

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 450 047**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **31.10.2006 E 06794950 (3)**

97 Fecha y número de publicación de la concesión europea: **11.12.2013 EP 1954818**

54 Título: **Método de preparación de bibliotecas de polinucleótidos molde**

30 Prioridad:

01.11.2005 GB 0522310
14.07.2006 US 486953

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
21.03.2014

73 Titular/es:

ILLUMINA CAMBRIDGE LIMITED (100.0%)
Chesterford Research Park Little Chesterford
Saffron Walden
Essex CB10 1XL, GB

72 Inventor/es:

GORMLEY, NIALL ANTHONY;
SMITH, GEOFFREY PAUL;
BENTLEY, DAVID y
RIGATTI, ROBERTO

74 Agente/Representante:

ZEA CHECA, Bernabé

ES 2 450 047 T3

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

DESCRIPCIÓN

Método de preparación de bibliotecas de polinucleótidos molde

5 **Campo de la invención**

10 **[0001]** La invención se refiere a un método de preparación de una biblioteca de polinucleótidos molde y también al uso de la biblioteca de moldes en métodos de amplificación de ácidos nucleicos en fase sólida. En particular, la invención se refiere a un método de preparación de una biblioteca de polinucleótidos molde que tienen secuencias comunes en sus extremos 5' y en sus extremos 3'.

Antecedentes de la invención

15 **[0002]** El desarrollo de fármacos de biología molecular y farmacológica actual hace un uso intensivo de análisis de ácidos nucleicos. Las áreas más conflictivas son las de secuenciación del genoma completo, detección de polimorfismo de un solo nucleótido, exploración y control de la expresión de genes, que típicamente requieren análisis de grandes cantidades de ácido nucleico.

20 **[0003]** Un área de la tecnología que revolucionó el estudio de los ácidos nucleicos fue el desarrollo de las técnicas de amplificación de ácidos nucleicos, tales como la reacción en cadena de la polimerasa (PCR). Las reacciones de amplificación, tales como PCR, pueden permitir al usuario amplificar específica y selectivamente un ácido nucleico diana particular de interés a partir de una mezcla compleja de ácidos nucleicos. Sin embargo, existe una necesidad constante de técnicas de amplificación de ácidos nucleicos que permitan la amplificación simultánea de mezclas complejas de moldes de diversa secuencia, tales como fragmentos de ADN genómico (por ejemplo, amplificación de "genoma completo") o bibliotecas de ADNc, en una única reacción de amplificación.

30 **[0004]** La amplificación por PCR no puede producirse en ausencia de hibridación de cebadores de amplificación directos e inversos para secuencias de unión a cebador en el molde a amplificar en las condiciones de las etapas de hibridación de la reacción PCR, es decir, si hay insuficiente complementariedad entre cebadores y molde. Un cierto conocimiento previo de la secuencia del molde se requiere por tanto antes de poder realizarse una reacción de PCR para amplificar un molde *específico*, salvo que se usen cebadores al azar con una pérdida consecuente de especificidad. El usuario debe normalmente conocer la secuencia de al menos los sitios de unión a cebador en el molde por adelantado para que puedan diseñarse los cebadores apropiados, aunque la secuencia restante del molde sea desconocida. La necesidad de conocimiento previo de la secuencia del molde aumenta la complejidad y el coste de la amplificación por PCR de mezclas complejas de moldes, tales como fragmentos de ADN genómico.

40 **[0005]** Los documentos WO 98/44151 y WO 00/18957 describen métodos de formación de series de polinucleótidos basadas en amplificación de ácidos nucleicos en "fase sólida", que es análoga a una reacción en cadena de la polimerasa en la que los productos de amplificación están inmovilizados sobre un soporte sólido para formar series comprendidas por agrupaciones de ácido nucleico o "colonias". Cada agrupación o colonia de dicha serie se forma a partir de una pluralidad de cadenas de polinucleótidos idénticas inmovilizadas y una pluralidad de cadenas de polinucleótidos complementarias inmovilizadas idénticas. Las series así formadas se denominan generalmente en el presente documento "series agrupadas" y sus características generales se entenderán haciendo referencia a los documentos WO 98/44151 o WO 00/18957.

50 **[0006]** Como se ha indicado anteriormente, los métodos de amplificación en fase sólida de los documentos WO 98/44151 y WO 00/18957 son esencialmente una forma de la reacción en cadena de la polimerasa realizada sobre un soporte sólido. Como cualquier reacción PCR, estos métodos requieren el uso de cebadores de amplificación directos e inversos (que pueden ser idénticos o diferentes) que pueden hibridarse con un molde a amplificar. En los métodos de los documentos WO 98/44151 y WO 00/18957 ambos cebadores se inmovilizan sobre el soporte sólido en el extremo 5'. Se conocen otras formas de amplificación en fase sólida en las que solo un cebador está inmovilizado y el otro está presente en la solución libre (Mitra, R. D y Church, G. M., Nucleic Acids Research, 1999, Vol.27, N° 24).

55 **[0007]** En común con todas las técnicas de PCR, la amplificación por PCR en fase sólida requiere el uso de cebadores de amplificación directos e inversos que incluyen secuencias de nucleótidos "específicas de molde" que pueden hibridarse con secuencias en el molde a amplificar, o su complemento, en condiciones de las etapas de hibridación de la reacción PCR. Las secuencias en el molde con el que los cebadores se hibridan en condiciones de la reacción PCR pueden denominarse en el presente documento secuencias "de unión a cebador".

60 **[0008]** Determinadas realizaciones de los métodos descritos en los documentos WO 98/44151 y WO 00/18957 utilizan cebadores "universales" para amplificar moldes que comprenden una parte molde variable que se desea amplificar flanqueada en 5' y 3' por secuencias de unión a cebador común o "universal". Los cebadores directo e inverso "universales" incluyen secuencias que pueden hibridarse con las secuencias de unión a cebador "universal" en la construcción molde. La parte molde variable puede en sí misma ser de secuencia conocida, desconocida o

parcialmente desconocida. Esta estrategia tiene la ventaja de que no es necesario diseñar un par específico de cebadores para cada molde a amplificar; pueden utilizarse los mismos cebadores para la amplificación de diferentes moldes siempre que cada molde se modifique por la adición de las mismas secuencias de unión a cebador universal en sus extremos 5' y 3'. La secuencia molde variable puede por lo tanto ser cualquier fragmento de ADN de interés.

5 Una estrategia análoga puede usarse para amplificar una mezcla de moldes, tal como una pluralidad o biblioteca de moléculas de ácido nucleico molde (por ejemplo, fragmentos de ADN genómico), usando un solo par de cebadores directos e inversos universales siempre que cada molécula molde en la mezcla se modifique por la adición de las mismas secuencias de unión de cebador universal.

10 **[0009]** Dichas estrategias de "cebador universal" para amplificación por PCR, y en particular amplificación por PCR en fase sólida, son ventajosas ya que permiten que puedan realizarse múltiples moléculas molde de secuencia igual o diferente, conocida o desconocida, a amplificar en una única reacción de amplificación, que puede realizarse en un soporte sólido que lleva un solo par de cebadores "universales". La amplificación simultánea de una mezcla de moldes de diferentes secuencias por PCR requeriría por otro lado una pluralidad de pares de cebadores, siendo
15 cada par complementario a cada molde exclusivo en la mezcla. La generación de una pluralidad de pares de cebadores para cada molde individual no es una acción viable para mezclas complejas de moldes.

[0010] La adición de secuencias de cebador universal sobre los extremos de los moldes a amplificar por PCR puede conseguirse mediante diversos métodos conocidos por los expertos en la técnica. Por ejemplo, un cebador universal que consiste en una secuencia universal en su extremo 5' y una secuencia degenerada en su extremo 3' puede usarse en una PCR (DOP-PCR, por ejemplo PNAS 1996 vol. 93 páginas 14676-14679) para amplificar fragmentos al azar a partir de un molde complejo o una mezcla compleja de moldes. La parte 3' degenerada del cebador se hibrida en posiciones al azar sobre ADN y puede extenderse para generar una copia del molde que tiene la secuencia universal en su extremo 5'.
20

[0011] Como alternativa, pueden ligarse adaptadores que contengan secuencias de cebador universal en los extremos de los moldes. Los adaptadores pueden ser mono o bicatenarios. Si son bicatenarios, estos pueden tener extremos solapantes que son complementarios a los extremos solapantes de las moléculas molde que se han generado con una endonucleasa de restricción. Como alternativa, los adaptadores bicatenarios pueden ser romos, en cuyo caso los moldes también tienen extremos romos. Los extremos romos de los moldes pueden haberse formado durante un proceso para cizallar el ADN en fragmentos, o pueden haberse formado mediante una reacción de reparación de extremos, como conocen los expertos en la técnica.
25
30

[0012] Pueden usarse un solo adaptador o dos adaptadores diferentes en una reacción de ligación con moldes. Si un molde se ha manipulado de tal manera que sus extremos son iguales, es decir, ambos son romos o ambos tienen el mismo saliente, entonces la ligación de un solo adaptador compatible generará un molde con ese adaptador en ambos extremos. Sin embargo, si se usan dos adaptadores compatibles, adaptador A y adaptador B, entonces se forman tres permutaciones de productos ligados: molde con adaptador A en ambos extremos, molde con adaptador B en ambos extremos, y molde con adaptador A en un extremo y adaptador B en el otro extremo. Este último producto es, en algunas circunstancias, el único producto deseado de la reacción de ligación y por consiguiente se requieren etapas de purificación adicionales después de la reacción de ligación para purificarlo de los productos de ligación que tengan el mismo adaptador en ambos extremos.
35
40

[0013] La presente invención que se presenta en el presente documento es un método que usa un solo adaptador en una reacción de ligación para generar una biblioteca de polinucleótidos molde, cada uno de los cuales tiene secuencias de cebador universal comunes, pero diferentes, en sus extremos 5' y 3'. El método puede aplicarse para preparar mezclas simples o complejas de moldes para amplificación, por ejemplo, una superficie sólida, usando secuencias de cebador, sin conocimiento previo de secuencias molde. La invención puede aplicarse para la preparación de moldes a partir de muestras complejas tales como genomas completos o mezclas de ADNc, así como aplicaciones monomolde.
45
50

Sumario de la invención

[0014] La invención proporciona un método como se define en la reivindicación 1. Se proporciona un método para generar una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3', comprendiendo el método:
55

ligar polinucleótidos adaptadores con apareamiento erróneo idénticos en los dos extremos de cada uno de uno o más dúplex de polinucleótido diana para formar una o más construcciones adaptador-diana, en el que cada adaptador con apareamiento erróneo se forma a partir de dos cadenas polinucleotídicas hibridadas que forman un complejo biomolecular que comprende al menos una región bicatenaria y una región no apareada, y realizar una reacción de extensión de cebador inicial en la que un oligonucleótido cebador se hibrida con una parte de adaptador de cada una de las construcciones adaptador-diana y se extiende por adición secuencial de nucleótidos para formar productos de extensión complementarios a al menos una cadena de cada una de las construcciones adaptador-diana,
60
65

en el que los productos de extensión, y opcionalmente productos de amplificación derivados de los mismos, proporcionan en su conjunto una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3'; en el que los dúplex de polinucleótido diana que van a ligarse son una mezcla compleja de fragmentos de ADN genómico que representan un genoma completo o sustancialmente completo, y en el que la biblioteca de moléculas polinucleotídicas molde generadas es representativa del genoma completo o sustancialmente completo.

[0015] Un segundo aspecto de la invención refiere al uso de una biblioteca de moléculas polinucleotídicas molde preparadas de acuerdo con el método del primer aspecto de la invención como un molde para la amplificación por PCR en fase sólida. Por tanto, en una realización particular la invención proporciona un método de amplificación de ácidos nucleicos en fase sólida de moléculas polinucleotídicas molde que comprende:

preparar una biblioteca de moléculas polinucleotídicas de molde que tienen secuencias comunes en sus extremos 5' y 3' utilizando el procedimiento de acuerdo con el primer aspecto de la invención y realizar una amplificación de ácidos nucleicos en fase sólida, en el que dichas moléculas polinucleotídicas molde se amplifican.

[0016] Un tercer aspecto de la invención se refiere al uso de una biblioteca de moléculas polinucleotídicas molde preparadas de acuerdo con el método del primer aspecto de la invención como un molde para la amplificación del genoma completo. Por lo tanto, en una realización particular la invención proporciona un método de amplificación del genoma completo que comprende:

utilizar el método de acuerdo con el primer aspecto de la invención para preparar una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y 3' comenzando a partir de una mezcla compleja de fragmentos de genoma completo y realizar una reacción de amplificación de genoma completo en el que dichas moléculas polinucleotídicas molde se amplifican.

[0017] En un cuarto aspecto la invención proporciona un kit para su uso en la preparación de una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y 3' en el que la secuencia común en el extremo 5' de cada molde individual en la biblioteca no es idéntica y no complementaria del todo a la secuencia común en el extremo 3' de dicho molde, comprendiendo el kit polinucleótidos adaptadores de apareamiento erróneo y cebadores como se define en la reivindicación 15 que pueden hibridarse con los polinucleótidos adaptadores con apareamiento erróneo.

Breve descripción de los dibujos

[0018]

La Figura 1 ilustra diversos ejemplos de adaptadores con apareamiento erróneo bifurcados para su uso en el método de la invención, específicamente representan diferentes estructuras de extremo como o solapantes permisibles en el extremo "ligable" del adaptador. La Figura 1 (e) ilustra esquemáticamente los componentes de secuencia de dos cadenas parcialmente complementarias (denominadas oligo A y oligo B) que forman el adaptador bifurcado cuando están hibridadas. El extremo 5' del oligo B es complementario (COMP) a una parte de la secuencia SEC CEBADOR en el oligo A. El oligo A incluye un solo saliente de nucleótido "T" en el extremo 3'. El extremo 5' del oligo A está fosforilado. P representa un grupo fosfato terminal; X e Y representan funcionalidades de captura de superficie.

La Figura 2 ilustra una realización del método de la invención basado en el uso de los adaptadores bifurcados ilustrados en la Figura 1. La Figura 2(a) representa las etapas de fragmentación de una muestra compleja tal como ADN genómico para generar una pluralidad de fragmentos dúplex diana, ligación de fragmentos dúplex diana con adaptadores con apareamiento erróneo (bifurcados) para generar construcciones adaptador-molde y eliminar adaptadores no unidos. El adaptador bifurcado incluye un grupo biotina en el extremo 5' que no está ligado al fragmento diana para facilitar la captura en fase sólida de las construcciones adaptador-diana, por ejemplo, sobre perlas magnéticas de estreptavidina. La Figura 2(b) representa una reacción de extensión con cebador inicial en la que los cebadores se hibridan con regiones de adaptador con apareamiento erróneo en cada cadena de una construcción adaptador-diana y se extienden para generar productos de extensión complementarios a cada cadena de la construcción adaptador-diana. Para simplificar la ligación y las etapas de extensión con cebador se ilustran para una sola construcción adaptador-diana.

La Figura 3 ilustra una realización alternativa de la invención en la que las construcciones adaptador-diana se someten a rondas múltiples de hibridación con cebador y extensión para generar copias múltiples monocatenarias de cada construcción adaptador-diana. Para simplificar, las etapas de extensión con cebador se ilustran para una sola construcción adaptador-diana.

La Figura 4 ilustra otra realización adicional de la invención en la que las construcciones adaptador-diana se someten a amplificación por PCR para generar múltiples copias bicatenarias de cada construcción adaptador-diana. Para simplificar, la amplificación por PCR se ilustra para una sola construcción adaptador-diana.

5 La Figura 5 ilustra una realización de la invención, que representa etapas de fragmentación de una muestra compleja tal como ADN genómico para generar una pluralidad de fragmentos diana, ligación de los fragmentos diana con adaptadores con apareamiento erróneo (bifurcados) para generar construcción adaptador-molde y posteriormente eliminar los adaptadores no unidos, en el que los adaptadores no incluyen un grupo biotina en el extremo 5'. Las construcciones adaptador-diana resultantes pueden someterse a amplificación por PCR para
10 generas múltiples copias bicatenarias de cada construcción adaptador-diana. Para simplificar las etapas de ligación se ilustran para una sola construcción adaptador-diana.

La Figura 6 ilustra ejemplos adicionales de adaptadores con apareamiento erróneo bifurcados para su uso, en el método de la invención, representando de nuevo los formatos permisibles como y solapantes en el extremo "ligable" del adaptador". La Figura 6(e) ilustra esquemáticamente las secuencias componentes presentes en las dos cadenas (denominadas oligo C y oligo B) que forman el adaptador cuando están hibridadas. P representa un grupo fosfato terminal; X e Y representan funcionalidades de captura superficiales.

20 La Figura 7 ilustra una realización adicional más de la invención basada en el uso de los adaptadores bifurcados ilustrados en la Figura 6. La Figura 7(a) representa etapas de fragmentación y ligación sustancialmente similares a las ilustradas en la Figura 5. La Figura 7(b) representa posterior amplificación por PCR utilizando cebadores PCR "con cola" que tienen secuencias terminales 3' complementarias a una secuencia del extremo 5' del adaptador, e ilustra esquemáticamente la composición de secuencias de los productos de amplificación bicatenarios formados en la reacción PCR. Para simplificar, las etapas de ligación y amplificación por PCR se
25 ilustran para una sola construcción adaptador-diana.

La Figura 8 ilustra realizaciones alternativas de adaptadores con apareamiento erróneo para su uso en el método de la invención. P representa un grupo fosfato terminal; X e Y representan funcionalidades de captura superficiales; X y Z representan modificaciones para impedir la ligación.

30 La Figura 9a ilustra una realización adicional más de la invención basada en el uso de adaptadores alternativos ilustrados en la Figura 8. La Figura 9(a) representa la fragmentación, ligación y posterior retirada de adaptadores no unidos. La Figura 9(b) representa la hibridación de cebadores de amplificación idénticos con una región bicatenaria del adaptador en cada cadena de la construcción adaptador-diana. Las construcciones adaptador-diana pueden amplificarse mediante PCR usando esta única especie de cebador. Para simplificar las etapas de ligación e hibridación con cebador se ilustran para una sola construcción adaptador-diana.

Descripción detallada de la invención

40 **[0019]** La invención proporciona un método para generar una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y 3'. En este contexto, el término "común" se interpreta que significa común a todos los moldes en la biblioteca. Como se explica adicionalmente con detalle a continuación, todos los moldes dentro de la biblioteca contendrán regiones de secuencia común en (o próximas a) sus extremos 5' y 3', no siendo idéntica la secuencia común en el extremo 5' de cada molde individual en la biblioteca y no complementaria
45 del todo a una secuencia común en el extremo 3' de dicho molde.

[0020] El término "biblioteca" se refiere únicamente a una colección o pluralidad de moléculas molde que comparten secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3'. El uso del término "biblioteca" para referirse a una colección o pluralidad de moléculas molde no debería considerarse que implica que los moldes que constituyen la biblioteca deriven de una fuente particular, o que la "biblioteca" tenga una composición particular. A modo de ejemplo, el uso del término "biblioteca" no debe considerarse que implica que los moldes individuales dentro de la biblioteca deban ser de diferente secuencia de nucleótidos o que los moldes estén relacionados en cuanto a secuencia y/o fuente.

55 **[0021]** En sus diversas realizaciones la invención incluye la formación de bibliotecas "complejas" en las que muchas, si no todas, de las moléculas molde individuales comprenden diferentes secuencias diana (como se define más adelante), aunque todas compartan secuencias comunes en sus extremos 5' y 3'. Dichas bibliotecas molde complejas pueden prepararse usando el método de la invención comenzando a partir de una mezcla compleja de polinucleótidos diana tal como (pero sin limitación) fragmentos de ADN genómico al azar, bibliotecas de ADNc, etc.
60 En realizaciones preferidas más del 50%, o más del 60%, o más del 70%, o más del 80%, o más del 90%, o más del 95% de los moldes polinucleotídicos individuales en una biblioteca compleja puede comprender diferentes secuencias diana, aunque todos los moldes en una biblioteca determinada compartirán secuencia común en sus extremos 5' y una secuencia común en sus extremos 3'.

65

[0022] El uso del término “molde” para referirse a moléculas polinucleotídicas individuales en la biblioteca indica meramente que una o dos cadenas de los polinucleótidos en la biblioteca pueden actuar como moldes para la polimerización de ácidos nucleicos dependiente de molde catalizada por una polimerasa. El uso de este término no debe considerarse como limitante del alcance de la invención con respecto a bibliotecas de polinucleótidos que se usan en realidad como moldes en una reacción de polimerización catalizada por enzimas posterior.

[0023] La biblioteca se forma ligando moléculas polinucleotídicas idénticas adaptadoras idénticas (“adaptadores con apareamiento erróneo”, cuyas características generales se refieren a continuación) en los extremos 5' y 3' de uno o más dúplex polinucleotídicos diana (que pueden ser de secuencia conocida, parcialmente conocida o desconocida) para formar construcciones adaptador-diana y después realizar una reacción de extensión inicial con cebador en la que se forman productos de extensión complementarios a ambas cadenas de cada construcción adaptador-diana individual. Los productos de extensión con cebador resultantes y opcionalmente copias amplificadas de los mismos, proporcionan en su conjunto una biblioteca de polinucleótidos molde.

[0024] Cada cadena de cada molécula molde en la biblioteca formada en la reacción de extensión con cebador tendrá por lo tanto la siguiente estructura, cuando se observa como una sola cadena:

$$5' \text{-[secuencia común I]-[secuencia diana]-[secuencia común II]-3'}$$

representando “secuencia común I” una secuencia derivada de copiar una primera cadena del adaptador con apareamiento erróneo y siendo común a todas las moléculas molde en la biblioteca generada en la reacción de extensión con cebador inicial;

“diana” representa una secuencia derivada de una cadena del dúplex polinucleotídico diana y puede ser diferente en moléculas molde individuales diferentes dentro de la biblioteca; y

“secuencia común II” representa una secuencia derivada de copiar una segunda cadena de adaptador con apareamiento erróneo y es también común a todas las moléculas molde en la biblioteca generada en la reacción de extensión con cebador inicial.

[0025] Puesto que “secuencia común I” y “secuencia común II” son comunes a todas las cadenas molde en la biblioteca pueden incluir secuencias de unión a cebador “universal”, que permiten que todos los moldes en la biblioteca se amplifiquen finalmente en un procedimiento de PCR en fase sólida usando cebadores universales.

[0026] Esta es una característica clave de la invención, sin embargo, las secuencias comunes terminales 5' y 3' denominadas “secuencia común I” y “secuencia común II” no son *completamente* complementarias entre sí, lo que significa que cada cadena molde individual puede contener diferentes secuencias de cebador universal (y no complementarias) en sus extremos 5' y 3'.

[0027] Esto es generalmente ventajoso para bibliotecas complejas de moldes a amplificar por PCR (por ejemplo amplificación del genoma completo) ya se realice en solución o en un soporte sólido, incluir regiones de secuencia “diferente” en sus extremos 5' y 3', que sin embargo son comunes a todas las moléculas molde en la biblioteca, especialmente si los productos de amplificación se secuencian finalmente. Por ejemplo, la presencia de una secuencia única común en un extremo solamente de cada molde en la biblioteca puede proporcionar un sitio de unión para un cebador de secuenciación, lo que permite que una cadena de cada molde en la forma amplificada de la biblioteca se secuencie en una sola reacción de secuenciación usando un solo tipo de cebador de secuenciación.

[0028] Típicamente, “secuencia común I” y “secuencia común II” consistirá en no más de 100, o no más de 50 o no más de 40 nucleótidos consecutivos en los extremos 5' y 3', respectivamente, de cada cadena de cada polinucleótido molde. La longitud exacta de las dos secuencias puede o no ser idéntica. Las secuencias de nucleótidos de la “secuencia común I” y “secuencia común II” en los polinucleótidos molde se determinará en parte por las secuencias de las cadenas adaptadoras ligadas a los polinucleótidos diana y en parte por la secuencia del cebador usado en la reacción de extensión con cebador inicial y cualquier ronda posterior de amplificación de ácidos nucleicos.

[0029] En realizaciones en las que el producto de extensión con cebador inicial se somete a amplificación adicional mediante PCR convencional, los productos de la reacción de amplificación serán polinucleótidos bicatenarios, una cadena de los cuales tiene la estructura:

$$5' \text{-[secuencia común I]-[secuencia diana]-[secuencia común II]-3'}$$

[0030] Se apreciará que “secuencia común II” en los productos de amplificación pueden diferir algo con respecto a la “secuencia común II” presente en los productos de la región de extensión con cebador inicial, ya que esta última se determinará en parte por la secuencia del cebador de PCR usado para cebar la síntesis de una cadena polinucleotídica complementaria al producto de extensión cebador inicial, mientras que la última se determinará exclusivamente copiando las secuencias adaptadoras en los extremos 3' de las construcciones adaptador-molde en la extensión con cebador inicial. Sin embargo, dado que el cebador de PCR se diseña para hibridarse con una

secuencia en los productos de extensión iniciales que es complementaria al adaptador 3', las dos formas de "secuencia común II" contendrán idéntica secuencia, al menos en el extremo 3'. Puede incluirse una secuencia adicional en el extremo 5' de la "secuencia común II" en los productos amplificados, por ejemplo mediante el uso de cebadores de PCR "con cola" como se describe con detalle más adelante. En otras realizaciones las secuencias comunes presentes en los productos de amplificación pueden realmente ser más cortas que las secuencias comunes incluyendo en los adaptadores originalmente ligados a la diana.

[0031] Las secuencias de nucleótidos exactas de las regiones comunes de las moléculas molde en la biblioteca no son generalmente esenciales para la invención y el usuario puede seleccionarlas. Las secuencias comunes deben comprender al menos secuencias "de unión a cebador" que permiten la hibridación específica de cebadores de amplificación cuando se usan los moldes en una reacción de amplificación en fase sólida. Las secuencias de unión a cebador se determinan por tanto mediante la secuencia de los cebadores que va a usarse finalmente para la amplificación en fase sólida. La secuencia de estos cebadores a su vez se selecciona ventajosamente para impedir o minimizar la unión de los cebadores con las partes diana de los moldes dentro de la biblioteca en las condiciones de la reacción de amplificación, pero no está particularmente de otro modo. A modo de ejemplo, si las partes diana de los moldes derivan de ADN genómico humano, entonces las secuencias de los cebadores a usar en amplificación en fase sólida idealmente deben seleccionarse para minimizar la unión no específica a cualquier secuencia genómica humana.

[0032] Los polinucleótidos adaptadores utilizados en el método de la invención se denominan en el presente documento adaptadores "con apareamiento erróneo" porque, como se explicará con detalle en el presente documento, es esencial que los adaptadores incluyan una región de apareamiento erróneo de secuencia, es decir, no deben formarse por hibridación de cadenas polinucleotídicas completamente complementarias.

[0033] Los adaptadores con apareamiento erróneo para su uso en la invención se forman hibridando dos cadenas polinucleotídicas parcialmente complementarias para proporcionar, cuando las dos cadenas están hibridadas, al menos una región bicatenaria y al menos una región no apareada.

[0034] La "región bicatenaria" del adaptador es una región bicatenaria corta, comprendiendo típicamente 5 o más pares de bases consecutivos, formados por hibridación de las dos cadenas polinucleotídicas parcialmente complementarias. Este término simplemente se refiere a una región bicatenaria de ácido nucleico en la que las dos cadenas están hibridadas y no implica ninguna conformación estructural particular.

[0035] Generalmente es ventajoso que la región bicatenaria sea lo más corta posible sin pérdida de función. En este contexto "función" se refiere a que la región bicatenaria forma un dúplex estable en condiciones de reacción convencionales para una reacción del ligación de ácidos nucleicos catalizada por enzimas, que conocerá bien el experto lector (por ejemplo incubación a una temperatura en el intervalo de 4 °C a 25 °C en un tampón de ligación apropiado para la enzima), de tal manera que las dos cadenas que forman el adaptador continúen parcialmente hibridadas durante la ligación del adaptador a una molécula diana. No es absolutamente necesario que la región bicatenaria sea estable en condiciones típicamente usadas en las etapas de hibridación de la extensión con cebador o reacciones PCR.

[0036] Dado que se ligan adaptadores idénticos en ambos extremos de cada molécula molde, la secuencia diana en cada construcción adaptador-diana estará flanqueada por secuencias complementarias derivadas de la región bicatenaria de los adaptadores. Cuanto más larga sea la región bicatenaria, y por tanto las secuencias complementarias derivadas de la misma en las construcciones adaptador-diana, mayor será la posibilidad de que la construcción adaptador-diana sea capaz de volverse a plegar y formar pares de bases consigo misma en estas regiones de autocomplementariedad interna en las condiciones de hibridación utilizadas en la extensión con cebador y/o PCR. Generalmente se prefiere que la región bicatenaria tenga 20 o menos, 15 o menos, o 10 o menos pares de bases de longitud para reducir este efecto. La estabilidad de la región bicatenaria puede aumentarse, y por tanto su longitud posiblemente reducirse, mediante la inclusión de nucleótidos no naturales que presentan formaciones de pares de bases más fuertes que los pares de bases de Watson-Crick convencionales.

[0037] Se prefiere, pero no es absolutamente esencial que las dos cadenas del adaptador tengan una complementariedad del 100% en la región bicatenaria. Se apreciará que pueden tolerarse uno o más apareamientos erróneos de nucleótidos dentro de la región bicatenaria, siempre que las dos cadenas sean capaces de formar un dúplex estable en condiciones de ligación convencionales.

[0038] Los adaptadores para su uso en la invención generalmente incluirán una región bicatenaria adyacente al extremo "ligable" del adaptador, es decir el extremo que está unido a un polinucleótido diana en la reacción de ligación. El extremo ligable del adaptador puede ser romo o, en otras realizaciones, salientes en 5' o 3' cortos de uno o más nucleótidos presentes para facilitar/promover la ligación. El nucleótido terminal 5' en el extremo ligable del adaptador debe estar fosforilado para permitir la unión fosfodiéster a un grupo hidroxilo 3' en el polinucleótido diana.

[0039] La expresión “región no apareada” se refiere una región del adaptador en el que las secuencias de las dos cadenas polinucleotídicas que forman el adaptador presentan un grado de no complementariedad de tal manera que las dos cadenas no pueden hibridarse entre sí en condiciones de hibridación convencionales para una extensión con cebador o reacción PCR. Las dos cadenas en la región no apareada pueden presentar cierto grado de hibridación en condiciones de reacción convencionales para una reacción de ligación catalizada por enzimas, siempre que las dos cadenas vuelvan a una forma monocatenaria en condiciones de hibridación.

[0040] Las condiciones encontradas durante las etapas de hibridación de una reacción PCR son generalmente conocidas por un experto en la técnica, aunque las condiciones de hibridación exactas variarán de reacción a reacción (véase Sambrook *et al.*, 2001, Molecular Cloning, A Laboratory Manual, 3ª Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel *et al.*). Típicamente dichas condiciones pueden comprender, pero sin limitación, (después de una etapa de desnaturalización a una temperatura de aproximadamente 94 °C durante aproximadamente 1 minuto) exposición a una temperatura en el intervalo de desde 40 °C hasta 72 °C (preferentemente 50-68 °C) durante un periodo de aproximadamente 1 minuto en un tampón de reacción PCR convencional.

[0041] Pueden utilizarse diferentes condiciones de hibridación para una sola reacción de extensión con cebador que no forma parte de una reacción PCR (de nuevo véase Sambrook *et al.*, 2001, Molecular Cloning, A Laboratory Manual, 3ª Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel *et al.*). Las condiciones para la hibridación con cebador en una sola extensión con cebador incluyen, por ejemplo, exposición a una temperatura en el intervalo de 30 a 37 °C en un tampón de extensión con cebador convencional. Se apreciará que diferentes enzimas, y por tanto diferentes tampones de reacción, pueden usarse para una sola reacción de extensión con cebador contrariamente a lo que ocurre con una reacción PCR. No es un requisito el uso de una polimerasa termoestable para una sola reacción de extensión con cebador.

[0042] Debe entenderse que la “región no apareada” se proporciona por diferentes partes de las mismas dos cadenas polinucleotídicas que forman la región (o regiones) bicatenaria(s). Sin embargo, las partes de las dos cadenas que forman la región no apareada no están hibridadas en condiciones en las que otras partes de las mismas dos cadenas se hibridan para formar una o más regiones bicatenarias. Para evitar la duda debe entenderse que un saliente de base única o monocatenario en el extremo 5' o 3' de un dúplex polinucleotídico no constituye una “región no apareada” en el contexto de la presente invención.

[0043] Las partes de las dos cadenas que forman la región bicatenaria típicamente comprenden al menos 10, o al menos 15 o al menos 20 nucleótidos consecutivos en cada cadena. El límite inferior sobre la longitud de la región no apareada típicamente se determinará por la función, por ejemplo, la necesidad de proporcionar una secuencia adecuada para la unión de un cebador para la extensión con cebador, PCR y/o secuenciación. Teóricamente no existe límite superior sobre la longitud de la región no apareada, excepto que generalmente sea ventajoso minimizar la longitud completa del adaptador, por ejemplo para facilitar la separación de adaptadores no unidos de construcciones adaptador-diana después de la etapa de ligación. Por lo tanto, se prefiere que la región no apareada sea menor de 50, o menor de 40, o menor de 30 o menor de 25 nucleótidos consecutivos de longitud en cada cadena.

[0044] La longitud total de las dos cadenas que forman el adaptador típicamente estará en el intervalo de desde 25 hasta 100 nucleótidos, más típicamente desde 30 hasta 55 nucleótidos.

[0045] Las partes de las dos cadenas que forman la región no apareada deben preferentemente ser de similar longitud, aunque no es absolutamente esencial, siempre que la longitud de cada parte sea suficiente para cumplir la función deseada (por ejemplo unión a cebador). Los inventores han demostrado mediante experimentos que las partes de las dos cadenas que forman la región no apareada pueden diferir hasta en 25 nucleótidos sin afectar negativamente a la función del adaptador.

[0046] Con la mayor preferencia las partes de las dos cadenas polinucleotídicas que forman la región no apareada se encontrarán completamente con apareamiento erróneo, o tendrán un 100% de no complementariedad. Sin embargo, los lectores expertos apreciarán que algunas “coincidencias” de secuencia, es decir un menor grado de no complementariedad puede tolerarse en esta región sin afectar la función en un grado esencial. Como se ha indicado anteriormente, el grado de apareamiento erróneo de secuencia o no complementariedad de secuencia debe ser tal que las dos cadenas en la región no apareada permanezcan en forma monocatenaria en condiciones de hibridación como se define anteriormente.

[0047] La secuencia de nucleótidos exacta de los adaptadores no es generalmente esencial para la invención y puede seleccionarse por el usuario de tal manera que los elementos de secuencia deseados se incluyan finalmente en las secuencias comunes de la biblioteca de moldes derivados de los adaptadores, por ejemplo para proporcionar sitios de unión para conjuntos particulares de cebadores de amplificación universales y/o cebadores de secuenciación. Pueden incluirse elementos de secuencia adicionales, por ejemplo para proporcionar sitios de unión para cebadores de secuenciación que finalmente se usarán en la secuenciación de moléculas molde en la biblioteca,

o productos derivados de amplificación de la biblioteca molde, por ejemplo en un soporte sólido. Los adaptadores pueden incluir adicionalmente secuencias "etiqueta", que pueden usarse para etiquetar o marcar moléculas molde derivadas de una fuente particular. Las características y el uso generales de dichas secuencias etiqueta se describen en la publicación en trámite del solicitante publicada como WO 05/068656.

5 [0048] Aunque la secuencia de nucleótidos exacta del adaptador es generalmente no limitante para la invención, las secuencias de las cadenas individuales en la región no apareada debe ser de tal manera que ninguna cadena individual presente ninguna autocomplementariedad interna que pudiera conducir a una autohibridación, formación de estructuras de horquilla, etc. en condiciones de hibridación convencionales. La autohibridación de una cadena en la región no apareada debe impedirse ya que esto puede evitar o reducir la unión específica de un cebador de amplificación a esta cadena.

10 [0049] Los adaptadores con apareamiento erróneo se forman preferentemente a partir de dos cadenas de ADN, pero pueden incluir mezclas de nucleótidos naturales y no naturales (por ejemplo, uno o más ribonucleótidos) unidos mediante una mezcla de enlaces de estructura fosfodiéster y no fosfodiéster. Otras modificaciones no nucleotídicas pueden incluirse tales como, por ejemplo, restos biotina, grupos bloqueantes y restos de captura para la unión a una superficie sólida, como se analiza con detalle más adelante.

15 [0050] El uno o más "dúplex de polinucleótido diana" al los que se ligan los adaptadores puede ser cualquier molécula de polinucleótidos que se desea amplificar mediante PCR en fase sólida, generalmente con una vista a la secuenciación. Los dúplex de polinucleótido diana pueden originarse en forma de ADN bicatenario (por ejemplo fragmentos de ADN genómico) o pueden haberse originado en forma monocatenaria, tal como ADN o ARN, y haberse convertido a ADNbc antes de la ligación. A modo de ejemplo, las moléculas de ARNm pueden copiarse en ADNbc bicatenarios adecuados para su uso en el método de la invención utilizando técnicas convencionales bien conocidas en la técnica. La secuencia exacta de las moléculas diana no es generalmente esencial para la invención, y puede ser conocida o desconocida. Las moléculas de ADN modificadas que incluyen enlaces de estructura no natural y/o nucleótidos no naturales pueden servir como diana, siempre que las modificaciones no excluyan la ligación con adaptador y/o copiando en una reacción de extensión con cebador.

20 [0051] Aunque en teoría el método puede aplicarse a un único dúplex diana (es decir una molécula bicatenaria individual), se usa una mezcla o una pluralidad de dúplex de polinucleótido diana. El método de la invención se aplica a una mezcla de diferentes moléculas diana que difieren entre sí con respecto a la secuencia de nucleótidos sobre toda o una parte de su longitud, por ejemplo, una mezcla compleja de moldes. El método puede aplicarse a una pluralidad de moléculas diana derivadas de una fuente común, por ejemplo, una biblioteca de fragmentos de ADN genómico derivada de un individuo particular. En una realización preferida los polinucleótidos diana comprenderán fragmentos al azar de ADN genómico humano. Los fragmentos pueden derivar de un genoma completo o formar parte de un genoma (por ejemplo un solo cromosoma o subfracción del mismo) y de un individuo o diversos individuos. Las moléculas diana de ADN pueden tratarse química o enzimáticamente antes de o posteriormente de la ligación de las secuencias adaptadoras. Las técnicas para la fragmentación de ADN genómico incluyen, por ejemplo, digestión enzimática o cizalladura mecánica.

25 [0052] La "ligación" de adaptadores en extremos 5' y 3' de cada polinucleótido diana implica la unión de dos cadenas polinucleotídicas del adaptador a un polinucleótido diana bicatenario de tal manera que se forman enlaces covalentes entre ambas cadenas de las dos moléculas bicatenarias. En este contexto "unión" significa unión covalente de dos cadenas polinucleotídicas que previamente no estaban unidas covalentemente. Preferentemente dicha "unión" se realizará por formación de un enlace fosfodiéster entre las dos cadenas polinucleotídicas pero pueden utilizarse otros medios de unión covalente (por ejemplo, enlaces de estructura no fosfodiéster). Sin embargo, es un requisito esencial que los enlaces covalentes formados en las reacciones de ligación permitan una lectura (por ejemplo enlaces de estructura no fosfodiéster). Sin embargo, es un requisito esencial que los enlaces covalentes formados en las reacciones de ligación permitan una lectura a través de una polimerasa, de tal manera que la construcción resultante pueda copiarse en una reacción de extensión con cebador usando cebadores que se unen a secuencias en las regiones de la construcción adaptador-diana que derivan de las moléculas adaptadoras.

30 [0053] Las reacciones de ligación preferentemente se catalizarán enzimáticamente. La naturaleza de la enzima ligasa usada para la ligación enzimática no está particularmente limitada. También pueden usarse técnicas de ligación no enzimática (por ejemplo ligación química), siempre que la ligación no enzimática conduzca a la formación de un enlace covalente que permita la lectura a través de una polimerasa, de tal manera que la construcción resultante pueda copiarse en una reacción de extensión con cebador.

35 [0054] Los productos deseados de la reacción de ligación son construcciones adaptador-diana en las que adaptadores idénticos están ligados en ambos extremos de cada polinucleótido diana, dada la estructura adaptador-diana-adaptador. Las condiciones de la reacción de ligación deben por lo tanto optimizarse para maximizar la formación de este producto, en preferencia para dianas que tienen un adaptador solo en un extremo.

[0055] Los productos de la reacción de ligación pueden someterse a etapas de purificación para retirar las moléculas adaptadoras no unidas antes de procesarse adicionalmente las construcciones adaptador-diana. Puede usarse cualquier técnica adecuada para eliminar el exceso de adaptadores no unidos, a continuación se detallan ejemplos preferidos de los mismos.

[0056] Las construcciones adaptador-diana formadas en la reacción de ligación se someten después a una reacción de extensión con cebador inicial en la que un oligonucleótido cebador se hibrida con una parte adaptadora de cada una de las construcciones adaptador-diana y se extiende por adición secuencial de nucleótidos en el extremo libre 3' hidroxilo del cebador para formar productos de extensión complementarios a al menos una cadena en cada una de las construcciones adaptador-diana.

[0057] El término reacción de extensión con cebador "inicial" se refiere a una reacción de extensión con cebador en la que los cebadores se hibridan directamente con las construcciones adaptador-diana, en oposición a cualquiera de sus cadenas complementarias formadas por extensión con cebador usando la construcción adaptador-diana como un molde o copias amplificadas de la construcción adaptador-diana. Es una característica clave del método de la invención que la reacción de extensión con cebador inicial se realice usando un cebador "universal" que se una *específicamente* a una secuencia afín dentro de una parte adaptadora de la construcción adaptador-diana, y no se realice usando un cebador específico de diana o una mezcla de cebadores al azar. El uso de un cebador específico de adaptador para la reacción de extensión con cebador inicial es clave para la formación de una biblioteca de moldes que tienen secuencia común en el extremo 5' y secuencias comunes en el extremo 3'.

[0058] Los cebadores usados para la reacción de extensión con cebador inicial podrán hibridarse con cada cadena individual de construcciones adaptador-diana que tengan adaptadores ligados en ambos extremos y puede extenderse para obtener dos productos de extensión con cebador individuales, uno complementario a cada cadena de la construcción. Por tanto, en la realización más preferida la reacción de extensión con cebador inicial dará como resultado la formación de productos de extensión con cebador complementarios a cada cadena de cada adaptador-diana.

[0059] En una realización preferida el cebador utilizado en la reacción de extensión con cebador inicial se hibridará con una secuencia de unión a cebador (en una cadena) en la región no apareada del adaptador.

[0060] El término "hibridación" como se usa en este contexto se refiere a una unión/hibridación específica de secuencia del cebador con una secuencia de unión a cebador en una región adaptadora de la construcción adaptador-diana en las condiciones que van a usarse para la etapa de hibridación con cebador de la reacción de extensión con cebador inicial.

[0061] Los productos de la reacción de extensión con cebador pueden someterse a condiciones desnaturalizantes convencionales para separar los productos de extensión de las cadenas de las construcciones adaptador-diana. Opcionalmente las cadenas de las construcciones adaptador-diana pueden retirarse en esta fase. Los productos de extensión (con o sin las cadenas originales de las construcciones adaptador-diana) forman en su conjunto una biblioteca de polinucleótidos molde que puede usarse como moldes para la PCR en fase sólida.

[0062] Si se desea, la reacción de extensión con cebador inicial puede repetirse una o más veces, a través de rondas de hibridación con cebador, extensión y desnaturalización, para formar múltiples copias de los mismos productos de extensión complementarios a las construcciones adaptador-diana.

[0063] En otras realizaciones los productos de extensión iniciales pueden amplificarse mediante PCR en fase solución convencional, como se describe con detalle más adelante. Los productos de dicha amplificación por PCR adicional pueden recogerse para formar una biblioteca de moldes que comprenden "productos de amplificación derivados de" los productos de extensión con cebador iniciales. En una realización preferida ambos cebadores usados para la amplificación por PCR adicional se hibridarán con diferentes secuencias de unión a cebador sobre cadenas opuestas en la región no apareada del adaptador. Otras realizaciones pueden, sin embargo, basarse en el uso de un solo tipo de cebador de amplificación que hibrida con una secuencia de unión a cebador, en la región bicatenaria del adaptador. En realizaciones del método basado en amplificación por PCR la reacción de extensión con cebador "inicial" se produce en el primer ciclo de PCR.

[0064] La inclusión de la etapa de extensión con cebador inicial (y opcionalmente rondas posteriores de amplificación por PCR) para formar copias complementarias de las construcciones adaptador-diana (antes de genoma completo o PCR en fase sólida) es ventajosa, por diversas razones. En primer lugar, la inclusión de la etapa de extensión con cebador y posterior amplificación por PCR, actúa como una etapa de enriquecimiento para seleccionar construcciones adaptador-diana, con adaptadores ligados en ambos extremos. Solo construcciones diana con adaptadores ligados en ambos extremos proporcionan moldes eficaces para el genoma completo o PCR en fase sólida usando cebadores comunes o universales específicos para secuencias de unión a cebador en los adaptadores, por tanto es ventajoso producir una biblioteca de moldes que comprenda solo dianas ligadas doblemente antes de la amplificación en fase sólida o amplificación de todo el genoma.

5 [0065] En segundo lugar, la inclusión de la etapa de extensión con cebador inicial y posterior amplificación por PCR, permite que la longitud de las secuencias comunes en los extremos 5' y 3' de la diana aumente antes de la PCR en fase sólida o de genoma completo. Como se ha indicado anteriormente, es generalmente ventajoso que la longitud de las moléculas adaptadoras sea lo más corta posible, para maximizar la eficacia de la ligación y posterior retirada de los adaptadores no unidos. Sin embargo, para los fines del PCR del genoma completo o en fase sólida puede ser ventajoso tener secuencias más largas comunes o secuencias "universales" en los extremos 5' y 3' de los moldes a amplificar. La inclusión de las etapas de extensión con cebador (y posterior amplificación) significa que la longitud de las secuencias comunes en un extremo (o en ambos extremos) de los polinucleótidos en la biblioteca de moldes puede aumentarse después de la ligación por inclusión de secuencias adicionales en los extremos 5' de los cebadores usados para la extensión con cebador (y posterior amplificación). El uso de dichos cebadores "con cola" se describe con detalle más adelante.

15 [0066] Ahora se describirán con más detalle diversas realizaciones específicas no limitantes del método de la invención con referencia a los dibujos adjuntos. Salvo que se especifique de otra manera las características descritas como preferidas en relación con una realización específica de la invención se aplica con los cambios que correspondan a otras realizaciones específicas de la invención.

20 [0067] La Figura 1 ilustra diversas realizaciones de un tipo particular de adaptador con apareamiento erróneo para su uso en el método de la invención. El adaptador se forma hibridando dos oligonucleótidos monocatenarios, en el presente documento denominados "oligo A" y "oligo B". El oligo A y el oligo B pueden prepararse mediante técnicas de síntesis de oligonucleótidos automatizadas convencionales en el uso rutinario en la técnica. Los oligonucleótidos son parcialmente complementarios de tal manera que el extremo 3' del oligo A es complementario al extremo 5' del oligo B. El extremo 5' del oligo A y el extremo 3' del oligo B no son complementarios entre sí. Cuando las dos cadenas hibridan, la estructura resultante es bicatenaria en un extremo (la región bicatenaria) y monocatenaria en el otro extremo (la región no apareada) y se denomina en el presente documento un "adaptador bifurcado" (Fig. 1a). La región bicatenaria del adaptador bifurcado puede ser de extremos romos (Fig. 1b) o puede tener un saliente. En el último caso, el saliente puede ser un saliente 3' (Fig. 1c) o un saliente 5' (Fig. 1d), y puede comprender un solo nucleótido o más de un nucleótido.

30 [0068] El extremo 5' de la parte bicatenaria del adaptador bifurcado está fosforilada, es decir, el extremo 5' del oligo B (Fig. 1a-d). La presencia del grupo fosfato 5' identifica a este como el extremo "ligable" del adaptador. El extremo 5' del oligo A puede estar biotinilado o llevar otra funcionalidad (representada por X) que le permita capturar en una superficie, tal como una perla. Dichas funcionalidades alternativas distintas de las de la biotina se conocen por los expertos en la técnica. El extremo 3' del oligo B también puede estar biotinilado o llevar otra funcionalidad (representada por Y) que lo permite capturar en una superficie (Fig. 1d).

40 [0069] Los enlaces fosfodiéster que comprenden la estructura de los oligonucleótidos puede reemplazarse con enlaces no enzimáticamente escindibles tales como enlaces fosforotioato. Preferentemente solo el último, o el último y penúltimo, enlaces fosfodiéster en ambos extremos 3' y 5' de los oligonucleótidos se sustituirán con enlaces fosforotioato. En la realización más preferida de la invención, el oligo A contiene un grupo biotina en su extremo 5', y el oligo B está fosforilado en su extremo 5' y la parte bicatenaria del dúplex contiene un saliente en 3' de una sola base que comprende un nucleótido "T". El oligo A consiste en dos secuencias: una secuencia en el extremo 5' que es idéntica a la del cebador universal a usar para la amplificación por PCR, denominada en el presente documento secuencia "CEBADOR 1", y en su extremo 3' una secuencia idéntica a la de un cebador de secuenciación universal, denominada en el presente documento secuencia "SEC CEBADOR", más un nucleótido "T" adicional en el extremo 3'. El oligo B también consiste en dos secuencias: una secuencia en su extremo 5' es complementaria a solo parte del extremo 3' de la secuencia SEC CEBADOR en el oligo A, excluyendo el saliente "T" del oligo A, y una secuencia complementaria a la de un cebador de amplificación por PCR universal, en el presente documento denominado "CEBADOR 2-comp" en su extremo 3' (Fig. 1e).

50 [0070] La Figura 2 ilustra una realización del método de la invención basada en el uso de los adaptadores bifurcados ilustrados en la Figura 1. Puede prepararse una mezcla de moléculas de ADN diana de secuencia diferente mezclando un número, más grande que uno, de moléculas de ADN individuales. En el procedimiento preferido, el ADN genómico está fragmentado en moléculas pequeñas, preferentemente menores de 1000 pares de bases de longitud. La fragmentación del ADN puede conseguirse mediante diversos métodos incluyendo: digestión enzimática, escisión química, sonicación, nebulización o hidrocizalla, preferentemente nebulización.

60 [0071] El ADN fragmentado puede fabricarse con extremos romos mediante diversos métodos conocidos por los expertos en la técnica. En el método preferido, los extremos del ADN fragmentado se reparan en el extremo con ADN polimerasa de T4 y Klenow polimerasa, un procedimiento bien conocido por los expertos en la técnica, y después se fosforila con una enzima polinucleótido quinasa. Después se añade un solo desoxinucleótido "A" en ambos extremos 3' de las moléculas de ADN usando la enzima Taq polimerasa, produciendo un saliente en 3' de una base que es complementaria al otro saliente "T" de una base en 3' en el extremo bicatenario del adaptador bifurcado.

65

[0072] Después se realiza una reacción de ligación entre el adaptador bifurcado y los fragmentos de ADN usando una enzima ligasa adecuada (por ejemplo, ADN ligasa de T4) que une dos copias del adaptador a cada fragmento de ADN, en cualquier extremo, para formar construcciones adaptador-diana. Los productos de esta reacción pueden purificarse a partir de adaptador no ligado mediante diversos medios incluyendo cromatografía de inclusión molecular, preferentemente por electroforesis a través de un gel de agarosa seguido por escisión de una parte de la agarosa que contiene el ADN de mayor tamaño que el tamaño del adaptador (Fig. 2a).

[0073] Después de haber eliminado el exceso de adaptador, el ADN diana no ligado permanece además de las construcciones adaptador-diana ligadas y este puede retirarse por captura selectiva de solo aquellas moléculas de ADN diana que tienen adaptador unido. La presencia de un grupo biotina en el extremo 5' de Oligo A del adaptador permite que cualquier ADN diana ligado al adaptador se capture sobre una superficie recubierta con estreptavidina, una proteína que une selectiva y estrechamente biotina. La estreptavidina puede aplicarse sobre una superficie por medios conocidos por los expertos en la técnica. En el método preferido, pueden usarse perlas magnéticas disponibles en el comercio que están revestidas de estreptavidina para capturar construcciones adaptador-diana ligadas. La aplicación de un imán al lado de un tubo que contiene estas perlas las inmoviliza de tal manera que pueden lavarse de las moléculas de ADN diana no ligadas (Fig. 2a).

[0074] Un oligonucleótido denominado en el presente documento CEBADOR 2, que se hibrida con la secuencia "CEBADOR 2-comp" sobre la cadena oligo B de las construcciones adaptador-diana puede usarse en una reacción de extensión con cebador inicial para generar una copia complementaria de la cadena adaptador-diana unida a la perla. El producto de extensión con cebador resultante forma un dúplex bicatenario con su cadena adaptador-diana complementaria unida a la perla y puede después aislarse y purificarse de la cadena adaptador-diana de la perla por desnaturalización (Fig. 2b).

[0075] Existen diversos métodos convencionales para separar la cadena de un dúplex de ADN por desnaturalización, incluyendo desnaturalización térmica, preferentemente desnaturalización química en solución de hidróxido de sodio 100 mM. El pH de una solución de ADN monocatenario en hidróxido sódico recogido del sobrenadante de una suspensión de perlas magnéticas puede neutralizarse ajustando con una solución de ácido apropiada, o preferentemente mediante un intercambio de tampón a través de una columna de cromatografía de exclusión molecular preequilibrada en una solución tamponada. La solución resultante contiene una biblioteca de moléculas molde de ADN monocatenario todas ellas comprendiendo en este orden: la secuencia 5' CEBADOR 2, ADN diana, el complemento de la secuencia SEC CEBADOR, después el complemento de la secuencia CEBADOR 1. Esta biblioteca de moldes puede después usarse para la amplificación por PCR en fase sólida usando los oligonucleótidos CEBADOR 1 y CEBADOR 2 inmovilizados. Sin embargo, se apreciará que la utilidad de la biblioteca no está limitada a PCR en fase sólida sino que se extiende a cualquier tipo de amplificación por PCR, incluyendo amplificación de todo el genoma realizado completamente en solución.

[0076] La Figura 3 ilustra una realización alternativa de la invención en la que las construcciones adaptador-diana preparadas como se ha descrito anteriormente con referencia a la Figura 2 se someten a rondas múltiples de hibridación con cebador y extensión para generar múltiples copias monocatenarias de cada construcción adaptador-diana. En esta realización de la invención, la reacción de extensión con cebador inicial sobre las moléculas adaptador-molde inmovilizadas en perlas con CEBADOR 2 se reemplaza en efecto con una amplificación por PCR asimétrica con el oligonucleótido CEBADOR 2 (Fig. 3), siendo esto equivalente a rondas múltiples de la misma reacción de extensión con cebador. En esta realización, se generan copias monocatenarias múltiples de las cadenas inmovilizadas en perlas en el sobrenadante de la suspensión con perlas debido al termociclado de PCR, dado que no es necesaria una etapa de desnaturalización independiente para recuperar las copias complementarias recién sintetizadas de las cadenas adaptador-diana inmovilizadas en perlas; las copias pueden purificarse del sobrenadante mediante métodos convencionales conocidos por los expertos en la técnica.

[0077] En otra realización de la invención, ilustrada en la Figura 4, la reacción de extensión con cebador inicial de las construcciones adaptador-diana inmovilizadas en perlas con CEBADOR 2 forma parte de una amplificación por PCR (simétrica) convencional con los oligonucleótidos CEBADOR 2 y CEBADOR 1. En esta realización, se generan copias bicatenarias múltiples de las cadenas inmovilizadas en perlas en el sobrenadante de la suspensión con perlas debido al termociclado de PCR, por tanto no es necesaria una etapa de desnaturalización independiente para recuperar las copias complementarias recién sintetizadas de las cadenas adaptador-diana inmovilizadas en perlas; las copias pueden purificarse del sobrenadante mediante métodos convencionales conocidos por los expertos en la técnica.

[0078] En otra realización de la invención, ilustrada en la Figura 5, el adaptador bifurcado no contiene un grupo con biotina en el extremo 5' de la cadena Oligo A. En esta realización, el ADN fragmentado puede tener los extremos romos mediante diversos métodos conocidos por los expertos en la técnica. En un método preferido, los extremos del ADN fragmentado están reparados en el extremo con una ADN polimerasa de T4 y Klenow polimerasa y después fosforilados con la enzima polinucleótido quinasa. Después se añade un solo desoxinucleótido "A" en ambos extremos 3' de las moléculas de ADN con la enzima Taq polimerasa, produciendo un saliente 3' de una base que es complementario a un saliente "T" en 3' de una base en el extremo "ligable" bicatenario del adaptador

bifurcado. Después se realiza una reacción de ligación entre el adaptador bifurcado y los fragmentos de ADN, por ejemplo, usando la enzima ADN ligasa de T4 que une dos copias del adaptador con cada molécula molde de ADN, una en cada extremo.

5 **[0079]** Los productos de la reacción de ligación pueden purificarse del adaptador no ligado mediante diversos medios, incluyendo cromatografía de inclusión molecular, preferentemente por electroforesis a través de un gel de agarosa seguido por escisión de una parte de la agarosa que contiene el ADN de mayor tamaño que el tamaño del adaptador. Una alícuota del ADN purificado se usa después en una amplificación por PCR con los oligonucleótidos CEBADOR 2 y CEBADOR 1 (Fig. 5). El primer ciclo de PCR implicará una reacción de extensión con cebador inicial con el cebador 2 (no ilustrado). Los cebadores amplifican selectivamente aquellas moléculas de ADN molde que tengan adaptadores ligados en ambos extremos. El producto de la reacción es una biblioteca de moléculas molde bicatenarias, cada una de las cuales comprende en orden uno o dos de las cadenas dúplex: secuencia CEBADOR 2 5', ADN diana, el complemento de la secuencia SEC CEBADOR, y después el complemento de la secuencia CEBADOR 1. Esta biblioteca puede después usarse en una plataforma de PCR en fase sólida que contenga los oligonucleótidos CEBADOR 1 y CEBADOR 2 inmovilizados.

20 **[0080]** La Figura 6 ilustra ejemplos adicionales de adaptadores con apareamiento erróneo bifurcados para su uso en el método de la invención. En esta realización el adaptador bifurcado se forma hibridando dos oligonucleótidos monocatenarios, denominados en el presente documento "oligo C" y "oligo B". Los oligonucleótidos son parcialmente complementarios de tal manera que el extremo 3' del oligo C es complementario al extremo 5' del oligo B. El extremo 5' del oligo C y el extremo 3' del oligo B no son complementarios entre sí. Cuando los dos oligos se hibridan la estructura resultante es bicatenaria en un extremo (región bicatenaria) y monocatenaria en el otro extremo (región no apareada) (Fig. 6a). La región bicatenaria del adaptador bifurcado puede tener extremos romos (Fig. 6d) o puede tener un saliente. En el último caso, el saliente puede ser un saliente 3' (Fig. 6c) o un saliente 5' (Fig. 6b) y puede comprender una sola base o más de una base.

30 **[0081]** El extremo 5' de la región bicatenaria del adaptador bifurcado está fosforilada, es decir, el extremo 5' del "oligo B" (Fig. 6a-d) para proporcionar un extremo "ligable". El extremo 5' del oligo C puede estar biotinilado o llevar otra funcionalidad (X) que lo permita capturarse sobre una superficie, tal como una perla. El extremo 3' del oligo B puede estar biotinilado o llevar otra funcionalidad (Y) que le permita capturarse sobre una superficie (Fig. 6d).

35 **[0082]** Los enlaces fosfodiéster que comprenden la estructura de los oligonucleótidos puede reemplazarse con enlaces no escindibles enzimáticamente, tales como enlaces fosforotioato. Preferentemente solo el último, o el último y penúltimo, enlaces fosfodiéster del extremo 3' y 5' de los oligonucleótidos se sustituirá por enlaces fosforotioato. El oligo C consiste en solo una secuencia: una secuencia idéntica a la del cebador de secuenciación universal denominado "SEC CEBADOR" (o idéntica a parte del extremo 3' de la secuencia "SEC CEBADOR") más un nucleótido "T" adicional en el extremo 3'. El oligo B consiste en dos secuencias: una secuencia en su extremo 5' que es complementaria a solo una parte del extremo 3' de la secuencia SEC CEBADOR en el oligo C, excluyendo el saliente "T" del "Oligo C" y una secuencia en su extremo 3' que es complementaria a la de un cebador de amplificación por PCR universal, denominado en el presente documento secuencia "CEBADOR 2-comp", (Fig. 6e).

45 **[0083]** La Figura 7 ilustra una realización adicional de la invención basada en el uso de adaptadores bifurcados ilustrados en la Figura 6. En esta realización, las construcciones adaptador-diana se preparan sustancialmente como se describe anteriormente con referencia a la Figura 5, excepto que se usan los adaptadores ilustrados en la Figura 6 (Fig. 7a).

50 **[0084]** Se usa una alícuota de la construcción adaptador-diana purificada en una amplificación por PCR en fase de solución convencional con oligonucleótidos cebadores "de cola". Los cebadores con cola son cebadores que solo se hibridan mediante su extremo 3' con una secuencia diana, dejando una cola 5' no hibridada. La longitud y secuencia exacta de la cola no hibridada es no limitante, pero típicamente puede comprender de 10 a 50 nucleótidos, más típicamente de 15 a 30 nucleótidos. Cuando se usan amplificaciones por PCR, la ronda de amplificación por PCR inicial (es decir, la primera y segunda reacción extensión con cebador) se basan en la unión de los extremos 3' de los cebadores con cola con las secuencias de unión a cebador afines en las regiones adaptadoras de las construcciones adaptador-diana. Las colas no hibridadas 5' derivadas de los cebadores con cola actúan como moldes en posteriores ciclos de PCR y por lo tanto se copian dando los productos de PCR bicatenarios resultantes.

60 **[0085]** En la presente realización, uno cualquiera o ambos de los cebadores usados en la reacción de amplificación pueden ser cebadores "con cola". En una realización, los cebadores usados se denominan CEBADOR 3 y CEBADOR 4, en el que el CEBADOR 3 consiste en una secuencia con cola 5', y una secuencia 3' que es complementaria a la secuencia "CEBADOR 2-comp" en el adaptador bifurcado; y el CEBADOR 4 consiste en una secuencia con cola 5' y una secuencia 3' que es idéntica al extremo 5' de la secuencia SEC CEBADOR presente en la región no apareada del adaptador bifurcado. Después de la amplificación por PCR, las secuencias con cola se incorporan en las copias de la construcción de ADN adaptador-diana.

65

[0086] En una realización de la invención, las secuencias con cola sobre CEBADOR 3 y CEBADOR 4 no son secuencias idénticas. La secuencia con cola del CEBADOR 3 y la secuencia con cola del CEBADOR 4 pueden después usarse para formar la secuencia de cebadores inmovilizados en superficie para usar en una plataforma de amplificación de ADN en fase sólida (Fig. 7b).

[0087] En otra realización de la invención, las secuencias con cola sobre el CEBADOR 3 y CEBADOR 4 son secuencias idénticas. Los productos de la PCR en fase solución tendrán por tanto la misma secuencia en sus extremos: la secuencia con cola común del CEBADOR 3 y CEBADOR 4. Esta secuencia de cola común puede después usarse para formar la secuencia de un solo cebador inmovilizado sobre la superficie en una plataforma de amplificación de ADN en fase sólida. La amplificación en superficie de la biblioteca de moldes puede por tanto realizarse usando un solo cebador PCR inmovilizado sobre la superficie.

[0088] La Figura 8 ilustra realizaciones alternativas de adaptadores con apareamiento erróneo para su uso en el método de la invención. Estos adaptadores bifurcados "modificados" pueden diseñarse para permitir la amplificación en fase sólida de moldes usando un solo cebador unido a superficie. El adaptador se forma hibridando dos oligonucleótidos monocatenarios, denominados en el presente documento "oligo D" y "oligo E". Los oligonucleótidos son parcialmente complementarios de tal manera que el extremo 3' del oligo D es complementario al extremo 5' del oligo E, y el extremo 5' del oligo D es complementario al extremo 3' del oligo E, sin embargo, las partes centrales del oligo D y oligo E no son complementarias. Cuando el oligo D y el oligo E se hibridan la estructura resultante es bicatenaria en ambos extremos (regiones bicatenarias) y monocatenaria en la región central (no apareada) y se denomina en el presente documento "adaptador bifurcado modificado" (Fig. 8a).

[0089] Un extremo del adaptador bifurcado modificado se modifica para impedir la ligación de una molécula de ADN en su extremo. Dichas modificaciones son conocidas por los expertos en la técnica y pueden incluir, por ejemplo, la presencia de un saliente en 5' o 3'. El otro extremo "ligable" puede ser un extremo romo (Fig. 8d) o puede tener un saliente. En el último caso, el saliente puede ser un saliente 3' (Fig. 8c) o un saliente 5' (Fig. 8b) y puede comprender una sola base o más de una base. La cadena 5' del extremo ligable está fosforilada es decir el extremo 5' del oligo E (Figura 8a-d). El extremo 5' del oligo D puede estar biotinilado o llevar otra funcionalidad que permita que se capture sobre una superficie, tal como una perla. El extremo 3' del oligo E puede estar biotinilado o llevar otra funcionalidad que permita que se capture en una superficie (Fig. 8d). Las modificaciones para impedir la ligación (Z, W) pueden ser las mismas que o diferentes de las funcionalidades de captura de superficie (X, Y).

[0090] Los enlaces fosfodiéster que comprenden la estructura de los oligonucleótidos pueden reemplazarse con enlaces no enzimáticamente escindibles tales como enlaces fosforotioato. Preferentemente solo el último, o el último y el penúltimo, enlaces fosfodiéster en ambos extremos 3' y 5' de los oligonucleótidos se sustituirán con enlaces fosforotioato.

[0091] En la realización preferida de la invención el oligo E está fosforilado en su extremo 5' y el extremo 3' del oligo D contiene un saliente 3' de una sola base, que comprende un nucleótido "T". El oligo D consiste en dos secuencias: una secuencia en su extremo 5' que es idéntica a la de un cebador de amplificación por PCR universal, denominado en el presente documento secuencia "CEBADOR 5", próxima a una secuencia idéntica a la de un cebador de secuenciación universal denominado secuencia "SEC CEBADOR" más el nucleótido "T" adicional en el extremo 3'. Oligo E consiste en tres secuencias: una secuencia en su extremo 5' que es complementaria a solamente parte del extremo 3' de la secuencia SEC CEBADOR en el oligo D, excluyendo el saliente T del Oligo D, una secuencia central no complementaria a ninguna parte del Oligo D y un extremo 3' que es complementario a la secuencia "CEBADOR 5" del Oligo D (Fig. 8e).

[0092] La Figura 9 ilustra otra realización adicional de la invención basada en el uso de adaptadores alternativos ilustrados en la Figura 8. En esta realización, pueden prepararse construcciones adaptador-diana sustancialmente como se describe anteriormente en relación con la Figura 5 excepto que se usan los adaptadores bifurcados modificados ilustrados en la Figura 8. Se usa una alícuota de las construcciones adaptador-diana purificadas en una amplificación por PCR en fase solución usando el oligonucleótido CEBADOR 5 para amplificar selectivamente aquellos productos de ligación que tengan el adaptador modificado en ambos extremos (Fig. 9b). El producto de la PCR en fase solución después puede purificarse y amplificarse en una plataforma de PCR en fase sólida con un solo cebador inmovilizado, por ejemplo, el CEBADOR 5. La inclusión de la secuencia con apareamiento erróneo en el oligo E garantiza que todos los productos de esta amplificación en fase sólida contendrán secuencias de unión a cebadores de secuenciación comunes solo en una cadena, permitiendo la secuenciación usando un cebador de secuenciación universal que se hibrida con esta secuencia común.

60 Uso de la biblioteca de moldes

[0093] Las bibliotecas de moldes preparadas de acuerdo con el método de la invención pueden usarse esencialmente en cualquier método de análisis de ácidos nucleicos que requiera amplificación adicional de los moldes y/o secuenciación de los moldes o productos de amplificación de los mismos. Usos ejemplares de bibliotecas molde incluyen, pero sin limitación, proporcionar moldes para amplificación de todo el genoma y también

amplificación por PCR en fase sólida (de bibliotecas molde complejas o monomolde). Un uso particularmente preferido es la amplificación del genoma completo realizada en un soporte sólido.

Amplificación del genoma completo

[0094] Las bibliotecas de moldes preparadas de acuerdo con el método de la invención comenzando a partir de una mezcla compleja de fragmentos de ADN genómico que representan un genoma completo o sustancialmente completo de moldes adecuados para amplificación denominada de "genoma completo". La expresión "amplificación de genoma completo" se refiere a una reacción de amplificación de ácidos nucleicos (por ejemplo PCR) en la que el molde a amplificar comprende una mezcla compleja de fragmentos de ácidos nucleicos representativa de un genoma completo (o sustancialmente completo).

Amplificación en fase sólida

[0095] Una vez formada, la biblioteca de moldes preparada de acuerdo con los métodos descritos anteriormente puede usarse para la amplificación de ácidos nucleicos en fase sólida.

[0096] Por tanto, en aspectos adicionales la invención proporciona un método de amplificación de ácidos nucleicos en fases sólida de moléculas polinucleotídicas molde que comprende:

preparar una biblioteca de moléculas polinucleotídicas molde que tengan secuencias comunes en sus extremos 5' y 3' usando un método de acuerdo con el primer aspecto de la invención descrita en el presente documento y realizar una reacción de amplificación de ácidos nucleicos en fase sólida en la que dichas moléculas polinucleótidos molde se amplifican.

[0097] La expresión "amplificación en fase sólida" como se usa en el presente documento se refiere a cualquier reacción de amplificación de ácidos nucleicos realizada en o en asociación con un soporte sólido de tal manera que toda o una parte de los productos amplificados se inmovilicen sobre el soporte sólido según se forman. En particular, el término abarca la reacción en cadena de la polimerasa en fase sólida (PCR en fase sólida) que es una reacción análoga a la PCR en fase en solución convencional, excepto que uno o ambos cebadores de amplificación directo o inverso están inmovilizados sobre el soporte sólido.

[0098] Aunque la invención abarca métodos de amplificación en "fase sólida" en los que solo un cebador de amplificación se inmoviliza (el otro cebador normalmente está presente en solución libre), se prefiere que el soporte sólido se proporcione con los cebadores tanto directo e inverso inmovilizados. En la práctica, habrá una "pluralidad" de cebadores directos idénticos y/o una "pluralidad" de cebadores inversos idénticos inmovilizados sobre el soporte sólido, ya que el proceso de la PCR requiere un exceso de cebadores para llevar a cabo la amplificación. Las referencias en el presente documento a cebadores directo e inverso deben interpretarse por consiguiente como que abarcan una "pluralidad" de dichos cebadores a menos que el contexto indique otra cosa.

[0099] Como apreciará un experto lector, cualquier reacción PCR determinada requiere al menos un tipo de cebador directo y al menos un tipo de cebador inverso específico para el molde a amplificar. Sin embargo, en determinadas realizaciones, los cebadores directo e inverso pueden comprender partes específicas de molde de secuencia idéntica o pueden tener estructura y secuencia de nucleótidos completamente idénticas (incluyendo cualquier modificación no nucleotídica). En otras palabras, es posible realizar amplificación en fase sólida usando solo un tipo de cebador, y los métodos de solo un cebador se incluyen dentro del ámbito de la invención. Otras realizaciones pueden usar cebadores directos e inversos que contengan idénticas secuencias específicas molde pero que difieran en algunas características estructurales. Por ejemplo un tipo de cebador puede contener una modificación no nucleotídica que no está presente en la otra.

[0100] En otras realizaciones de la invención los cebadores inverso y directo pueden contener partes específicas de molde de diferente secuencia.

[0101] En todas las realizaciones de la invención, los cebadores de amplificación para PCR en fase sólida se inmovilizan preferentemente por unión covalente al soporte sólido en o cerca del extremo 5' del cebador, dejando la parte específica de molde del cebador libre para hibridación con su molde afin y el grupo hidroxilo 3' libre para su extensión con cebador. Para esta finalidad puede usarse cualquier medio de unión covalente adecuado conocido en la técnica. La química de unión seleccionada dependerá de la naturaleza del soporte sólido y cualquier derivatización o funcionalización aplicada al mismo. El propio cebador puede incluir un resto, que puede ser una modificación química no nucleotídica, para facilitar la unión. En una realización particularmente preferida, el cebador puede incluir un nucleófilo que contiene azufre, tal como fosforotioato o tiosulfato, en el extremo 5'. En el caso de hidrogeles de poli(acrilamida) de soporte sólido (como se describe más adelante), este nucleófilo se unirá a un grupo "C" presente en el hidrogel. Los medios más preferidos de unión de cebadores y moldes a un soporte sólido se realiza mediante unión fosforotioato 5' a un hidrogel compuesto por acrilamida polimerizada y N-(5-bromoacetamidilpentil)acrilamida (BRAPA).

[0102] Se prefiere el uso de la biblioteca de moldes preparada de acuerdo con la invención para preparar series agrupadas de colonias de ácidos nucleicos, análogas en las descritas en los documentos WO 00/18957 y WO 98/44151, por amplificación por PCR en fase sólida. Los términos “agrupación” y “colonia” se usan indistintamente en el presente documento para referirse a un sitio diferenciado sobre un soporte sólido compuesto por una pluralidad de cadenas de ácido nucleico inmovilizadas idénticas y una pluralidad de cadenas de ácido nucleico complementarias inmovilizadas idénticas. La expresión “serie agrupada” se refiere a cualquier serie formada por dichos grupos o colonias. En este contexto el término “serie” no debe entenderse como que se requiere ninguna disposición ordenada de agrupaciones.

10 Uso en secuenciación/métodos de secuenciación

[0103] La invención también incluye métodos de secuenciación de ácidos nucleicos amplificados generados por amplificación de genoma completo o en fase sólida. Por tanto, la invención proporciona el método de secuenciación de ácidos nucleicos que comprende amplificar una biblioteca de moldes de ácidos nucleicos usando amplificación del genoma completo o en fase sólida como se describe anteriormente y realizar una reacción de secuenciación de ácidos nucleicos para determinar la secuencia de todo o parte de al menos una cadena de ácido nucleico amplificada producida en la reacción de amplificación del genoma completo o en fase sólida.

[0104] La secuenciación puede realizarse usando cualquier técnica de “secuenciación por síntesis” adecuada en la que los nucleótidos se añaden sucesivamente al grupo hidroxilo 3' libre, dando como resultado la síntesis de una cadena polinucleotídica en dirección 5' a 3'. La naturaleza del nucleótido añadido se determina preferentemente después de cada adición de nucleótido.

[0105] El punto de inicio para la reacción de secuenciación puede proporcionarse hibridando un cebador de secuenciación a un producto de la reacción de genoma completo o en fase sólida. En conexión con esto, uno o ambos adaptadores añadidos durante la formación de la biblioteca de moldes puede incluir una secuencia de nucleótidos que permita la hibridación de un cebador de secuenciación con productos amplificados derivados mediante amplificación del genoma completo o en fase sólida de la biblioteca de moldes.

[0106] Los productos de reacciones de amplificación en fase sólida en la que tanto los cebadores de amplificación directo como inverso se inmovilizan covalentemente sobre la superficie sólida también denominadas estructuras “puente” formadas por hibridación de pares de cadenas polinucleotídicas inmovilizadas y cadenas complementarias inmovilizadas, uniéndose ambas cadenas a un soporte sólido en el extremo 5'. Las series compuestas por dichas estructuras puente proporcionan moldes ineficaces para la secuenciación de ácidos nucleicos, ya que la hibridación de un cebador de secuenciación convencional con una de las cadenas inmovilizadas no está favorecida en comparación con la hibridación de esta cadena con su cadena complementaria inmovilizada en condiciones de hibridación convencionales.

[0107] Para proporcionar moldes más adecuados para la secuenciación de ácidos nucleicos se prefiere eliminar sustancialmente toda o al menos una parte de una de las cadenas inmovilizadas en la estructura “puente” para generar un molde que sea al menos parcialmente monocatenario. La proporción del molde que es monocatenario permitirá de este modo la hibridación con un cebador de secuenciación. El proceso de eliminación de toda o una parte de la cadena inmovilizada en una estructura de ácido nucleico bicatenaria “puente” puede denominarse en el presente documento “linealización”.

[0108] Las estructuras de molde “puente” pueden linealizarse escindiendo una o ambas cadenas con una endonucleasa de restricción o por escisión de una cadena con una endonucleasa de corte enzimático. Otros métodos de escisión pueden usarse como una alternativa a las enzimas de restricción o enzimas de corte enzimático, incluyendo, entre otras, escisión química (por ejemplo escisión de un enlace diol con peryodato), escisión de sitios abásicos por escisión con endonucleasa, o por exposición a calor o a sustancias alcalinas, escisión de ribonucleótidos incorporados en productos de amplificación de otra manera compuestos por desoxirribonucleótidos, escisión fotoquímica o escisión de un enlace peptídico.

[0109] Se apreciará que una etapa de linealización puede no ser esencial si la reacción de amplificación en fase sólida se realiza solo con un cebador inmovilizado covalentemente y el otro en solución libre.

[0110] Para generar un molde linealizado adecuado para la secuenciación es necesario retirar cantidades “distintas” de las cadenas complementarias en la estructura puente formada por amplificación para dejar después un molde linealizado para la secuenciación que es completa o parcialmente monocatenaria. Más preferentemente una cadena de la estructura puente está sustancialmente completamente eliminada.

[0111] Después de la etapa de escisión, independientemente del método utilizado para la escisión, el producto de reacción de escisión puede someterse a condiciones desnaturizantes para retirar la parte (o las partes) de la cadena (o las cadenas) escindida(s) que no están unidas a un soporte sólido. Las condiciones de desnaturización adecuadas serán obvias para el experto lector con referencia a protocolos de biología molecular convencionales

(Sambrook *et al.*, 2001, Molecular Cloning, A Laboratory Manual, 3ª Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel *et al.*).

5 [0112] La desnaturalización (y posterior rehibridación de las cadenas escindidas) da como resultado la producción de un molde de secuenciación que es parcial o sustancialmente monocatenario. Después puede iniciarse una reacción de secuenciación por hibridación de un cebador de secuenciación con la parte monocatenaria del molde.

10 [0113] Por tanto, la reacción de secuenciación de ácidos nucleicos puede comprender hibridar un cebador de secuenciación con una región monocatenaria de un producto de amplificación linealizado, incorporar secuencialmente uno o más nucleótidos en una cadena polinucleotídica complementaria a la región de la cadena molde amplificada a secuenciar, identificar la base presente en uno más de los nucleótidos incorporados y por lo tanto determinar la secuencia de una región de la cadena molde.

15 [0114] Un método de secuenciación preferido que puede usarse de acuerdo con la invención se basa en el uso de nucleótidos modificados que pueden actuar como terminadores de cadena. Una vez que el nucleótido modificado se ha incorporado en la cadena polinucleotídica en crecimiento complementaria a la región de un molde a secuenciar no hay grupo OH 3' libre disponible para dirigir la extensión de secuencia adicional y por lo tanto la polimerasa no puede añadir otros nucleótidos. Una vez que la naturaleza de la base incorporada en la cadena en crecimiento se ha determinado, el bloque 3' puede retirarse para permitir la adición del próximo nucleótido siguiente. Ordenando los productos derivados usando estos nucleótidos modificados es posible deducir la secuencia de ADN del molde de ADN. Dichas reacciones pueden realizarse en un solo experimento si cada uno de los nucleótidos modificados tiene una etiqueta diferente unida, que se sabe que corresponde a la base particular, para facilitar la discriminación entre las bases añadidas en cada etapa de incorporación. Como alternativa, puede realizarse una reacción independiente que contenga cada uno de los nucleótidos modificados por separado.

25 [0115] Los nucleótidos modificados pueden llevar una etiqueta para facilitar su detección. Preferentemente es una etiqueta fluorescente. Cada tipo de nucleótido puede llevar una etiqueta fluorescente diferente. Sin embargo la etiqueta detectable no tiene que ser una etiqueta fluorescente. Puede usarse cualquier etiqueta que permita la detección de un nucleótido incorporado.

30 [0116] Un método para detectar nucleótidos marcados de manera fluorescente comprende usar luz láser de una longitud de onda específica para los nucleótidos marcados, o el uso de fuentes adecuadas de iluminación. La fluorescencia del marcador sobre nucleótido puede detectarse mediante una cámara CCD o cualquier medio de detección adecuado.

35 [0117] La invención no pretende limitar el uso del método de secuenciación indicado anteriormente, ya que esencialmente puede utilizarse cualquier metodología de secuenciación que se base en la incorporación sucesiva de nucleótidos en una cadena polinucleotídica. Dichas técnicas alternativas incluyen, por ejemplo, Pyrosequencing™, FISSEQ (secuenciación de fluorescencia *in situ*), MPSS (secuenciación de marca masivamente paralela) y secuenciación por métodos basados en ligación .

40 [0118] El polinucleótido diana a secuenciar usando el método de la invención puede ser cualquier polinucleótido que se desee secuenciar. Usando el método de preparación de bibliotecas de moldes descrito con detalle en el presente documento es posible preparar bibliotecas de moldes comenzando a partir de un polinucleótido diana monocatenario o bicatenario de secuencia conocida, desconocida o parcialmente conocida. Con el uso de series agrupadas preparadas por amplificación en fase sólida es posible secuenciar dianas múltiples de la misma o diferente secuencia en paralelo.

50 Kits

[0119] La invención también se refiere a kits como se define en las reivindicaciones 15 y 16 para su uso en la preparación de bibliotecas de polinucleótidos molde usando el método del primer aspecto de la invención.

55 [0120] Las preparaciones preferidas de kits comprenden al menos un suministro de un adaptador con apareamiento erróneo como se define en el presente documento, más un suministro de al menos un cebador de amplificación que sea capaz de hibridarse con el adaptador con apareamiento erróneo e iniciar la síntesis de un producto de extensión, cuyo producto de extensión incluiría cualquier secuencia diana ligada al adaptador cuando el adaptador está en uso.

60 [0121] Las características preferidas de los adaptadores "con apareamiento erróneo" para la inclusión en el kit son como se describe en cualquier parte del presente documento en relación a los otros aspectos de la invención. La estructura y propiedades de cebadores de amplificación será conocida por los expertos en la técnica. Los cebadores adecuados de secuencia de nucleótidos apropiada para su uso con los adaptadores incluidos en el kit pueden prepararse fácilmente usando equipos de síntesis de ácidos nucleicos automatizados convencionales y reactivos de uso rutinario en la técnica. El kit puede incluir un suministro de un solo tipo de cebador o suministros distintos (o

incluso una mezcla) de dos cebadores diferentes, por ejemplo, un par de cebadores de PCR adecuados para amplificación por PCR de moldes modificados con el adaptador con apareamiento erróneo en fase solución y/o un soporte sólido adecuado (es decir, PCR en fase sólida).

5 **[0122]** En una realización el kit puede incluir suministros de diferentes pares de cebadores para su uso en PCR en fase solución y fase sólida. En este contexto los pares de cebadores “diferentes” pueden ser de secuencia de nucleótidos sustancialmente idéntica pero diferir con respecto a alguna otra característica o modificación, tal como por ejemplo restos de captura de superficie, etc. En otras realizaciones el kit puede incluir un suministro de cebadores para su uso en una reacción de extensión con cebador inicial y un par (o pares) de cebador diferente para amplificación por PCR en solución y/o en fase sólida.

15 **[0123]** Los adaptadores y/o cebadores pueden suministrarse en los kits listos para su uso o más preferentemente como concentrados que requieren dilución antes de su uso, o incluso en una forma liofilizada o seca que requiere reconstrucción antes de su uso. Si se requiere, los kits pueden incluir adicionalmente un suministro de un diluyente adecuado para dilución o reconstitución de los cebadores. Opcionalmente los kits pueden comprender adicionalmente suministros de reactivos, tampones, enzimas, dNTP, etc. para su uso en la realización de amplificación por PCR. Ejemplos adecuados (pero no limitantes) de dichos reactivos, se describe en la sección materiales y métodos de los ejemplos adjuntos. Otros componentes que pueden proporcionarse opcionalmente en el kit incluyen cebadores de secuenciación “universales” adecuados para secuenciar moldes preparados usando los cebadores y adaptadores con apareamiento erróneo.

[0124] La invención se entenderá adicionalmente con referencia a los siguientes ejemplos experimentales no limitantes:

25 Ejemplo

Visión de conjunto experimental

30 **[0125]** Los siguientes detalles experimentales describen la exposición completa de una realización de la invención como se ha descrito anteriormente. La fuente de ADN usada es ADN de la línea celular humana purificada proporcionada por Coriell Cell Repositories, Camden, NJ 08103 USA, catálogo nº NA07055. El ADN se prepara primero para la ligación a adaptadores bifurcados por: fragmentación del ADN por nebulización, preparación de los extremos de ADN para hacer los extremos romos y fosforilación, después adición de un solo nucleótido “A” en los extremos 3’ de los fragmentos de ADN humano. La reacción de ligación se realiza con el ADN fragmentado preparado y los adaptadores preformados por hibridación “Oligo A” y “Oligo B” (secuencias proporcionadas más adelante). El producto de la reacción se aísla/purifica del adaptador no ligado por electroforesis en gel. Finalmente el producto de la reacción de ligación se somete a ciclos de PCR para amplificar selectivamente el producto ligado que contiene el adaptador en ambos extremos de los fragmentos.

40 Materiales y métodos

Nebulización

Materiales:

45 **[0126]**

- ADN genómico humano (1 mg/ml) Coriell NA07055
- Tampón (glicerol 53,1 ml, agua 42,1 ml, TrisHCl 1 M pH 7,5 3,7 ml, EDTA 0,5 M 1,1 ml)
- Nebulizador Invitrogen (Nº K7025-05)
- Kit de purificación de PCR de columnas Qiagen (Nº 28104)

Mezcla: 25 µl (5 microgramos) de ADN
Tampón 725 µl

55 Procedimiento:

60 **[0127]** La solución de ADN enfriada se fragmentó en el nebulizador sobre hielo durante 5 a 6 minutos bajo al menos 32 psi de presión. El volumen recuperado (normalmente en algún punto entre 400 y 600 µl) se dividió en 3 alícuotas y se purificó con un kit de purificación de PCR de Qiagen, usando solo una columna y finalmente se eluyó en 30 µl de EB (Qiagen).

Reparación de extremos

65 Materiales:

[0128]

- ADN Polimerasa de T4 NEB N° M0203S
- 10xNEB 2 tampón NEB N° M7002S
- 5 • 100xBSANEb N° M9001S
- mezcla de dNTP (10 mM cada una) NEB N° N0447S
- fragmento largo Poli I de ADN de *E. coli* (Klenow, NEB N° M0210S)
- polinucleótido quinasa de T4 NEB N° M0201S
- tampón T4 PNK NEB N° M0201S
- 10 • ATP 100 mM
- kit de purificación de PCR de columnas Qiagen (N° 28104)

[0129] La mezcla de reparación de extremos se acopló de la siguiente manera:

ADN	30 µl
Agua	12 µl
10xNEB2	5 µl
100xBSA	0,5 µl
10 mM dNTP	2 µl
ADN pol T4 (3U/µl)	5 µl
	50 µl total

15 **[0130]** La reacción se incubó durante 15 minutos a temperatura ambiente, y después se añadió 1 µl de fragmento grande de Pol I de ADN de *E. coli* (Klenow) y la reacción se incubó durante 15 minutos más a temperatura ambiente. El ADN se purificó de las enzimas, tampón, etc. cargando la mezcla de reacción en una columna Qiagen, finalmente eluyendo en 30 µl de EB. Los extremos 5' del ADN se fosforilaron después usando polinucleótido quinasa de la siguiente manera:

ADN	30 µl
Agua	9,5 µl
Tampón 10xPNK	5 µl
ATP 100 mM	0,5 µl
PNK T4 (10U/µl)	5 µl
	50 µl total

25 **[0131]** La reacción se incubó durante 30 minutos a 37 °C, después se inactivó con calor a 65 °C durante 20 minutos. Después el ADN se purificó de las enzimas, tampón, etc. cargando la mezcla de reacción en una columna Quiagen, finalmente eluyendo en EB 30 µl. Se agruparon tres grupos distintos para dar 90 µl total.

Reacción de cola A

Materiales:

[0132]

- Taq ADN polimerasa NEB N° M0267L
- 10 x tampón termopol NEB N° B9004S
- 35 • dATP 1 mM Amersham-Pharmacia N° 272050
- kit de purificación de PCR de columna Qiagen (N° 28004)

[0133] Se ensambló la siguiente mezcla de reacción:

ADN	30 µl
10x tampón termopol	5 µ
dATP 1 mM	10 µl
Taq pol (5U/µl)	3 µl
	~50 µl total

40 **[0134]** La reacción se incubó durante 30 minutos a 70 °C, después el ADN se purificó de las enzimas, tampón, etc. cargando la mezcla de reacción en una columna Qiagen MinElute, eluyendo finalmente en 10 µl EB.

Hibridación del adaptador bifurcado

Materiales:

45

[0135]

- “Oligo A” y “OligoB”
- Tris 50 mM/NaCl 50 mM pH 7
- máquina de PCR

Oligo A 100 µM 20 µl
 Oligo B 100 µM 20 µl
 Tris/NaCl 10 µl
 50 µl a 40 µM de dúplex en Tris 10 mM /NaCl 10 mM pH 7,5

Oligo A: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATC_xT (x = enlace fosforotioato) (SEC ID N°: 1)
 Oligo B: 5'Fosfato-GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG (SEC ID N°: 2)

[0136] Las cadenas adaptadoras se hibridaron en una máquina de PCR programada de la siguiente manera:

Rampa a 0,5 °C/segundo hasta 97,5 °C
 Mantener a 97,5 °C durante 150 segundos

[0137] Después una etapa de 97,5 °C durante 2 segundos con una caída de temperatura de 0,1 °C/ciclo durante 775 ciclos

Reacción de ligación

Materiales:

[0138]

- adaptador bifurcado 40 µM
- ADN genómico con cola A
- Quick Ligasa NEB N° M2200L
- tampón Quick Ligasa 2x NEB N° M2200L
- máquina de PCR
- kit de purificación de PCR de columnas Qiagen (N° 28104)

[0139] La mezcla de reacción se ensambló de la siguiente manera:

ADN	10 µl
2x tampón	25 µl
Adaptador 40 µM	10 µl
Quick Ligasa	5 µl
	~50 µl total

[0140] La reacción se incubó durante 20 minutos a temperatura ambiente y después el ADN se purificó de las enzimas, tampón, etc. cargando la mezcla de reacción en una columna Qiagen, finalmente eluyendo en EB 30 µl.

Purificación en gel

Materiales:

[0141]

- Agarosa Biorad N° 161-3101
- escalera de 100 pares de bases NEB N° N3231L
- TAE
- Tampón de carga (Tris 50 mM pH 8, EDTA 40 mM, sacarosa 40% p/v)
- Bromuro de etidio
- Bandejas y tanque de gel. Unidad de electroforesis

[0142] Toda la muestra de la reacción de ligación purificada se cargó en un carril de un gel de agarosa al 2% que contenía bromuro de etidio y se desarrolló a 120 V durante 50 minutos. Después el gel se observó en una caja de “luz blanca” y los fragmentos de por encima de 300 pb a al menos 750 pb se escindieron y purificaron con un kit de purificación de un Gel de Qiagen, eluyendo en 30 µl de EB. Para tortas de gel grandes se usaron dos columnas de minElute, eluyendo cada una en 15 µl de EB y se agruparon.

Amplificación por PCR

Materiales:

5 **[0143]**

- ADN ligado
- CEBADOR 1: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA (SEC ID N°: 3)
- CEBADOR 2: CAAGCAGAAGACGGCATACGA (SEC ID N°: 4)
- 10 • mezcla 2x Jump Start Red Taq PCR Sigma N° P0982
- máquina de PCR
- columnas MinElute de QiagenQiagen (N° 28004)

15 **[0144]** El ADN ligado purificado se eluyó 25 veces, y después se preparó una mezcla de reacción PCR de la siguiente manera:

ADN	1 µl
Mezcla 2x Jump Start Red Taq	25 µl
Cebador 1 100 µM	0,5 µl
Cebador 2 100 µM	0,5 µl
Agua	23 µl
	~50 µl total

[0145] El termociclado se realizó en una máquina de PCR bajo las siguientes condiciones:

- 20
- 2 min a 94 °C
 - [45 segundos a 94 °C, 45 segundos a 65 °C, 2 minutos a 70 °C] 16 ciclos
 - 5 minutos a 70 °C
 - mantener a 4 °C

25 **[0146]** Los productos de PCR se purificaron de las enzimas, tampón, etc. en una columna MinElute de Qiagen eluyendo en 10 µl de EB. La biblioteca de ADN resultante está lista para amplificación sobre una plataforma de PCR de superficie.

Validación de bibliotecas

30 **[0147]** 1 µl de la biblioteca de ADN se clonó en un vector plasmídico y se sembró en placas de agar. Se seleccionaron 19 colonias se sometieron a miniprep y los insertos clonados se secuenciaron mediante secuenciación Sanger convencional. Los datos de secuencia obtenidos fueron los siguientes:

35 Clon 1 (SEC ID N°: 5)

```
TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGG
GGACCGTCCTGTGCATTGTAGGGTGTTC AACAGCATCCCTGACCTCCACCTACAAGATGC
CAGTAGCGAATCCCCTCAGCCCTCATCTCCTTGCCATAGTTGTGTCAACCAAATCATCT
CCACACATTGTTAGATGTTTACTGGGAGGCAGACTCACTCCCACCTGAGAACCACTGTAC
TAGAAATATCACCAAGAGAATGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG
```

ES 2 450 047 T3

Clon 2 (SEC ID N°: 6)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGT
GGGCTTTGTTCTTTGAGAGGTTGCAGTCAACATGATTCTTTAAGACCAGAACCCTGCACA
CTTCTGGGCTGTATTTCTTACATTCCTTTTCTATTTTAACCATATCCCATCTTACCTAC
TTCCAGCATAGTGGTCATATTTAATTTTACAAAACCATTTTGCCACTTGCTGCCAACTA
TGTTCTTTATAAAGCAGACTTTGAGATGGAGGCTAGTGTTGAGAGGGGATGCTTAGGAGA
ACTTTGGAGATTAATACTTATGGCAGGTAAGGGAAGGAAGCAGGATTAGACAGAAATATT
GAACTGTGATACAAAGTCAGCAAAGACTTTAGTCAATAGATCGGAAGAGCTCGTATGCCG
TCTTCTGCTTG

5 Clon 3 (SEC ID N°: 7)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTT
CGATTCCCTTCAATGATTATTCCATTCGAGTACATTCGATGATTCCATTCGATTCTATAT
GATGATGATTGCATTCGAGTCCGTGTATFATTCCATTCATTCATTAGATGATTCCATT
CGAGTCCATTCGATGATTCTCTTCGATTCCGTTGATAATTACGCTTGATTCCGTTTGAT
GTTGATTCCATTCGAGTCCATTCAATGTTAATTCCATTCGATTCTAAGCGATGATTCCAT
TCCTTTCCATTAGAAGATGATTCCATTCGAGACCATTCGATGATTGCATTCAACTCATTC
GATGACGATTCCATTCAATTCTGTTCAATGATTCCATTAGATTCCATTTGATGATGAGTC
CATTGATTCCATTTGATGATGATTCCATGCGATTCCATTAGATGATGACCCCTTTCATT
TCCAAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

10 Clon 4 (SEC ID N°: 8)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTT
AAATGCTAGGCATATTGTGTACCCACATTGGTTTGTAGCCAGCTCTATGTCATAGGGCC
CTTACCCTTTACCTATTTATTGTTAGTATAATGTCCATAAACAAGCCAATGGCTCAGCAT

GAACTGATGCTAAAGAAAGCTCATGCCTGAGTGATAAATTAAGTGACCTCAGCTATTTCT
CTTCAGTGTGTGAAAGTTATTTTAAACAGTAGGTTTCCCTGGTAGATTCTCTAACCCTC
GGTATTTACATGGCCCAACTTGGTAACTCGACTGGTTACGGCAAATGCTGAAGATCGG
AAGAGCTCGTATGCCGTCTTCTGCTTG

Clon 5 (SEC ID N°: 9)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGAAGCTCTCCGATCTTA
AGGAAGTTAGGTAGATAATTTTTGTTTAGGCCATATAGCTTTGATTTTCTGATAACAATT
TTATAAACTTAGAAATTTTCATGTAAGATACAGGAATACTGGAAGCAAAAAAAGAAGGT
GCTTTAACCTTAGGGATTGAAAAATAGTAATTTAGGTTGAAAATGCTGCTTGAAAGTTA
ATGCTGATAGCATTACTACACATGATGATTTTTTCTGGAAGGAAAGCTTTATCTGGGCCT
TCAATTTAGGAATTTTTCTCTTTGGTTTTTAAAAGCTGCCATATCACTTGAGCTTCATG
GGAAAGATGCAAATAACTAAAACAAATGAACAAAACCATGTTGAGGTCAGGAACTTATT
TCAAGAAAGCAAGTTCTAGGTTTTCTTTTAAAGTGACAGTAGAGCCTTAGGCCTCAAACC
ATCTACAACCATGTTAACAGTAAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

5 Clon 6 (SEC ID N°: 10)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCAGATCTTC
CTGCCTCAGCCTCCCGAGTAGTTGGGATTGCAGGCATGTGCCACCATGCCCTGCTAATTT
TTGTATTTTTAACTAAAGGAGGGGTTTTGCCATGTTGGCCAGGCTGGTCTTGAATGCCTG
ACCTCAGGTGATCCGCCACCTCAGCCTCCCAGTGCTGGGATTACAGGTGTGAGCCACTG
CGCCCAGCTGAGGGTAACTATTTTTAATGTGGCTGATGAATGTAACTATCCTGTCCCATG
TCTCTGTCCCAGCTGCAGAGCCCTCGTCGAGATCGGAAGAGCTCGTATGCCGTCTTCTG
CTTG

10 Clon 7 (SEC ID N°: 11)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTAA
GGTTTAAATTTTCTATATAGCCTGAAAAGTTTGAATGTTTAAATCAAACATAATTTATTGA
GCAATGGCATTTAAGAAAATGGAAAGATACAAAGGGACTTTCATCAGATGATAAGTGGAT
AAGAGAGAAAAATGCAGACAGATGAGCCAGAGTTGTGTAAAAGCTGGGAGGCTAGCAGGG
CCTTGTAGATAGCCAAGCTGACTGGGGAACAGAGATAAATGGGAGCAGAGATCAGAAAGT
TCATCCTTACCCTGCTGCCCGTGGTGAAAGGAGACTTGCAAGATCGGAAGAGCTCGTATG
CCGTCTTCTGCTTG

Clon 8 (SEC ID N°: 12)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTAA
 GGAATCTTTATTTCTACATTTGAGTTTGGAAAACCTGAGCTAGCACATCTAAATCCATCT
 AATTTTGGTCATTGGTTTAAACAAGTTCATCTTATTTTTTAAACATCTGATCTTTATTT
 TATAGAATAGACTACACAAAGTCTTTTGGAAAATAAAATATTTAACTTCCAACAATTT
 TCAGATTTTACTTATAAAAAAATTTAAAATCCTCTACTTTACTCGCATCTTATTATTTC
 TGACTTTCTAGCTACTTAAAGTTAAGGAGGAAAATTAACCTCTCTAAGATCGGAAGAGCTC
 GTATGCCGTCTTCTGCTTG

5 Clon 9 (SEC ID N°: 13)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTCG
 CCATAACCACAGCCCAGGCCTCCGTAGCCACAGCACCCATAGCCATGGCCACCATTGTAG
 TTTCTGTAGTAGCTGCCACACATGGTGTGGTTGTTGAGGTCATCCTTGGGTAGGAAGGA
 GTGTAGGTGACTTCAGTATGGACACTTCTCCGCAGAGGGCCTTTTATATGCCTCAGTGAA
 TCAGAACATAGCGTGCCCTGCAAAAATATCTCTAAAGGCCTTTCATTGTGCTGAGAAGT
 TCTGGCCCTTACGTATCTCTCTGATTTTCATATCCTGCTACTCTCCTCCATTTATCTATAA
 TGCTCAAACCTCTGCTGGCTTTTTGTCTTTTAAAATGCAGCAGGTTTCTTCTCACAATAAG
 GAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

10 Clon 10 (SEC ID N°: 14)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTAG
 AGAGAAGTTATTTAAGACAAGTAAGGTATCAGGTTGTGACTCAAATACCACAGAATCTAG
 CTATTGTTAGCAATATTAAGTATATTTTCTTAAATGAAGGATTCTCCATTTACCATATGC
 CCCTTGGATAATTTCCAGAGAATTTAATTTTTTAAAAGGAATTTTCACCAATTAATTTAT
 TGTTTTGATCAAAAGAGGACCCACTGAACACCTTATTCATTATTAAAATGTATCATAAAA
 CTTAATTATGGAGCTGGGTACAGGGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTG
 AGGCAGGAGGACTGCTTGAGTCCAGGAGTTTGAGACCAGCCTGGGTAACATGGTGAAACC
 CTGTCTCTACAAAAAATACAAAAAATTAGCCAGGTGTGGTGAGATCGGAAGAGCTCGTAT
 GCCGTCTTCTGCTTG

Clon 11 (SEC ID Nº: 15)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTCC
 CAGGGAAGCCAAAAGATTGGACACCCCTCTTCAAACCTATAAATTCCCTCCCATAGTTAGT
 TTGGCCTATGCTGGGAAATGAACAAGGGTGGCTTTGAGGTTAGAAGCAAATGGAGTCAG
 TTAGGTCAGACTTTTTTTCACCTATCATACTTTTTCTATGTCAGATTTATCTCACTTGTA
 TTTTTGCAAGGGTGGTTTCAGAGCCACTAAGCTTGTGGTAATTTTTTACTGCAATAGAAA
 ACTAATGCATTAGGTAACCCTCTTTTTTCCCTCTGATTGCTTGCTCTGGGGTAAGCCAG
 CTGCCATGTTGAATTTTTTCATTTTGTACTGAGCAATCAATGGGCTCACTGCCCGACGTG
 CATAGAGGCCAATACTGTGGCACTAGTTTTTGAGAAAAGATCGGAAGAGCTCGTATGCCG
 TCTTCTGCTTG

5 Clon 12 (SEC ID Nº: 16)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTTT
 TCAGATTTTATATGTATTAACCTCAGAACACACACCTCTTATCACACATATTTTTTCATGT
 AATTTATCTAAATCTTATAGAAAAGGGTCCATTTGCATTTTCTCTTATTAGACTCCTGAT
 TTCAAATAATATATTACTTATGAGTATTTTTCTGTGCTGTAGTTATTCATTCTTATAGAT
 ATGTAACATAATTCCTTTTGCAAAGGTAAAAATTGAGCTATCTCTTGTTGAGGATTTGTT
 GATCTCTGTCTAAAGTTTCAAAAATAAAGAACTTTAAAAGCAAATGTAAATTCCTTTCA
 AGTTTTAGTAAAATTACTTCAAACCTAGTAGCTTAAACAATACAGATTTATTATGTTACA
 GTTCTGTAAGACAGAAATCTGACTTGATCACACCATGGTAAAACCAAGATACTGCCAGGG
 TTGGTTTTTTCTTGGGGGGGGTCTGTGGGAAGAGTTTGTTCCTTTGGTTTTTCCACAGCC
 CAGAGGCTGCTTGCATTCCTTTGATCACTAGATCGGAAGAGCTCGTATGCCGTCTTCTGC
 TTG

10 Clon 13 (SEC ID Nº: 17)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTTA
 TTGTTTAGGGAATAATGACAAGGAAAAAAGTCTGTAGATATTCAGTACAGAGGCACCCA
 TCTTTTTAAATTTCTGAAGATTTTTTACTCATGCTTGGTTGAATCCACAGATGCAGAACC
 CATAGGTTTCAGAGGGCCAGCTGTGCTTTGAAAATATTAGCTTGTGTTTTTATTAGAAAGA
 AAACCTCTGAGGCCAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGAG
 GTGGCGGATCACAAGGTGAGGAGATCGAGACCATTCTGGCTAACATGGTGAACCCTGT
 CTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTAGTGAGCACCTAGATCGGAAGAG
 CTCGTATGCCGTCTTCTGCTTG

Clon 14 (SEC ID Nº: 18)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTGC
AATTGGTAAAACAGTAAGCAATGAAACAGACACTTCTCAAATATTCCAAGATGGTACACG
CTTTTCAGTGTGTATGATCCAATAAAGCCATTGGAAGTAGGCTTTAATAGTCAAAAAAGA
CTATTCAGTTAGATAGGAACTATTTGCCTATAACTATTGGCCAAAAATAGGTTAAAAAAT
TGTTTTAAATTTGTGCTTTACAAAACATGTGGACTTTTTTTAGAAAATGTGTCAAATTTCA
AAAGAAATATAGACATTATGGAAAGGTCAGTTAAGCACAGCCCTAATCCTGAAAACATAA
CTATGAAAGATACTAGCTGTTACTTGTAACCAAAGGAAAAAAGATATTAGTAACCAA
TAATTAGCAAACAATGCCCATATATTTCCCTTTTTTTTTTTTTTTTTTTTGGAGACAGGGGCT
TACTCAGGCTGGATGTGATCATGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

5 Clon 15 (SEC ID Nº: 19)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTTT
AGAAGCTCTATTATACTGGAAAAGAGATATGAGACCCTTCCTACTTTAAGAATCAATGAA
GCCGGGTGTGGTGGCTCACGCCTGTACTCCCAGCATTTTGGAGAGGCCAAGCTGGGCAGAT
CACCTGAGGTCGGGAGTTTGAGACCAGCCTAGCCAACATGATGAAATCCTGTTTCTACTA
ATAACACAAAAATTAGCCGGGTGTGGTGGCGCACATCTGGAATCCCAGCTACTCCAGAGG
CTGAGGCAGGAAAATTGCTTGAAGCTGGGAAGCAGAAGTTGCAGTGAAGATCGGAAGAGC
TCGTATGCCGTCTTCTGCTTG

10 Clon 16 (SEC ID Nº: 20)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTCCGATCTGG
CTGGACTGAATAGGATAGCCTTAGCTGTAAAATTGGGCTGATCTTTCAAATGGACTCATG
CTTGCCGAATGACTCACGCTCCTGTTTACAAATCAGCTCTGTGAAGAAATGCAGAGTGGG
AGGCTCTGCTTGCCAGACGGAGACCTTAGACCTCCAGGGGCGGAGAACGGAGTACTTCCT
CTGGTGCTCGGCTTCCCTTCCTGGGGGCAGATCTCTCAGCTTCTGGTTGGTGGCTCTCAA
AATCCAGACACAAGGTCAGCTGCAGCCAGCGTGGGCCCTGGAGTAGCTCCAGTTATGGGG
CAGCAATGGCCCCCTCTCATTTTGGAGAGCTCACTTTGCCTGTGGATGGTTTTAATCCATC
TGGATAAACTTGAGGCCCATGGGAATACCATATACTATGGTAACCATGTACACTGCTCTA

AAGATGTGGCTGCTGTTGTATAACTTTTTCCCTTATTTTTGTCAATTCCTATTTTCCAG
AGTCTTGCATACCCACTATGTCTACTGTGATAGTGAACGTAAAAACATACAAGATGTTGG
TGTTATCCTCAATCTCAGATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG

Clon 17 (SEC ID N°: 21)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCT
GTGAGAAATCAATGTCTGCTGTTTATAAGCCGCCGGGCTGTGATATCCTGTGAGAGTGGC
CCCAGTGGATGAAGACAGATGCTCTCAAGGAGCGCAGATGACGCGGGTCCGAAGGACTC
GGCACCCAGCCCGGAGGCCGGCAACATGGGCAAGGGGCCTCTCACGGCTGACCTGTTTCC
TCATCAGCACATCAGGACAATAAGAGCTCCCACCTCACAGGTGGTGAAGAGCCAACGTGG
TGAAGAATGAATAAAGCAGCTCGTGGAAAGTGTGTGCATGAGGCCTGGCAACCGGTCCC
TGCTCTGAGGTACCTGCCACGGAGCTGCTGACAGGACCATTAAAAACACAATTGTGCAA
GTGCTCACCCACATTCACAGCAGCAGAATCTCCACCAGCCAAGCATTGGAGACGATCCTT
GCATCCATAGACATGAACAGATGAGCAAAACGTGGTCTATACGGACGATGAAATAGCACT
CAGCCCTAAGAAGAAATAAAATCCCGACAGAGAGATCGGAAGAGCTCGTATGCCGTCTTC
TGCTT

5 Clon 18 (SEC ID N°: 22)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGC
TGAGAGGAGAGGAGAGGCTGGCAGGGGGCTGATGCAGGAGGTGATAGGGCTCCCGGTGAT
AAGAGGTGAGAAGAACAGTCTCTGTGTGCCTAGAGAAGAGATTACCAGAAGTCTGCTATC
TGTTTGTTCGCGGATGTCGGACAGGCAGGATCGGTGATGGCAGGTCTTGGGGGAAGGATT
ATCAGGAGCTAAAAGCTGTCTTCACCTTGGCTGCTAAGAACTCATCTCGGATCTTCTTAG
AATTCCAAATCGGACTTTTCTCCTAGCAGTGGCTACATCCTTAACCTCAAAAATACCCGT
ATTAGCAGATCTACCTCCATGAAATAGACAATTCTTGACAACTAAGATCGGAAGAGCTC
GTATGCCGTCTTCTGCTTG

10 Clon 19 (SEC ID N°: 23)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTGT
TATTTTTCATACTGATGATTCTAGGGCGGTTCCCTGGACTTGGCTGGGACCAGAACTGTC
TCACAGCTGCCAGCGCTCCCCACCCCATGAGAACCATCGTGCTGAAGATGAAGAGCCAA
AGCCAACTGACCAGCCTCTGCAACCCCAAGCATCACCGTGTGAGGACAGCGTGATGTGG

AGCCGCGGAGACCCACGCAGTGCTGCGAGGGGCAACCACATGCAGGGGGGCAGAGGTGT
 GGGGGAGCAAGCAGATGCCACCGGCATGCCTAGGAGCCACAGGGAGCATGCGGGGCTGGGC
 AGTGATGGGAATGAACGTGAACTCCAGTCCTGCCCCAAGAAGTCCTCCGGCCCCCTCCTTC
 TCAATTCAGGGCACAAAGTGGTAACTGCAGCATGCAGAGGAGTCGAGGAGTCTTCCCTCG
 TCCCAGGAGCAGCACCTCGGGCACAGTCTCGGTCCCACAGAACAGCCAAGTGTGGGTTGG
 TGGTCTAGAGACCTCCGAAGATCCAGTGGGGGAAGGATGGGCAGCAGAGGGTCTACTCTC
 TGAAAATAAGGGGAAGGGATTTTCCCTCCCCACTGCCAAGGTCCCAGCTACTGGACGTGG
 GTGGAGATCGGAAGAGCTCGTATGCCGTCTTCTGCCTT

[0148] Estos resultados confirman que el método de preparación de bibliotecas produce una biblioteca de moldes de ADN "secuenciados" que contienen una mezcla de fragmentos de diferente secuencia. El ADN inserto de cada una de las 19 colonias secuenciadas se encontró que se alineaba con una secuencia de referencia del genoma humano (alineamiento no mostrado), lo que ilustra que el método produce una biblioteca de moldes que refleja realmente la composición de secuencia de los fragmentos diana de partida (es decir los clones contienen fragmentos genómicos humanos en lugar de secuencia "redundante").

10 LISTADO DE SECUENCIAS

[0149]

- 15 <110> Solexa Limited
- <120> Método de preparación de bibliotecas de polinucleótidos molde
- <130> P81137W000
- 20 <160> 23
- <170> FastSEQ para Windows Versión 4.0
- 25 <210> 1
- <211> 33
- <212> ADN
- <213> Secuencia artificial
- 30 <220>
- <221> misc_feature
- <222> oligonucleótido sintético
- 35 <400> 1
- acactcttc cctacacgac gctctccga tct 33
- <210> 2
- <211> 33
- <212> ADN
- <213> Secuencia artificial
- 40 <220>
- <221> misc_feature
- <222> oligonucleótido sintético
- 45 <400> 2
- gatcggaaga gctcgtatgc cgtctctgc ttg 33
- <210> 3
- <211> 44
- <212> ADN
- <213> Secuencia artificial

<220>
 <221> misc_feature
 <222> oligonucleótido sintético

5 <400> 3
 aatgatacgg cgaccaccga gatctacact ctttccctac acga 44

<210> 4
 <211> 21
 10 <212> ADN
 <213> Secuencia artificial

<220>
 <221> misc_feature
 15 <222> oligonucleótido sintético

<400> 4
 caagcagaag acggcatacg a 21

<210> 5
 <211> 296
 <212> ADN
 <213> *Homo sapiens*

20

25 <400> 5

tgatacggcg	accaccgaga	tctacactct	ttccctacac	gacgctcttc	cgatctgtgg	60
ggaccgtcct	gtgcattgta	gggtgttcaa	cagcatccct	gacctccacc	tacaagatgc	120
cagtagcgaa	tcccctcage	cctcatctcc	ttgccatagt	tgtgtcaacc	aaaatcatct	180
ccacacattg	ttagatgttt	actggggagg	agactcactc	ccacttgaga	accactgtac	240
tagaaatato	accaagagaa	tgagatcggg	agagctcgtg	tgccgtcttc	tgcttg	296

<210> 6
 <211> 431
 <212> ADN
 <213> *Homo sapiens*

30

35 <400> 6

aatgatacgg	cgaccaccga	gatctacact	ctttccctac	acgacgctct	tccgatctgt	60
ggcctttggt	ctttgagagg	ttgcagtcaa	catgattctt	taagaccaga	accctgcaca	120
cttcttgggc	tgtatttctt	acattccttt	tctattttta	ccatatccca	tcttacctac	180
ttccagcata	gtggtcatat	ttaattttta	caaaaccatt	ttgccacttg	ctgccaacta	240
tgttctttat	aaagcagact	ttgagatgga	ggctagtgtt	cagaggggat	gcttaggaga	300
actttggaga	ttaatactta	tggcaggtaa	gggaagggaag	caggattaga	cagaaatatt	360
gaactgtgat	acaaagtcag	caaagacttt	agtcaataga	tcggaagagc	togtatgccg	420
tcttctgctt	g					431

<210> 7
 <211> 518
 <212> ADN
 <213> *Homo sapiens*

40

45 <400> 7

aatgatacgg	cgaccaccga	gatctacact	ctttccctac	acgacgctct	tccgatcttt	60
cgattccctt	caatgattat	tccattcgag	tacattcgat	gattccattc	gattctatat	120
gatgatgatt	gcattcgagt	cogtgtatta	ttccattcca	ttccattaga	tgattccatt	180
cgagtccatt	cgatgattct	cttcgattcc	gttcgataat	tacgcttgat	tccgtttgat	240
gttgattcca	ttcgagtcca	ttcaatgta	attccattcg	attctaagcg	atgattccat	300
tcctttccat	tagaagatga	ttccattcga	gaccattcga	tgattgcatt	caactcattc	360
gatgacgatt	ccattcaatt	ctgttcaatg	attccattag	attccatttg	atgatgagtc	420
cattcgattc	catttgatga	tgattccatg	cgattccatt	agatgatgac	cccttccatt	480
tccaagatcg	gaagagctcg	tatgccgtct	ctgcttg			518

5
 <210> 8
 <211> 387
 <212> ADN
 <213> *Homo sapiens*

<400> 8

aatgatacgg	cgaccaccga	gatctacaact	ctttccctac	acgacgctct	tccgatcttt	60
aaatgctagg	catatttgtg	accccacatt	ggtttgtagc	cagctctatg	tcataggggc	120
cttacccttt	acctatttat	tgtagtata	atgtccataa	acaagccaat	ggctcagcat	180
gaactgatgc	taaagaaagc	tcatgcctga	gtgataaatt	aagtgcctc	agctatttct	240
citcagtggt	gtgaaagtta	tttttaacag	taggtttcct	ggtagattct	ctaaccactc	300
ggtatttcac	atggcccaac	ttggttaact	cgactgggta	cggcaaatgc	tgaagatcgg	360
aagagctcgt	atgccgtcct	ctgcttg				387

10
 <210> 9
 <211> 536
 <212> ADN
 <213> *Homo sapiens*

15
 <400> 9

aatgatacgg	cgaccaccga	gatctacaact	ctttccctac	acgaagctct	tccgatctta	60
aggaagttag	gtagataaatt	tttgtttagg	ccatatagct	ttgattttct	gataacaatt	120
ttataaaact	agaaattttc	atgtaagata	caggaatact	ggaagcaaaa	aaaagaaggt	180
gctttaacct	tagggattga	aaaaatagta	athtaggtg	aaaatgctgc	ttgaaagtta	240
atgctgatag	cattactaca	catgatgatt	tttctggaa	ggaagcttt	atctgggctc	300
tcaatthtagg	aatthttctc	tttggttttt	aaaagctgcc	atattcactt	gagcttcatg	360
ggaagatgc	aaataactaa	aacaaatgaa	caaaaaccat	gttgaggta	ggaacttatt	420
tcaagaagc	aagttctagg	ttttctttta	aagtgcagct	agagccttag	gcctcaaacc	480
atctacaacc	atgthaacag	taagatcggg	agagctcgta	tgccgtcttc	tgcttg	536

20
 <210> 10
 <211> 364
 <212> ADN
 <213> *Homo sapiens*

25
 <400> 10

aatgatacgg	cgaccaccga	gatctacaact	ctttccctac	acgacgctct	tcagatcttc	60
ctgccctcagc	ctcccagta	gthgggattg	caggcatgtg	ccaccatgcc	ctgctaattt	120
ttgtatthttt	aactaaagga	ggggthttgc	catgttgccc	aggetggctg	tgaatgcttg	180
acctcaggtg	atccgcccac	ctcagcctcc	cagtgtggg	attacaggtg	tgagccactg	240
cgcccagctg	agggttaacta	tttttaatgt	ggctgatgaa	tgtaactatc	ctgtcccctg	300
tctctgtccc	cagctgcaga	gccctcgtcg	agatcgggag	agctcgtatg	ccgtcttctg	360
cttg						364

30
 <210> 11
 <211> 374
 <212> ADN
 <213> *Homo sapiens*

35
 <400> 11

aatgatacgg	cgaccaccga	gatctacaact	ctttccctac	acgacgctct	tccgatctaa	60
ggtthaaatt	ttctatatag	cctgaaaagt	ttgaatgtht	aattcaaact	aatttattga	120
gcaatggcat	ttaagaaaat	ggaaagatac	aaagggactt	tcatacagatg	ataagtggat	180
aagagagaaa	aatgcagaca	gatgagccag	agttgtgtaa	aagctgggag	gctagcaggg	240
ccttgtagat	agccaagctg	actggggaac	agagataaat	gggagcagag	atcagaaaagt	300
tcacctttac	cctgctgcc	gtggtgaaaag	gagacttgca	agatcgggag	agctcgtatg	360
ccgtcttctg	cttg					374

<210> 12

ES 2 450 047 T3

<211> 379
 <212> ADN
 <213> *Homo sapiens*

5 <400> 12

```

aatgatacgg cgaccaccga gatctaacact ctttccctac acgacgctct tccgatctaa 60
ggaatcttta tttctacat ttgagtttgg aaaactgagc tagcacatct aaatccatct 120
aattttggtc attggtttta acaagttcat cttatTTTTT taaacatctg atctttatTT 180
tatagaatag actacacaaa gtcttttTgga aaattaaaaT attttaactt ccaacaattT 240
tcagatTTta cttataaaaa aatttTaaat cctctactTT actcgcacot ttattatTTc 300
tgactTTtcta gctactTaaa gTTaaggagg aaattTaaact ctctaagatc ggaagagctc 360
gtatgcccgc ttctgcttg
    
```

<210> 13
 <211> 455
 <212> ADN
 <213> *Homo sapiens*

10 <400> 13

```

aatgatacgg cgaccaccga gatctaacact ctttccctac acgacgctct tccgatctcg 60
ccataaccac agcccaggcc tccgtagcca cagcaccat agccatggcc accattgtag 120
tttctgtagt agctgccaca catgggtgtg gttggtgagg tcatccttgg gtaggaagga 180
gtgtaggTga cttoagtatg gacacttctc cgcagagggc cttttatatg cctcagtTaa 240
tcagaacata gcgtgccctt gcaaaaaatT ctctaaggc ctttcattgt gctgagaagt 300
tctggccctt acgtatctct ctgatttoat atcctgctac tctctccat ttatctataa 360
tgctcaaaact ctgctggctt tttgtctTTT aaaatgcagc aggtttcttc tcacaataag 420
gagatcggaa gagctcgtat gccgtcttct gcttg
    
```

<210> 14
 <211> 495
 <212> ADN
 <213> *Homo sapiens*

20 <400> 14

```

aatgatacgg cgaccaccga gatctaacact ctttccctac acgacgctct tccgatctag 60
agagaagtta tTtaagacaa gtaaggTatc aggttgTgac tcaaatacca cagaatctag 120
ctattgttag caatattaag tatatTTtct taaatgaagg attctccatt taccatattg 180
cccttggaTa atttccagag aattTaaTTT tTtaaaaagga atTTTcacca attaaattat 240
tgTtttgatc aaaagaggac ccaactgaaca ccttatTcat tattaaaaTg tatcataaaa 300
ctTaaTtatg gagctgggTg caggggctca tgccTgtaat cccagcactt tgggaggctg 360
aggcaggag actgcttgag tccaggagTt tgagaccagc ctgggTaaCa tggTgaaacc 420
ctgtctctac aaaaaataca aaaaattagc caggtgtggt gagatcggaa gagctcgtat 480
gccgtcttct gcttg
    
```

25 <400> 15

<210> 15
 <211> 491
 <212> ADN
 <213> *Homo sapiens*

30 <400> 15

```

aatgatacgg cgaccaccga gatctaacact ctttccctac acgacgctct tccgatctcc 60
cagggaaGCC aaaagattgg acacccctct tTcaaactat aaattcctcc catagTtagt 120
ttggcctatg ctgggaaatg acaagggTg gctttgaggt tagaagcaaa atggagTcag 180
ttaggtcaga cttttttTca ctatcactact ttttctatgt cagatTTatc tcactTgTaa 240
ttttTgcaag gTtggtTtca gagccactaa gctTgtggTa atTTTttact gcaatagaaa 300
actaatgcat taggTaaCCc tctTTttTtc cctctgattg ctTgctctgg gTaaagccag 360
ctgccatgtt gaattTTtca tttTgttact gagcaatcaa tgggctcact gccgacgtg 420
catagaggcc aatactgtgg cactagTTTT tgagaaaaga tccgaaagac tCGtatgccg 480
tcttctgctt g
    
```


ES 2 450 047 T3

aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatcttt 60
 agaagctcta ttatactgga aaagagatat gagacccttc ctactttaag aatcaatgaa 120
 gccgggtgtg gtggctcag cctgtactcc cagcattttg agaggccaag ctgggcagat 180
 cacctgaggf cgggagtttg agaccagcct agccaacatg atgaaatcct gtttctacta 240
 ataacacaaa aattagccgg gtgtgggtggc gcacatctgg aatcccagct actccagagg 300
 ctgaggcagg aaaattgctt gaagctggga agcagaagtt gcagtgaaga tccgaagagc 360
 tcgtatgccg tcttctgctt g 381

<210> 20
 <211> 650
 <212> ADN
 <213> *Homo sapiens*

5

<400> 20

aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctgg 60
 ctggactgaa taggatagcc ttagctgtaa aattgggctg atctttcaaa tggactcatg 120
 cttgccgaat gactcacgct cctgtttaca aatcagctct gtgaagaaat gcagagtggg 180
 aggctctgct tgcagacgg agacctaga cctccagggg cggagaacgg agtacttcc 240
 ctggtgctcg gcttcccttc ctgggggcag atctctcagc ttctgggtgg tggctctcaa 300
 aatccagaca caaggtcagc tgcagccagc gtgggccctg gactagctcc agttatgggg 360
 cagcaatggc cccctctcat tttgagagct cactttgcct gtggatggtt ttaatccatc 420
 tggataaact tgaggcccat gggataacca tatactatgg taacctatga cactgctcta 480
 aagatggggc tgctgttga taactttttc ctttattttt gtcaatttcc tattttccag 540
 agtcttgcat acccactatg tctactgtga tagtgaacgt aaaaacatac aagatggtgg 600
 tgttatcctc aatctcagat cgggaagagct cgtatgccgt cttctgcttg 650

10

<210> 21
 <211> 605
 <212> ADN
 <213> *Homo sapiens*

15

<400> 21

aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctct 60
 gtgagaaatc aatgtctgct gtttataagc cgccgggctg tgatatcctg tgagagtggc 120
 cccagtggtg gaagacagat gctctcaagg agcgcagatg acgcgggttc cgaaggactc 180
 ggcaaccagc ccggaggccg gcaacatggg caaggggctt ctcacggctg acctgtttcc 240
 tcatcagcac atcaggacaa taagagctcc cacttcacag gtggtgaaga gccaacgtgg 300
 tgaagaatga ataaagcagc tcgtggaaag tgctgtgcat gaggcctggc aaccggctcc 360
 tgctctgagg tcacctgcca cggagctgct gacaggacca ttaaaaacac aattgtgcaa 420
 gtgctcacc cacttcacag cagcagaatc tccaccagcc aagcattgga gacgatcctt 480
 gcatccatag acatgaacag atgagcaaaa cgtggtctat acggacgatg aaatagcact 540
 cagccctaag aagaaataaa atcccgacag agagatcggg agagctcgta tgcctctctc 600
 tgctt 605

20

<210> 22
 <211> 439
 <212> ADN
 <213> *Homo sapiens*

25

<400> 22

aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctgc 60
 tgagaggaga ggagaggtg gcagggggct gatgcaggag gtgatagggc tccgggtgat 120
 aagaggtgag aagaacagtc tctgtgtgcc tagagaagag attaccagaa gtctgtatc 180
 tgtttgttcg cggatgtcgg acaggcagga tcgggtgatg caggtcttgg gggaaggatt 240
 atcaggagct aaaagctgtc ttcaacctgg ctgctaagaa ctcatctcgg atcttcttag 300
 aattccaaat cggacttttc tcctagcagt ggctacatcc ttaacctcaa aaataccctg 360
 attagcagat ctacctecat gaaatagaca attcttgaca aactaagatc ggaagagctc 420
 gtatgcgctc ttctgcttg 439

30

ES 2 450 047 T3

<210> 23
<211> 698
<212> ADN
<213> *Homo sapiens*

5

<400> 23

```
aatgatacgg cgaccaccga gatetacact ctttccctac acgacgctct tccgatctgt . 60
tatttttcat actgatgatt ctagggcggt tccctggaact tggctgggac cagaactgtc 120
tcacagctgc cagcgtccc ccccccatg agaaccatcg tgctgaagat gaagagccaa 180
agccaaactg accagcctct gcaaccccaa gcatcacctg gtgaggacag cgtgatgtgg 240
agccgcggag accccacgca gtgctgcgag gggcaaccac atgcaggggg gcagaggtgt 300
gggggagcaa gcagatgcca ccggcatgcc taggagccac agggagcatg cgggctgggc 360
agtgatggga atgaacgtga actccagtcc tgcccaaga agtcctcgg cccctcctc 420
tcaattcagg gcacaaagtg gtaactgcag catgcagagg agtcgaggag tcttccctcg 480
tcccaggagc agcacctcgg gcacagtctc ggtcccacag aacagccaag tgtgggttgg 540
tggcttagag acctccgaag atccagtggg ggaaggatgg gcagcagagg gtctactctc 600
tgaaaataag gggaaaggat ttccctccc cactgccaag gtcccagcta ctggaogtgg 660
gtggagatcg gaagagctcg tatgccgtct tctgcctt . 698
```

10

REIVINDICACIONES

1. Un método para generar una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3', comprendiendo el método:

ligar polinucleótidos adaptadores con apareamiento erróneo idénticos en ambos extremos de cada uno de uno o más dúplex de polinucleótido diana para formar una o más construcciones adaptador-diana, en el que cada adaptador con apareamiento erróneo se forma a partir de dos cadenas polinucleotídicas hibridadas que forman un complejo biomolecular que comprende al menos una región bicatenaria y una región no apareada, y realizar una reacción de extensión con cebador inicial en la que un oligonucleótido cebador se hibrida con una parte de adaptador de cada una de las construcciones adaptador-diana y se extiende por adición secuencial de nucleótidos para formar productos de extensión complementarios a al menos una cadena de cada una de las construcciones adaptador-diana, en el que los productos de extensión, y opcionalmente productos de amplificación derivados de los mismos, proporcionan en su conjunto una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3'; en el que los dúplex de polinucleótido diana que van a ligarse son una mezcla compleja de fragmentos de ADN genómico que representan un genoma completo o sustancialmente completo, y en el que la biblioteca de moléculas polinucleotídicas molde generadas es representativa del genoma completo o sustancialmente completo.

2. El método de acuerdo con la reivindicación 1, en el que el oligonucleótido cebador es un cebador con cola que comprende una secuencia con cola no hibridante en el extremo 5'.

3. El método de acuerdo con la reivindicación 1 o 2, en el que la reacción de extensión con cebador inicial comprende:

- a) hibridar un primer oligonucleótido cebador con una parte de adaptador de cada una de las construcciones adaptador-diana, siendo el primer cebador un cebador con cola que comprende una cola 5' no hibridante,
- b) extender el primer cebador por adición secuencial de nucleótidos para formar productos de extensión complementarios a al menos una cadena de cada una de las construcciones adaptador-diana, y
- c) someter los productos obtenidos en la etapa b) a condiciones desnaturalizantes, separando de este modo los productos de extensión de cadenas de las construcciones adaptador-diana.

4. El método de la reivindicación 3, que comprende la etapa adicional de hibridar segundos oligonucleótidos cebadores con los productos de extensión y extender los segundos cebadores para formar cadenas complementarias a los productos de extensión.

5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en el que el primer cebador utilizado en la reacción de extensión con cebador inicial se hibrida con una región no apareada de los adaptadores.

6. El método de acuerdo con la reivindicación 4, en el que la reacción de extensión con cebador inicial se realiza como parte de una reacción en cadena de la polimerasa y los productos de amplificación de la reacción PCR se recogen para proporcionar una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y 3'.

7. Un método de acuerdo con la reivindicación 6, en el que la reacción en cadena de la polimerasa se realiza usando primeros cebadores de oligonucleótido que pueden hibridarse con una región no apareada de la parte de adaptador de una primera cadena de las construcciones adaptador-diana y segundos cebadores de oligonucleótido que pueden hibridarse con una región de las cadenas extendidas producidas por extensión de los primeros cebadores de oligonucleótido, siendo esta región complementaria a la región no apareada de la parte de adaptador en una segunda cadena de las construcciones adaptador-diana.

8. El método de acuerdo con la reivindicación 7, en el que los segundos cebadores también son cebadores con cola que comprenden una parte de cola en 5' que no hibrida con los adaptadores con apareamiento erróneo.

9. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 8, en el que los polinucleótidos adaptadores con apareamiento erróneo son adaptadores bifurcados formados por hibridación de cadenas polinucleotídicas primera y segunda parcialmente complementarias, en el que una secuencia de 5 o más nucleótidos consecutivos en el extremo 3' de la primera cadena es complementaria a una secuencia de 5 o más nucleótidos consecutivos en el extremo 5' de la segunda cadena de tal manera que una región bicatenaria de 5 o más pares de bases consecutivos se forma hibridando las dos cadenas y en el que una secuencia de al menos 10 nucleótidos consecutivos en el extremo 5' de la primera cadena y una secuencia de al menos 10 nucleótidos consecutivos en el extremo 3' de la segunda cadena no son complementarias de tal manera que la región no apareada de al menos 10 nucleótidos consecutivos en cada cadena permanece como una forma monocatenaria cuando la región bicatenaria se hibrida.

10. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 8, en el que los dúplex de polinucleótido diana son fragmentos de un genoma completo.
- 5 11. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 8, en el que los dúplex de polinucleótido diana son ADNc bicatenarios.
12. Un método de amplificación de ácidos nucleicos en fase sólida de moléculas polinucleotídicas molde que comprende:
- 10 preparar una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en sus extremos 5' y secuencias comunes en sus extremos 3' usando el método de acuerdo con una cualquiera de las reivindicaciones 1 a 8 y realizar una reacción de amplificación de ácido nucleico en fase sólida en el que dichas moléculas polinucleotídicas molde se amplifican.
- 15 13. Uso de una biblioteca de moléculas polinucleotídicas molde preparadas de acuerdo con el método de una cualquiera de las reivindicaciones 1 a 11 como un molde para amplificación en fase sólida.
14. Uso de una biblioteca de moléculas polinucleotídicas molde preparadas de acuerdo con el método de la reivindicación 10 como un molde para amplificación del genoma completo.
- 20 15. Un kit para su uso en la preparación de una biblioteca de moléculas polinucleotídicas molde que tienen secuencias comunes en su extremo 5' y secuencias comunes en sus extremos 3', comprendiendo el kit polinucleótidos adaptadores bifurcados como se define en la reivindicación 9 y uno o más cebadores oligonucleotídicos que pueden hibridarse con los polinucleótidos adaptadores con apareamiento erróneo en el que los cebadores oligonucleotídicos son cebadores con cola que comprenden una parte en cola 5' que no hibrida con el adaptador bifurcado.
- 25 16. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 12, o el uso de acuerdo con una cualquiera de las reivindicaciones 13 o 14, o el kit de acuerdo con la reivindicación 15, en el que al menos una de dichas secuencias comunes comprende la secuencia de la SEC ID N° 1, o la secuencia de la SEC ID N° 2.
- 30

Fig. 1

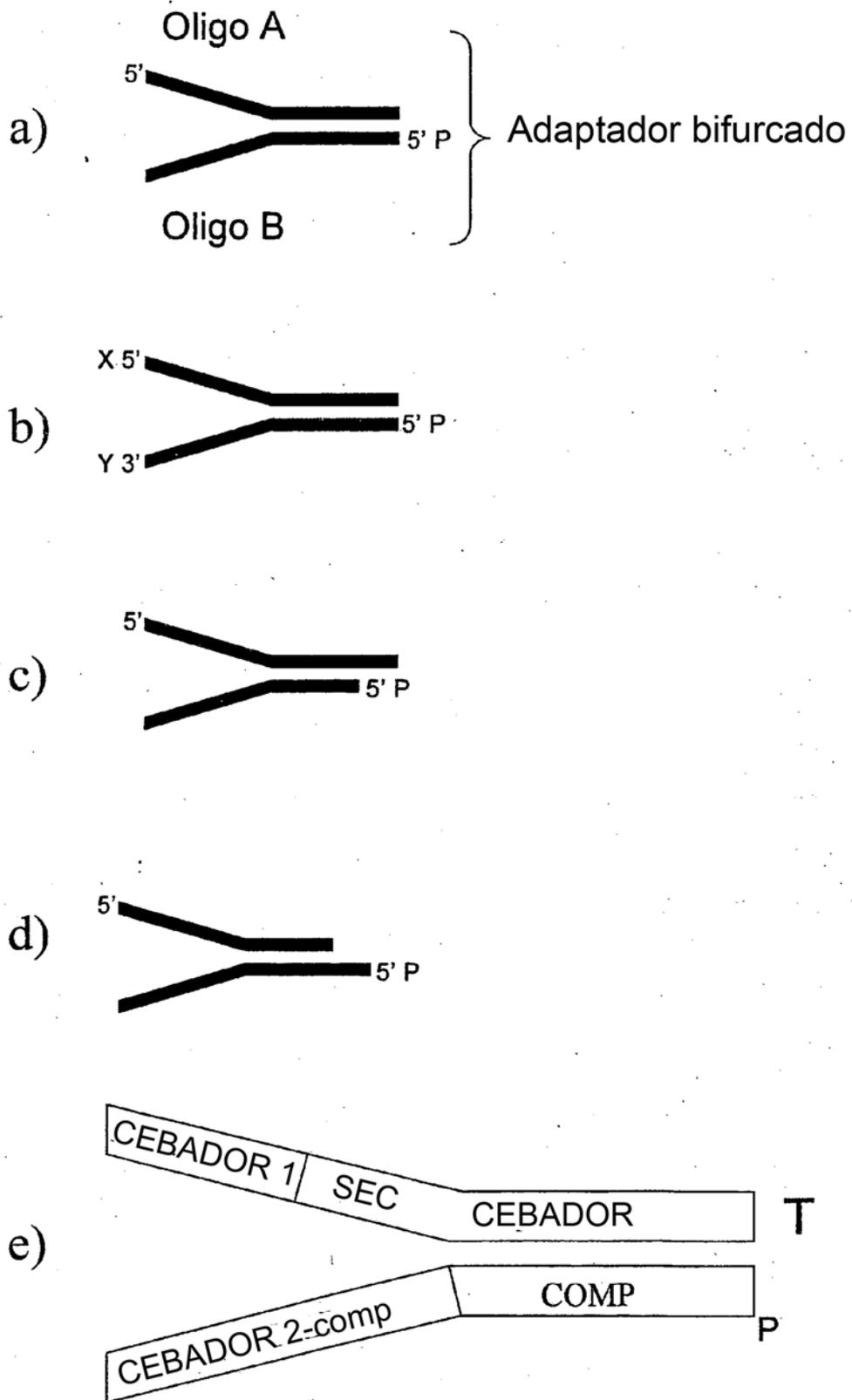


Fig. 2a

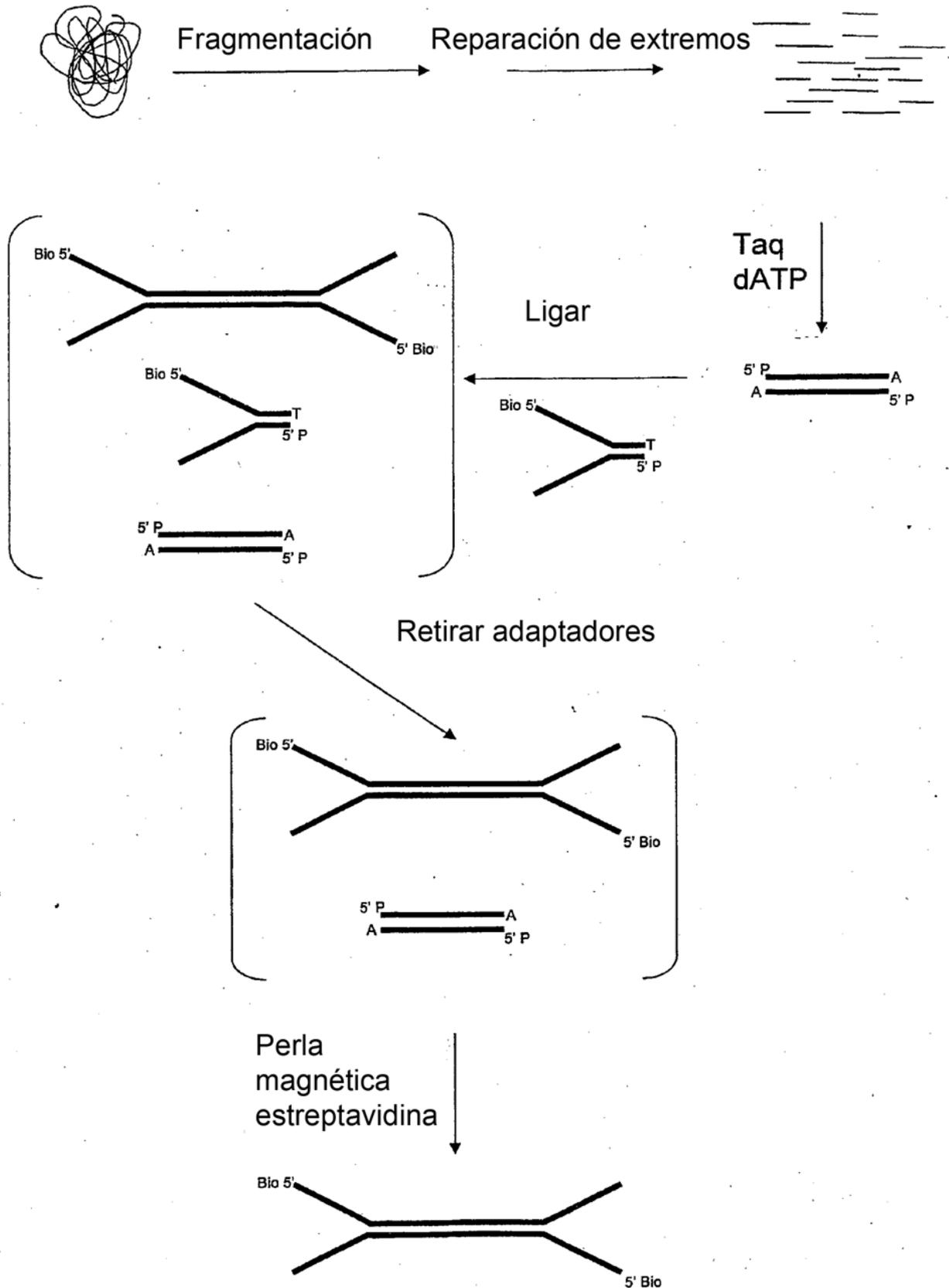


Fig. 2b

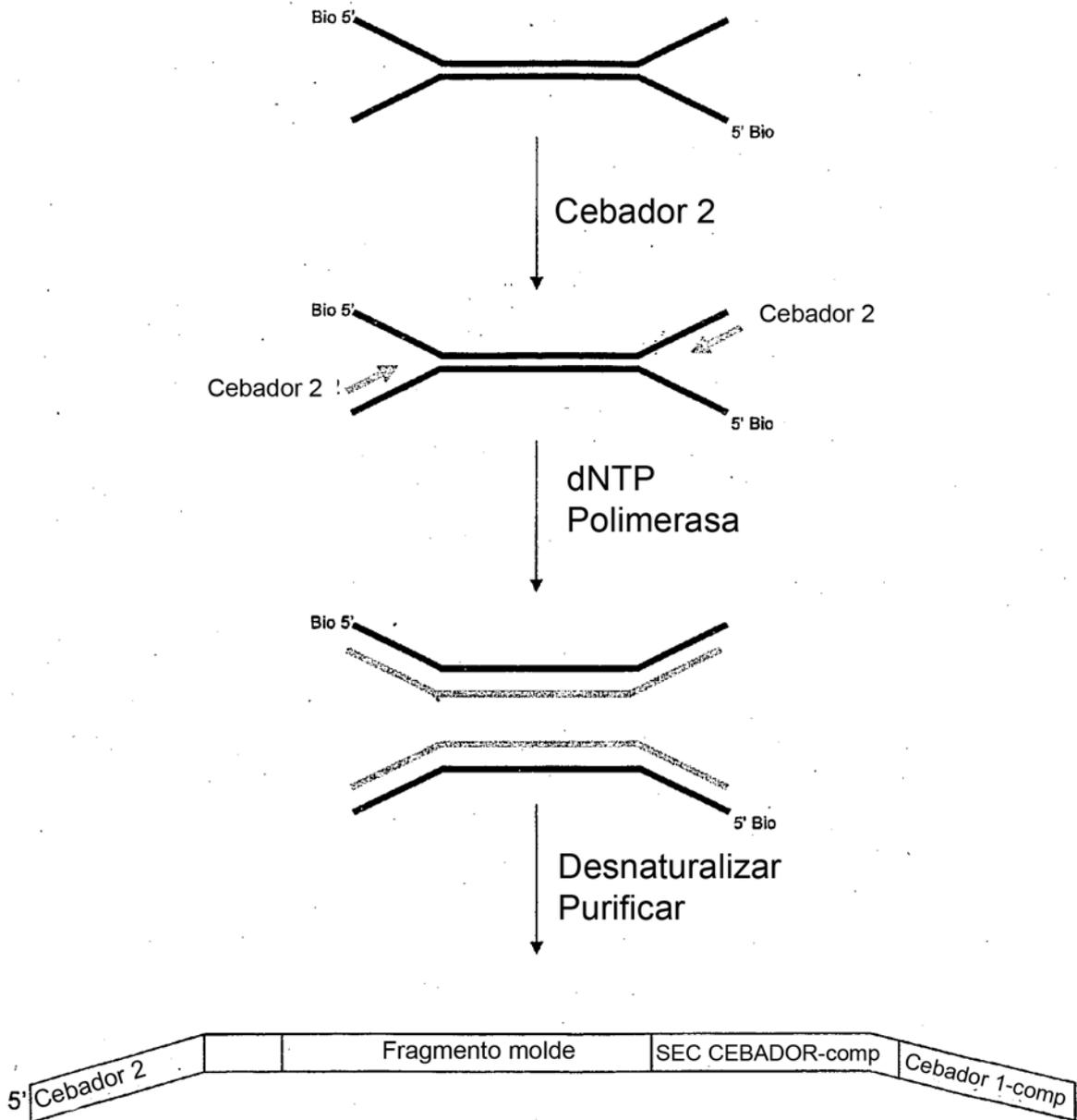


Fig. 3

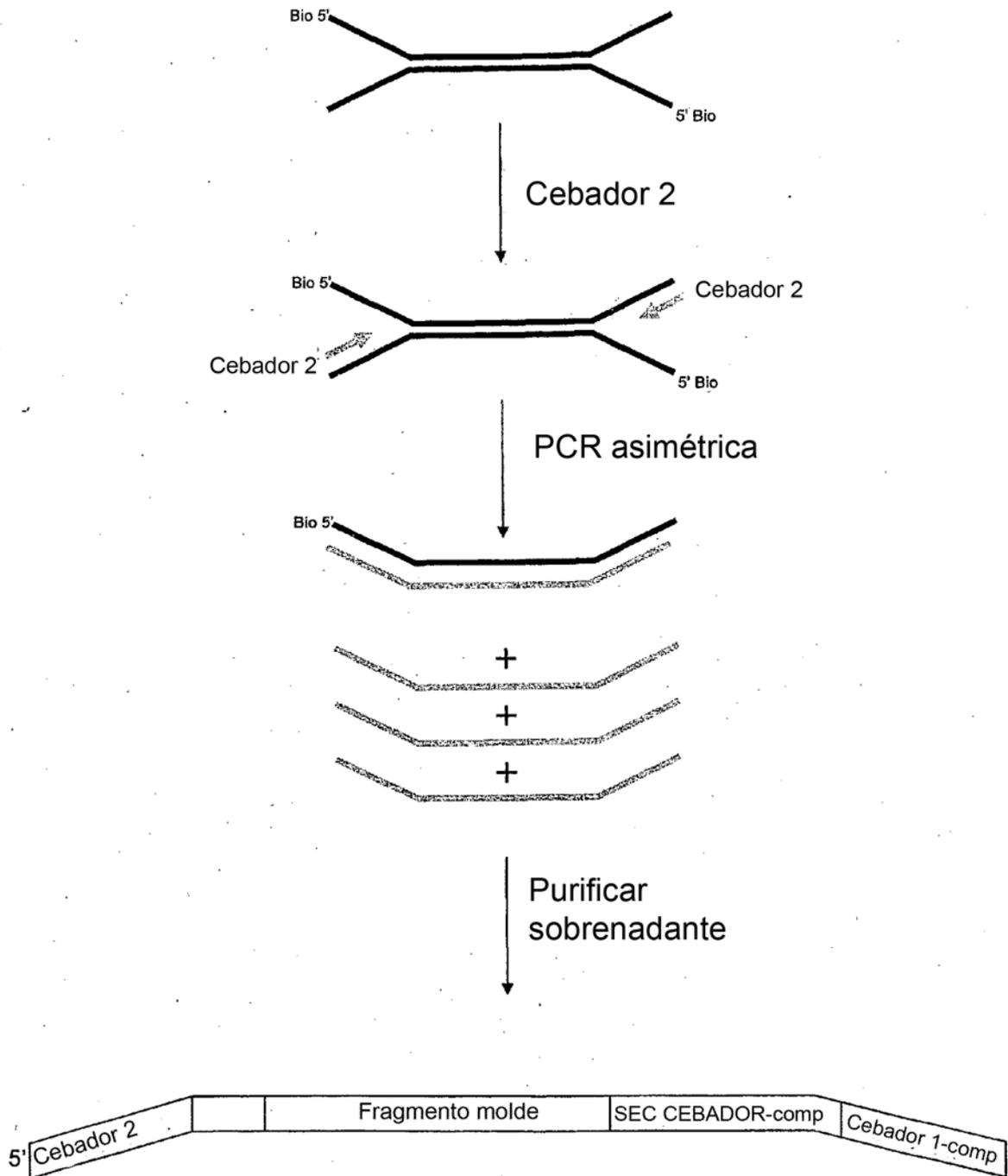


Fig. 4

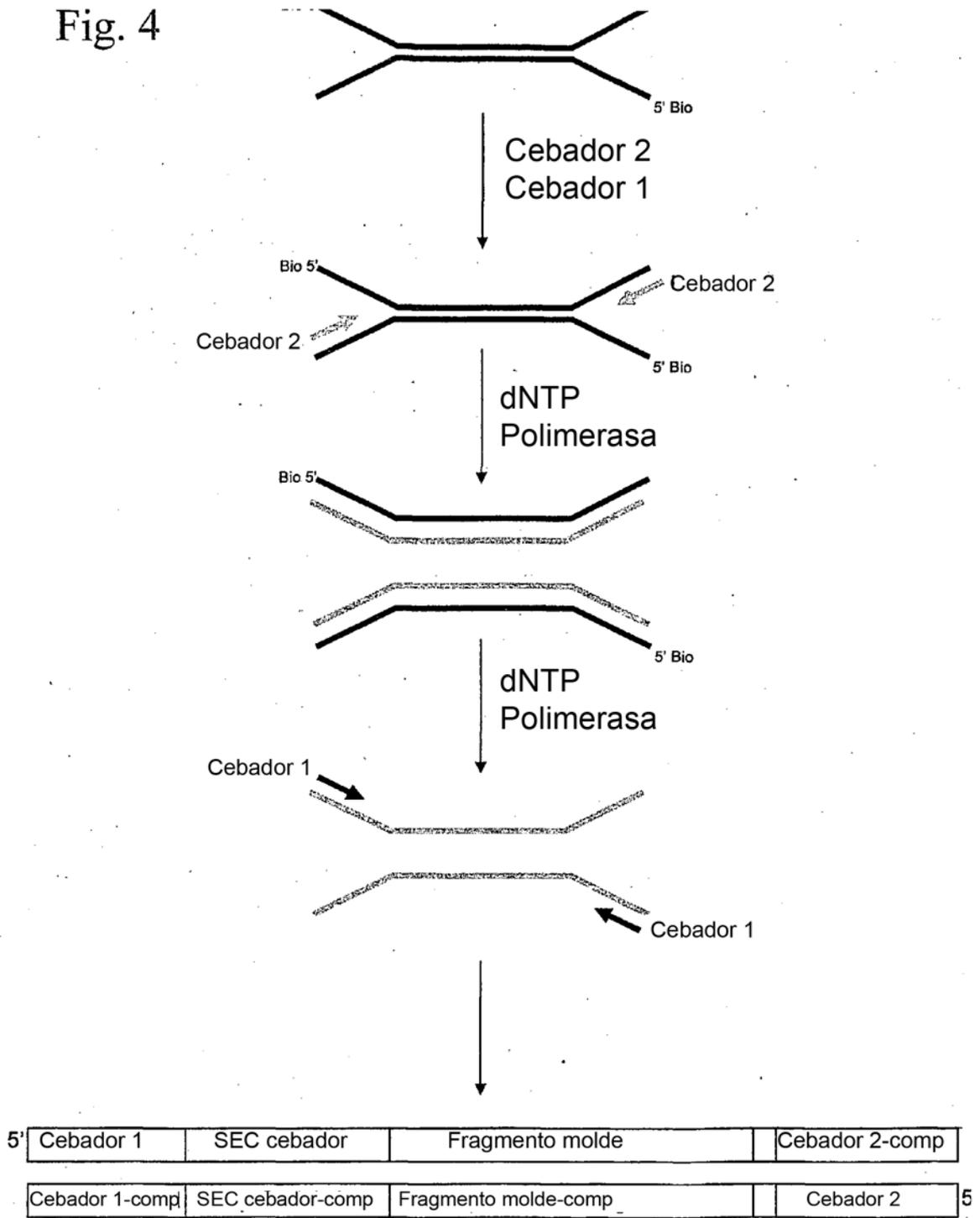
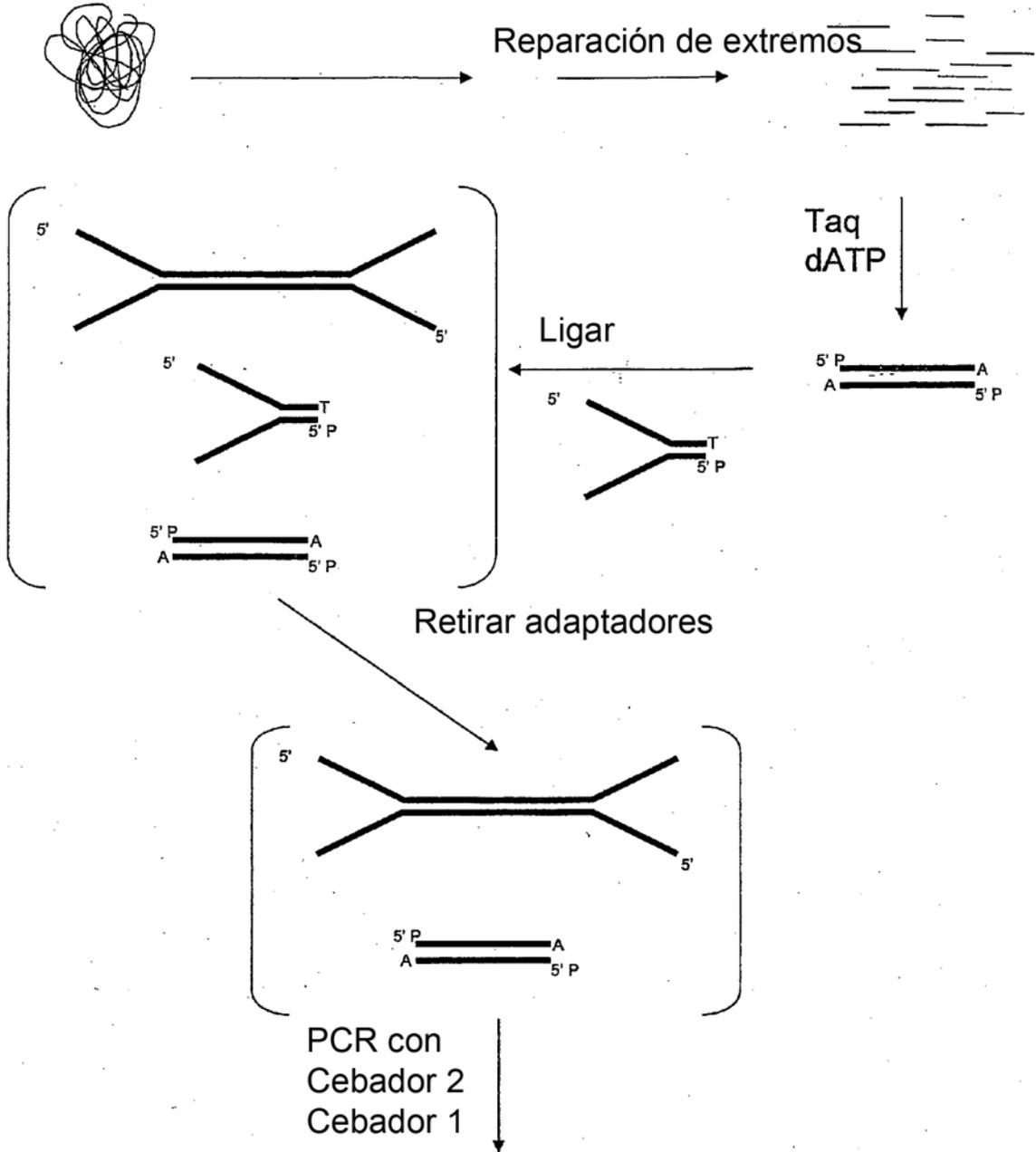


Fig. 5



5' Cebador 1	SEC cebador	Fragmento molde	Cebador 2-comp
Cebador 1-comp	SEC cebador-comp	Fragmento molde-comp	Cebador 2

Fig. 6

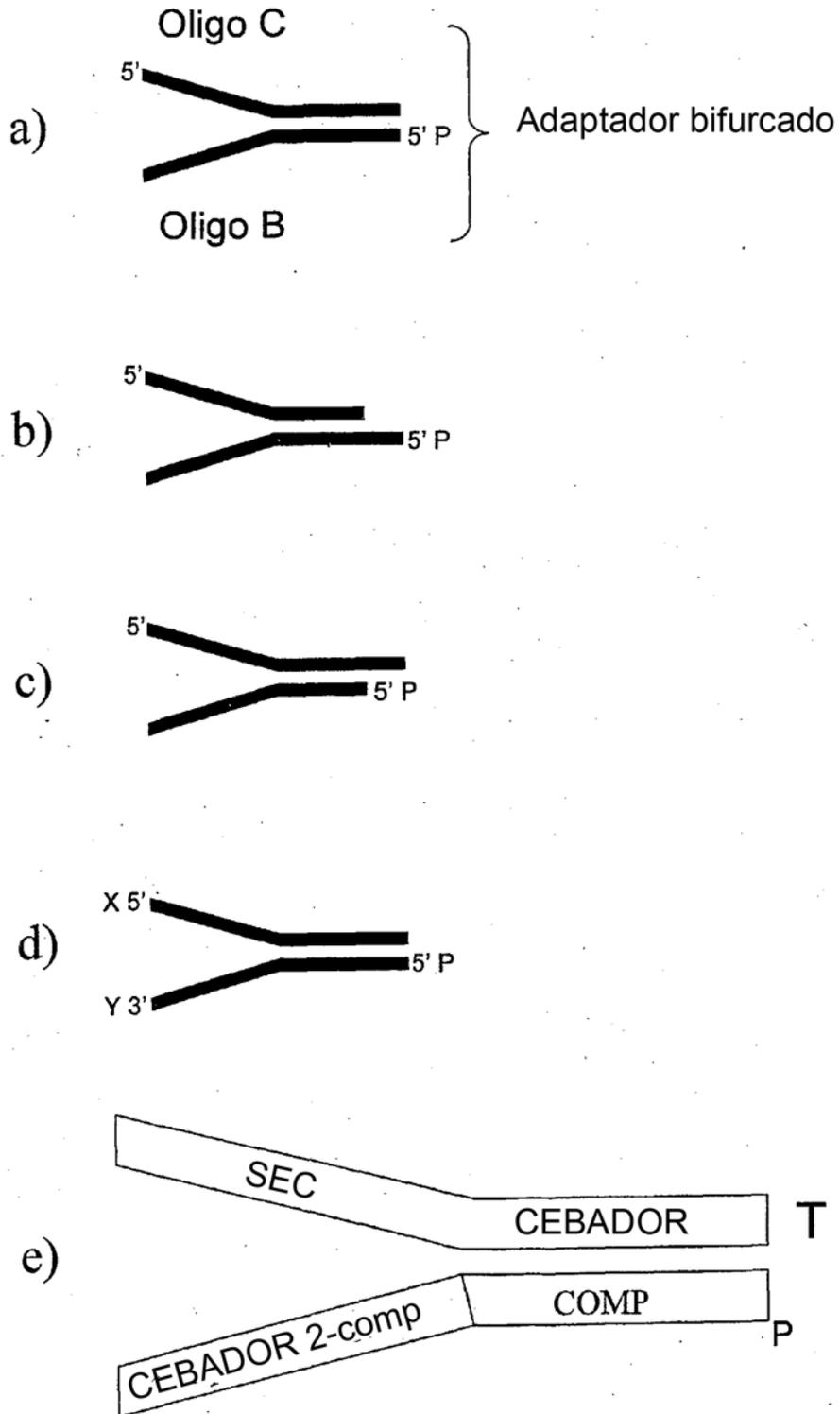


Fig. 7a

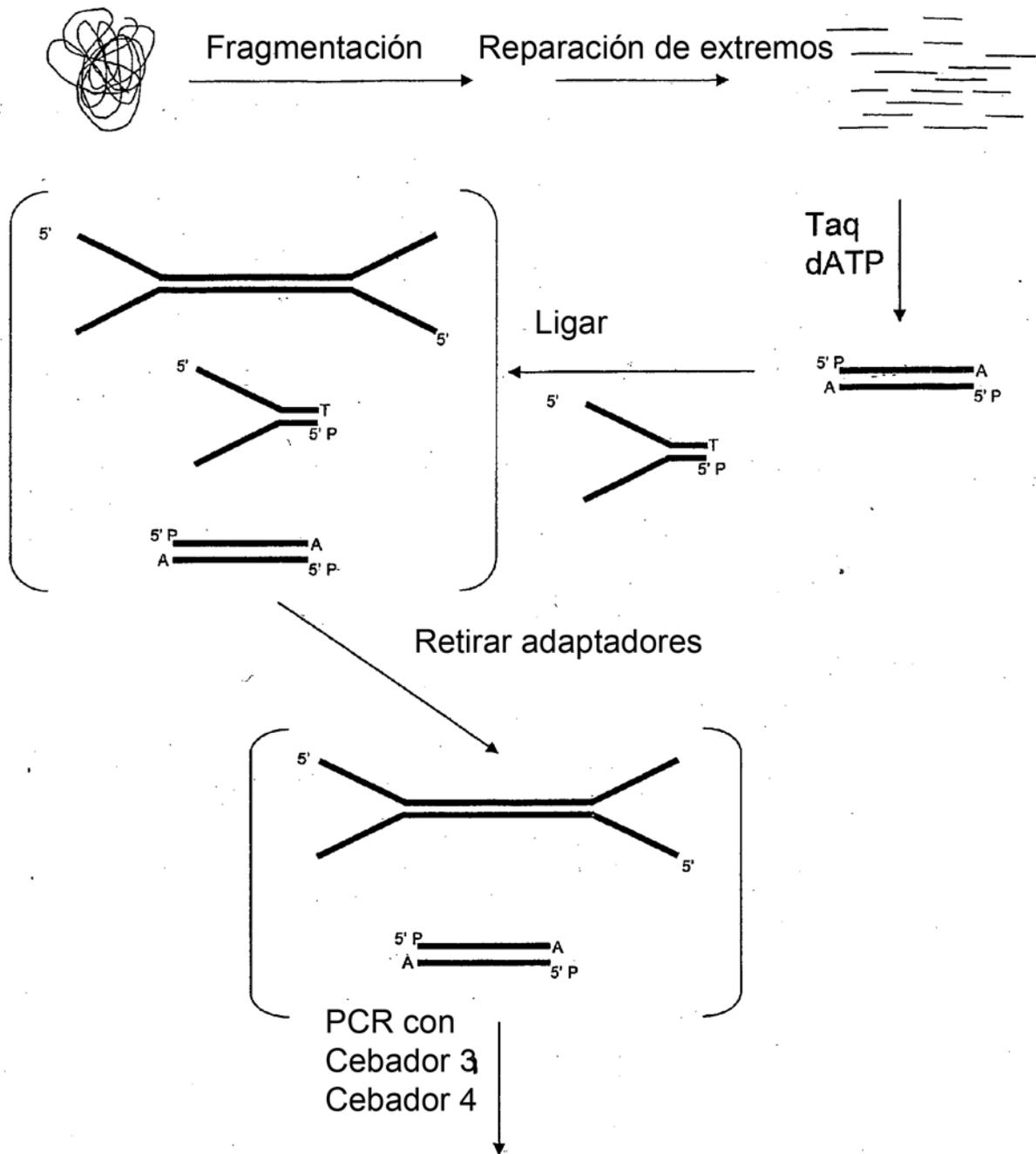


Fig. 7b

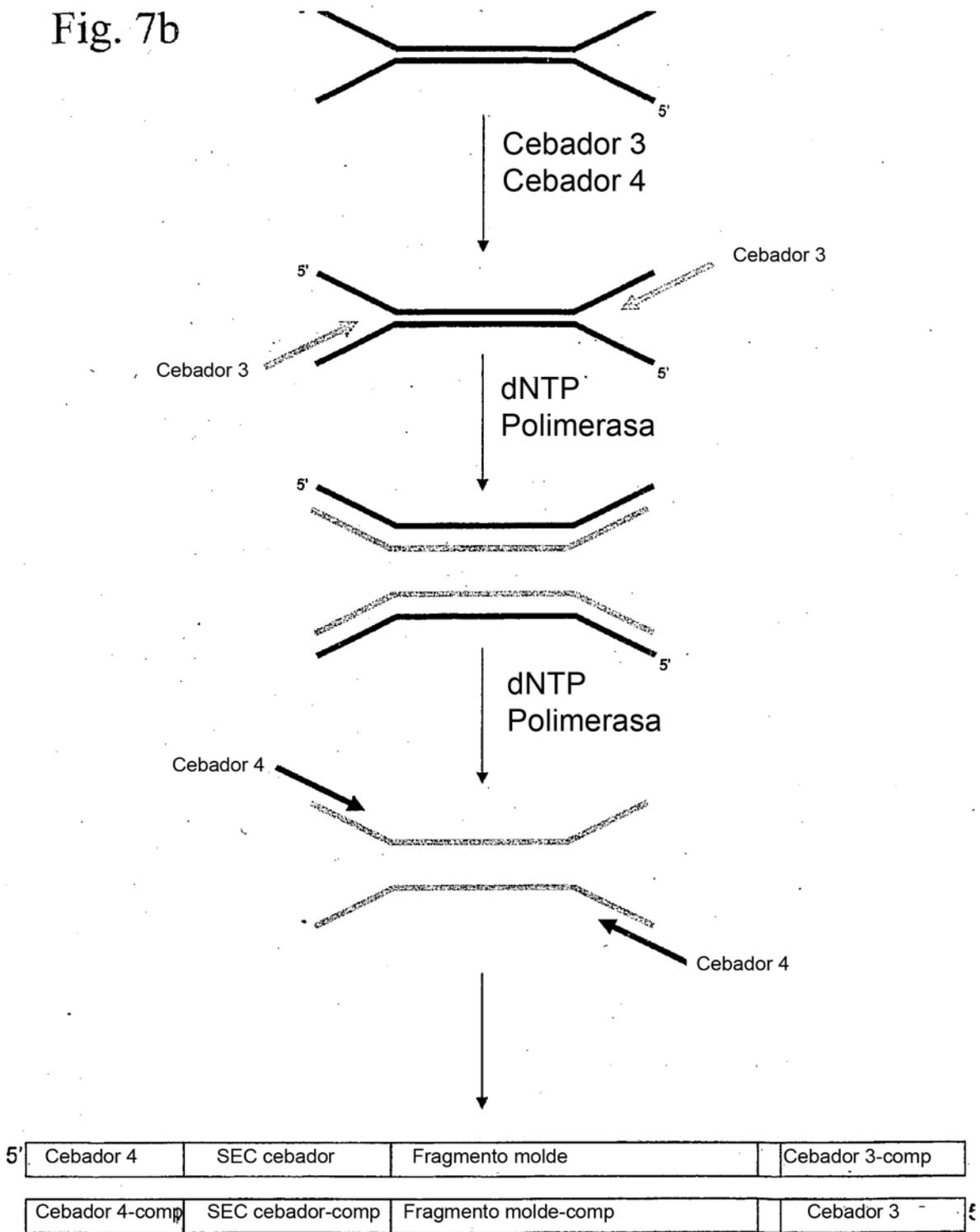


Fig. 8

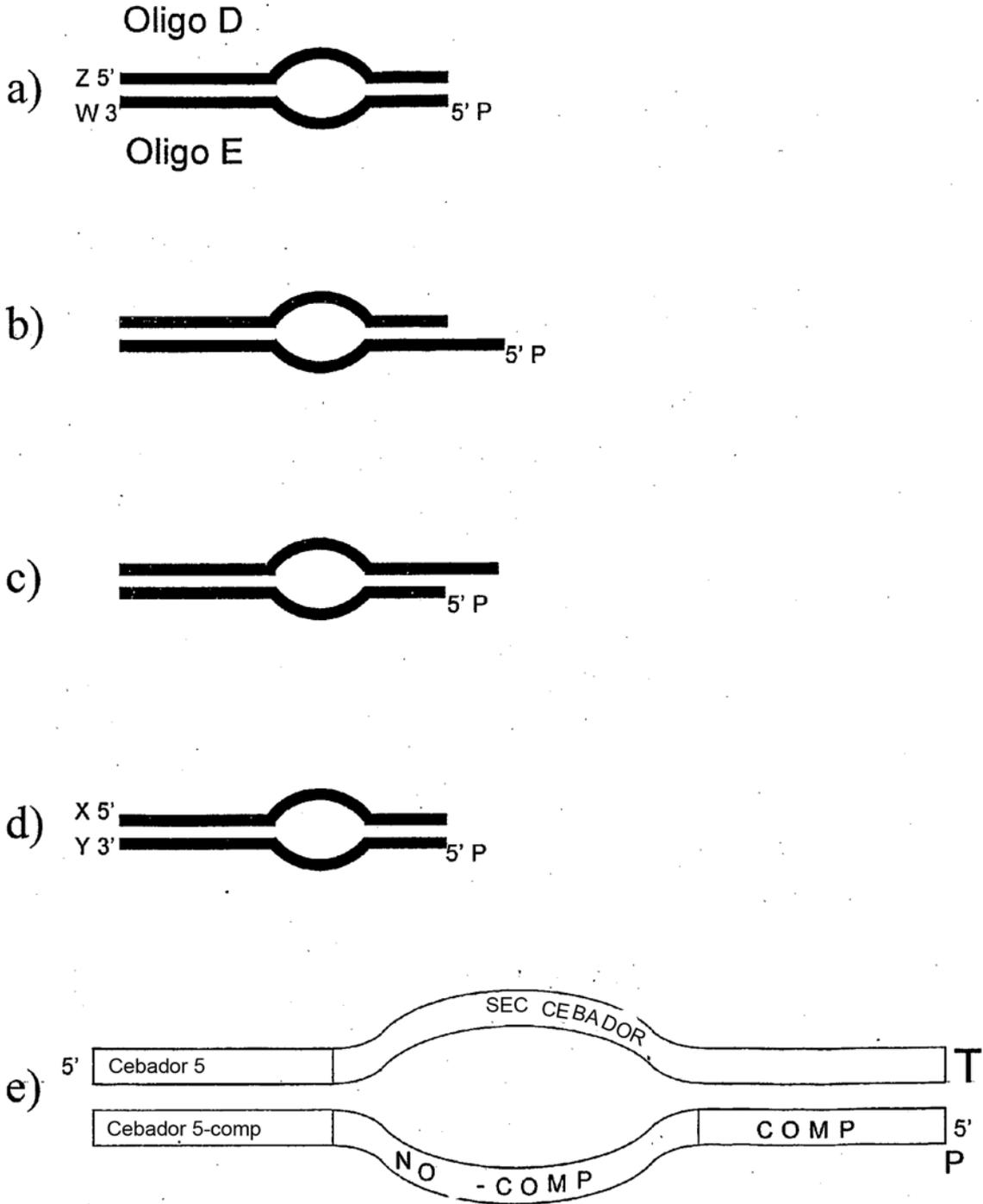


Fig. 9a

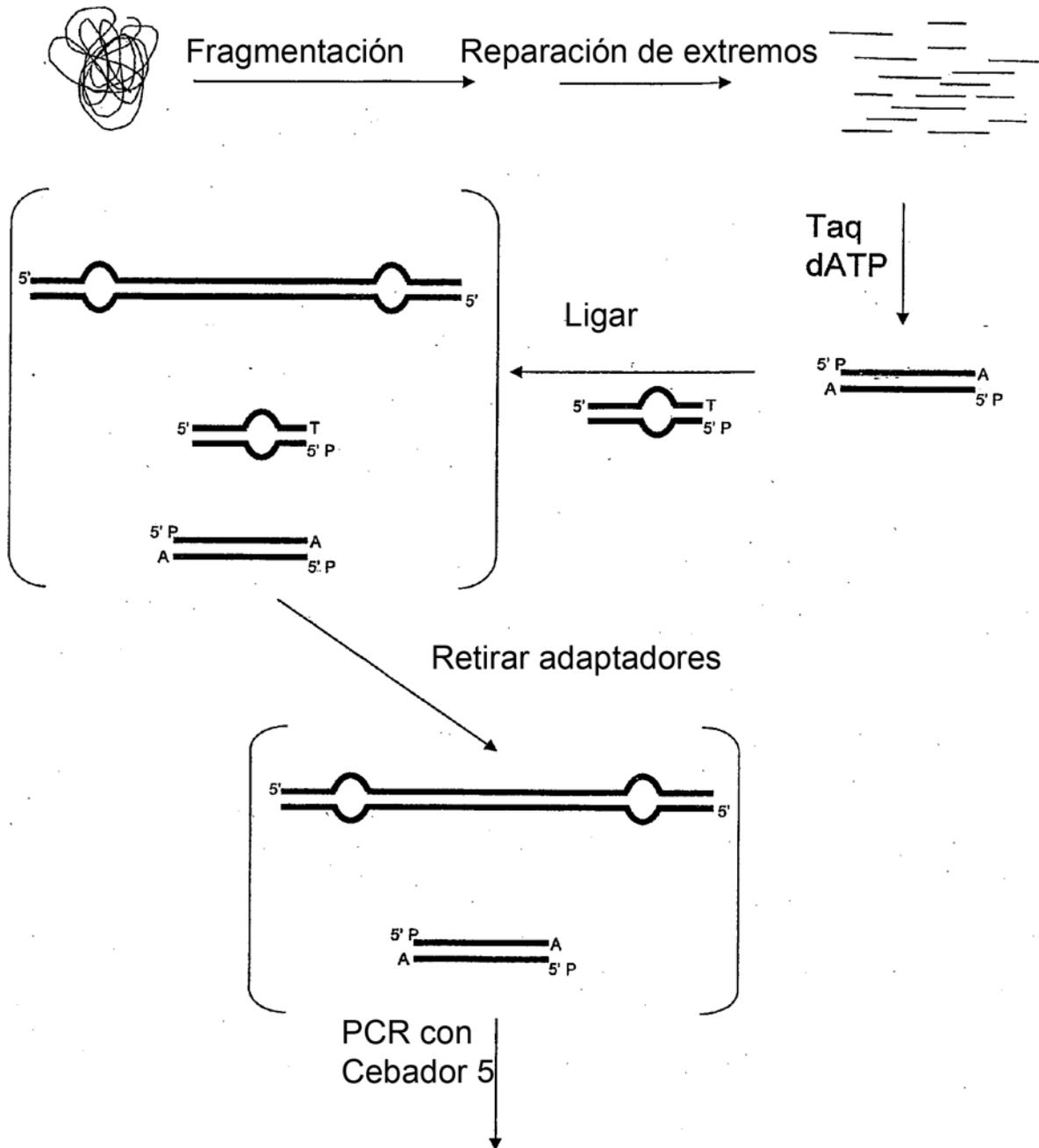
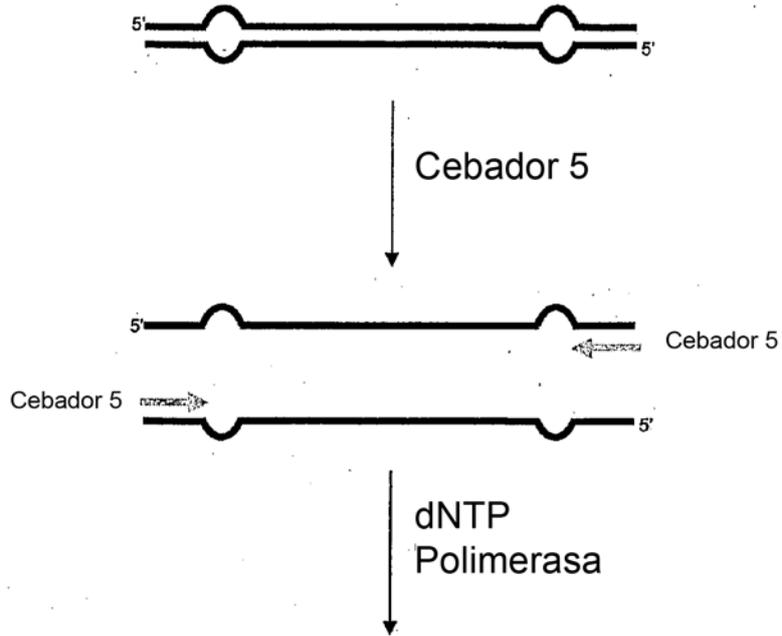


Fig. 9b



5'	Cebador 5	SEC cebador	Fragmento molde	Cebador 5-comp
	Cebador 5-comp	SEC cebador-comp	Fragmento molde-comp	Cebador 5