

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 917 400**

51 Int. Cl.:

C12Q 1/6844 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.02.2013** **E 18183934 (1)**

97 Fecha y número de publicación de la concesión europea: **06.04.2022** **EP 3428290**

54 Título: **Composiciones y métodos para la amplificación de ácidos nucleicos**

30 Prioridad:

26.07.2012 US 201261676153 P

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

08.07.2022

73 Titular/es:

**ILLUMINA, INC. (100.0%)
5200 Illumina Way
San Diego, CA 92122, US**

72 Inventor/es:

**FAN, JIAN-BING;
GUNDERSON, KEVIN L.;
STEEMERS, FRANK J.;
POKHOLOK, DMITRY K. y
KAPER, FIONA**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 917 400 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

Composiciones y métodos para la amplificación de ácidos nucleicos

Campo de la invención

La presente descripción se refiere en general a los campos de la genética y la medicina. Más específicamente, la presente descripción se refiere a la amplificación de bancos de ácidos nucleicos, incluidos los genomas completos.

Antecedentes

Las aplicaciones de genotipado de alto rendimiento se basan en una amplificación eficiente y relativamente imparcial, tal como la amplificación del genoma completo u hologenómica (WGA), del ADN genómico. La amplificación con cebadores aleatorios y la amplificación por desplazamiento múltiple (MDA) se pueden usar en una gran cantidad de aplicaciones diferentes, desde la amplificación de ADN hasta la creación de bancos para secuenciación genómica. Sin embargo, tales métodos pueden dar lugar a una amplificación sesgada que puede producir un conjunto de datos sesgado. La capacidad de amplificar el ADN deseado de una manera relativamente imparcial es importante en muchas aplicaciones, en concreto en la secuenciación. Sin embargo, sigue existiendo una gran necesidad de metodologías de amplificación que den como resultado una amplificación imparcial mejorada de los bancos de ácidos nucleicos deseados. Las solicitudes de patente internacional n.ºs WO2011/142861, WO2008/051928, WO2009/105531 y WO2006/020617 explican el uso de cebadores aleatorios y específicos en los métodos de amplificación.

Compendio de la invención

En la presente memoria se presentan métodos, composiciones y kits para la amplificación de muestras de ácidos nucleicos para generar bancos de ácidos nucleicos. Los métodos, las composiciones y los kits que se presentan en la presente memoria son sorprendentemente eficaces para reducir el sesgo que se produce cuando se amplifica una muestra de ácidos nucleicos utilizando cebadores aleatorios.

Por consiguiente, en la presente memoria se presenta un kit según se da a conocer en la reivindicación 1 y un método según la reivindicación 7. De forma idónea, se presenta un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: a) proporcionar un conjunto de cebadores para amplificación a una muestra de ácidos nucleicos, en donde el conjunto de cebadores para amplificación comprende numerosos cebadores aleatorios y numerosos cebadores específicos de locus, en donde los cebadores específicos de locus están configurados para amplificar numerosas regiones predeterminadas del banco de ácidos nucleicos, y en donde los cebadores aleatorios están en mayor abundancia que los cebadores específicos de locus; y b) amplificar el banco de ácidos nucleicos gracias al uso del conjunto de cebadores para amplificación, creando así un banco de ácidos nucleicos.

En algunos aspectos del método descrito anteriormente, el conjunto de cebadores para amplificación puede ser una mezcla de cebadores. En ciertos aspectos, los cebadores aleatorios tienen una longitud de aproximadamente 5 a aproximadamente 18 nucleótidos. En algunos otros aspectos, los cebadores aleatorios tienen una longitud de 9 nucleótidos. En algunos aspectos, los cebadores específicos de locus pueden tener la misma longitud que los cebadores aleatorios. En otros aspectos, los cebadores específicos de locus pueden ser más cortos que los cebadores aleatorios. En otros aspectos más, los cebadores específicos de locus pueden ser más largos que los cebadores aleatorios. En algunos aspectos, los cebadores específicos de locus están configurados para bloquear o reducir la amplificación de una o más regiones predeterminadas del banco de ácidos nucleicos. En algunos aspectos, los cebadores específicos de locus comprenden un bloqueo en 3' o carecen de un grupo OH en 3'.

En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos aspectos, el ADN genómico comprende ADN humano. En ciertos otros aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos comprende ADN de una o más plantas, hongos, protistas, bacterias y/o arqueas. En algunos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos comprende ADN de orgánulos celulares, tales como ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

En algunos aspectos del método descrito anteriormente, los cebadores para amplificación aleatoria comprenden uno o más cebadores casi aleatorios que se seleccionan del grupo que consiste en un conjunto de cebadores ricos en AT para la amplificación aleatoria; un conjunto de cebadores para amplificación aleatoria que comprenden extremos 5' ricos en AT; un conjunto de cebadores para amplificación aleatoria de longitud variable, en el que cada cebador comprende una porción aleatoria en 3' y un extremo 5' degenerado, en donde extremo 5' degenerado puede tener una longitud proporcional al contenido de A/T de la porción aleatoria en 3' del cebador; un conjunto de cebadores para amplificación con la T_m normalizada, en el que cada cebador del conjunto comprende uno o más análogos de base que pueden normalizar la T_m de cada cebador con respecto a la T_m de otros cebadores en el conjunto de cebadores; un conjunto de cebadores para amplificación aleatoria, en el que cada cebador comprende una parte 3' aleatoria y una parte cebadora en 5' constante; un conjunto de cebadores para amplificación aleatoria, en el que cada cebador comprende una porción en 3' aleatoria y una porción cebadora en 5' constante, y en el que la porción en 3' aleatoria

comprende ARN; un conjunto de cebadores para amplificación aleatoria, en el que cada cebador comprende una porción en 3' aleatoria y una porción de cebado en 5' constante, y en el que la porción en 3' aleatoria comprende al menos una base no natural seleccionada del grupo que consiste en ácidos nucleicos que incluyen 2'-desoxi-2-tiotimidina (2-tío-dT), 2-aminopurina-2'-desoxirribósido (2-amino-dA), N4-etil-2'-desoxicitidina (N4-Et-dC), N4-metildesoxicitidina (N4-Me-dC), 2'-desoxiinosina, 7-desazaguanina (7-desaza-G), 7-yodo-7-desazaguanina (I-desaza-G), 7-metil-7-desazaguanina, (MecG), 7-etil-7-desazaguanina (EtcG) y cualquier combinación de los conjuntos de cebadores anteriores. Los cebadores casi aleatorios presentados anteriormente se describen con mayor detalle en la presente memoria. En algunas realizaciones descritas en la presente memoria, los cebadores casi aleatorios se dan a conocer en parejas o conjuntos.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende numerosos cebadores aleatorios y numerosos cebadores específicos de locus configurados para amplificar numerosas regiones predeterminadas de un banco de ácidos nucleicos. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para usar los cebadores aleatorios y los cebadores específicos de locus en un conjunto de reacciones de amplificación, en donde los cebadores aleatorios son más abundantes que los cebadores específicos de locus. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos.

En algunos aspectos del kit de amplificación descrito en la presente memoria, el kit comprende además una ADN polimerasa. En ciertos aspectos, los cebadores aleatorios pueden tener una longitud de aproximadamente 5 a 18 nucleótidos. En algunos otros aspectos, los cebadores aleatorios tienen una longitud de 9 nucleótidos. En algunos aspectos, los cebadores específicos de locus pueden tener la misma longitud que los cebadores aleatorios. En otros aspectos, los cebadores específicos de locus pueden ser más cortos que los cebadores aleatorios. En otros aspectos más, los cebadores específicos de locus pueden ser más largos que los cebadores aleatorios.

Además del método anterior, también se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: a) amplificar una muestra de ácidos nucleicos con un conjunto de cebadores ricos en AT para la amplificación aleatoria. En ciertos aspectos, el conjunto de cebadores ricos en AT para la amplificación aleatoria es una mezcla de cebadores.

En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos otros aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos comprende ADN de una o más plantas, hongos, protistas, bacterias y/o arqueas. En algunos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos comprende ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

En algunos aspectos del método descrito anteriormente, la composición global del conjunto de cebadores ricos en AT para la amplificación aleatoria es mayor que el 25 % de A y el 25 % de T. En ciertos aspectos, el conjunto de cebadores ricos en AT para la amplificación aleatoria comprende el 30 % de A, 20 % de dC, 20 % de G y 30 % de T. En algunos otros aspectos, el conjunto de cebadores ricos en AT para la amplificación aleatoria comprende el 35 % de A, 15 % de C, 15 % de G y 35 % de T. En otros aspectos más, el conjunto de cebadores ricos en AT para la amplificación aleatoria tiene una longitud de 5 a 18 nucleótidos.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores ricos en AT para la amplificación aleatoria. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos otros aspectos, el kit comprende además una ADN polimerasa. En otros aspectos más, el conjunto de cebadores ricos en AT para la amplificación aleatoria es una mezcla de cebadores.

También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: a) amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en donde los cebadores para amplificación aleatoria comprenden colas en 5' ricas en AT. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos otros aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos comprende ADN de una o más plantas, hongos, protistas, bacterias y/o arqueas. En algunos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos de orgánulos celulares tales como ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

En algunos aspectos del método descrito anteriormente, la cola en 5' rica en AT comprende el 30 % de A, 20 % de C, 20 % de G y 30 % de T. En ciertos aspectos, la cola en 5' rica en AT comprende el 35 % de A, 15 % de C, 15 % de G y 35 % de T. En algunos otros aspectos, la cola en 5' rica en AT comprende el 40 % de A, 10 % de C, 10 % de G y 40 %

de T. En otros aspectos más, la cola en 5' rica en AT comprende el 50 % de A y el 50 % de T.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación aleatoria, en donde los cebadores para amplificación aleatoria comprenden colas en 5' ricas en AT. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos otros aspectos, el kit comprende además una ADN polimerasa. En otros aspectos más, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria de longitud variable, en el que cada cebador para amplificación aleatoria de longitud variable comprende una porción en 3' aleatoria y una cola en 5' degenerada, en donde la cola en 5' degenerada es proporcional en longitud al contenido de A/T de la porción aleatoria en 3' del cebador. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria de longitud variable es una mezcla de cebadores.

En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos otros aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos comprende ADN de una o más plantas, hongos, protistas, bacterias y/o arqueas. En algunos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos comprende ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

En ciertos aspectos del método descrito anteriormente, la cola en 5' del cebador para amplificación aleatoria de longitud variable comprende al menos un nucleótido degenerado por cada dos nucleótidos A o T en la porción en 3' aleatoria.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación aleatoria de longitud variable, en donde cada cebador para amplificación aleatoria de longitud variable comprende una porción en 3' aleatoria y una cola en 5' degenerada, en donde la cola en 5' degenerada es proporcional en longitud al contenido de A/T de la porción aleatoria en 3' del cebador. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos otros aspectos, el kit comprende además una ADN polimerasa. En otros aspectos más, el conjunto de cebadores para amplificación aleatoria de longitud variable es una mezcla de cebadores.

También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación con la T_m normalizada, en donde cada cebador del conjunto de cebadores para amplificación con la T_m normalizada comprende uno o más análogos de bases que normalizan la T_m de cada cebador a la T_m de otros cebadores en el conjunto de cebadores. En ciertos aspectos, el conjunto de cebadores para amplificación con la T_m normalizada es una mezcla de cebadores.

En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos otros aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos comprende ADN de una o más plantas, hongos, protistas, bacterias y/o arqueas. En algunos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos comprende ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

En ciertos aspectos del método descrito anteriormente, el uno o más análogos de bases se seleccionan del grupo que consiste en 2-tio-dT, 2-amino-dA, N4-Et-dC y 7-desaza-G.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en el que el kit comprende un conjunto de cebadores para amplificación con la T_m normalizada, en el que cada cebador del conjunto de cebadores para amplificación con la T_m normalizada comprende uno o más análogos de bases que normalizan la T_m de cada cebador a la T_m de otros cebadores para amplificación en el kit. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores de amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos otros aspectos, el kit comprende además una ADN polimerasa. En otros aspectos más, el conjunto de cebadores para amplificación con la T_m normalizada es una mezcla de cebadores.

También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: a) amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en donde cada cebador comprende una porción en 3' aleatoria y una porción en 5' constante de cebado, con lo que se producen así productos de amplificación, en donde cada producto de amplificación comprende la porción en 5' de cebado constante; b) circularizar los productos de amplificación; y c) amplificar los productos de amplificación circularizados con el uso de cebadores que se hibridan con la porción de cebado en 5' constante. En ciertos aspectos, la amplificación de la etapa (c) comprende realizar una amplificación por desplazamiento múltiple. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos otros aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos comprende ADN de una o más plantas, hongos, bacterias protistas y/o arqueas.

5 En algunos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos comprende ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

En ciertos aspectos del método descrito anteriormente, los cebadores para amplificación comprenden al menos una base no natural entre la porción en 3' aleatoria y la porción cebadora en 5' constante. En ciertos aspectos, la base no natural es isoC.

10 También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación aleatoria que comprenden una porción en 3' aleatoria y una porción cebadora en 5' constante. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos otros aspectos, el kit comprende además una ADN polimerasa. En otros aspectos más,

15 el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: a) amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en donde cada cebador comprende una porción en 3' aleatoria y una porción en 5' constante de cebado, y en donde la porción en 3' aleatoria comprende ARN, con lo que se producen

20 así los productos de amplificación, en donde cada producto de amplificación comprende la porción de cebado en 5' constante. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos otros aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos comprende ADN de una o más plantas, hongos, bacterias protistas y/o arqueas.

25 En algunos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos comprende ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

En ciertos aspectos, el método descrito anteriormente comprende, además: b) circularizar los productos de amplificación; y c) amplificar los productos de amplificación circularizados con el uso de cebadores que se hibridan con la porción de cebado en 5' constante. En ciertos aspectos, la amplificación de la etapa c) comprende realizar una amplificación por desplazamiento múltiple.

30 En ciertos aspectos, el método descrito anteriormente comprende, además: b) circularizar los productos de amplificación; y c) amplificar los productos de amplificación circularizados con el uso de cebadores que se hibridan con la porción de cebado en 5' constante. En ciertos aspectos, la amplificación de la etapa c) comprende realizar una amplificación por desplazamiento múltiple.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación aleatoria, en donde los cebadores comprenden una porción aleatoria en 3' y una porción de cebado constante en 5', en donde la porción aleatoria en 3' comprende ARN. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos otros aspectos, el kit comprende además una ADN polimerasa. En otros aspectos más, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

35 También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en donde cada cebador del conjunto comprende una porción en 3' aleatoria y una porción constante de cebado en 5', y en donde la porción aleatoria en 3' comprende al menos una base no natural seleccionada del grupo que consiste en: 2-tío-dT y 2-amino-dA, con lo que se producen así los productos de amplificación, en donde cada producto de amplificación comprende la porción de cebado en 5' constante.

40 También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en donde cada cebador del conjunto comprende una porción en 3' aleatoria y una porción constante de cebado en 5', y en donde la porción aleatoria en 3' comprende al menos una base no natural seleccionada del grupo que consiste en: 2-tío-dT y 2-amino-dA, con lo que se producen así los productos de amplificación, en donde cada producto de amplificación comprende la porción de cebado en 5' constante.

45 En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos otros aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos comprende ADN de una o más plantas, hongos, bacterias protistas y/o arqueas.

50 En algunos aspectos, la muestra de ácidos nucleicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos comprende ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

En ciertos aspectos, el método descrito anteriormente comprende, además: b) circularizar los productos de amplificación; y c) amplificar los productos de amplificación circularizados con el uso de cebadores que se hibridan con la porción de cebado en 5' constante. En ciertos aspectos, la amplificación de la etapa c) comprende realizar una amplificación por desplazamiento múltiple.

55 En ciertos aspectos, el método descrito anteriormente comprende, además: b) circularizar los productos de amplificación; y c) amplificar los productos de amplificación circularizados con el uso de cebadores que se hibridan con la porción de cebado en 5' constante. En ciertos aspectos, la amplificación de la etapa c) comprende realizar una amplificación por desplazamiento múltiple.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende cebadores para amplificación aleatoria, en donde los cebadores para amplificación aleatoria comprenden una porción en 3' aleatoria y una porción de cebado en 5' constante, en donde la porción en 3' aleatoria comprende al menos

una base no natural seleccionada del grupo que consiste en: 2-tío-dT y 2-amino-dA. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos otros aspectos, el kit comprende además una ADN polimerasa. En otros aspectos más, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

- 5 También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos genómicos que comprende: a) proporcionar un conjunto de cebadores para amplificación a una muestra de ácidos nucleicos genómicos, en donde el conjunto de cebadores para amplificación comprende una primera pluralidad de cebadores de secuencia aleatoria, y proporcionar una segunda pluralidad de cebadores de secuencia específica de especie configurados para amplificar regiones genómicas definidas en la muestra de ácidos nucleicos, en donde los cebadores de secuencia específica de especie tienen una abundancia igual o mayor que los cebadores aleatorios, y b) amplificar la muestra de ácidos nucleicos genómicos usando el conjunto de cebadores para amplificación, con lo que se crea así un banco de ácidos nucleicos.

- 10 En algunos aspectos del método descrito anteriormente, la muestra de ácidos nucleicos genómicos comprende un ADN genómico. En ciertos otros aspectos, la muestra de ácidos nucleicos genómicos comprende numerosos ADN genómicos. En otros aspectos más, la muestra de ácidos nucleicos genómicos comprende ADN de una o más especies económicamente importantes. En algunos aspectos, la muestra de ácidos nucleicos genómicos comprende ADN de una o más plantas, hongos, bacterias, protistas y/o arqueas. En algunos aspectos, la muestra de ácidos nucleicos genómicos comprende ácidos nucleicos distintos del ADN. En ciertos aspectos, la muestra de ácidos nucleicos genómicos comprende ADN mitocondrial, ADN cloroplástico y/o ADN de otros orgánulos celulares.

- 15 En algunas realizaciones del método descrito anteriormente, la muestra de ácidos nucleicos genómicos es de un mamífero, tal como un ser humano u otro animal económicamente relevante. En determinadas realizaciones, la muestra de ácidos nucleicos genómicos procede de una planta económicamente relevante.

- 20 En ciertas realizaciones, la segunda pluralidad de cebadores de secuencia específica de especie comprende secuencias humanizadas. En ciertas realizaciones, las regiones definidas en una muestra de ácidos nucleicos genómicos comprenden una o más regiones no repetitivas y/o regiones altamente representadas de una muestra de ácidos nucleicos genómicos. En ciertas realizaciones, las secuencias de la pluralidad de cebadores específicos de especie se distribuyen esencialmente de manera uniforme por el ácido nucleico genómico, de modo que la hibridación de los cebadores específicos de especie con las regiones de ácidos nucleicos genómicos definidas no está sesgada o, al menos, está menos sesgada que con los métodos de amplificación previamente conocidos.

- 25 En algunas realizaciones del método descrito anteriormente, los numerosos cebadores de secuencia específica de especie tienen aproximadamente 6 nucleótidos, 7 nucleótidos, 8 nucleótidos, 9 nucleótidos, 10 nucleótidos, 11 nucleótidos, 12 nucleótidos, 13 nucleótidos, 14 nucleótidos, 15 nucleótidos, 16 nucleótidos, 17 nucleótidos, 18 nucleótidos, 19 nucleótidos o 20 nucleótidos de longitud. En ciertas realizaciones, los cebadores de secuencia específica de especie tienen una longitud de aproximadamente 6 nucleótidos, 7 nucleótidos, 8 nucleótidos o 9 nucleótidos.

- 30 En algunas realizaciones del método descrito anteriormente, la segunda pluralidad de cebadores está presente con la misma abundancia que la primera pluralidad de cebadores. En otras realizaciones, la segunda pluralidad de cebadores está presente en mayor abundancia que la primera pluralidad de cebadores. En algunas realizaciones, los cebadores de secuencia específica de especie comprenden secuencias de cebado específicas de especie para al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, al menos 1000, al menos 1200, al menos 1400, al menos 1600, al menos 1800, al menos 2000, al menos 2200, al menos 2400, al menos 2600, al menos 2800, al menos 3000, al menos 3500, al menos 4000, al menos 4500, al menos 5000, al menos 5500, al menos 6000, al menos 6500, al menos 7000, al menos 7500, al menos 8000, al menos 8500, al menos 9000, al menos 9500 o al menos 10000 regiones genómicas definidas.

Breve descripción de los dibujos

- 35 La figura 1 es un diagrama esquemático que muestra la amplificación de una región de ADN genómico que contiene un SNP utilizando la amplificación hologenómica (WGA) iniciada por cebadores específicos de locus.

- La figura 2 es un gráfico que muestra la intensidad de la señal de 120 sondas en una matriz de perlas de genotipado para muestras de ADN amplificadas con y sin cebadores específicos de locus diseñados para las regiones que rodean las secuencias de las sondas de captura. Se muestra un aumento evidente en la intensidad de la señal, con un promedio de 1,8 veces, para la muestra de ADN amplificada con el conjunto de cebadores específicos de locus.

- 40 La figura 3 es un gráfico de SNP para un sitio de SNP seleccionado para dos réplicas de una muestra de ADN amplificada con o sin un conjunto de cebadores específicos de locus diseñados para la región que rodea la secuencia del SNP.

- 45 En la figura 4 se muestra la intensidad de la señal por todo el cromosoma 2 humano para las muestras amplificadas con o sin nonámeros específicos de locus a alta o baja concentración.

En la figura 5 se muestra una comparación de la T_m de dos mezclas de cebadores, lo que demuestra la normalización

de la T_m mediante la adición de un número de bases en el extremo 5' proporcionales al contenido de AT de un hexámero.

La figura 6 es un diagrama que muestra un ejemplo de amplificación de ADN genómico con el uso de cebadores quiméricos.

- 5 Las figuras 7A-D son diagramas que muestran un ejemplo de amplificación de ADN genómico con el uso de cebadores aleatorios que contienen ARN.

La figura 8 es un gráfico de barras que muestra el efecto de la composición de los nonámeros sobre el sesgo de amplificación por cebado aleatorio (RPA).

Descripción detallada

- 10 Las aplicaciones de genotipado de alto rendimiento se basan en la amplificación hologenómica (WGA) eficiente y relativamente imparcial del ADN genómico analizado. La amplificación por cebado aleatorio y la amplificación por desplazamiento múltiple (MDA) se pueden usar en una gran cantidad de aplicaciones diferentes, desde la amplificación de ADN, tal como en la WGA, hasta la creación de bancos para secuenciación genómica. La capacidad de amplificar el ADN deseado de una manera relativamente imparcial es importante en muchas aplicaciones, sobre todo en la secuenciación.

15 La amplificación por cebado aleatorio (RPA) y la MDA emplean típicamente n-meros aleatorios (n es de aproximadamente 5 a aproximadamente 18). Los n-meros aleatorios pueden exhibir una degeneración cuádruple en cada posición. La metodología general comprende cebadores n-mericos aleatorios utilizados en presencia de una polimerasa de desplazamiento de cadena, nucleótidos, tampón y diana para amplificar la secuencia deseada de ácido nucleico original.

20 En la práctica, existe una diferencia significativa en la eficiencia de amplificación de diferentes regiones del genoma debido, por ejemplo, a las propiedades estructurales y de composición locales del ADN genómico. Por lo tanto, las regiones del genoma pueden amplificarse a velocidades insuficientes, lo que daría como resultado, por ejemplo, el fracaso del genotipado o datos de genotipado confundidos, como, por ejemplo, discrepancias en la variación del número de copias.

25 Además, la RPA y la MDA de ADN y ARN a menudo dan como resultado la introducción de uno o más artefactos. Por ejemplo, se puede introducir un sesgo de secuencia debido a la eficiencia de cebado diferencial entre las secuencias de nucleótidos ricas en AT (n-meros ricos en AT) en comparación con las secuencias de nucleótidos ricas en GC (n-meros ricos en GC). Además, la formación de quimeras puede ocurrir debido al cebado incorrecto de las cadenas del producto en otras cadenas del producto, o al cebado incorrecto en la secuencia deseada original, por ejemplo. Además, pueden surgir artefactos de extensión entre los cebadores.

30 En la presente memoria se presentan estrategias novedosas para reducir o eliminar estos artefactos durante la amplificación de ácidos nucleicos por cebado aleatorio. Como se describe con mayor detalle a continuación, estas estrategias conducen sorprendentemente a una amplificación mejorada y a una reducción del sesgo en los bancos de ácidos nucleicos.

Cebadores específicos de locus

35 En la presente memoria se presenta una estrategia para amplificar y/o mejorar la eficiencia de amplificación de locus seleccionados complementando una reacción de WGA con oligonucleótidos específicos de locus. Los numerosos oligonucleótidos diseñados en uno o dos lados (p. ej., delante y detrás, 5' y 3', etc.) de un solo sitio o múltiples locus se usan como cebadores en las reacciones de amplificación de ADN por desplazamiento de cadena. Se puede determinar empíricamente un número suficiente de cebadores que rodean las regiones de interés y se puede optimizar la longitud de los cebadores específicos de locus para garantizar una hibridación eficaz de los cebadores durante la WGA. Los cebadores específicos de locus pueden ser de igual longitud, más cortos, o más largos, que los cebadores aleatorios. Los cebadores se pueden diseñar variándoles la longitud para proporcionarles una especificidad de hibridación óptima.

40 Un experto en la materia apreciará los métodos de la técnica con respecto al diseño de cebadores. Los protocolos de amplificación genómica estándar utilizan cebadores aleatorios, tales como hexámeros o nonámeros aleatorios. Sin embargo, para cada genoma específico de interés, las secuencias de 6 o 9 unidades NO se distribuyen por igual por todo el genoma; algunas de las secuencias pueden aparecer en el genoma con más frecuencia que otras. Como tal, ciertas secuencias de cebadores pueden estar infrarrepresentadas o sobrerrepresentadas en el grupo de cebadores aleatorios.

45 Por lo tanto, en ciertas realizaciones, las secuencias específicas de locus adicionales, como, por ejemplo, un conjunto de nonámeros que están altamente representados en el genoma humano, se pueden agregar a un grupo de cebadores aleatorios para constituir un grupo de cebadores para amplificación mejorada (EAPP) para disminuir al mínimo la naturaleza estocástica del cebado y amplificar todo el genoma por igual. Como resultado, esta estrategia puede conducir a aumentos de la tasa éxito del genotipado y la representación del genoma en la secuenciación.

50 Un descubrimiento sorprendente presentado en la presente memoria es que los cebadores específicos de locus pueden ser funcionales cuando hay mucho ruido de fondo y un exceso de cebadores aleatorios. Estos cebadores

específicos pueden potenciar la amplificación de regiones específicas al amplificar la región directamente a partir de ADN genómico, a partir de ADN genómico copiado por cebado aleatorio, o una combinación de los mismos.

En consecuencia, en la presente memoria se presentan métodos para amplificar una muestra de ácidos nucleicos que utilizan cebadores específicos de locus en una mezcla de cebadores aleatorios. En algunas realizaciones, los métodos pueden comprender proporcionar una mezcla de cebadores para amplificación a un banco de ácidos nucleicos, en donde la mezcla comprende numerosos cebadores aleatorios y numerosos cebadores específicos de locus configurados para amplificar numerosas regiones predeterminadas del banco de ácidos nucleicos, y amplificar el banco de ácidos nucleicos con la mezcla de cebadores para amplificación.

También se presenta en la presente memoria un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos genómicos que comprende: a) proporcionar un conjunto de cebadores para amplificación a una muestra de ácidos nucleicos genómicos, en donde el conjunto de cebadores para amplificación comprende una primera pluralidad de cebadores de secuencia aleatoria y una segunda pluralidad de cebadores de secuencia específica de especie configurados para amplificar regiones genómicas definidas en la muestra de ácidos nucleicos, en donde los cebadores de secuencia específica de especie tienen una abundancia igual, menor o mayor que los cebadores aleatorios, y b) amplificar la muestra de ácidos nucleicos genómicos con el conjunto de cebadores para amplificación, con lo que se crea así un banco de ácidos nucleicos. En algunas realizaciones, los cebadores de secuencia específica de especie están diseñados para hibridarse con secuencias con repeticiones de un genoma deseado para disminuir al mínimo la amplificación excesiva de las secuencias repetitivas. En algunas realizaciones, los cebadores de secuencia específica de especie están diseñados para disminuir al mínimo la hibridación cruzada entre los propios cebadores. Por ejemplo, pueden diseñarse cebadores de secuencia específica de especie para disminuir al mínimo la complementariedad entre los cebadores, sobre todo en sus extremos 3'. En algunas realizaciones, los cebadores de secuencia específica de especie están diseñados para evitar o disminuir al mínimo la hibridación con sitios polimórficos en un genoma. En algunas realizaciones, los cebadores de secuencia específica de especie están diseñados para distribuirse uniformemente por todo un genoma. En algunas realizaciones, los cebadores de secuencia específica de especie están diseñados para distribuirse uniformemente por todo un cromosoma. En algunas realizaciones, los cebadores de secuencia específica de especie están diseñados para distribuirse uniformemente por toda una región genómica deseada, tal como una porción de un genoma. En algunas realizaciones, cada secuencia de cebador específica de especie concreta está presente a la misma concentración que otras secuencias de cebador específicas de especie en la segunda población. En otras realizaciones, cada secuencia de cebador específica de especie concreta puede estar presente a una concentración diferente que otras secuencias de cebador específicas de especie en la segunda población.

En ciertas realizaciones, como se analiza con mayor detalle a continuación, la muestra de ácidos nucleicos genómicos puede ser de un mamífero, tal como un ser humano u otro animal económicamente relevante. En ciertas realizaciones, la muestra de ácidos nucleicos genómicos puede ser de una planta económicamente relevante. Tal como se usa en la presente memoria, el término económicamente relevante se refiere a las especies animales o vegetales utilizadas en la agricultura, la medicina, la investigación de laboratorio o cualquier otra actividad que genere una relevancia económica para la especie animal o vegetal. Las plantas económicamente relevantes se conocen en la técnica e incluyen, pero no se limitan a ellas, la brasicácea de Thal (*Arabidopsis thaliana*), maíz (*Zea mays*), sorgo (*Sorghum bicolor*), avena (*Avena sativa*), trigo (*Triticum aestivum*), arroz (*Oryza sativa*), nabo silvestre (*Brassica campestris*), soja (*Glycine max*) y cultivares, variantes e híbridos de las mismas. Los animales económicamente relevantes son conocidos en la técnica e incluyen, entre otros, mamíferos tales como roedores, ratones, ratas, conejos, conejillos de Indias, ungulados, caballos, ovejas, cerdos, cabras, vacas, gatos, perros, primates, primate humano o no humano; un nemátodo, tal como *Caenorhabditis elegans*; un insecto, tales como *Drosophila melanogaster*, mosquito, mosca de la fruta, abeja melífera o araña; un pez como el pez cebra (*Danio rerio*); un reptil; o un anfibio, tal como una rana o *Xenopus laevis*, así como todos los demás organismos descritos en la presente memoria.

En ciertas realizaciones, la segunda pluralidad de cebadores de secuencia específica de especie comprende secuencias humanizadas. En ciertas realizaciones, las regiones definidas en una muestra de ácidos nucleicos genómicos comprenden una o más regiones no repetitivas y regiones repetitivas altamente representadas de una muestra de ácidos nucleicos genómicos. En ciertas realizaciones, las secuencias de los numerosos cebadores específicos de especie se distribuyen de manera esencialmente uniforme por el ácido nucleico genómico de manera que la hibridación de los cebadores específicos de especie con las regiones definidas del ácido nucleico genómico es sustancialmente imparcial. En algunas realizaciones, los numerosos cebadores específicos de especie se distribuyen de una manera relativamente imparcial. En otras realizaciones, los numerosos cebadores específicos de especie se distribuyen de manera significativamente imparcial. En aún otras realizaciones, los numerosos cebadores específicos de especie se distribuyen de una manera esencialmente imparcial. En otras realizaciones más, los numerosos cebadores específicos de especie se distribuyen de manera imparcial.

En ciertas realizaciones, los numerosos cebadores de secuencia específica de locus o específica de especie comprenden secuencias de al menos 100, al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos al menos 900, al menos 1000, al menos 1200, al menos 1400, al menos 1600, al menos 1800, al menos 2000, al menos 2200, al menos 2400, al menos 2600, al menos 2800, al menos 3000, al menos 3500, al menos 4000, al menos 4500, al menos 5000, al menos 5500, al menos 6000, al menos 6500, al menos 7000, al menos 7500, al menos 8000, al menos 8500, al menos 9000, al menos 9500 o al menos 10000 regiones

genómicas definidas. En realizaciones específicas, los cebadores aleatorios tienen una abundancia igual, menor o mayor que los cebadores específicos de locus. Tal como se usa en la presente memoria, el término "abundancia" se refiere a una cantidad de una entidad. La cantidad se puede describir en términos de concentración, un término de la técnica bien conocido por los expertos en la materia. La abundancia relativa de cebadores en una reacción se puede describir, por ejemplo, utilizando la concentración de los cebadores. La concentración puede ser relativa a un estándar conocido o puede ser absoluta. Por lo tanto, en algunas realizaciones, la concentración de cebadores aleatorios es mayor que la concentración de cebadores específicos de locus. En algunas realizaciones, la concentración de cebadores aleatorios es aproximadamente de 1,1 veces, 1,2 veces, 1,3 veces, 1,4 veces, 1,5 veces, 1,6 veces, 1,7 veces, 1,8 veces, 1,9 veces, 2,0 veces, 2,5 veces, 3,0 veces, 3,5 veces, 4,0 veces, 4,5 veces, 5,0 veces, 5,5 veces, 6,0 veces, 6,5 veces, 7,0 veces, 7,5 veces, 8,0 veces, 8,5 veces, 9,0 veces, 9,5 veces, 10,0 veces, 11 veces, 12 veces, 13 veces, 14 veces, 15 veces, 16 veces, 17 veces, 18 veces, 19 veces, 20 veces, 30 veces, 40 veces, 50 veces, 60 veces, 70 veces, 80 veces, 90 veces, 100 veces, 200 veces, 300 veces, 400 veces, 500 veces, 600 veces, 700 veces, 800 veces, 900 veces, 1000 veces, 2000 veces, 3000 veces, 4000 veces o aproximadamente 5000 veces o más que la concentración de cebadores específicos de locus. En realizaciones típicas, la concentración de cebadores aleatorios es de aproximadamente 50 veces a aproximadamente 5000 veces mayor que la concentración de cebadores específicos de locus. Por lo tanto, en una realización en la que la concentración de hexámeros aleatorios en una reacción de amplificación es de 50 micromolar, la concentración final de cebadores específicos de locus en la reacción podría ser, por ejemplo, aproximadamente 0,01 micromolar, aproximadamente 0,02 micromolar, aproximadamente 0,03 micromolar, aproximadamente 0,04 micromolar, aproximadamente 0,05 micromolar, aproximadamente 0,06 micromolar, aproximadamente 0,07 micromolar, aproximadamente 0,08 micromolar, aproximadamente 0,09 micromolar, aproximadamente 0,1 micromolar, aproximadamente 0,2 micromolar, aproximadamente 0,3 micromolar, aproximadamente 0,4 micromolar o aproximadamente 0,5 micromolar, por oligonucleótido.

En algunas realizaciones, los numerosos cebadores aleatorios comprenden cebadores de ácidos nucleicos que son cualquiera de una variedad de longitudes de secuencia aleatoria, como se conoce en la técnica. Por ejemplo, los numerosos cebadores aleatorios pueden comprender una secuencia aleatoria que tiene 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más nucleótidos de largo. En ciertas realizaciones, los numerosos cebadores aleatorios pueden comprender cebadores aleatorios de varias longitudes. En ciertas realizaciones, los numerosos cebadores aleatorios pueden comprender cebadores aleatorios que tienen la misma longitud. En ciertas realizaciones, los numerosos cebadores aleatorios pueden comprender una secuencia aleatoria que tiene una longitud de aproximadamente 5 a aproximadamente 18 nucleótidos. En algunas realizaciones, los numerosos cebadores aleatorios comprenden hexámeros aleatorios. Los cebadores aleatorios, y en concreto los hexámeros aleatorios, están disponibles comercialmente y se usan ampliamente en las reacciones de amplificación, tal como la amplificación por desplazamiento múltiple (MDA), como lo ejemplifican los kits de amplificación hologenómica REPLI-g (QIAGEN, Valencia, CA). Se apreciará que puede usarse cualquier longitud adecuada de cebadores aleatorios en los métodos y composiciones presentados de la presente memoria.

En algunas realizaciones, los cebadores específicos de locus comprenden cebadores de ácidos nucleicos que tienen cualquiera de una variedad de longitudes de secuencia, como se conoce en la técnica. Por ejemplo, los cebadores específicos de locus pueden comprender una secuencia específica de locus que tiene 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más nucleótidos de largo. En ciertas realizaciones, los cebadores específicos de locus pueden comprender cebadores específicos de locus de distintas longitudes. En ciertas realizaciones, los cebadores específicos de locus pueden comprender cebadores específicos de locus que tienen la misma longitud. En ciertas realizaciones, los cebadores específicos de locus pueden comprender una secuencia específica de locus que tiene una longitud de aproximadamente 5 a aproximadamente 18 nucleótidos. Se apreciará que se puede usar cualquier longitud adecuada de cebadores específicos de locus en los métodos y las composiciones presentados en la presente memoria. En ciertas realizaciones preferidas, los cebadores específicos de locus tienen una longitud de 9 nucleótidos. En algunas realizaciones, los cebadores específicos de locus tienen la misma longitud que los cebadores aleatorios. En algunas realizaciones, los cebadores específicos de locus son más cortos que los cebadores aleatorios. Aún en otras realizaciones, los cebadores específicos de locus son más largos que los cebadores aleatorios. En algunas realizaciones, los cebadores específicos de locus pueden tener una combinación de longitudes de cebadores, como una mezcla de cebadores específicos de locus que comprenden cebadores más cortos, iguales y más largos que los cebadores aleatorios.

En algunas realizaciones, los cebadores aleatorios comprenden cebadores resistentes a las exonucleasas. En algunas realizaciones, los cebadores específicos de locus comprenden cebadores resistentes a las exonucleasas. Los cebadores resistentes a las exonucleasas son especialmente útiles cuando las reacciones de amplificación comprenden una polimerasa con actividad exonucleásica, tal como la ADN polimerasa de $\phi 29$, que tiene una actividad exonucleásica correctora 3'→5' inherente. Los cebadores resistentes a las exonucleasas se conocen en la técnica y pueden incluir, por ejemplo, enlaces fosforotioato y similares. En algunas realizaciones, por ejemplo, un cebador puede poseer uno, dos, tres o más enlaces de tipo fosforotioato entre los nucleótidos en el extremo 3' de la secuencia del cebador. En algunas realizaciones, el cebador puede comprender uno o más nucleótidos modificados. El nucleótido modificado puede ser un nucleótido con fosforotioato. En algunas realizaciones, el nucleótido modificado es un nucleótido en el extremo 3'. En algunas realizaciones, el nucleótido modificado está ubicado en una posición distinta de la posición del extremo 3'. Cuando el nucleótido modificado está ubicado en posiciones distintas del extremo 3' de una secuencia de cebador, el nucleótido del extremo 3' del cebador puede ser eliminado por la actividad exonucleasa

3'→5'. Por ejemplo, los cebadores aleatorios que comprenden secuencias tales como NNNN*N*N se pueden utilizar en los métodos y las composiciones dados a conocer en la presente memoria (* representa un enlace de tipo fosforotioato entre los nucleótidos N). En algunas realizaciones, al menos un subconjunto de los cebadores aleatorios comprende un grupo de bloqueo en 3' o carece de un grupo OH en 3' y no puede cebar la extensión de una polimerasa.

Por ejemplo, en ciertas situaciones, regiones concretas de un genoma se sobreamplifican durante la amplificación hologenómica, mientras que otras regiones están subrepresentadas. Por lo tanto, en una realización, se añade en exceso una diversidad limitada de cebadores bloqueados en 3' resistentes a las exonucleasas a una mezcla de cebadores aleatorios. Los cebadores bloqueados en 3' pueden diseñarse para que se hibriden con cualquier región o regiones en las que se desee una reducción de la sobreamplificación. Los cebadores bloqueados en 3' se pueden diseñar para que se hibriden con regiones que con frecuencia se encuentran sobrerepresentadas en la amplificación hologenómica. Los cebadores bloqueados en 3' se pueden diseñar para que se hibriden con regiones altamente repetitivas en el genoma que, de lo contrario, imposibilitarían que se mapearan ciertas lecturas. Por ejemplo, los hexámeros bloqueados en 3' pueden superar a sus contrapartes no bloqueados que están presentes en un conjunto de hexámeros aleatorios y, por lo tanto, bloquear la hibridación y la extensión de estos hexámeros. El resultado es que se reducen las regiones de sobreamplificación y que se reduce la representación de las regiones altamente repetitivas que imposibilitan que se mapeen ciertas lecturas de secuenciación.

El término "aleatorio", tal como se utiliza en la presente descripción con respecto a cebadores, oligonucleótidos, polinucleótidos y similares, debería entenderse que se refiere a la degeneración en una o más posiciones de un nucleótido. Por lo tanto, en ciertas realizaciones, el término aleatorio puede referirse a una distribución puramente aleatoria en todas las posiciones, donde cada posición puede tener la misma probabilidad de ser cualquiera de los cuatro nucleótidos estándar. En ciertas realizaciones, el término aleatorio puede referirse a una distribución aleatoria en no todas las posiciones de nucleótidos en el polinucleótido. En algunas realizaciones, el término aleatorio puede referirse a una distribución aleatoria ponderada en una o más posiciones, o en todas las posiciones de un polinucleótido, en donde la degeneración no se distribuye por igual entre los cuatro nucleótidos estándar en cada posición definida como aleatoria. En algunas realizaciones, el término aleatorio puede referirse a una distribución sesgada de los cuatro nucleótidos estándar en una o más posiciones. También se puede entender que los polinucleótidos aleatorios incluyen uno o más análogos de base como parte de los nucleótidos disponibles para distribución en una o más posiciones. Por lo tanto, en algunas realizaciones, los numerosos cebadores aleatorios pueden ser completamente aleatorios con respecto a cada posición del cebador. En algunas realizaciones, la secuencia aleatoria de un cebador aleatorio puede incluir posiciones intercaladas que tienen un nucleótido fijo o regiones que tienen una secuencia fija de dos o más nucleótidos, si se desea.

Métodos de amplificación y reactivos

Los métodos dados a conocer en la presente memoria comprenden amplificar una muestra de ácidos nucleicos con el uso de la mezcla de cebadores para la amplificación. Tal y como se emplea en la presente memoria, los términos "amplificado", "amplificar", "amplificación" y términos similares se refieren a producir una o más copias de un ácido nucleico monocatenario o bicatenario, o una porción del mismo. En algunas realizaciones, los métodos dados a conocer en la presente memoria pueden incluir una etapa de producción de un banco de ácidos nucleicos amplificada en condiciones isotérmicas o de temperatura variable. Los ejemplos de métodos de amplificación isotérmica que se pueden usar en un método de la presente descripción incluyen, entre otros, la amplificación por desplazamiento múltiple (MDA) tal como se ejemplifica, por ejemplo, en Dean et al., *Proc Natl. Acad Sci USA* 99:5261-66 (2002) o amplificación isotérmica de ácidos nucleicos por desplazamiento de cadena, ejemplificada por, por ejemplo, la patente de los EE. UU. n.º 6 214 587. Otros métodos que no se basan en la PCR que se pueden usar en la presente descripción incluyen, por ejemplo, la amplificación por desplazamiento de cadena (SDA) que se describe, por ejemplo, en Walker et al., *Molecular Methods for Virus Detection*, Academic Press, Inc., 1995; patentes de los EE. UU. n.ºs 5 455 166, y 5 130 238, y en Walker et al., *Nucl. Acids Res.* 20:1691-96 (1992) o la amplificación por desplazamiento de cadena hiperramificada que se describe, por ejemplo, en Lage et al., *Genome Research* 13:294-307 (2003). Los métodos de amplificación isotérmicos se pueden utilizar con la polimerasa de ϕ 29 de desplazamiento de cadena o el fragmento grande de la ADN polimerasa de Bst, 5'→3' exo⁻ para la amplificación aleatoria de cebadores de ADN genómico. El uso de estas polimerasas aprovecha su alta procesividad y actividad de desplazamiento de hebras. La alta procesividad permite que las polimerasas produzcan fragmentos de 10 a 20 kb de longitud. Tal como se establece anteriormente, se pueden producir fragmentos más pequeños en las condiciones isotérmicas con el uso de polimerasas poco procesivas y con actividad de desplazamiento de hebras, tal como la polimerasa Klenow. La descripción adicional de las reacciones de amplificación, las condiciones y los componentes se presentan en detalle en la descripción de la patente de EE. UU. n.º 7 670 810.

Otro método de amplificación de ácidos nucleicos que es útil en la presente descripción es la PCR marcada que utiliza una población de cebadores con dos dominios que tienen una región 5' constante seguida de una región 3' aleatoria como se describe, por ejemplo, en Grothues et al. *Nucleic Acids Res.* 21(5):1321-2 (1993). Las primeras rondas de amplificación se llevan a cabo para permitir una multitud de iniciaciones en el ADN desnaturalizado por calor en función de cada hibridación de la región 3' sintetizada aleatoriamente. Debido a la naturaleza de la región 3', se contempla que los sitios de iniciación sean aleatorios por todo el genoma. A partir de entonces, los cebadores no unidos pueden eliminarse y puede tener lugar la replicación con cebadores complementarios a la región 5' constante.

Otra estrategia que se puede usar para amplificar el ADN_g en relación con los métodos de la presente descripción es

la reacción en cadena de la polimerasa cebada con oligonucleótidos degenerados (DOP-PCR) en condiciones tales como, entre otras, las descritas por Cheung et al., *Proc. Natl. Acad. Sci. USA*, 93:14676-79 (1996) o la patente de los EE. UU. n.º 5 043 272. Pueden amplificarse cantidades pequeñas de ADN_g, por ejemplo, 15 pg de ADN_g humano, hasta niveles que se detectan con comodidad en los métodos de la presente descripción. Las condiciones de reacción utilizadas en los métodos de Cheung et al. puede seleccionarse para producir una población representativa amplificada de fragmentos de genoma que tienen una cobertura casi completa del genoma humano. Además, las versiones modificadas de DOP-PCR, tal como las descritas por Kittler et al. en un protocolo conocido como LL-DOP-PCR (DOP-PCR de productos largos a partir de cantidades bajas de ADN) se puede usar para amplificar el ADN_g de acuerdo con la presente descripción (Kittler et al., *Anal. Biochem.* 300:237-44 (2002)).

La reacción en cadena de la polimerasa de preamplificación con extensión de cebador (PEP-PCR) también se puede usar en un método de la presente descripción para amplificar el ADN_g. Las condiciones útiles para la amplificación del ADN_g utilizando la PEP-PCR incluyen, por ejemplo, las descritas en Casas et al., *Biotechniques* 20:219-25 (1996).

Los presentes métodos no se limitan a ninguna técnica de amplificación en concreto y las técnicas de amplificación descritas en la presente memoria son ejemplares solo con respecto a los métodos y realizaciones de la presente descripción.

En general, la actividad de la polimerasa, que incluye, por ejemplo, la procesividad y la actividad de desplazamiento de cadena, puede verse influida por factores tales como el pH, la temperatura, la fuerza iónica y la composición del tampón. Los expertos en la materia sabrán qué tipos de polimerasas y condiciones se pueden utilizar para obtener fragmentos que tengan una longitud deseada en vistas de lo que se sabe sobre la actividad de las polimerasas, tal como se describe, por ejemplo, en Eun, H.M., *Enzymology Primer for Recombinant DNA Technology*, Academic Press, San Diego (1996)). Además, un experto en la materia podrá determinar las polimerasas y las condiciones apropiadas mediante pruebas sistemáticas utilizando los ensayos conocidos, tales como electroforesis en gel o espectrometría de masas, para medir la longitud de los fragmentos amplificados.

La Pol I de *E. coli* o su fragmento Klenow se pueden utilizar para la amplificación isotérmica de un genoma para producir pequeños fragmentos de ADN genómico, por ejemplo, en una reacción con poca sal ($I = 0,085$) incubada a una temperatura entre aproximadamente 5 °C y 37 °C. Los tampones ejemplares y las condiciones de pH que se pueden usar para amplificar ADN_g con el fragmento Klenow incluyen, por ejemplo, Tris HCl a 50 mM (pH 7,5), MgCl₂ a 5 mM, NaCl a 50 mM, 50 µg/ml de seroalbúmina bovina (SAB), 0,2 mM de cada dNTP, 2 µg de cebador aleatorio ($n = 6$), 10 ng de ADN_g de molde y 5 unidades de Klenow exo^- incubadas a 37 °C durante 16 h. Se pueden realizar unas condiciones de reacción similares en las que se omiten o sustituyen uno o más componentes de la reacción. Por ejemplo, el tampón se puede reemplazar por fosfato a 50 mM (pH 7,4) o se pueden usar otros valores de pH en el margen de aproximadamente 7,0 a 7,8. Se puede proporcionar una plantilla de ADN_g para amplificar en cualquiera de una variedad de cantidades que incluyen, sin limitación, las presentadas anteriormente en la presente memoria. En una realización alternativa, las condiciones para la amplificación pueden incluir, por ejemplo, 10 ng de plantilla de ADN genómico, dNTP a 2 mM, MgCl₂ a 10 mM, 0,5 U/µl de polimerasa, cebador aleatorio a 50 µM ($n = 6$) e incubación isotérmica a 37 °C durante 16 h.

En realizaciones concretas, una reacción de amplificación se puede llevar a cabo en dos etapas que incluyen, por ejemplo, una etapa inicial de hibridación seguida de una etapa de extensión. Por ejemplo, se pueden hibridar 10 ng de ADN_g con cebadores aleatorios a 100 µM ($n = 6$) en 30 µl de Tris-Cl a 10 mM (pH 7,5) mediante una breve incubación a 95 °C. La reacción puede enfriarse a temperatura ambiente y llevarse a cabo una etapa de hibridación al añadirle un volumen igual de Tris-Cl a 20 mM (pH 7,5), MgCl₂ a 20 mM, ditiotretitol a 15 mM, dNTP a 4 mM y Klenow exo^- a 1 U/µl e incubación a 37 °C durante 16 h. Aunque se ejemplifica para la amplificación con Klenow, los expertos en la técnica reconocerán que se pueden utilizar etapas independientes de hibridación y extensión para las reacciones de amplificación llevadas a cabo con otras polimerasas, tales como las que se presentan a continuación.

En realizaciones concretas, los cebadores que tienen regiones de hibridación aleatoria de diferentes longitudes (n) pueden sustituirse en los métodos de amplificación basados en Klenow. Por ejemplo, los cebadores aleatorios $n = 6$ en las condiciones ejemplares anteriores se pueden reemplazar con cebadores que tienen otras longitudes de secuencia aleatoria que incluyen, sin limitación, $n = 7, 8, 9, 10, 11$ o 12 nucleótidos. Nuevamente, aunque ejemplificado para la amplificación basada en Klenow, los expertos en la técnica reconocerán que pueden usarse cebadores aleatorios que tienen diferentes longitudes de secuencia aleatoria (n) para las reacciones de amplificación llevadas a cabo con otras polimerasas, tales como las que se presentan a continuación.

La ADN polimerasa de T4 se puede utilizar para la amplificación de ADN_g monocatenario o desnaturalizado, por ejemplo, en HEPES a 50 mM a pH 7,5, Tris-HCl a 50 mM y pH 8,6, o glicinato a 50 mM y pH 9,7. Una mezcla de reacción típica también puede contener KCl a 50 mM, MgCl₂ a 5 mM, ditiotretitol (DTT) a 5 mM, ADN_g a 40 µg/ml, dNTP a 0,2 mM cada uno, SAB a 50 µg/ml, cebador aleatorio a 100 µM ($n = 6$) y 10 unidades de polimerasa de T4 incubadas a 37 °C durante al menos 1 h. El ciclo de temperatura se puede utilizar para desplazar las cadenas replicadas para múltiples rondas de amplificación.

La polimerasa de T7 suele ser muy procesiva, lo que permite la polimerización de miles de nucleótidos antes de disociarse de un ADN molde. Las condiciones de reacción típicas en las cuales la polimerasa de T7 es muy procesiva

son Tris-HCl a 40 mM y pH 7,5, MgCl₂ a 15 mM, NaCl a 25 mM, DTT a 5 mM, 0,25 mM de cada dNTP, 50 µg/ml de ADNg monocatenario, cebadores aleatorios a 100 µM (n = 6) y de 0,5 a 1 unidad de polimerasa de T7. Sin embargo, a temperaturas inferiores a 37°C, la procesividad de la polimerasa de T7 se reduce considerablemente. La procesividad de la polimerasa de T7 también se puede reducir a altas fuerzas iónicas, por ejemplo, por encima de NaCl a 100 mM. La polimerasa de T7 de forma II no suele ser capaz de amplificar el ADN de doble cadena. Sin embargo, la forma I de la polimerasa de T7 y la polimerasa de T7 modificada (SEQUENASE™ versión 2.0 que carece de la región de 28 aminoácidos Lys 118 a Arg 145) pueden catalizar la replicación por desplazamiento de cadena. Por consiguiente, pueden amplificarse pequeños fragmentos del genoma en un método de la presente descripción con el uso de una polimerasa de T7 modificada o condiciones modificadas, tales como las presentadas anteriormente. En realizaciones concretas, SEQUENASE™ puede usarse en presencia de la proteína de unión a monocadenas (SSB) de *E. coli* para aumentar el desplazamiento de la cadena. La SSB también se puede utilizar para aumentar la procesividad de la SEQUENASE™, si se desea.

La polimerasa Taq es muy procesiva a una temperatura de aproximadamente 70 °C cuando reacciona con un exceso molar de 10 veces de molde y de cebadores aleatorios (n = 6). Una reacción de amplificación realizada en estas condiciones puede incluir además un tampón tal como Tris-HCl a aproximadamente 20 mM, pH de aproximadamente 7, aproximadamente 1 a 2 mM de MgCl₂, y 0,2 mM de cada dNTP. Además, se puede agregar un agente estabilizador, tal como glicerol, gelatina, DMSO, betaína, SAB o un detergente no iónico. La polimerasa de Taq tiene poca procesividad a temperaturas por debajo de 70 °C. En consecuencia, se pueden obtener pequeños fragmentos de ADNg con el uso de la polimerasa de Taq a baja temperatura en un método de la presente descripción, o en otra condición en la que Taq tiene poca procesividad. En otra realización, el fragmento de Stoffel, que carece de los 289 residuos de aminoácidos del extremo amino de la polimerasa de Taq y tiene poca procesividad a 70 °C, puede usarse para generar fragmentos de ADNg relativamente pequeños en un método de la presente descripción. La Taq puede usarse para amplificar moldes de ADN monocatenario o desnaturalizado en un método de la presente descripción. El ciclo de temperaturas se puede utilizar para desplazar las cadenas replicadas durante las numerosas rondas de amplificación.

Los expertos en la técnica reconocerán que las condiciones para la amplificación con las diversas polimerasas que se presentaron anteriormente son ejemplares. Por lo tanto, se pueden realizar cambios menores que no alteran sustancialmente la actividad. Además, las condiciones se pueden cambiar sustancialmente para lograr una actividad de amplificación deseada o para adaptarse a una aplicación concreta de la presente descripción.

Los métodos y las composiciones de la presente descripción también se pueden usar con variantes de las polimerasas descritas anteriormente, siempre que conserven la actividad polimerasa. Las variantes ejemplares incluyen, sin limitación, las que tienen menos actividad exonucleasa, más fidelidad, más estabilidad o más afinidad por los análogos de nucleósidos. Las variantes de ejemplo, así como otras polimerasas que son útiles en un método de la presente descripción, incluyen, sin limitación, la ADN polimerasa del bacteriófago φ29 (patentes de los EE. UU. n.ºs 5 198 543 y 5 001 050), exo(-)Bca ADN polimerasa (Walker y Linn, *Clinical Chemistry* 42:1604-1608 (1996)), ADN polimerasa del fago M2 (Matsumoto et al., *Gene* 84:247 (1989)), ADN polimerasa del fago φ29 PRD 1 (Jung y col., *Proc. Natl. Acad. Sci. USA* 84: 8287 (1987)), ADN polimerasa exo(-)VENT™ (Kong y col., *J. Biol. Chem.* 268:1965-1975 (1993)), ADN polimerasa de T5 (Chatterjee et al., *Gene* 97:13-19 (1991)) y la ADN polimerasa PRD1 (Zhu y col., *Biochim. Biophys. Acta* 1219:267-276 (1994)).

Otra variante de la polimerasa que es útil con los métodos y composiciones de la presente descripción es una polimerasa modificada que, en comparación con su versión no modificada de tipo silvestre, tiene una capacidad reducida, o eliminada, para agregar al extremo 3' de un ácido nucleico nucleótidos que no vienen dirigidos por la plantilla. Las variantes ejemplares incluyen las que pueden afectar la actividad de la polimerasa por la adición de todos los tipos de nucleótidos o uno o más tipos de nucleótidos, tales como nucleótidos de pirimidina, nucleótidos de purina, A, C, T, U o G. Las modificaciones pueden incluir la modificación química de los grupos de aminoácidos en la polimerasa o mutaciones de secuencia tales como delecciones, adiciones o reemplazos de aminoácidos. Los ejemplos de polimerasas modificadas que tienen una capacidad reducida, o eliminada, para agregar nucleótidos que no vienen dirigidos por la plantilla al extremo 3' de un ácido nucleico se describen, por ejemplo, en la patente de los EE. UU. n.º 6 306 588 o Yang et al., *Nucl. Acids Res.* 30:4314-4320 (2002)). En una realización concreta, dicha variante de polimerasa puede usarse en un método de detección SBE o ASPE descrito en la presente memoria.

El banco de ácidos nucleicos amplificado por los métodos y composiciones descritos en la presente memoria puede comprender esencialmente cualquier tipo de ácido nucleico. En algunas realizaciones de los presentes métodos y composiciones descritos en la presente memoria, la muestra de ácidos nucleicos comprende un genoma, un fragmento de genoma, una mezcla de genomas o una mezcla de fragmentos de genoma. El término fragmento de genoma pretende significar una molécula de ácido nucleico aislada que tiene una secuencia que es sustancialmente idéntica a una porción de un cromosoma. Se entiende que un cromosoma es un cuerpo lineal o, a veces, circular que contiene ácido nucleico de un virus, organismo procariótico o núcleo eucariótico que contiene la mayoría o la totalidad de los genes replicados. Una población de fragmentos de genoma puede incluir secuencias idénticas a sustancialmente un genoma completo o una parte del mismo. Un fragmento genómico puede tener, por ejemplo, una secuencia que sea sustancialmente idéntica a por lo menos aproximadamente 25, 50, 70, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 o más nucleótidos de un cromosoma. Un fragmento de genoma puede ser ADN, ARN o un análogo del mismo. Los expertos en la técnica entenderán que una secuencia de ARN y una secuencia de ADN cromosómico que difieren

por la presencia de uracilos en lugar de timinas tienen una secuencia sustancialmente idéntica.

En algunas realizaciones, la muestra de ácidos nucleicos comprende ARN. En algunas realizaciones de los presentes sistemas y métodos descritos en la presente memoria, la muestra de ácidos nucleicos comprende ADN. El ADN utilizado en la presente memoria no está limitado por el tipo, sino que típicamente comprende ADN genómico (ADNg) o ADNc. El ADN genómico puede referirse al material de ácido nucleico real aislado de un organismo o, como alternativa, una o más copias de partes del genoma de un organismo o una o más copias del genoma completo de un organismo. Por ejemplo, el ADN genómico puede referirse a una copia de un fragmento de ADN genómico que se ha aislado de un organismo. En algunas realizaciones, el ADN genómico se aísla de una célula u otro material y se fragmenta. A continuación, los fragmentos se copian o si no se amplifican. Aunque este material amplificado puede contener secuencias replicadas en lugar de moléculas de ácidos nucleicos aisladas directamente del organismo, este material se sigue denominando en la presente memoria ADN genómico o ADN obtenido o derivado del genoma de un organismo. Como tal, el ADN genómico descrito en la presente memoria puede incluir fragmentos o copias de fragmentos de secuencias de ADN genómico. En algunas realizaciones, la muestra es una muestra forense que contiene solo vestigios del material de ácido nucleico. En algunas realizaciones, la muestra de ADN genómico convertido con bisulfito y los cebadores específicos de secuencia están diseñados para amplificar específicamente las secuencias convertidas con bisulfito.

Por ejemplo, una muestra de ADN que se amplifica de acuerdo con los métodos dados a conocer en la presente memoria puede ser un genoma, tal como los expuestos anteriormente u otras plantillas de ADN, tal como el ADN mitocondrial o algún subconjunto de ADN genómico. Un ejemplo no limitante de un subconjunto de ADN genómico es un cromosoma concreto o una región de un cromosoma concreto.

Los métodos y composiciones de la presente descripción son útiles en una serie de aplicaciones que incluyen, por ejemplo, análisis de haplotipos de cada uno de los espermatozoides, genotipado de un gran número de individuos en un formato de alto rendimiento, identificación de nuevos haplotipos, eliminación gradual de haplotipos, secuenciación *de novo*, medicina forense, determinación de enfermedades, atención médica, determinaciones de eficacia terapéutica, diagnóstico y pronóstico de enfermedades, por nombrar solo algunas. Además, la presente descripción reduce la cantidad de muestra de ADN o ARN requerida en muchos ensayos de matrices actuales. Aún más, la mejoría de la sensibilidad de la matriz disponible con la presente descripción puede conducir, por ejemplo, a que se necesite menos muestra, mejorar la capacidad de puntuación de LOD y mayor margen dinámico.

Los métodos y composiciones de la presente descripción pueden usarse para amplificar ADN genómico (ADNg) de cualquier organismo. Los métodos son ideales para la amplificación y el análisis de genomas grandes, tales como los que se encuentran típicamente en los organismos eucariotas unicelulares y multicelulares. Los ejemplos de ADNg eucariótico que se pueden usar en un método de la presente descripción incluyen, sin limitación, el de un mamífero, tal como un roedor, ratón, rata, conejo, conejillo de Indias, ungulado, caballo, oveja, cerdo, cabra, vaca, gato, perro, primate, primate humano o no humano; una planta, tal como *Arabidopsis thaliana*, maíz (*Zea mays*), sorgo, avena, trigo, arroz (*Oryza sativa*), colza o soja; un alga, tal como *Chlamydomonas reinhardtii*; un nematodo, tal como *Caenorhabditis elegans*; un insecto, tal como *Drosophila melanogaster*, mosquito, mosca de la fruta, abeja melífera o araña; un pez, tal como el pez cebra (*Danio rerio*); un reptil; un anfibio, tal como una rana o *Xenopus laevis*; un hongo, tal como *Dictyostelium discoideum*, *Pneumocystis carinii*, *Takifugu rubripes*, *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe*; o un organismo parásito, causante de enfermedades, tal como *Plasmodium falciparum*. Un método de la presente descripción también se puede usar para detectar locus tipificables de genomas más pequeños, tal como los de un procarionta, tal como una bacteria, *Escherichia coli*, *Staphylococcus aureus* o *Mycoplasma pneumoniae*; una arquea; un virus, tal como el virus de la hepatitis C o el virus de la inmunodeficiencia humana (VIH); o un viroide.

Un ADN genómico utilizado en relación con los métodos y composiciones de la presente descripción puede tener uno o más cromosomas. Por ejemplo, se puede usar un ADN genómico procariontario que incluye un cromosoma. Como alternativa, un ADN genómico eucariota que incluye numerosos cromosomas puede usarse en un método de la presente descripción. Por lo tanto, los métodos se pueden usar, por ejemplo, para amplificar o detectar locus tipificables de un ADN genómico que tiene n igual a 2 o más, 4 o más, 6 o más, 8 o más, 10 o más, 15 o más, 20 o más, 23 o más, 25 o más, 30 o más, o 35 o más, cromosomas, donde n es el número haploide de cromosomas y el recuento diploide de cromosomas es $2n$. El tamaño de un ADN genómico usado en un método de la presente descripción también se puede medir según el número de parejas de bases o la longitud de nucleótidos del complemento cromosómico. Las estimaciones de tamaño ejemplares para algunos de los genomas que son útiles en la presente descripción son de aproximadamente 3,1 Gbp (humano), 2,7 Gbp (ratón), 2,8 Gbp (rata), 1,7 Gbp (pez cebra), 165 Mbp (mosca de la fruta), 13,5 Mbp (*S. cerevisiae*), 390 Mbp (fugu), 278 Mbp (mosquito) o 103 Mbp (*C. elegans*). Los expertos en la técnica reconocerán que los genomas que tienen tamaños distintos a los ejemplificados anteriormente, incluidos, por ejemplo, genomas más pequeños o más grandes, pueden usarse en un método de la presente descripción.

El ADN genómico se puede aislar de una o más células, líquidos corporales o tejidos. Pueden usarse los métodos conocidos para obtener un líquido corporal, tal como sangre, sudor, lágrimas, linfa, orina, saliva, semen, líquido cefalorraquídeo, heces o líquido amniótico. De manera similar, se pueden usar los métodos de biopsia conocidos para obtener células o tejidos tales como frotis bucales, enjuagues bucales, extirpación quirúrgica, biopsia por aspiración o similares. El ADN genómico también se puede obtener de una o más células o tejidos en cultivo primario, en una línea celular propagada, una muestra de archivo fijada, una muestra forense o una muestra arqueológica.

Los tipos de células ejemplares a partir de los cuales se puede obtener ADNg para usar en relación con los métodos y composiciones de la presente descripción incluyen, sin limitación, una célula sanguínea, tal como un linfocito B, un linfocito T, un leucocito, un eritrocito, un macrófago o un neutrófilo; una célula muscular, tal como una célula esquelética, una célula de músculo liso o una célula de músculo cardíaco; célula germinal, tal como un espermatozoide o un óvulo; célula epitelial; célula de tejido conjuntivo, tal como un adipocito, fibroblasto u osteoblasto; neurona; astrocito; células del estroma; célula renal; célula pancreática; hepatocito; o queratinocito. Una célula de la que se obtiene ADNg puede estar en un nivel de desarrollo concreto que incluye, por ejemplo, una célula madre hematopoyética o una célula que surge de una célula madre hematopoyética, tal como un glóbulo rojo, un linfocito B, un linfocito T, un linfocito citolítico natural, neutrófilo, basófilo, eosinófilo, monocito, macrófago o plaqueta. Otras células incluyen una célula del estroma de la médula ósea (célula madre mesenquimatosas) o una célula que se desarrolla a partir de ella, como una célula ósea (osteocitos), células de cartilago (condrocitos), célula grasa (adipocito) u otros tipos de células del tejido conjuntivo, como una encontrada en los tendones; célula madre neural o una de las células que da lugar, que incluye, por ejemplo, una célula nerviosa (neurona), astrocito u oligodendrocito; célula madre epitelial o una célula que surge de una célula madre epitelial, tal como una célula absorbente, célula caliciforme, célula de Paneth o célula enteroendocrina; células madre de la piel; célula madre epidérmica; o célula madre folicular. Generalmente se puede usar cualquier tipo de célula madre incluida, sin limitación, una célula madre embrionaria, una célula madre adulta o una célula madre pluripotente.

Una célula de la que se obtiene una muestra de ADNg para su uso en relación con los métodos y composiciones de la presente descripción puede ser una célula normal o una célula que muestre uno o más síntomas de una enfermedad o afección concreta. Por lo tanto, un ADNg usado en un método de la presente descripción se puede obtener de una célula cancerosa, una célula neoplásica, una célula necrótica o similares. Los expertos en la técnica conocerán o podrán determinar fácilmente los métodos para aislar el ADNg de una célula, líquido o tejido utilizando los métodos conocidos en la técnica, tal como los descritos en Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3.^a edición, Cold Spring Harbor Laboratory, Nueva York (2001)) o en Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1998).

Se puede preparar un ADNg para su uso en un método de la presente descripción mediante la lisis de una célula que contiene el ADN. Normalmente, una célula se lisa en condiciones que preservan sustancialmente la integridad del ADNg de la célula. En concreto, la exposición de una célula a un pH alcalino se puede usar para lisar una célula en un método de la presente descripción mientras se causa un daño relativamente pequeño al ADNg. Cualquiera de una serie de compuestos básicos puede usarse para la lisis, entre ellos, por ejemplo, hidróxido de potasio, hidróxido de sodio y similares. Además, se puede obtener ADNg relativamente intacto a partir de una célula lisada por una enzima que degrada la pared celular. Las células que carecen de una pared celular, ya sea de forma natural o debido a la eliminación enzimática, también pueden lisarse mediante exposición a estrés osmótico. Otras condiciones que se pueden usar para lisar una célula incluyen la exposición a detergentes, ruptura mecánica, calor con sonicación, diferencial de presión tal como en un dispositivo de prensa French u homogeneización Dounce. Los agentes que estabilizan el ADNg se pueden incluir en un lisado celular o en una muestra aislada de ADNg que incluyen, por ejemplo, inhibidores de nucleasas, quelantes, reductores, tampones salinos y similares. Los métodos para lisar una célula para obtener ADNg se pueden llevar a cabo en las condiciones conocidas en la técnica como se describe, por ejemplo, en Sambrook et al., véase más arriba (2001) o en Ausubel et al., véase más arriba, (1998).

En realizaciones concretas de la presente descripción, un lisado celular crudo que contiene ADNg puede amplificarse o detectarse directamente sin aislamiento adicional del ADNg. Como alternativa, un ADNg puede aislarse adicionalmente de otros componentes celulares antes de la amplificación o detección. Por consiguiente, se puede llevar a cabo en el ADNg purificado o parcialmente purificado un método de detección o amplificación de la presente descripción. El ADN genómico se puede aislar con los métodos conocidos que incluyen, por ejemplo, extracción en fase líquida, precipitación, extracción en fase sólida, cromatografía y similares. Dichos métodos a menudo se denominan minipreparaciones y se describen, por ejemplo, en Sambrook et al., véase más arriba, (2001) o en Ausubel et al., véase más arriba, (1998) o están disponibles de varios proveedores comerciales, incluidos, por ejemplo, Qiagen (Valencia, California) o Promega (Madison, Wisconsin).

Puede usarse un método o composición de la presente descripción para producir un genoma amplificado a partir de un pequeño número de copias del genoma. En consecuencia, las muestras de tejido pequeñas u otras muestras que tienen relativamente pocas células, por ejemplo, debido a la poca abundancia, las limitaciones de la biopsia o el alto coste, pueden secuenciarse, genotiparse o evaluarse a escala hologenómica. Los métodos y composiciones de la presente descripción también se pueden usar para producir un genoma amplificado a partir de una sola copia del genoma nativo obtenida, por ejemplo, de una sola célula. En otras realizaciones ejemplares de la presente descripción, se puede producir un genoma amplificado a partir de un mayor número de copias de un genoma nativo que incluye, entre otras, aproximadamente 1000 copias (para un genoma humano, aproximadamente 3 nanogramos de ADN) o menos, 10 000 copias o menos, 1×10^5 copias (para un genoma humano, aproximadamente 300 nanogramos de ADN) o menos, 5×10^5 copias o menos, 1×10^6 copias o menos, 1×10^8 copias o menos, 1×10^{10} copias o menos, o 1×10^{12} copias o menos.

Cebadores

Un cebador utilizado en relación con un método o composición de la presente descripción puede tener cualquiera de

una serie de composiciones o tamaños, siempre que tenga la capacidad de hibridarse a un ácido nucleico molde con especificidad de secuencia y pueda participar en la replicación del molde. Por ejemplo, un cebador puede ser un ácido nucleico que tenga una estructura nativa o un análogo del mismo. Un ácido nucleico con una estructura nativa generalmente tiene un esqueleto que contiene enlaces fosfodiéster y puede ser, por ejemplo, ácido desoxirribonucleico o ácido ribonucleico. Una estructura análoga puede tener un esqueleto alternativo que incluye, entre otros, fosforamida (véase, por ejemplo, Beaucage et al., *Tetrahedron* 49(10):1925 (1993) y referencias en el mismo; Letsinger, *J. Org. Chem* 35:3800 (1970)); Sprinzl et al., *Eur. J. Biochem.* 81:579 (1977)); Letsinger et al., *Nucl. Acids Res.* 14:3487 (1986)); Sawai et al., *Chem. Lett.* 805 (1984)), Letsinger et al., *J. Am. Chem. Soc.* 110:4470 (1988)); y Pauwels et al., *Chemica Scripta* 26:141 91986)), fosforotioato (véase, por ejemplo, Mag et al., *Nucleic Acids Res.* 19:1437 (1991)); y patente de los EE. UU. n.º 5 644 048), fosforoditioato (véase, por ejemplo, Briu et al., *J. Am. Chem. Soc.* 111:2321 (1989)), enlaces O-metilfosforoamidita (véase, por ejemplo, Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press) y enlaces y cadenas principales de ácidos peptidonucleicos (véase, por ejemplo, Egholm, *J. Am. Chem. Soc.* 114:1895 (1992)); Meier et al., *Chem. Int. Ed. Engl.* 31:1008 (1992)); Nielsen, *Nature*, 365:566 (1993); Carlsson et al., *Nature* 380:207 (1996)). Otras estructuras análogas incluyen aquellas con esqueletos positivos (véase, por ejemplo, Denpcy et al., *Proc. Natl. Acad. Sci. USA* 92:6097 (1995); esqueletos no iónicos (véase, por ejemplo, patentes de los EE. UU. n.ºs 5 386 023, 5 637 684, 5 602 240, 5 216 141 y 4 469 863; Kiedrowski et al., *Angew. Chem. Intl. Ed. English* 30:423 (1991); Letsinger et al., *J. Am. Chem. Soc.* 110:4470 (1988)); Letsinger et al., *Nucleoside & Nucleotide* 13:1597 (1994); capítulos 2 y 3, Serie de simposios 580 de ASC, "Carbohydrate Modifications in Antisense Research", ed. Y. S. Sanghui y P. Dan Cook; Mesmaeker et al., *Bioorganic & Medicinal Chem. Lett.* 4:395 (1994)); Jeffs et al., *J. Biomolecular NMR* 34:17 (1994); *Tetrahedron Lett.* 37:743 (1996)) y esqueletos sin ribosa, incluidos, por ejemplo, los descritos en las patentes de los EE. UU. n.ºs 5 235 033 y 5 034 506, y capítulos 6 y 7, *ASC Symposium Series 580*, "Modifications in Antisense Research" ed. Y. S. Sanghui y P. Dan Cook. Las estructuras análogas que contienen uno o más azúcares carbocíclicos también son útiles en los métodos y se describen, por ejemplo, en Jenkins et al., *Chem. Soc. Rev.* (1995) págs. 169-176. Otras varias estructuras análogas que son útiles en la presente descripción se describen, por ejemplo, en Rawls, *C & E News*, 2 de junio de 1997, página 35.

Un ácido nucleico útil en relación con los métodos y/o las composiciones de la presente descripción puede contener un resto de azúcar no natural en el esqueleto. Los ejemplos de modificaciones de azúcar incluyen, pero no se limitan a ellas, modificaciones en 2' tales como la adición de halógeno, alquilo, alquilo sustituido, alcarilo, aralquilo, O-alcarilo u O-aralquilo, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂ CH₃, ONO₂, NO₂, N₃, NH₂, heterocicloalquilo, heterocicloalcarilo, aminoalquilamino, polialquilamino, sililo sustituido y similares. También se pueden realizar modificaciones similares en otras posiciones del azúcar, en concreto la posición 3' del azúcar en el nucleótido del extremo 3' o en oligonucleótidos enlazados 2'-5' y la posición 5' del nucleótido del extremo 5'.

Un ácido nucleico utilizado en relación con los métodos y/o composiciones de la presente descripción también puede incluir bases nativas o no nativas. A este respecto, un ácido desoxirribonucleico nativo puede tener una o más bases seleccionadas del grupo que consiste en adenina, timina, citosina o guanina y un ácido ribonucleico puede tener una o más bases seleccionadas del grupo que consiste en uracilo, adenina, citosina o guanina. Ejemplos de bases no nativas que se pueden incluir en un ácido nucleico, ya sea que tenga un esqueleto nativo o una estructura análoga, incluyen, entre otros, inosina, xantina, hipoxantina, isocitosina, isoguanina, 5-metilcitosina, 5-hidroximetilcitosina, 2-aminoadenina, 6-metiladenina, 6-metilguanina, 2-propilguanina, 2-propiladenina, 2-tiouracilo, 2-tiotimina, 2-tiocitosina, 15-halouracilo, 15-halocitosina, 5-propiniluracilo, 5-propinilcitosina, 6-azouracilo, 6-azocitosina, 6-azotimina, 5-uracilo, 4-tiouracilo, 8-halo adenina o -guanina, 8-amino adenina o -guanina, 8-tiol adenina o -guanina, 8-tioalquil adenina o -guanina, 8-hidroxil adenina o -guanina, uracilo o citosina sustituidos con halo en 5, 7-metilguanina, 7-metiladenina, 8-azaguanina, 8-azaadenina, 7-desazaguanina, 7-desazaadenina, 3-desazaguanina, 3-desazaadenina, 8-oxoguanina, o similares. En una realización concreta se puede utilizar isocitosina e isoguanina en un ácido nucleico para reducir la hibridación inespecífica, como se describe en líneas generales en la patente de los EE. UU. n.º 5 681 702.

Una base no nativa utilizada en un ácido nucleico en relación con los métodos y/o las composiciones de la presente descripción puede tener una actividad de emparejamiento de bases universal, en donde es capaz de emparejarse con cualquier otra base natural. Las bases ejemplares que tienen una actividad de emparejamiento de bases universal incluyen, pero no se limitan a ellas, 3-nitropirrol y 5-nitroindol. Otras bases que se pueden usar incluyen las que tienen actividad de emparejamiento de bases con un subconjunto de las bases naturales, tales como la inosina, cuyas bases se emparejan con citosina, adenina o uracilo.

Un ácido nucleico que tiene una estructura análoga o modificada se puede usar en relación con los métodos y/o composiciones de la presente descripción, por ejemplo, para facilitar la adición de marcaciones o para aumentar la estabilidad o la semivida de la molécula en las condiciones de amplificación u otras condiciones usadas de acuerdo con la presente descripción. Como apreciarán los expertos en la técnica, uno o más de los ácidos nucleicos descritos anteriormente se pueden usar en relación con los métodos y/o composiciones de la presente descripción, que incluyen, por ejemplo, una mezcla que incluye moléculas con estructuras nativas o análogas. Además, un cebador de ácido nucleico utilizado en la presente descripción puede tener una estructura deseada para una técnica de amplificación concreta utilizada en la presente descripción, como las que se presentan a continuación.

Los expertos en la técnica reconocerán que la especificidad de hibridación suele aumentar a medida que aumenta la longitud del cebador de ácido nucleico. Por lo tanto, se puede usar un cebador de ácido nucleico más largo, por ejemplo, para aumentar la especificidad o la reproducibilidad de la replicación, si se desea. En consecuencia, un ácido

nucleico utilizado en un método de la presente descripción puede tener al menos 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 o más nucleótidos de longitud. Los expertos en la técnica reconocerán que una sonda de ácido nucleico utilizada en la presente descripción también puede tener cualquiera de las longitudes ejemplares expuestas anteriormente.

En realizaciones concretas, una población de ácidos nucleicos utilizada en relación con los métodos y/o composiciones de la presente descripción puede incluir miembros cuyas secuencias están diseñadas mediante algoritmos o procesos racionales. De manera similar, una población de ácidos nucleicos puede incluir miembros que tengan cada uno al menos una parte de su secuencia diseñada mediante algoritmos o procesos racionales. Los algoritmos o procesos de diseño racional se pueden usar para dirigir la síntesis de un producto de ácido nucleico que tiene una secuencia determinada o para dirigir la síntesis de una mezcla de ácidos nucleicos que está sesgada para contener preferiblemente unas secuencias concretas.

Al usar los métodos de diseño racional, las secuencias de ácidos nucleicos en una población pueden seleccionarse, por ejemplo, en función de las secuencias conocidas en el ADN_g que se va a amplificar o detectar. Las secuencias se pueden seleccionar de modo que la población incluya preferiblemente secuencias que se hibridan con el ADN_g con una cobertura deseada. Por ejemplo, se puede diseñar una población de cebadores para que incluya preferiblemente miembros que se hibridan con un cromosoma concreto o con una porción de un ADN_g, tal como regiones codificantes o regiones no codificantes. También se pueden seleccionar otras propiedades de una población de ácidos nucleicos para lograr una hibridación preferencial entre sí en posiciones a lo largo de una secuencia de ADN_g que tienen una longitud promedio, mínima o máxima deseada. Por ejemplo, la longitud del cebador se puede seleccionar para que se hibride y cebe al menos aproximadamente cada 64, 256, 1 000, 4 000, 16 000 o más bases entre sí a lo largo de una secuencia de ADN_g.

Los ácidos nucleicos útiles en relación con los métodos y/o las composiciones de la presente descripción también pueden diseñarse para omitir o reducir preferiblemente las secuencias que se hibridan con secuencias concretas en un ADN_g que se va a amplificar o detectar, tal como repeticiones conocidas o elementos repetitivos que incluyen, por ejemplo, las repeticiones Alu. En consecuencia, se puede diseñar una sola sonda o cebador como el que se usa en la amplificación de cebadores arbitrarios para incluir o excluir una secuencia concreta. De manera similar, se puede sintetizar una población de sondas o cebadores, tal como una población de cebadores utilizados para la amplificación aleatoria de cebadores, para excluir o incluir preferiblemente secuencias concretas, tal como las repeticiones Alu. También se puede sintetizar una población de cebadores aleatorios para que incluya preferiblemente un mayor contenido de nucleótidos G y/o C en comparación con los nucleótidos A y T. La población de cebadores aleatorios resultante será rica en GC y, por lo tanto, tendrá una mayor probabilidad de hibridarse con regiones ricas en GC de un genoma, tal como las regiones codificantes de genes de un genoma humano, que normalmente tienen un contenido de GC más alto que las regiones de ADN_g no codificantes. Por el contrario, los cebadores ricos en AT se pueden sintetizar para amplificar o hibridarse preferiblemente con regiones ricas en AT, tal como las regiones no codificantes de un genoma humano. Otros parámetros que se pueden usar para influir en el diseño de los ácidos nucleicos incluyen, por ejemplo, la eliminación preferencial de secuencias que hacen que los cebadores sean autocomplementarios, propensos a la formación de dímeros de cebadores o propensos a la formación de horquillas, o la selección preferencial de secuencias que tienen una T_m máxima, mínima o promedio deseada. Los métodos y algoritmos ejemplares que se pueden usar en la presente descripción para diseñar sondas incluyen los descritos en la patente de los EE. UU. US 2003/0096986A1.

Los cebadores en una población de cebadores aleatorios pueden tener una región de secuencia idéntica, tal como una cola universal. Una cola universal puede incluir un sitio de cebado universal para un paso de amplificación posterior o un sitio que se hibrida con un agente de unión concreto útil para aislar o detectar las secuencias amplificadas. Los métodos para hacer y usar una población de cebadores aleatorios con colas universales se describen, por ejemplo, en Singer et al., *Nucl. Acid. Res.* 25:781-786 (1997)) o Grothues et al., *Nucl. Acids Res.* 21:1321-2 (1993).

Los expertos en la técnica reconocerán que cualquiera de una serie de ácidos nucleicos utilizados en relación con los métodos y/o composiciones de la presente descripción, tales como sondas, pueden tener una o más de las propiedades, o pueden producirse, como se presentó anteriormente, incluidos los ejemplos dados a conocer en relación con los cebadores.

Cebadores ricos en AT

El sesgo de secuencia debido a la eficiencia de cebado diferencial entre los cebadores aleatorios (n-meros) ricos en AT y los ricos en GC puede reducirse o disminuirse al mínimo de acuerdo con cualquiera de una variedad de estrategias novedosas presentadas en la presente memoria. Una de estas estrategias es sesgar la composición de una mezcla aleatoria de n-meros de modo que las secuencias ricas en AT estén presentes en una concentración más alta que la que se encontraría en una verdadera mezcla aleatoria de n-meros. En consecuencia, en la presente memoria se encuentra un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: amplificar una muestra de ácidos nucleicos con un conjunto de cebadores ricos en AT para la amplificación aleatoria. En ciertos aspectos, el conjunto de cebadores ricos en AT para la amplificación aleatoria es una mezcla de cebadores. En ciertos aspectos, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos aspectos, el ADN genómico comprende ADN humano. En ciertos aspectos, la muestra de ácidos

nucleicos comprende numerosos ADN genómicos.

Por lo general, durante la síntesis de oligonucleótidos de n-meros aleatorios, la degeneración final de bases es aproximadamente de proporciones iguales entre las cuatro bases: ~25 % de A, ~25 % de C, ~25 % de G y ~25 % de T. Sin embargo, según los métodos y/o composiciones presentados en la presente memoria, durante la síntesis aleatoria de los oligonucleótidos se puede ajustar cualquier cantidad de degeneración para una base dada. Los métodos para ajustar la cantidad de degeneración de una base dada durante la síntesis de oligonucleótidos se conocen en la técnica y pueden lograrse, por ejemplo, mediante el ajuste al alza o a la baja de la concentración de una o más soluciones de nucleósido con fosforamidita durante la síntesis de oligonucleótidos por lotes. Se conocen en la técnica otras estrategias para modificar la proporción de bases durante la síntesis de oligonucleótidos aleatorios o para obtener una mezcla de n-meros con una proporción de bases sesgada.

Por lo tanto, en ciertos aspectos, la composición global del conjunto de cebadores ricos en AT para la amplificación aleatoria es de más del 25 % de A y 25 % de T. En ciertas realizaciones, el conjunto de cebadores ricos en AT para la amplificación aleatoria comprende la base 'A' en una proporción de aproximadamente el 25 %, 26 %, 27 %, 28 %, 29 %, 30 %, 31 %, 32 %, 33 %, 34 %, 35 %, 36 %, 37 %, 38 %, 39 %, 40 %, 41 %, 42 %, 43 %, 44 %, 45 %, 46 %, 47 %, 48 %, 49 %, 50 % o más que aproximadamente el 50 % del total de bases en la mezcla aleatoria de cebadores para amplificación. En ciertas realizaciones, el conjunto de cebadores ricos en AT para la amplificación aleatoria comprende la base 'T' en una proporción de aproximadamente el 25 %, 26 %, 27 %, 28 %, 29 %, 30 %, 31 %, 32 %, 33 %, 34 %, 35 %, 36 %, 37 %, 38 %, 39 %, 40 %, 41 %, 42 %, 43 %, 44 %, 45 %, 46 %, 47 %, 48 %, 49 %, 50 % o superior a aproximadamente el 50 % de las bases totales en la mezcla aleatoria de cebadores para amplificación. En ciertas realizaciones típicas, el conjunto de cebadores ricos en AT para la amplificación aleatoria comprende aproximadamente el 30 % de A, aproximadamente el 20 % de C, aproximadamente el 20 % de G y aproximadamente el 30 % de T. En ciertas realizaciones típicas, el conjunto de cebadores ricos en AT para la amplificación aleatoria comprende aproximadamente el 35 % de A, aproximadamente el 15 % de C, aproximadamente el 15 % de G y aproximadamente el 35 % de T. En algunas otras realizaciones, el conjunto de cebadores ricos en AT para la amplificación aleatoria es de aproximadamente 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o más de aproximadamente 30 nucleótidos de longitud. Se apreciará que la composición exacta de cada base en una mezcla de n-meros se puede ajustar según sea necesario para generar un nivel deseado de uniformidad de amplificación a lo largo del ácido nucleico deseado, tal como ADN genómico. Se presentan más ejemplos en el ejemplo 6 que viene a continuación.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores ricos en AT para la amplificación aleatoria. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos aspectos, el kit comprende además una ADN polimerasa. En ciertos aspectos, el conjunto de cebadores ricos en AT para la amplificación aleatoria es una mezcla de cebadores.

Colas en 5' ricas en AT

Otra estrategia novedosa para reducir o disminuir al mínimo el sesgo de secuencia debido a la eficiencia de cebado diferencial entre los cebadores aleatorios (n-meros) ricos en AT y los ricos en GC es agregar colas en 5' ricas en AT degeneradas a una mezcla de cebadores para amplificación aleatoria. Por lo tanto, aunque la cola en 5' es rica en AT, la porción 3' restante de los cebadores para amplificación en la mezcla está parcial o totalmente degenerada y puede o no ser rica en AT. En consecuencia, en la presente memoria se presenta un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: a) amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en donde los cebadores para amplificación aleatoria comprenden colas en 5' ricas en AT. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

La longitud de la cola en 5' rica en AT puede ser cualquier longitud apropiada para generar un nivel deseado de uniformidad de amplificación a lo largo del ácido nucleico deseado, tal como ADN genómico. En ciertos aspectos, la cola en 5' rica en AT puede tener aproximadamente 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o más de aproximadamente 30 nucleótidos de longitud. La cola en 5' puede ser cualquier porción en 5' de un oligonucleótido. Por ejemplo, la cola en 5' puede comprender la porción en 5' que tiene aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o más de aproximadamente 30 nucleótidos de un oligonucleótido que tiene aproximadamente 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o más de aproximadamente 30 nucleótidos de longitud. Por lo tanto, en ciertos aspectos, la composición global del conjunto de cebadores ricos en AT para la amplificación aleatoria es de más del 25 % de A y del 25 % de T. En ciertas realizaciones, la cola en 5' rica en AT comprende la base 'A' en una proporción de aproximadamente el 25 %, 26 %, 27 %, 28 %, 29 %, 30 %, 31 %, 32 %, 33 %, 34 %, 35 %, 36 %, 37 %, 38 %, 39 %, 40 %, 41 %, 42 %, 43 %, 44 %, 45 %, 46 %, 47 %, 48 %, 49 %, 50 % o más que aproximadamente el 50 % del total de bases en la mezcla aleatoria de cebadores para amplificación. En ciertas realizaciones, la cola en 5' rica en AT comprende la base 'T' en una proporción de aproximadamente el 25 %, 26 %, 27 %, 28 %, 29 %, 30 %, 31 %, 32 %, 33 %, 34 %, 35 %, 36 %, 37 %, 38 %, 39 %, 40 %, 41 %, 42 %, 43 %, 44 %, 45 %, 46 %, 47 %, 48 %, 49 %, 50 % o mayor que aproximadamente el 50 % de las bases totales en la mezcla aleatoria de cebadores para amplificación. En ciertas realizaciones típicas, la cola en 5' rica en AT comprende el 30 % de A, 20 % de C, 20 % de G y 30 % de T. En ciertas

realizaciones típicas, la cola en 5' rica en AT comprende el 35 % de A, 15 % de C, 15 % de G y 35 % de T. En ciertas realizaciones típicas, la cola en 5' rica en AT comprende el 40 % de A, 10 % de C, 10 % de G y 40 % de T. En ciertas realizaciones típicas, la cola en 5' rica en AT comprende el 50 % de A y el 50 % de T.

Como ejemplo, una secuencia degenerada de 9 unidades podría ser $[W]_m[N]_9$, donde W = degeneración de [A/T], m = un número entero de 0 a 20 y N = cualquier base. Así, por ejemplo, un caso concreto puede ser: 5'-WWWNNNNNNNN-3'. Este n-mero aleatorio en concreto tiene 9 posiciones totalmente degeneradas en el extremo 3' del cebador y 3 bases oscilantes [A/T] en el extremo 5'. Se presentan ejemplos adicionales en el ejemplo 6 que viene a continuación.

En ciertos aspectos, la muestra de ácidos nucleicos comprende un ADN genómico. En ciertos aspectos, el ADN genómico comprende ADN humano. En ciertos aspectos, la muestra de ácidos nucleicos comprende numerosos ADN genómicos.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación aleatoria, en donde los cebadores para amplificación aleatoria comprenden colas en 5' ricas en AT. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos aspectos, el kit comprende además una ADN polimerasa. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

Cola en 5' degenerada

Otra estrategia novedosa para reducir o disminuir al mínimo el sesgo de secuencia debido a la eficiencia de cebado diferencial entre los cebadores aleatorios (n-meros) ricos en AT y los ricos en GC es usar n-meros de longitud variable con colas en 5' degeneradas que son proporcionales en longitud al contenido de A/T. En consecuencia, en la presente memoria se presenta un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria de longitud variable, en donde cada cebador comprende una porción en 3' aleatoria y una cola en 5' degenerada, en donde la cola en 5' degenerada es proporcional en longitud al contenido de A/T de la porción aleatoria en 3' del cebador. En ciertas realizaciones, esto utiliza la síntesis individual de todas las secuencias de n-meros, en lugar de la síntesis aleatoria por lotes de una mezcla de oligonucleótidos. Por ejemplo, en una realización, utilizando bases degeneradas W(A/T) y S(C/G), un n-mero hexamérico de longitud variable requeriría 2^6 o 64 síntesis diferentes. Asimismo, un heptámero requeriría 128 síntesis distintas.

La estrategia descrita anteriormente permite la normalización de la T_m de cada oligonucleótido en una mezcla de oligonucleótidos. En la siguiente tabla se presenta una estrategia ejemplar para una mezcla de n-meros.

0 W	1 W	2 W	3 W	4 W	5 W	6 W
0	N	N	NN	NN	NNN	NNN

Tal como se presenta en la tabla anterior, un n-mero que tenga un contenido de 0 W (A/T) requeriría la adición de 0 bases degeneradas en el extremo 5' del n-mero. Asimismo, un n-mero que tenga un contenido de 3 W (A/T) requeriría la adición de 2 bases degeneradas (NN) en el extremo 5' del n-mero. De igual forma, un n-mero que tenga un contenido de 6 W (A/T) requeriría la adición de 3 bases degeneradas (NNN) en el extremo 5' del n-mero. Así, por ejemplo, a WWWWWW (6W) se le asigna una cola de 3 bases en el n-mero: 5'-NNNWWWWWW, y a WWSSSWW (4W) se le asigna una cola de 2 bases en el n-mero: 5' NNWSSWW, y de manera similar, a SSSSWW (2W) se le asigna una cola de 1 base en el n-mero: 5'NSSSWW. Este concepto se aplica a un heptámero aleatorio en el ejemplo 7 que viene a continuación.

Como resultado de la adición de bases degeneradas al extremo 5' de un n-mero como se expuso anteriormente, se puede lograr la normalización de las T_m de una mezcla de cebadores, como se establece, por ejemplo, en la figura 5.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación aleatoria de longitud variable, en donde cada cebador comprende una porción en 3' aleatoria y una cola en 5' degenerada, en donde la cola en 5' degenerada es proporcional en longitud al contenido de A/T de la porción aleatoria en 3' del cebador. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos aspectos, el kit comprende además una ADN polimerasa. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria de longitud variable es una mezcla de cebadores.

Cebadores con análogos de bases para normalizar la T_m y evitar las interacciones entre los cebadores

Otra estrategia novedosa para reducir o disminuir al mínimo el sesgo de secuencia debido a la eficiencia de cebado diferencial entre los cebadores aleatorios (n-meros) ricos en AT y los ricos en GC es emplear análogos de bases para normalizar la T_m de los n-meros que difieren en el contenido de AT. En consecuencia, en la presente memoria se presenta un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde

el método comprende: amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación con la T_m normalizada, en donde cada cebador del conjunto de cebadores para amplificación con la T_m normalizada comprende uno o más análogos de bases que normalizan la T_m de cada cebador a la T_m de otros cebadores del conjunto de cebadores. En ciertos aspectos, el conjunto de cebadores para amplificación con la T_m normalizada es una mezcla de cebadores.

El uso de parejas de bases pseudocomplementarias, tales como 2-tío-dT y 2-amino-dA que muestran poca afinidad por su nucleótido complementario en las reacciones de hibridación, es útil en los métodos descritos en la presente memoria. En las reacciones de RPA y MDA, el uso de cebadores aleatorios con bases pseudocomplementarias ayuda a evitar los artefactos de amplificación entre los cebadores. Los artefactos de amplificación entre cebadores son uno de los principales obstáculos para obtener una WGA eficaz a partir de poca cantidad de ácidos nucleicos, ya que la amplificación de los dímeros entre cebadores puede competir con la amplificación del ADN diana. Por lo tanto, la incorporación de bases pseudocomplementarias es en concreto ventajosa cuando la amplificación se realiza con poca cantidad de ADN de entrada, por ejemplo, durante la amplificación de una sola célula. Las ventajas adicionales de la incorporación de bases pseudocomplementarias incluyen, por ejemplo, la normalización de las T_m como se describe más arriba. Se apreciará que los análogos de bases que forman más o menos puentes de hidrógeno durante la hibridación se pueden incorporar en un conjunto de cebadores para ajustar la T_m global del conjunto de cebadores. En algunas realizaciones, el análogo de base es una base no natural en el cebador que puede o no formar puentes de hidrógeno con su complemento, pero puede evitar la amplificación de la base no natural al mantener intacta la funcionalidad del cebador. Las bases no naturales adicionales que son adecuadas en los métodos y composiciones dados a conocer en la presente memoria las conocen los expertos en la técnica, como se ejemplifica en Hoshika et al., *Angew. Chem. Int. Ed.* (2010) 49:5554. Otros ejemplos de análogos de nucleótidos o análogos de bases que se pueden usar para estos fines incluyen, por ejemplo, nucleótidos abásicos que crean un hueco, de modo que los puentes de hidrógeno se reducen durante la hibridación.

Por ejemplo, la 2-amino-dA forma un puente de hidrógeno adicional con la timidina, lo que conduce a la estabilización del dúplex con un aumento de la temperatura de fusión de 3 °C. Como otro ejemplo, los análogos N4-etilados de dC (N4-Et-dC) se hibridan específicamente con la guanina natural, pero la estabilidad de la pareja de bases se reduce a aproximadamente el nivel de una pareja de bases AT. De esta manera, un n-mero aleatorio que usa N4-Et-dC para reemplazar la G en una o más posiciones puede disminuir el sesgo de cebado de GC. Del mismo modo, un n-mero aleatorio que usa la 2-amino-dA para reemplazar la A en una o más posiciones puede disminuir el sesgo de cebado de GC. De igual forma, la N4-Et-dC podría usarse junto con la 2-amino-dA para reemplazar A y C en una o más posiciones para disminuir el sesgo de cebado de GC. Se apreciará que cualquier base análoga que reduzca o aumente la estabilidad de una pareja de bases puede usarse para ajustar la T_m de un juego de cebadores. Por lo tanto, por ejemplo, los análogos de bases que incluyen, entre otros, 2-tío-dT, 2-amino-dA, N4-Et-dC y 7-desaza-G se pueden usar en relación con los métodos y/o las composiciones presentadas en la presente memoria. Los expertos en la técnica conocen otros análogos de bases adecuados, como, por ejemplo, los presentados en otra parte de la presente memoria.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación con la T_m normalizada, en donde cada cebador del conjunto de cebadores para amplificación con la T_m normalizada comprende uno o más análogos de bases que normalizan la T_m de cada cebador con respecto a la T_m de otros cebadores para amplificación del kit. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos aspectos, el kit comprende además una ADN polimerasa. En ciertos aspectos, el conjunto de cebadores para amplificación con la T_m normalizada es una mezcla de cebadores.

Cebadores quiméricos

Otra estrategia novedosa para reducir el sesgo durante la amplificación con cebadores aleatorios es utilizar cebadores quiméricos con una parte de cebado aleatoria en 3' y una secuencia constante en 5'. En consecuencia, en la presente memoria se presenta un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: a) amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en donde cada cebador comprende una porción en 3' aleatoria y una porción de cebado constante en 5', con lo que se producen así productos de amplificación, y en donde cada producto de amplificación comprende la porción de cebado en 5' constante; b) circularizar los productos de amplificación; y c) amplificar los productos de amplificación circularizados con el uso de cebadores que se hibridan con la porción de cebado en 5' constante. En ciertos aspectos, la amplificación en el paso c) comprende realizar una amplificación por desplazamiento múltiple. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

Un ejemplo de cómo se puede utilizar este método para generar un banco de ácidos nucleicos con mayor uniformidad y menor sesgo se presenta a continuación en el ejemplo 8.

Con el fin de eliminar o reducir la formación de horquillas, los cebadores quiméricos presentados en la presente memoria pueden incluir opcionalmente una base no natural entre la porción en 3' aleatoria y la porción cebadora constante en 5'. Se apreciará que se puede utilizar cualquier base no natural adecuada que sirva para reducir o eliminar la formación de horquillas. Por ejemplo, en ciertas realizaciones típicas, la base no natural es isoC.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación aleatoria que comprenden una porción en 3' aleatoria y una porción cebadora en 5' constante. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos aspectos, el kit comprende además una ADN polimerasa. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

Cebadores de ARN aleatorios

En la presente memoria se presenta el descubrimiento de que los artefactos de extensión entre cebadores pueden reducirse o disminuirse al mínimo empleando cebadores aleatorios que no interaccionan. Tal como se usa en la presente memoria, el término «que no interacciona» se refiere a los cebadores diseñados para no hibridarse con otros cebadores en las condiciones normales de amplificación. Se puede utilizar cualquier tipo de cebador que no interaccione para reducir los artefactos de extensión entre cebadores normalmente asociados a la amplificación aleatoria del cebador. En consecuencia, una estrategia consiste en utilizar cebadores aleatorios que comprenden ARN en al menos la porción en 3' del cebador. Por lo tanto, en la presente memoria se presenta un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: a) amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en donde cada cebador comprende una porción en 3' aleatoria y una porción de cebado constante en 5', y en donde la porción aleatoria en 3' comprende ARN, con lo que se producen así productos de amplificación, en donde cada producto de amplificación comprende la porción de cebado en 5' constante. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores. En ciertos aspectos, el método comprende, además: b) circularizar los productos de amplificación; y c) amplificar los productos de amplificación circularizados mediante el uso de cebadores que se hibridan con la porción de cebado en 5' constante. En ciertos aspectos, la amplificación en el paso c) comprende realizar una amplificación por desplazamiento múltiple.

Un ejemplo de un uso de este método se presenta a continuación en el ejemplo 9 y en la figura 7.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende un conjunto de cebadores para amplificación aleatoria, en donde los cebadores comprenden una porción en 3' aleatoria y una porción de cebado en 5' constante, en donde la porción en 3' aleatoria comprende ARN. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos aspectos, el kit comprende además una ADN polimerasa. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

Cebadores de ADN aleatorios con 2-tío-dT o 2-amino-dA

Otra estrategia para evitar los artefactos de extensión entre cebadores es usar cebadores de ADN aleatorios que comprenden 2-tío-dT y 2-amino-dA en al menos la porción en 3' del cebador. Por lo tanto, en la presente memoria se presenta un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos, en donde el método comprende: amplificar una muestra de ácidos nucleicos con un conjunto de cebadores para amplificación aleatoria, en el que cada cebador comprende una porción en 3' aleatoria y una porción de cebado en 5' constante, y en donde la porción aleatoria en 3' comprende al menos una base no natural seleccionada del grupo que consiste en: 2-tío-dT y 2-amino-dA, con lo que se producen así los productos de amplificación, en donde cada producto de amplificación comprende la porción constante de cebado en 5'. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores. La incorporación de 2-tío-dT y 2-amino-dA en los oligonucleótidos se puede realizar de acuerdo con cualquiera de las numerosas formas conocidas en la técnica, como se ejemplifica en Kutiavin et al. (1996) *Biochemistry*, 35:11170-11176.

En ciertos aspectos, el método comprende, además: b) circularizar los productos de amplificación; y c) amplificar los productos de amplificación circularizados mediante el uso de cebadores que se hibridan con la porción de cebado en 5' constante. En ciertos aspectos, la amplificación de la etapa c) comprende realizar una amplificación por desplazamiento múltiple.

También se presenta en la presente memoria un kit para amplificar una muestra de ácidos nucleicos, en donde el kit comprende cebadores para amplificación aleatoria, en donde los cebadores comprenden una porción aleatoria en 3' y una porción constante de cebado en 5', en donde la porción aleatoria en 3' comprende al menos una base no natural seleccionada del grupo que consiste en: 2-tío-dT y 2-amino-dA. En ciertos aspectos, el kit comprende además un conjunto de instrucciones para combinar el conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar el banco de ácidos nucleicos. En ciertos aspectos, el kit comprende además una ADN polimerasa. En ciertos aspectos, el conjunto de cebadores para amplificación aleatoria es una mezcla de cebadores.

Ejemplos

Ejemplo 1

Amplificación hologenómica utilizando la amplificación por desplazamiento múltiple que incluye cebadores

específicos de locus

Las muestras de ADN genómico se obtuvieron de Coriell Cell Repositories (Camden, NJ). Cada concentración de ADN genómico se midió con el kit de reactivos de detección de RNasa P con TaqMan y la mezcla Master Universal para PCR con TaqMan (Life Technologies, Foster City, CA) y el reactivo Quant-iT PicoGreen de dsDNA (Life Technologies, Foster City, CA).

El ADN genómico se diluyó a varias diluciones que oscilaban entre números de copias haploides de 0,1n a 2n por 3 µl de agua. Se tomaron alícuotas de 3 µl del ADN genómico diluido en varios tubos. A cada tubo se le añadieron 3 µl del tampón REPLI-g UltraFast D2 de Qiagen seguido de 10 min de incubación a 4 °C (Qiagen, Valencia, CA). Se le añadieron 3 µl de la solución de parada ultrarrápida REPLI-g, seguidos de 33 µl de la mezcla Master. La mezcla Master contenía 30 µl de tampón de reacción ultrarrápida REPLI-g, 2 µl de la ADN polimerasa ultrarrápida REPLI-g y, opcionalmente, 1 µl de un conjunto de nonámeros que contenía oligonucleótidos específicos de locus diseñados específicamente para el genoma humano (para una concentración final de 0,03 µM por oligonucleótido). La mezcla Master contiene hexámeros aleatorios resistentes a las exonucleasa en una concentración tal que la concentración final en la mezcla de reacción era de 50 µM de hexámero. Las reacciones se incubaron durante 90 min a 30 °C seguido de la inactivación por calor durante 3 min a 65 °C. Los productos de amplificación por desplazamiento múltiple (MDA) se purificaron con columnas de centrifugación de DNA Clean & Concentrator-5 (Zymo Research, Irvine, CA) según el protocolo del fabricante, y se eluyeron en 12 µl de agua. La concentración de ADN de los productos amplificados se determinó con el reactivo Quant-iT™ PicoGreen para dsDNA (Life Technologies, Foster City, CA).

Ejemplo 2**Detección de snp mediante la amplificación de adn genómico con cebadores específicos de locus**

Un sitio de polimorfismo mononucleotídico (SNP), que se muestra en la figura 1 como "A/G", está ubicado en un locus de ADN ilustrado por líneas finas. Los cebadores, que se muestran en la figura 1 como flechas, están diseñados para que se hibriden con el ADN en ambos lados del sitio del SNP. La amplificación hologenómica (WGA) se inicia tal como se describe en el ejemplo 1. La realización de la WGA con cebadores específicos de locus genera una serie de productos de amplificación de diferentes longitudes, que se muestran en la figura 1. Estos productos de amplificación se amplifican aún más durante la WGA con el uso de cebadores específicos de locus y también con los cebadores aleatorios presentados en la mezcla de reacción.

Ejemplo 3**Uso de cebadores específicos de locus para aumentar la intensidad de la señal del ensayo de genotipado**

La amplificación hologenómica (WGA) se inició como se describe en el ejemplo 1, con o sin la adición de cebadores específicos de locus diseñados para las regiones que rodean las secuencias de 120 sondas de captura. A continuación, los productos amplificados se analizaron con una matriz de perlas para genotipado INFINIUM™ (Illumina, San Diego, CA). La intensidad de señal de las 120 sondas se representó gráficamente y se expone en la figura 2. Como se muestra en la figura 2, se observó un aumento evidente en la intensidad de la señal, con un promedio de 1,8 veces para la muestra de ADN amplificada con el conjunto de cebadores específicos de locus.

Ejemplo 4**Detección de snp mediante amplificación de adn genómico con cebadores específicos de locus**

La amplificación hologenómica (WGA) se inició tal como se describe en el ejemplo 1, con o sin la adición de cebadores específicos de locus diseñados para la región que rodea un SNP de interés. Se realizó un análisis de SNP para un sitio de SNP seleccionado con dos réplicas de la muestra de ADN amplificada en la reacción de WGA. En la figura 3 se muestra un gráfico del análisis de SNP. En la figura 3 se muestra que el SNP seleccionado se genotipó con éxito en la muestra de ADN amplificada en presencia de cebadores específicos de locus.

Ejemplo 5**Uso de cebadores específicos de locus en el ensayo de genotipado de células sueltas**

La amplificación hologenómica (WGA) se inició en una muestra de ADN de una única célula, usando los parámetros descritos en el ejemplo 1, con o sin la adición de aproximadamente 6000 nonámeros específicos que están muy representados en el genoma humano. Con la adición de estos cebadores nonaméricos humanizados, se logró una amplificación más robusta, como se describe a continuación.

En la figura 4 se muestra un resultado típico, que muestra la intensidad de la señal a lo largo del cromosoma 2 humano. Como se muestra en la figura 4, las reacciones de amplificación que incluyeron los cebadores nonaméricos dieron como resultado una amplificación homogénea más robusta por todo el cromosoma, lo que redujo en gran medida el sesgo a lo largo del cromosoma amplificado, en comparación con la amplificación sin los nonámeros. La tasa de éxito de la secuenciación de los nucleótidos y la precisión de la detección también mejoraron significativamente en las reacciones de amplificación que incluían los nonámeros, en comparación con la amplificación sin los nonámeros.

Ejemplo 6**Uso de cebadores ricos en at y cebadores con colas en 5' ricas en at en la amplificación genómica**

Se diseñaron y sintetizaron cuatro conjuntos diferentes de cebadores como se indica en la siguiente tabla. En la tabla, los dos primeros conjuntos tienen colas de AT de 5'. Los segundos tres conjuntos tienen una degeneración de bases sesgada. W = (A/T). El formato de la composición de las bases se especifica como [A-C-G-T]. Una mezcla equimolar se especifica como [25-25-25-25].

Mezcla	Secuencia (5' – 3')
N9-3AT	[W] [W] [W] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [N]
N9-6AT	[W] [W] [W] [W] [W] [W] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [25-25-25-25] [N]
N9-AT-rich	[35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [N]
N12-AT-rich	[35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [N]
N15-AT-rich	[35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [35-15-15-35] [N]

La amplificación por cebado aleatorio (RPA) usando un nonúmero aleatorio rico en AT se evaluó por la amplificación del ADN de Coriell con el uso de una composición estándar de nonúmeros ([25-25-25-25]) frente a nonúmeros ricos en AT ([35-15-15-35]). El producto amplificado se obtuvo en un ensayo de genotipado Infinium™ estándar y el efecto de la intensidad de las señales sin procesar se representó en función del contenido de GC de la sonda. Los resultados se exponen en la figura 8.

Como se muestra en la figura 8, el nonúmero sesgado en AT (35-15-15-35) hizo aumentar claramente la intensidad de las sondas ricas en AT e hizo disminuir la intensidad de las sondas ricas en GC en relación con el estándar.

Ejemplo 7**Normalización de la t_m global de una mezcla de cebadores con colas degeneradas en 5'**

Se diseñó la síntesis de un conjunto de hexámeros completamente degenerados (NNNNNN). Asimismo, se diseñó un conjunto distinto de hexámeros completamente degenerados para que incluyeran una cola adicional en 5' cuya longitud fuera proporcional al contenido de A/T del hexámero. Se calculó la T_m global de cada mezcla de cebadores. Los resultados se exponen en la figura 5. Como se muestra en la figura 5, la adición de colas a la mezcla de hexámeros da como resultado un aumento de la T_m global y una disminución de la desviación estándar de la T_m global para la mezcla.

El mismo concepto se aplica a un heptámero aleatorio. Se diseñó un conjunto de heptámeros completamente degenerados para que incluyeran colas en 5' de longitud variable cuya longitud es proporcional al contenido de A/T de los heptámeros. El conjunto completo de heptámeros se expone en la siguiente tabla.

NNNNWWWWWWW	NNWSSSWWW	NNWSSW'WWW	NNSWW'SWWW	NNSSWWWWWW	NSSSSWWW
NNNNWWWWWWS	NNWSSSWWS	NNWSSW'WWS	NNSWW'SWWS	NNSSWWWWWS	NSSSSWWS
NNNNWWWWWWSW	NNWSSSWSW	NNWSSW'WSW	NNSWW'SWSW	NNSSWWWWSW	NSSSSWSW
NNWWW'WSSS	NWWSSWSS	NWSSW'WSS	NSWW'SWSS	NSSWW'WSS	SSSSWSS
NNNNWWW'SWW	NNWSSS'WW	NNWSSW'SWW	NNSWW'SS'WW	NNSSWWS'WW	NSSSSS'WW
NNWWW'WSWS	NWWSSSWS	NWSSW'SWS	NSWW'SSWS	NSSW'SWS	SSSSWS
NNWWW'WSSW	NWWSSSSW	NWSSW'SSW	NSWW'SSSW	NSSWW'SSW	SSSSSSW
NNWWW'WSSS	NWWSSSSS	NWSSW'SSS	NSWW'SSSS	NSSWW'SSS	SSSSSSS
NNNNWWSWWW	NNNWSWWWWW	NNWSSS'WWW	NNSWS'WWW	NNSSWS'WW	
NNWWW'SWWS	NNWSWW'WS	NWSSS'WWS	NNSWS'WWS	NSSWS'WS	
NNWWW'SWSW	NNWSWW'WSW	NWSSS'WSW	NNSWS'WSW	NSSWS'WSW	
NNWWW'SWSS	NNWSWW'WSS	NWSSS'WSS	NSWS'WSS	NSSWS'WSS	
NNWWW'SSWW	NNWSWW'SWW	NWSSS'SWW	NNSWS'S'WW	NSSWSS'WW	
NNWWW'SSWS	NNWSWWS'WS	NWSSS'SWS	NSWS'WSWS	NSSWSS'WS	
NNWWW'SSSW	NNWSWW'SSW	NWSSS'SSW	NSWS'WSSW	NSSWSS'SW	
NWWW'SSSS	NWSW'WSSS	WSSSSS	NSWS'WSSS	SSW'SSSS	
NNNWSW'WWW	NNWSWS'WWW	NNNSW'WWWWW	NNSWSS'WW	NNSSW'WWW	
NNW'WSWWW	NNWSWS'WWS	NNSW'WWWW	NSWSS'WWS	NSSSW'WWS	
NNWWSW'WSW	NNWSWS'WSW	NNSW'WWWWS	NSWSS'WSW	NSSSW'WSW	
NNWWSW'WSS	NWSWS'WSS	NNSW'WWWSS	NSWSS'WSS	NSSSW'WSS	
NNWWSWS'W	NNWSWS'SWW	NNSW'WWWSW	NSWSS'SWW	NSSWS'WW	
NNWWSWS'WS	NWSWS'SWS	NNSW'WWWWS	NSWSS'SWS	NSSWS'WS	
NNWWSW'SSW	NWSW'SSSW	NNSW'WWSSW	NSWSS'SSW	NSSW'SSW	
NWWSW'SSS	NWSW'SSSS	NSWW'WSSS	SW'SSSS	SSW'SSS	

Ejemplo 8

Uso de cebadores quiméricos para reducir el sesgo de amplificación

Este ejemplo describe el uso de un cebador quimérico aleatorio que comprende una secuencia constante para la amplificación de ácidos nucleicos. Tal como se describe a continuación, la parte aleatoria del cebador se usa para cebar el ácido nucleico fuente, en este caso el ADN genómico. A continuación, la porción constante se utiliza para la amplificación exponencial imparcial de las secuencias amplificadas con el uso del cebador aleatorio.

Tal como se ilustra en la figura 6, la amplificación desde el ADN genómico (ADNg) se ceba inicialmente gracias a la porción aleatoria del cebador quimérico constante-aleatorio que contiene la secuencia aleatoria en 3' y la secuencia constante en 5' (porción A-B). Después de un cebado aleatorio inicial limitado, el producto se circulariza con CircLigase™ y luego se ceba en una reacción de amplificación por desplazamiento múltiple (MDA) con las parejas de cebadores de secuencia constante A' y B, o A y B'. Esto conduce a una amplificación exponencial imparcial del ADN circular. En las realizaciones típicas ilustradas en la figura 6, la secuencia constante (A-B) es infrecuente en el genoma que se amplifica. En la siguiente tabla se muestra un ejemplo de secuencias quiméricas utilizadas en este procedimiento, en donde cA y cB indican cebadores complementarios para las secuencias constantes A y B del cebador quimérico AB-N9.

A	5' -TCGCGAGTTAA
B	5' -ATTGCGAACGA
cA	5'-TTAACTCGCGA
cB	5'-TCGTTCGCAAT
AB-N9	5'-TCGCGAGTTAAATTGCGAACGANNNNNNNNNN

En otro ejemplo, los cebadores quiméricos comprenden una base no natural (es decir, isoC) entre la porción constante y la porción aleatoria del cebador. Durante la etapa de MDA, la inclusión de isoC e isoG junto con otros nucleótidos permite su amplificación.

Ejemplo 9

5 *Uso de cebadores quiméricos de arn-adn para reducir los artefactos de extensión entre cebadores*

En este ejemplo se describe la amplificación aleatoria de cebadores basada en ARN para reducir los artefactos de extensión entre cebadores. Tal como se muestra en la figura 7, el cebado aleatorio con cebadores que contienen ARN permite una circularización eficiente por parte de CircLigase™ (figura 7A). El uso de cebadores n-meros de ARN (cebadores quiméricos de ARN-ADN) no admitirá la formación de dímeros de cebadores de ARN (figura 7B), pero
 10 admite la extensión del ARN sobre el ADN. También permite que se escinda la porción de n-meros aleatorios de ARN y se elimine de las etapas posteriores de generación de bancos. Finalmente, cuando se usa un cebador que contiene una porción constante y una porción aleatoria, un cebador de ARN evita la formación de productos de RPA que pueden formar "mancuernas" (figura 7C y D) que pueden inhibir la circularización por la CircLigase™ mostrada en la figura 7A. Esto se debe a que las reacciones de amplificación del ADN no pueden cebarse sobre el ARN, excepto si se utiliza
 15 una polimerasa con actividad retrotranscriptasa.

Ejemplo 10

Genotipado de adn amplificado a partir de lisados celulares directos de 1 célula, 5 células y 25 células

La amplificación hologenómica se realizó mediante (1) MDA estándar (con todos los hexámeros aleatorios) o hexámeros aleatorios estándares más grupos de cebadores mejorados, correspondientes a las 3000 secuencias de
 20 nonámeros altamente representadas en el genoma humano, en concentraciones de (2) 40 PM o (3) 400 PM, respectivamente. La inclusión de nonámeros específicos de secuencia en la mezcla de reacción mejoró la amplificación en todos los locus que se genotiparon.

REIVINDICACIONES

1. Un kit para amplificar una muestra de ácidos nucleicos, en donde dicho kit comprende un conjunto de cebadores para amplificación para dicha muestra de ácidos nucleicos, en donde dicho conjunto comprende una mezcla de:
 - 5 numerosos cebadores específicos de locus configurados para amplificar al menos 100 regiones predeterminadas de dicha muestra de ácidos nucleicos,
 - opcionalmente, en donde las al menos 100 regiones predeterminadas están distribuidas uniformemente por dicha muestra de ácidos nucleicos y,
 - numerosos cebadores aleatorios, en donde dichos cebadores aleatorios son más abundantes en comparación con dichos cebadores específicos de locus, y
 - 10 numerosos cebadores bloqueados diseñados para que se hibriden con una región sobreamplificada de dicho ácido nucleico, opcionalmente en donde dicha región sobreamplificada es una región altamente repetitiva,
 - un conjunto de instrucciones para usar dichos cebadores aleatorios y dichos cebadores específicos de locus en una reacción de amplificación, y
 - 15 un conjunto de instrucciones para combinar dicho conjunto de cebadores para amplificación con un banco de ácidos nucleicos y amplificar dicho banco de ácidos nucleicos,
 - en donde la combinación garantiza la reducción del sesgo de amplificación con respecto a la amplificación con cebadores aleatorios.
2. El kit de acuerdo con la reivindicación 1, en el que dichos cebadores aleatorios tienen una longitud de 5 a 18 nucleótidos; opcionalmente, en el que dichos cebadores aleatorios tienen una longitud de 9 nucleótidos.
- 20 3. El kit de acuerdo con la reivindicación 1, en el que dichos cebadores específicos de locus comprenden las 1 000 secuencias nonaméricas más altamente representadas en el genoma humano.
4. El kit de acuerdo con la reivindicación 1, en el que:
 - (a) dicha muestra de ácidos nucleicos comprende un ADN genómico; en donde opcionalmente (i) dicho ADN genómico comprende ADN humano; o (ii) en el que las regiones predeterminadas en la muestra de ácidos nucleicos genómicos comprenden una o más regiones no repetitivas y regiones repetitivas altamente representadas de una muestra de ácidos nucleicos genómicos; o
 - 25 (b) dicha muestra de ácidos nucleicos comprende numerosos ADN genómicos.
5. El kit de acuerdo con la reivindicación 1, en el que dichos cebadores para amplificación aleatoria se seleccionan del grupo que consiste en un conjunto de cebadores ricos en AT para amplificación aleatoria; un conjunto de cebadores para amplificación aleatoria que comprenden colas en 5' ricas en AT; un conjunto de cebadores para amplificación aleatoria de longitud variable, en donde cada cebador comprende una porción en 3' aleatoria y una cola en 5' degenerada, en donde dicha cola en 5' degenerada es proporcional en longitud al contenido de A/T de la porción en 3' aleatoria del cebador; un conjunto de cebadores para amplificación con la T_m normalizada, en donde cada cebador de dicho conjunto comprende uno o más análogos de bases que normalizan la T_m de cada cebador por la T_m de otros cebadores en dicho conjunto de cebadores; un conjunto de cebadores para amplificación aleatoria, en el que cada cebador comprende una parte en 3' aleatoria y una parte cebadora en 5' constante; un conjunto de cebadores para amplificación aleatoria, en donde cada cebador comprende una porción en 3' aleatoria y una porción cebadora en 5' constante, y en donde dicha porción en 3' aleatoria comprende ARN; un conjunto de cebadores para amplificación aleatoria, en donde cada cebador comprende una porción aleatoria en 3' y una porción constante de cebado en 5', y en donde dicha porción aleatoria en 3' comprende al menos una base no natural seleccionada del grupo que consiste en 2-tío-dT y 2-amino-dA; y cualquier combinación de los conjuntos de cebadores anteriores.
- 30 6. El kit de acuerdo con la reivindicación 1, en el que los cebadores para amplificación comprenden un conjunto de cebadores con la T_m normalizada, en donde cada cebador de dicho conjunto de cebadores para amplificación con la T_m normalizada comprenden uno o más análogos de bases seleccionados del grupo que consiste en 2-tío-dT, 2-amino-dA, N4-Et-dC y 7-desaza-G.
- 45 7. Un método para crear un banco de ácidos nucleicos a partir de una muestra de ácidos nucleicos genómicos que comprende:
 - a) dar a conocer un conjunto de cebadores para amplificación a una muestra de ácidos nucleicos genómicos, en donde dicho conjunto de cebadores comprende una mezcla de:
 - 50 (i) numerosos cebadores de secuencia aleatoria que comprenden cebadores seleccionados del grupo que consiste en:

- (A) un conjunto de cebadores para amplificación aleatoria de longitud variable, en donde cada cebador comprende una porción en 3' aleatoria y una cola en 5' degenerada cuya longitud es proporcional al contenido de A/T de la porción en 3' aleatoria del cebador,
- 5 (B) un primer conjunto de cebadores para amplificación aleatoria, en donde cada cebador comprende una porción en 3' aleatoria que comprende ARN y una porción cebadora en 5' constante, y
- (C) un segundo conjunto de cebadores para amplificación aleatoria, en donde cada cebador comprende una porción aleatoria en 3' y una porción constante de cebado en 5', en donde la porción aleatoria en 3' comprende al menos una base no natural seleccionada del grupo que consiste en 2- tío-dT y 2-amino-dA, y
- 10 (ii) numerosos cebadores específicos de locus configurados para amplificar al menos 100 regiones genómicas predeterminadas de ácido nucleico de la muestra de ácidos nucleicos, y
- (iii) numerosos cebadores bloqueados complementarios a regiones de un genoma que comprende secuencias repetitivas; en donde cada pluralidad en el conjunto de cebadores para amplificación es una mezcla de cebadores; y
- 15 b) amplificar dicha muestra de ácidos nucleicos genómicos usando dicho conjunto de cebadores para amplificación, con lo que se crea así un banco de ácidos nucleicos,
- en donde la amplificación proporciona una amplificación más homogénea de la muestra de ácidos nucleicos que la amplificación en ausencia de los numerosos cebadores específicos de locus.
8. El método de acuerdo con la reivindicación 7, en el que:
- 20 (a) dicha muestra de ácidos nucleicos genómicos es de un mamífero,
- opcionalmente en donde (1) dicho mamífero es un ser humano, opcionalmente en donde dicha pluralidad de cebadores específicos de locus comprende secuencias humanizadas; o (2) dicho mamífero es un animal económicamente relevante; o
- (b) dicha muestra de ácidos nucleicos genómicos procede de una planta económicamente relevante.
- 25 9. El método de acuerdo con la reivindicación 7, en el que
- (a) dichas regiones definidas en una muestra de ácidos nucleicos genómicos comprenden una o más regiones no repetitivas y regiones altamente representadas de una muestra de ácidos nucleicos genómicos; o
- (b) las secuencias de dicha pluralidad de cebadores específicos de locus se distribuyen esencialmente de manera uniforme por el ácido nucleico genómico de tal manera que la hibridación de dichos cebadores específicos de
- 30 especie con dichas regiones definidas de ácidos nucleicos genómicos no está sesgada.
10. El método de acuerdo con la reivindicación 7, en el que los numerosos cebadores específicos de locus tienen una longitud de aproximadamente 6 a 20 nucleótidos, opcionalmente en el que dichos cebadores específicos de locus tienen una longitud de aproximadamente 6 a 9 nucleótidos.
11. El kit de acuerdo con la reivindicación 1 o el método de la reivindicación 7, en el que dichos cebadores aleatorios son más abundantes que dichos cebadores específicos de locus; o en donde dichos cebadores aleatorios están presentes en igual abundancia que dichos cebadores específicos de locus; o en donde dichos cebadores aleatorios son menos abundantes que dichos cebadores específicos de locus.
- 35 12. El método de acuerdo con la reivindicación 7, en el que
- dicha pluralidad de cebadores bloqueados comprende un grupo bloqueante en 3'; o
- 40 dicha pluralidad de cebadores bloqueados carece de 3' OH; o
- dicha pluralidad de cebadores específicos de locus comprende una secuencia específica para una o más regiones repetitivas de un genoma; o
- dichos cebadores específicos de locus comprenden secuencias que son específicas de una o más regiones de un genoma que normalmente están sobreamplificadas.

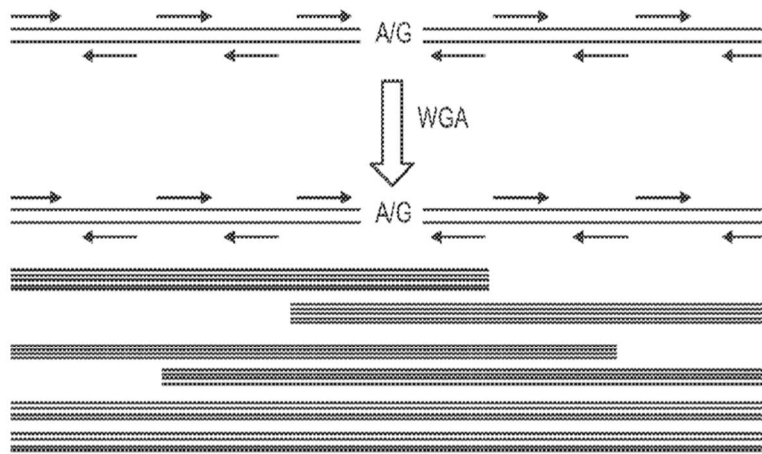


Fig. 1

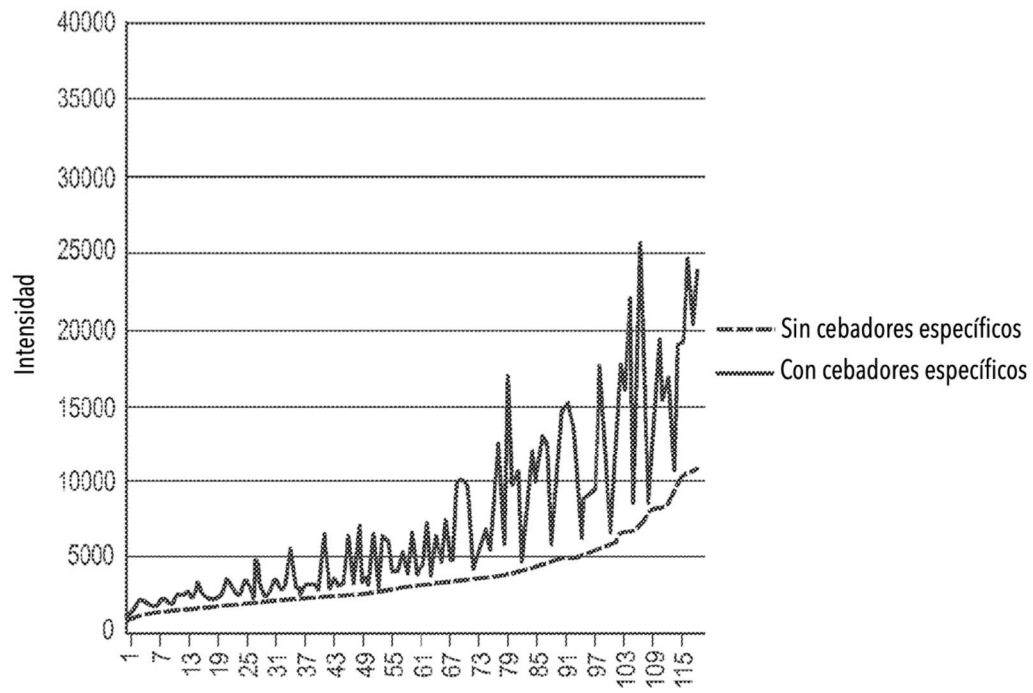


Fig. 2

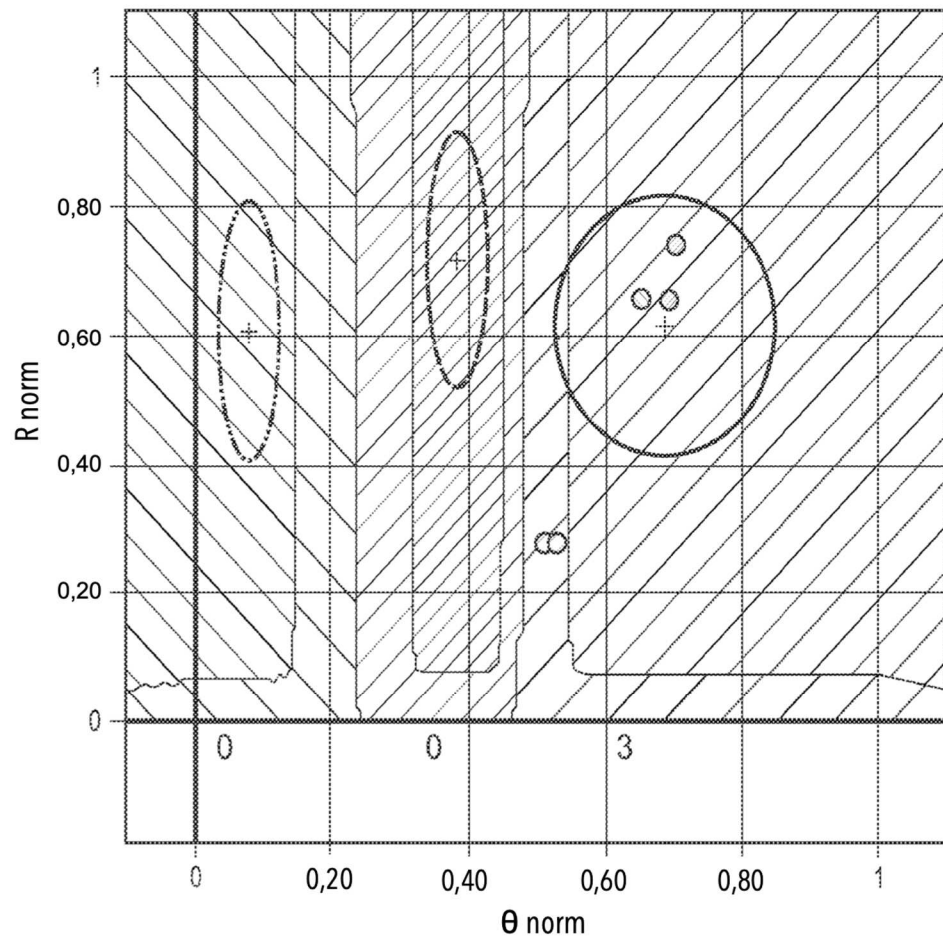


Fig. 3

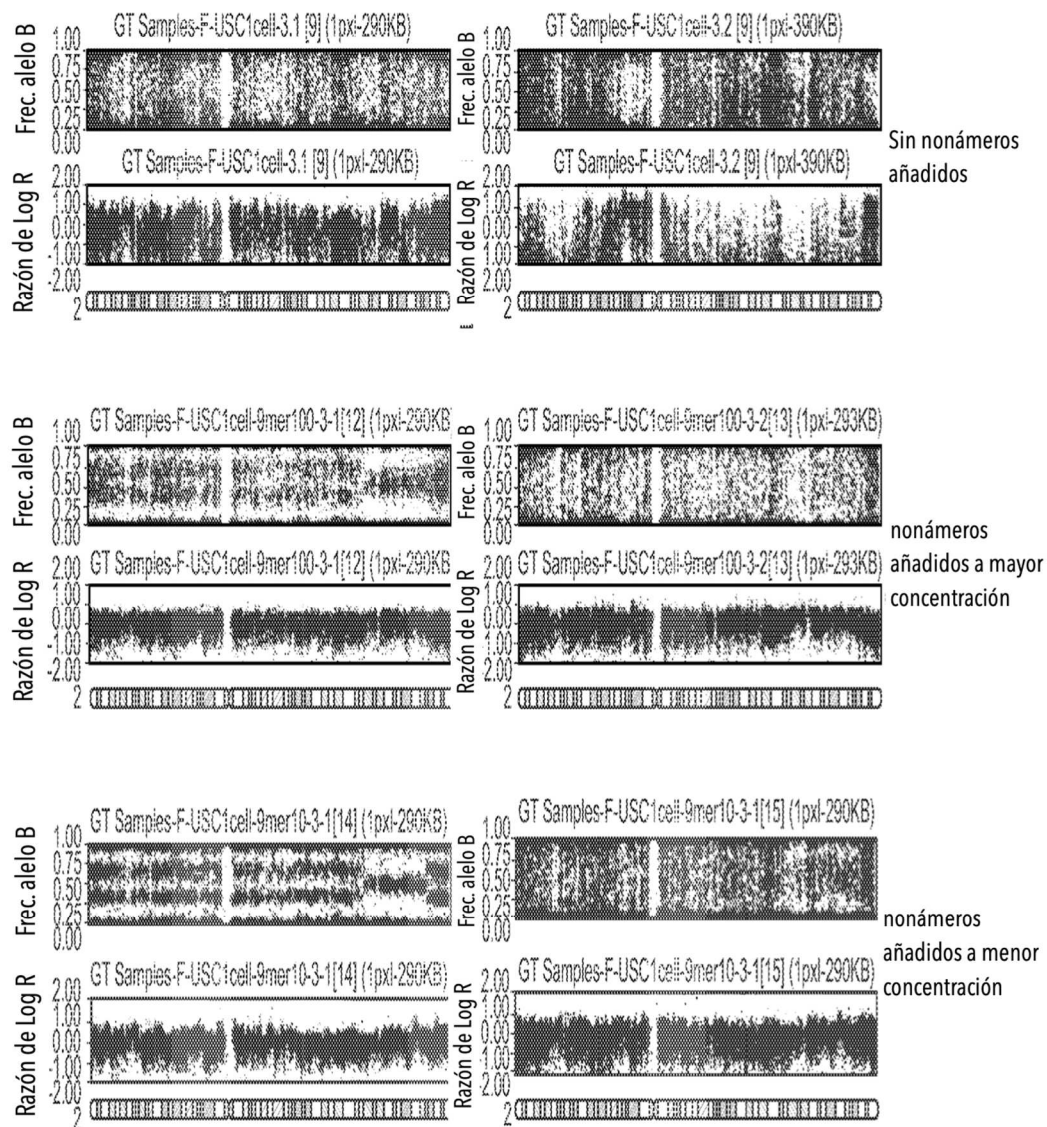


Fig. 4

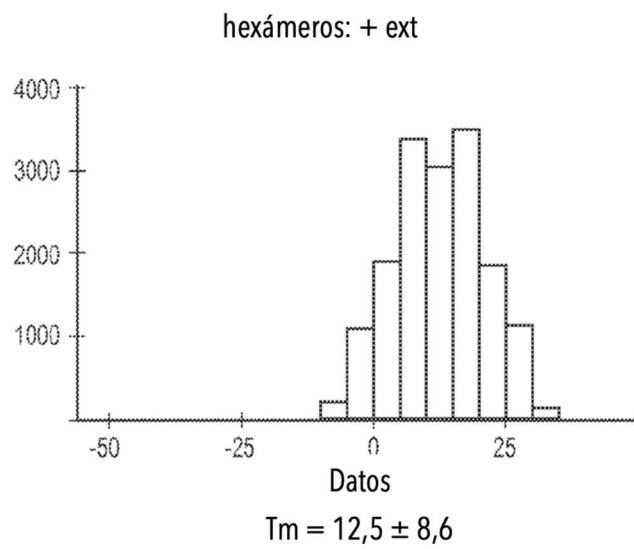
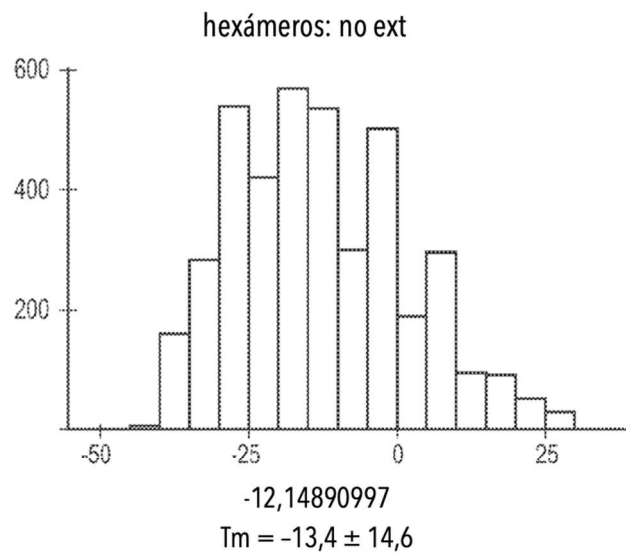


Fig. 5

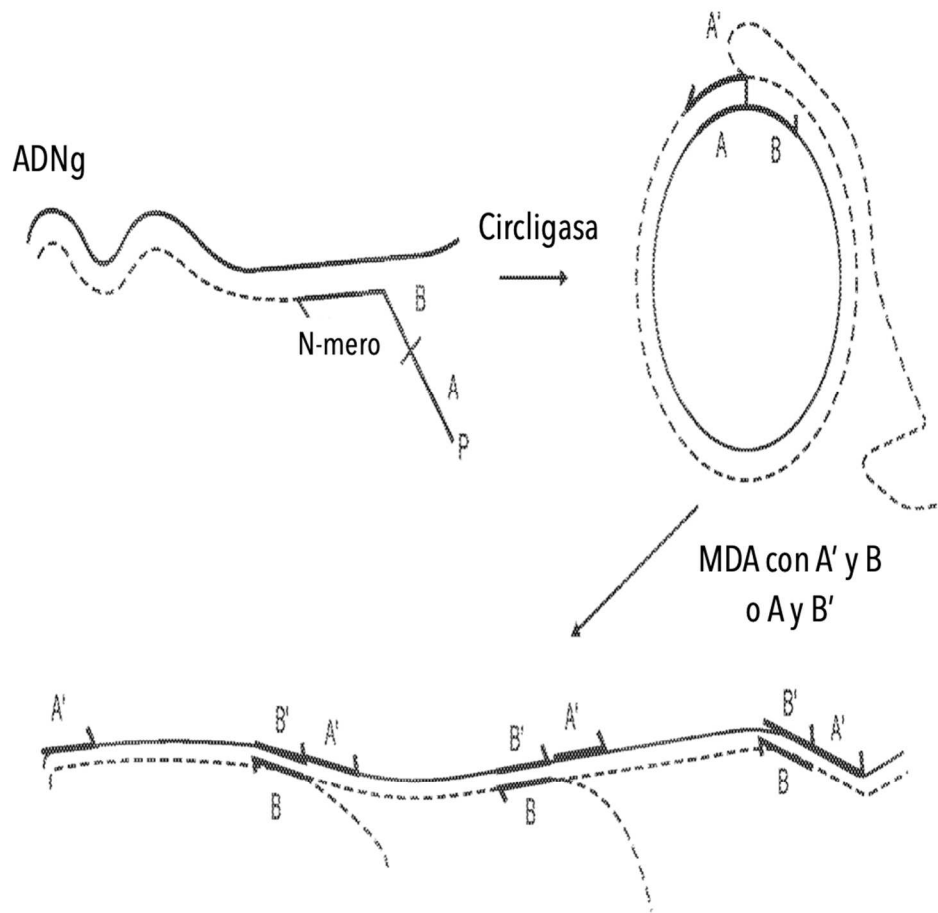


Fig. 6

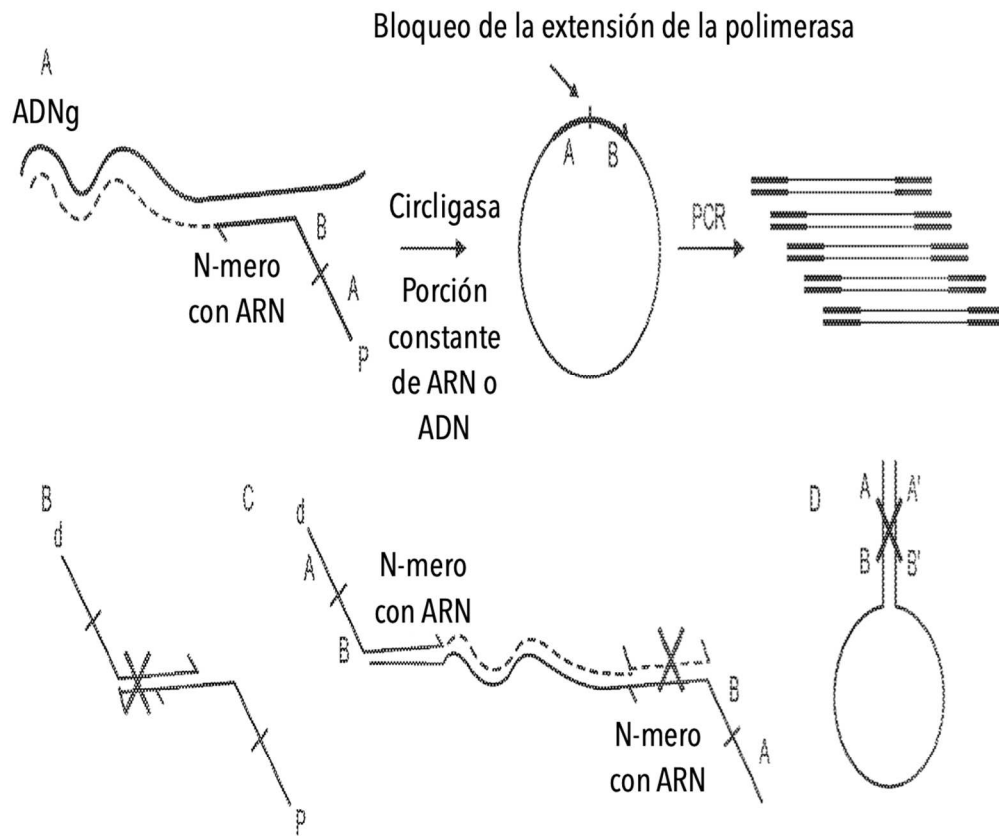


Fig. 7

Efecto de la composición de nonúmeros sobre el sesgo de la RPA

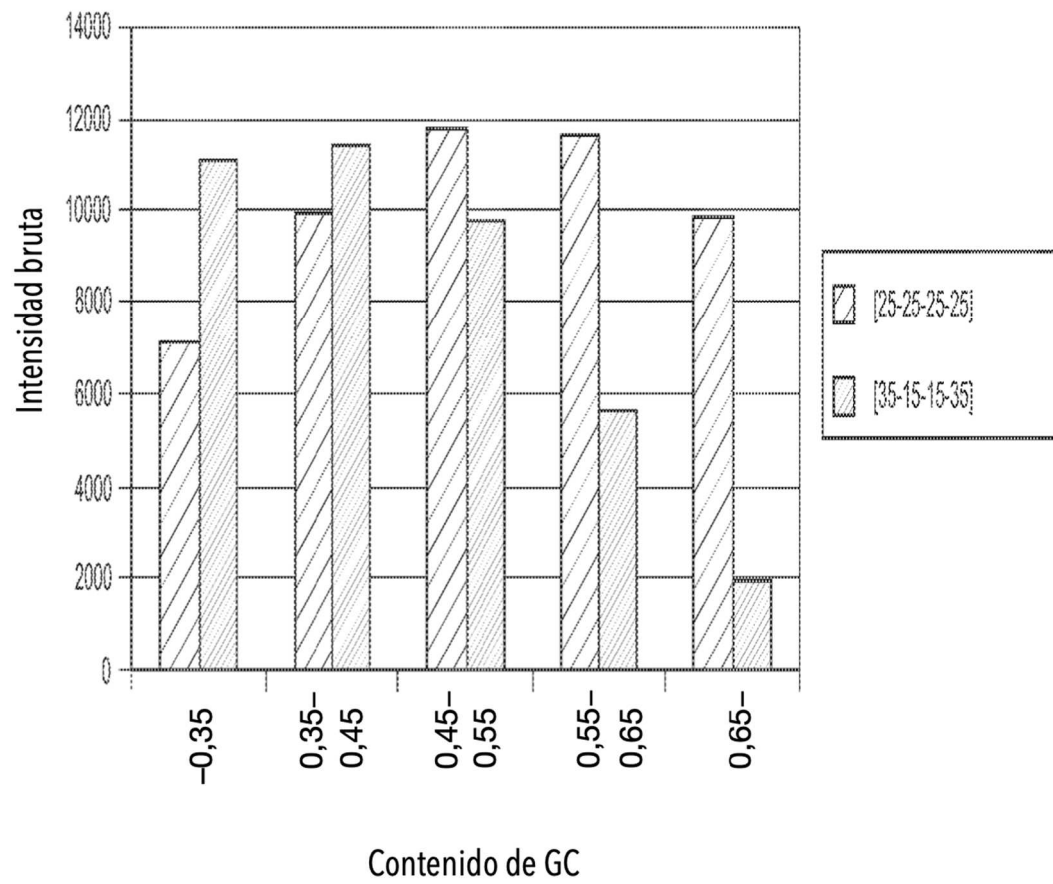


Fig. 8