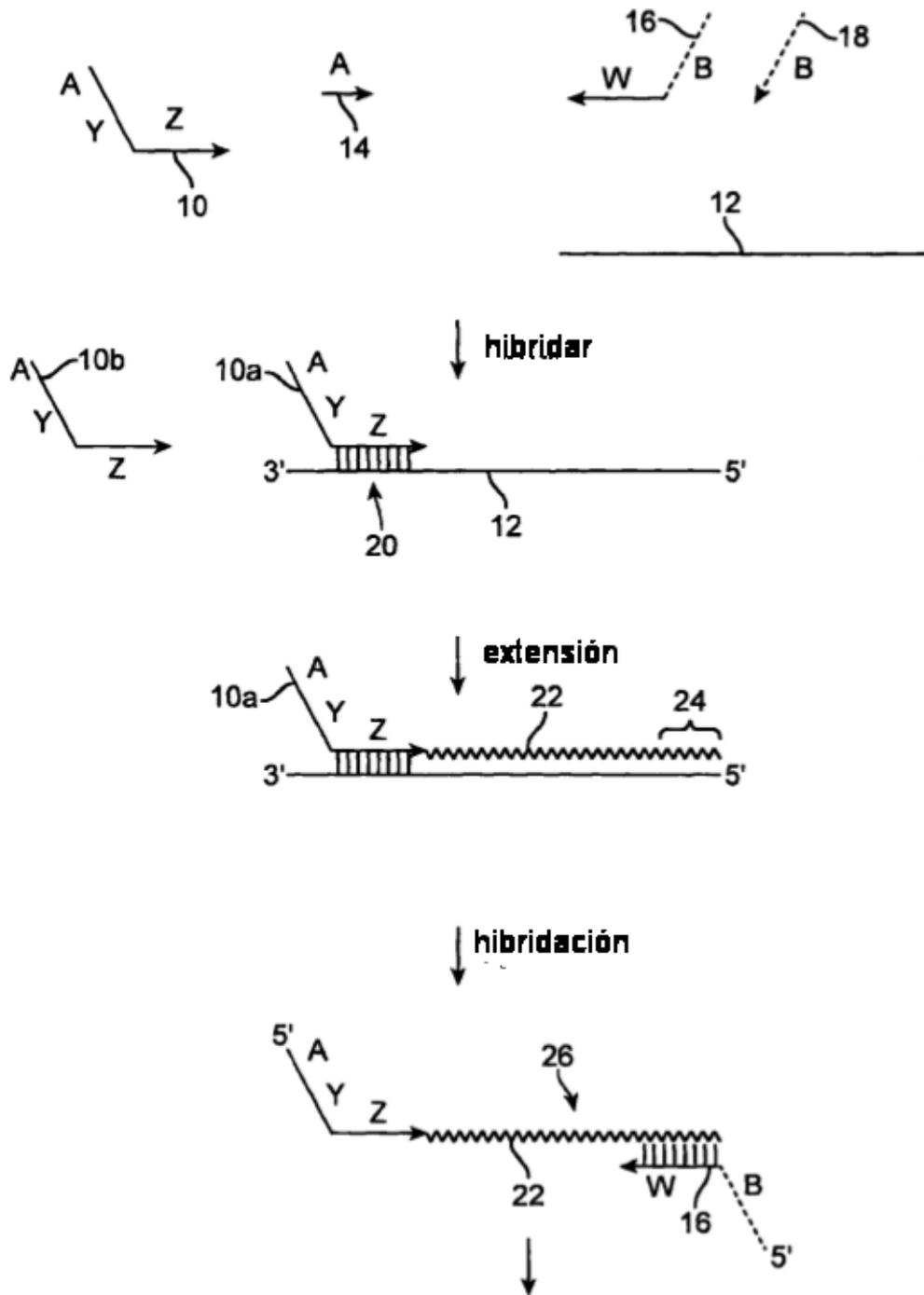


FIG. 3

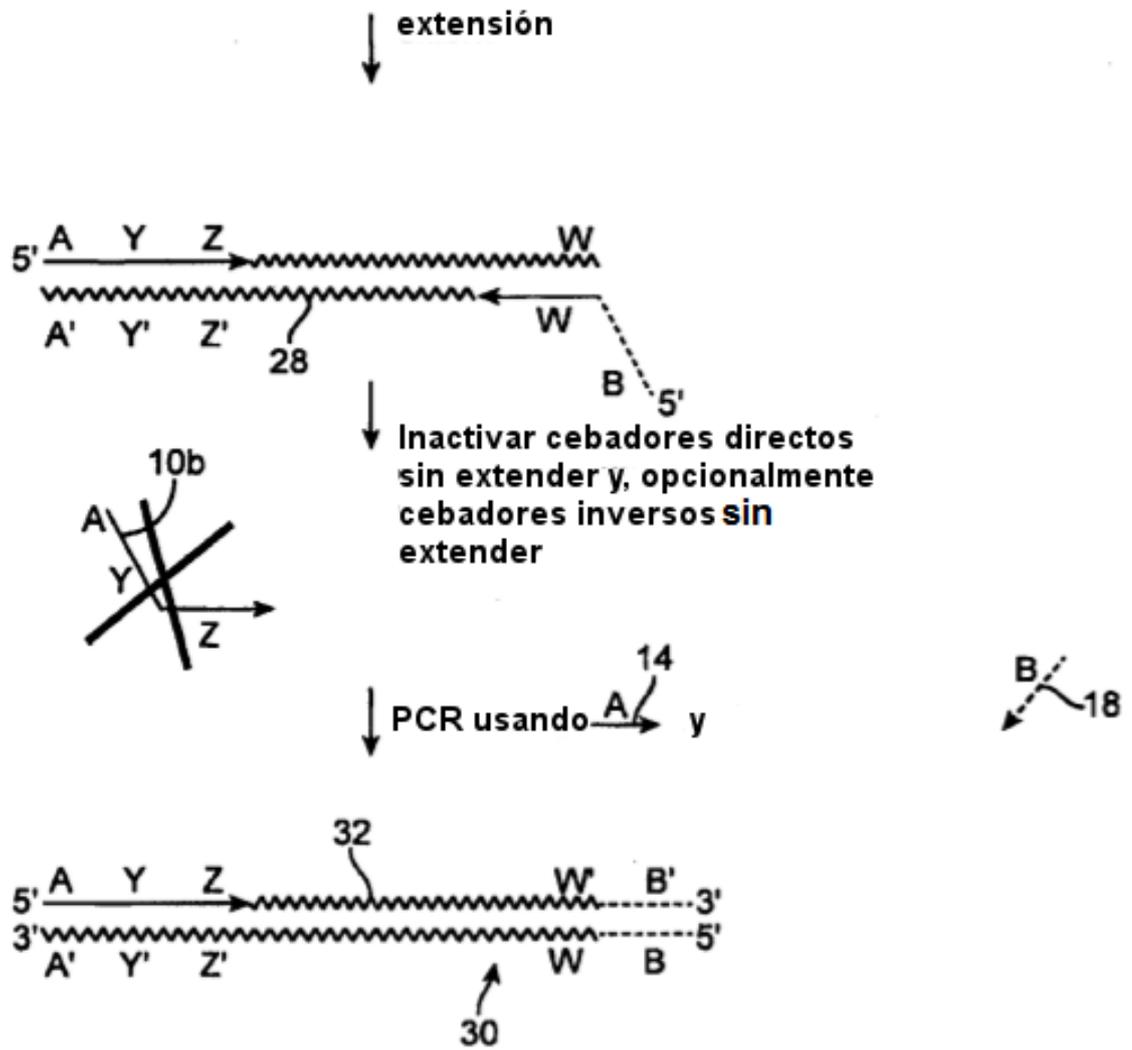


FIG. 3 (Cont.)

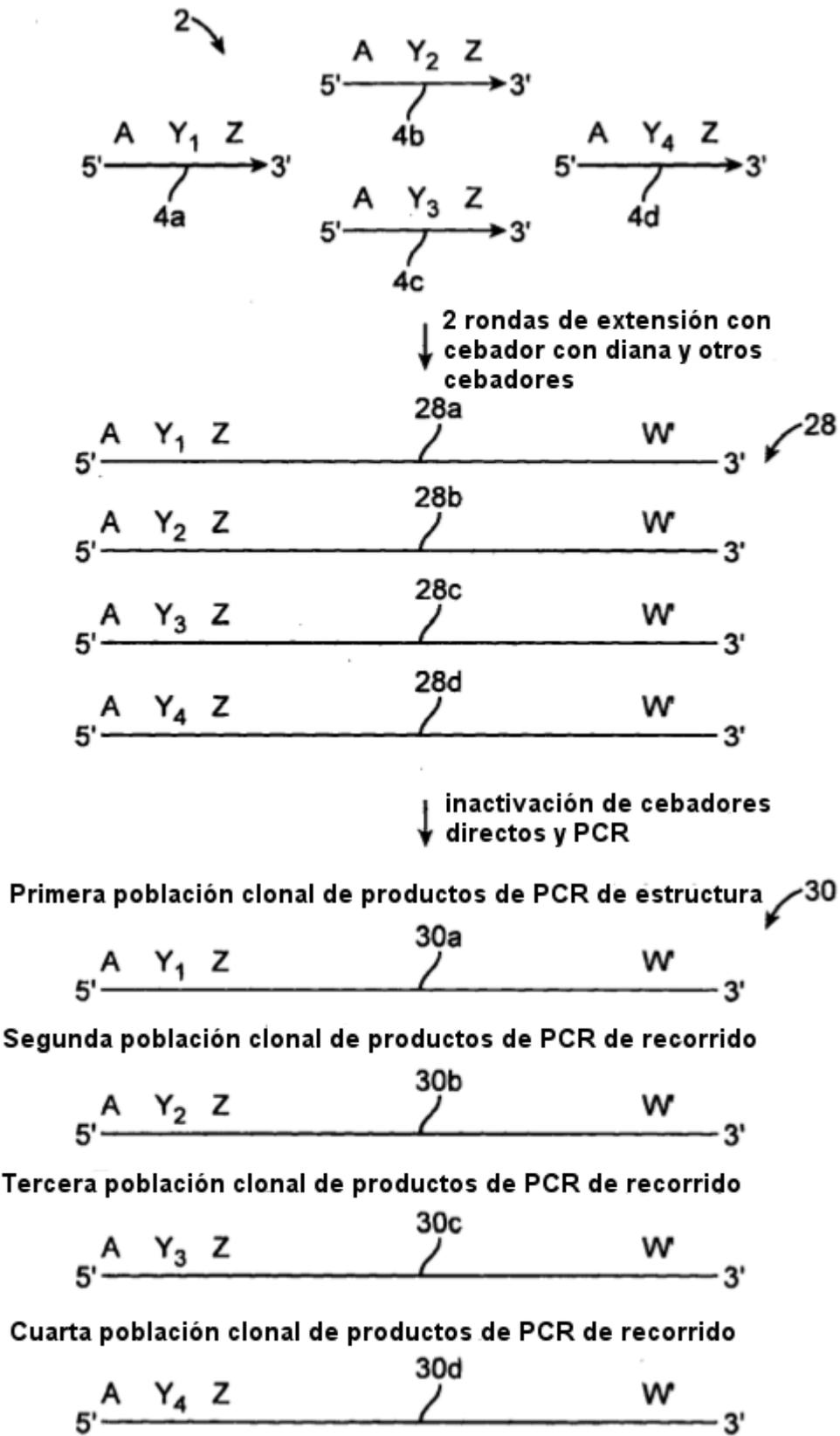


FIG. 4