

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 626 058**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

**C12P 19/34** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **21.12.2012 PCT/US2012/071308**

87 Fecha y número de publicación internacional: **27.06.2013 WO13096802**

96 Fecha de presentación y número de la solicitud europea: **21.12.2012 E 12859854 (7)**

97 Fecha y número de publicación de la concesión europea: **12.04.2017 EP 2794927**

54 Título: **Cebadores y métodos de amplificación**

30 Prioridad:

**22.12.2011 US 201161578976 P**

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

**21.07.2017**

73 Titular/es:

**IBIS BIOSCIENCES, INC. (100.0%)  
2251 Faraday Avenue Suite 150  
Carlsbad, CA 92008, US**

72 Inventor/es:

**GRAY, PHILLIP N. y  
ESHOO, MARK W.**

74 Agente/Representante:

**IZQUIERDO BLANCO, María Alicia**

ES 2 626 058 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).

**Cebadores y métodos de amplificación****DESCRIPCIÓN****5 Campo de la invención**

La presente invención se refiere a procedimientos para realizar la amplificación (por ejemplo, amplificación del genoma completo) empleando cebadores que tienen un sitio de restricción en 5', una secuencia aleatoria en 3' (por ejemplo, un hexámero aleatorio) y una secuencia de código de barras identificable.

10

**Antecedentes**

En muchos campos de investigación, tales como diagnóstico genético, investigación del cáncer o medicina forense, la escasez de ADN genómico puede ser un factor muy limitante del tipo y la cantidad de pruebas genéticas que se pueden realizar en una muestra. Un enfoque diseñado para superar este problema es la amplificación del genoma completo. El objetivo es amplificar una muestra limitada de ADN de una manera no específica para generar una nueva muestra que sea indistinguible del original pero con una concentración de ADN más alta. El objetivo de una técnica típica de amplificación del genoma completo es amplificar una muestra hasta un nivel de microgramos respetando la representación de la secuencia original.

15

20

Los primeros procedimientos de amplificación del genoma completo se describieron en 1992 y se basaron en los principios de la reacción en cadena de la polimerasa. Zhang y col. (Zhang, L., y col., Proc. Natl. Acad. Sci. USA, 1992, 89: 5847–5851) Desarrollaron la técnica de PCR de extensión por cebadores (PEP) y Telenius y col., (Telenius y col., Genomics. 1992, 13(3):718–25) diseñaron el procedimiento de PCR cebado con oligonucleótidos degenerados (DOP-PCR). La PEP implica un número alto de ciclos de PCR, generalmente usando Taq polimerasa y cebadores aleatorios de 15 bases que hibridan a una temperatura de rigurosidad baja. La DOP-PCR es un procedimiento que generalmente usa Taq polimerasa y oligonucleótidos semidegenerados que se unen a una temperatura de hibridación baja a aproximadamente un millón de sitios dentro del genoma humano. A los primeros ciclos les sigue un gran número de ciclos con una temperatura de hibridación más alta, lo que permite solamente la amplificación de los fragmentos que se marcaron en la primera etapa.

25

30

La amplificación de desplazamiento múltiple (MDA, también conocida como amplificación por desplazamiento de cadena, SDA) es un procedimiento isotérmico no basado en PCR fundamentado en la hibridación de hexámeros aleatorios al ADN desnaturalizado, seguido de una síntesis por desplazamiento de cadena a temperatura constante (Blanco y col., 1989, J. Biol. Chem. 264:8935–40; Dean, F.B. y col., (2002) Comprehensive human genome amplification using multiple displacement amplification; Proc. Natl. Acad. Sci. USA 99:5261; y Van, J. y col., (2004) Assessment of multiple displacement amplification in molecular epidemiology. Biotechniques 37, 136). Se ha aplicado a muestras pequeñas de ADN genómico, dando lugar a la síntesis de ADN de peso molecular alto con una desviación de la representación de secuencia limitada (Lizardi y col., Nature Genetics 1998, 19, 225–232; Dean y col., Proc. Natl. Acad. Sci. U.S.A. 2002, 99, 5261–5266). A medida que el ADN se sintetiza mediante desplazamiento de cadena, se produce un número gradualmente creciente de acontecimientos de cebado, formando una red de estructuras de ADN hiperramificadas. La reacción puede ser catalizada mediante la ADN polimerasa de Phi29 o mediante el fragmento grande de la Bst ADN polimerasa. La ADN polimerasa de Phi29 posee una actividad de corrección que da como resultado tasas de error 100 veces más bajas que la Taq polimerasa. Sin embargo, los procedimientos de tipo MDA requieren muchas horas (por ejemplo, 6 horas) para generar una amplificación suficiente.

35

40

45

El documento WO 2011/151777 (Tullo y col.) se refiere a un procedimiento para preparar y amplificar bibliotecas de ADNc representativas para secuenciación de alto rendimiento de próxima generación y a kits y cartuchos para kits de automatización que utilizan el procedimiento.

50

**Sumario de la invención**

La invención proporciona un procedimiento para generar ácido nucleico amplificado a partir de ARN que comprende:

a) exponer una secuencia molde de ARN a un conjunto de cebadores en condiciones de transcripción inversa, de tal manera que se genere una población mixta de las primeras cadenas de ADNc, en la que dicho conjunto de cebadores comprende cebadores individuales, comprendiendo cada uno: (i) un sitio de secuencia de restricción en 5', (ii) una secuencia hexamérica aleatoria en 3' y (iii) una secuencia de código de barras localizada entre el sitio de secuencia de restricción en 5' y la secuencia hexamérica aleatoria en 3', en la que dicho sitio de secuencia de restricción en cada uno de dichos cebadores individuales es idéntico y en la que dicho conjunto de cebadores comprende todas o casi todas las posibles secuencias hexaméricas aleatorias y en la que cada uno de dichas primeras cadenas del ADNc tiene uno de dichos cebadores individuales en su extremo 5'; (b) exponer dicha población mixta de las primeras cadenas del ADNc a dicho conjunto de cebadores en condiciones de polimerización, de tal manera que se genera una población mixta de moléculas de ADNc de doble cadena, (c) digerir dicha población mixta de moléculas de ADNc de doble cadena con una enzima de restricción específica de dicho sitio de secuencia de restricción en 5', (d) tratar dicha población mixta de moléculas de ADNc de doble cadena con un

55

60

65

agente de unión, de tal manera que las moléculas individuales de ADNc de doble cadena se ligan entre sí para formar una población mixta de concatémeros; (e) exponer dicha población mixta de concatémeros a cebadores aleatorios en condiciones de amplificación del genoma completo de tal manera que se genere ácido nucleico amplificado; (f) digerir, al menos parcialmente, dicho ácido nucleico amplificado con la misma enzima de restricción específica de dicho sitio de secuencia de restricción en 5', generando de este modo una pluralidad de secuencias digeridas; y (g) ligar secuencias adaptadoras de la secuenciación a los extremos de dicha pluralidad de secuencias digeridas para generar una población mixta de moldes de secuenciación ligados al adaptador, en los que dichas secuencias adaptadoras de la secuenciación contienen un sitio de enzima de restricción que es idéntico al sitio de secuencia de restricción en 5' en el conjunto de cebadores.

En el presente documento se divulgan procedimientos, composiciones y kits para realizar amplificación (por ejemplo, amplificación del genoma completo) empleando cebadores que tienen un sitio de restricción en 5', una secuencia aleatoria en 3' (por ejemplo, un hexámero aleatorio) y una secuencia de código de barras identificable. En ciertas realizaciones, la amplificación genera secuencias amplificadas individuales que se ligan entre sí para formar concatémeros que contienen al menos dos secuencias amplificadas (por ejemplo, no contiguas en la secuencia diana original) que están separadas por las secuencias de códigos de barras. En realizaciones particulares, se secuencia una pluralidad de concatémeros y se alinean con un algoritmo de alineación que utiliza las secuencias de códigos de barras para identificar uniones artificiales entre secuencias amplificadas.

En algunas realizaciones, en el presente documento se divulgan procedimientos para generar ácido nucleico amplificado a partir de ARN, que comprende: a) exponer una secuencia molde de ARN a un conjunto de cebadores en condiciones de transcripción inversa, de tal manera que se genere una población mixta de las primeras cadenas de ADNc, en la que el conjunto de cebadores comprende cebadores individuales, comprendiendo cada uno: i) un sitio de secuencia de restricción en 5', ii) una secuencia aleatoria en 3' (por ejemplo, una secuencia pentamérica, una secuencia hexamérica o una secuencia aleatoria más larga) y iii) una secuencia de código de barras (por ejemplo, secuencias de 2-15 bases identificables o secuencias de 5-10 bases identificables), en la que el conjunto de cebadores comprende todas o casi todas las posibles secuencias aleatorias, pentaméricas, hexaméricas o más largas, y en la que cada una de las primeras cadenas del ADNc tiene uno de los cebadores individuales en su extremo 5'; b) exponer la población mixta de primeras cadenas del ADNc al conjunto de cebadores en condiciones de polimerización, de manera que se genera una población mixta de moléculas de ADNc de doble cadena, c) digerir la población mixta de moléculas de ADNc de doble cadena con una enzima de restricción específica del sitio de la secuencia de restricción en 5', d) tratar la población mixta de moléculas de ADNc de doble cadena con un agente de unión, de forma tal que moléculas individuales de ADNc de doble cadena se ligan entre sí para formar una población mixta de concatémeros; y e) exponer la población mixta de concatémeros a cebadores aleatorios en condiciones de amplificación del genoma completo de tal manera que se genera ácido nucleico amplificado. En ciertas realizaciones, los concatémeros contienen dos, tres, cuatro, cinco, seis o más de las moléculas de ADNc de doble cadena individuales.

En ciertas realizaciones, el molde de ARN tiene una longitud inferior a 2.000 bases. En otras realizaciones, la población mixta de concatémeros comprende concatémeros individuales que tienen aproximadamente 2.000 bases de longitud o más. En realizaciones adicionales, las secuencias adaptadoras de secuenciación contienen un sitio de enzima de restricción que es idéntico al sitio de secuencia de restricción en 5' en los cebadores. En realizaciones adicionales, los procedimientos comprenden además digerir, al menos parcialmente, el ácido nucleico amplificado con una enzima de restricción específica del sitio de secuencia de restricción en 5', generando de este modo una pluralidad de secuencias digeridas.

En otras realizaciones, los procedimientos comprenden además ligar secuencias adaptadoras de la secuenciación a los extremos de la pluralidad de secuencias digeridas para generar una población mixta de moldes de secuenciación ligadas al adaptador. En ciertas realizaciones, las secuencias adaptadoras de la secuenciación contienen un sitio de enzima de restricción. En realizaciones concretas, el sitio de restricción en los adaptadores de la secuenciación es el mismo que el sitio de restricción presente en los cebadores. En otras realizaciones, las secuencias adaptadoras de la secuenciación son secuencias en horquilla.

En algunas realizaciones, la pluralidad de secuencias digeridas comprende secuencias digeridas individuales que contienen, cada una, las secuencias de bases de solo una de las moléculas de ADNc de doble cadena de la población mixta de moléculas de ADNc de doble hebra. En otras realizaciones, la pluralidad de secuencias digeridas comprende secuencias digeridas individuales que contienen cada una de ellas las secuencias de bases de dos o más de las moléculas de ADNc de doble cadena de la población mixta de moléculas de ADNc de doble hebra, en la que las secuencias de bases están separadas entre sí por las secuencias de código de barras (por ejemplo, si hay dos secuencias, hay un código de barras que las separa, si hay tres secuencias (o más) hay una secuencia de códigos de barras entre cada secuencia). En realizaciones concretas, los procedimientos comprenden además secuenciar al menos una de las secuencias digeridas individuales para generar información de secuencia electrónica y procesar la información de secuencia electrónica con un algoritmo de alineación en el que las secuencias de códigos de barras se usan para identificar uniones artificiales entre las secuencias de base de dos o más de las moléculas de ADNc de doble cadena.

En ciertas realizaciones, la secuenciación se lleva a cabo mediante un procedimiento seleccionado del grupo que consiste en: secuenciación dideoxi de Sanger, 454-pirosecuenciación, secuenciación Solexa/Illumina, secuenciación de moléculas verdaderas Helicos, secuenciación SMRT de Pacific Biosciences o secuenciación de Ion Torrent. En realizaciones concretas, el sitio de secuencia de restricción en 5' en cada uno de los cebadores individuales es idéntico. En realizaciones adicionales, la secuencia de código de barras en cada uno de los cebadores individuales es idéntica.

En algunas realizaciones, en el presente documento se divulgan procedimientos para generar ácido nucleico amplificado a partir de ADN, que comprende: a) tratar una población mixta de secuencias molde de ADN con un agente de unión, de forma tal que las secuencias molde de ADN individuales (por ejemplo, dos, tres, cuatro o más) se ligan entre sí para formar una población mixta de concatémeros, en la que la población mixta de secuencias molde de ADN comprende diferentes secuencias molde de ADN individuales; y b) exponer los concatémeros a un conjunto de cebadores en condiciones de amplificación del genoma completo de manera que se genere una población mixta de moléculas de ADN de doble cadena amplificadas, en la que el conjunto de cebadores comprende cebadores individuales que comprenden cada uno: i) un sitio de secuencia de restricción 5', ii) una secuencia aleatoria en 3' (por ejemplo, pentámero o hexámero o secuencia más larga) y iii) una secuencia de código de barras, en la que el conjunto de cebadores comprende todas o casi todas las posibles secuencias pentaméricas o hexaméricas aleatorias (por ejemplo, todas o el 90 %... 95 % ... 99 % ... de cada hexámero aleatorio posible).

En ciertas realizaciones, los procedimientos comprenden además: c) digerir al menos parcialmente la población mixta de moléculas de ADN de doble cadena amplificadas con una enzima de restricción específica para el sitio de secuencia de restricción en 5', generando de este modo una pluralidad de secuencias digeridas. En realizaciones adicionales, los procedimientos comprenden además: d) ligar secuencias adaptadoras de la secuenciación a los extremos de la pluralidad de secuencias digeridas para generar una población mixta de moldes de secuenciación ligadas al adaptador. En otras realizaciones, las diferentes secuencias molde de ADN individuales son de menos de 2.000 bases de longitud (por ejemplo, menos de 2.000 bases, menos de 1.500 bases, menos de 1.000 bases, menos de 500 bases, menos de 250 bases o menos de 150 bases; o entre 100-1000 bases o entre 250-1.500 bases). En realizaciones adicionales, la población mixta de concatémeros comprende concatémeros individuales que tienen aproximadamente 2.000 bases de longitud o más (por ejemplo, 2.000 bases..., 2.500 bases ... 3.000 bases, 4.000 bases o más). En otras realizaciones, las secuencias adaptadoras de la secuenciación contienen un sitio de enzima de restricción (por ejemplo, el mismo sitio que el presente en los cebadores). En otras realizaciones, las secuencias adaptadoras de la secuenciación son secuencias en horquilla.

En realizaciones particulares, la pluralidad de secuencias digeridas comprende secuencias digeridas individuales que contienen cada una de las secuencias de bases de solo una de las diferentes secuencias molde de ADN individuales. En otras realizaciones, la pluralidad de secuencias digeridas comprende secuencias digeridas individuales que contienen cada una las secuencias de bases de dos o más (por ejemplo, 2, 3, 4, 5, 6 o más) de las diferentes secuencias molde de ADN individuales, en las que las secuencias de bases están separadas entre sí por las secuencias de códigos de barras. En realizaciones adicionales, la población mixta de moldes de secuenciación ligadas a adaptadores comprende moldes de secuenciación ligadas a adaptadores individuales, en los que el procedimiento comprende además secuenciar al menos uno (o la mayoría o la totalidad) de los moldes de secuenciación ligados a adaptadores individuales para generar información de secuencia electrónica y procesar la información de secuencia electrónica con un algoritmo de alineación en el que las secuencias de códigos de barras se usan para identificar uniones artificiales entre las secuencias de bases de dos o más de las secuencias de molde de ADN individuales. En ciertas realizaciones, el sitio de secuencia de restricción en 5' en cada uno de los cebadores individuales es idéntico. En realizaciones adicionales, la secuencia de código de barras en cada uno de los cebadores individuales es idéntica.

En algunas realizaciones, en el presente documento se divulgan composiciones que comprenden un conjunto de cebadores, en el que el conjunto de cebadores comprende cebadores individuales que comprenden cada uno: i) un sitio de secuencia de restricción 5', ii) una secuencia aleatoria pentamérica, hexamérica o mayor aleatoria y iii) una secuencia de código de barras, en la que el conjunto de cebadores comprende todas o casi todas las posibles secuencias aleatorias pentaméricas, hexaméricas o mayores. En ciertas realizaciones, el sitio de secuencia de restricción en 5' en cada uno de los cebadores individuales es idéntico. En otras realizaciones, la secuencia de código de barras en cada uno de los cebadores individuales es idéntica.

En algunas realizaciones, en el presente documento se divulgan kits y sistemas que comprenden: a) una composición que comprende un conjunto de cebadores, en la que el conjunto de cebadores comprende cebadores individuales que comprenden cada uno: i) un sitio de secuencia de restricción 5', ii) una secuencia aleatoria pentamérica, hexamérica o mayor aleatoria y iii) una secuencia de código de barras, en la que el conjunto de cebadores comprende todas o casi todas las posibles secuencias aleatorias pentaméricas, hexaméricas o mayores; y b) una polimerasa adecuada para realizar amplificación del genoma completa. En ciertas realizaciones, la polimerasa comprende Phi29 o el fragmento grande de la Bst ADN polimerasa, o enzima de funcionamiento similar.

En ciertas realizaciones, en el presente documento se divulgan procedimientos para generar alineaciones de secuencias, que comprenden: a) secuenciar una población mixta de concatémeros para generar información de

secuencia, en la que la población mixta de concatémeros comprende una pluralidad de concatémeros individuales, comprendiendo cada uno: i) al menos dos secuencias de bibliotecas diferentes del genoma de un organismo, en las que las al menos dos secuencias de biblioteca diferentes no son contiguas en el genoma; y ii) al menos una secuencia de códigos de barras situada entre las al menos dos secuencias de biblioteca diferentes; e b) introducir la información de secuencia en un sistema, en el que el sistema comprende: i) un procesador informático para recibir, procesar y comunicar datos, ii) un programa informático, incrustado en el procesador informático, que está configurado para procesar la información de secuencia para formar alineaciones de secuencia; c) procesar la información de secuencia con el programa informático de manera que las secuencias de códigos de barras se usen para formar alineaciones de secuencia identificando uniones artificiales entre las al menos dos secuencias de biblioteca diferentes; y d) comunicar el resultado del programa informático a un usuario. En realizaciones particulares, los procedimientos comprenden además una etapa antes de la etapa a) de generación de la población mixta de concatémeros mediante amplificación del genoma completo y unión de los productos de amplificación del genoma completo.

En algunas realizaciones, en el presente documento se divulgan composiciones que comprenden: un conjunto de secuencias de biblioteca, en el que dicho conjunto de secuencias de biblioteca comprende secuencias de biblioteca individuales que comprenden: i) al menos una secuencia de códigos de barras; ii) al menos dos insertos de ADN, en los que cada uno de dichos insertos de ADN es una porción amplificada de una secuencia diana y en los que dichos al menos dos insertos de ADN están separados por una de dichas secuencias de códigos de barras; iii) un sitio de secuencia de restricción, iv) una secuencia aleatoria pentamérica, hexamérica o una secuencia más larga y v) secuencias adaptadoras (por ejemplo, una en cada extremo), en las que dicho conjunto de secuencias de biblioteca comprende todas o casi todas las secuencias de una secuencia diana. En ciertas realizaciones, el sitio de la secuencia de restricción es adyacente a dicha al menos una secuencia de código de barras. En realizaciones adicionales, la secuencia aleatoria pentamérica o hexamérica más larga es adyacente a la al menos una secuencia de código de barras. En realizaciones adicionales, la secuencia diana tiene más de 2.000 bases de longitud. En realizaciones adicionales, la secuencia diana tiene más de 20.000 bases de longitud.

Los presentes procedimientos no están limitados por el sitio de la enzima de restricción endonucleasa o de la enzima que se utilice. En ciertas realizaciones, el sitio de restricción es reconocido por, o la enzima empleada es: BamH1, EcoR1, EcoR1I, HindII, HindIII, HinfI, HpaI, MspI y SmaI. En la materia se conocen muchos otros sitios de restricción y enzimas.

### Descripción de las figuras

Las figuras 1A-C muestran un diagrama de flujo de ejemplo de la utilización de los cebadores descritos en el presente documento para amplificar ARN y, en combinación con la amplificación de WGA y la unión de productos, para generar una biblioteca de secuenciación adaptada a cebadores. Las figuras 2A-B muestran un diagrama de flujo de ejemplo de la utilización de los cebadores descritos en el presente documento para amplificar ADN y generar una biblioteca de secuenciación adaptada a cebadores.

### Descripción detallada

En el presente documento se divulgan procedimientos, composiciones y kits para realizar amplificación (por ejemplo, amplificación del genoma completo) empleando cebadores que tienen un sitio de restricción en 5', una secuencia aleatoria en 3' (por ejemplo, un hexámero aleatorio) y una secuencia de código de barras identificable. En ciertas realizaciones, la amplificación genera secuencias amplificadas individuales que se ligan entre sí para formar concatémeros que contienen al menos dos secuencias amplificadas (por ejemplo, no contiguas originalmente en la secuencia diana original) que están separadas por las secuencias de códigos de barras. En realizaciones particulares, se secuencia una pluralidad de concatémeros y se alinean con un algoritmo de alineación que utiliza las secuencias de códigos de barras para identificar uniones artificiales entre secuencias amplificadas. En realizaciones concretas, en el presente documento se divulgan procedimientos y composiciones para la preparación de una biblioteca de moldes cortos de ARN y ADN para la secuenciación de próxima generación usando amplificación del genoma completo y enzimas de restricción.

La amplificación del genoma completo (WGA) de los virus ARN (u otras moléculas pequeñas de ARNsc) utilizando Phi29 se ha descrito en la bibliografía, así como un proceso muy ineficiente. Se ha informado de que la ADN polimerasa de Phi29 no puede amplificar ARN o fragmentos pequeños de ADNc de menos de 2.000 bases generadas a partir de la transcriptasa inversa. Berthet y col., BMC Molecular Biology, 2008, 9:77 describen un enfoque de este problema, generando fragmentos de ADNc de virus de ARN usando hexámeros aleatorios y uniendo los fragmentos de ADNc de doble cadena de una manera aleatoria mediante unión de extremos romos. Las moléculas de cana concatenadas se utilizaron como sustrato para Phi29. La principal desventaja de este enfoque es que crea fragmentos de ADN quiméricos con uniones artificiales en la biblioteca de WGA. Estos fragmentos de ADN quiméricos plantean un reto para los algoritmos de alineación creados para los datos de secuenciación de próxima generación y típicamente se descartan durante el proceso de alineación. El presente procedimiento incorpora secuencias de ADN específicas en el ADN concatenado, lo que permite una fácil identificación de las uniones artificiales. Como resultado, se pueden recuperar lecturas de secuencias individuales de la molécula de ADN

concatenada.

En ciertas realizaciones, en el presente documento se divulgan procedimientos que usan cebado aleatorio de ácido nucleico diana usando cebadores que contienen un sitio de enzima de restricción en el extremo 5'. Para el ARN, las moléculas de ADNc de doble cadena que son el resultado de los cebadores aleatorios y la transcriptasa inversa (u otra polimerasa capaz de realizar transcripción inversa), seguido de la síntesis de la segunda cadena, se digieren con una enzima de restricción que cortará los extremos de los hexámeros aleatorios y creará moléculas de ADNc con un sitio de enzima de restricción específico en los extremos 5' y 3' de cada molécula (véase, por ejemplo, la figura 1). A continuación, estas moléculas de ADNc se ligan para formar moldes grandes de ADNc concatenado para la amplificación del genoma completo. Después de la amplificación del genoma completo, los productos amplificados se digieren con la misma enzima de restricción, de modo que se producen fragmentos de ADNc flanqueados con el sitio de la enzima de restricción. Los cebadores aleatorios con el sitio de restricción tienen una secuencia de códigos de barras para distinguirlos de un sitio de restricción natural en el genoma de interés. A continuación, se diseñan adaptadores específicos de una plataforma de secuenciación en particular (por ejemplo, Pacific Biosciences, Illumina, Ion Torrent, 454, SOLiD, etc.) de modo que contengan el sitio de la enzima de restricción que coincide con el sitio contenido en la biblioteca de ADNc amplificada del genoma completo. Los adaptadores que contienen los sitios de restricción ("RE-adaptadores") se mezclan después con los fragmentos de ADNc amplificados del genoma completo que contienen los sitios de restricción ("RE-ADNc") y se ligan para producir una biblioteca de ADNc flanqueada con los adaptadores y lista para la secuenciación (Adaptador-RE-ADNc-RE-adaptador). Este procedimiento puede utilizarse, por ejemplo, con ADN o ARN, o ácidos nucleicos degradados de muestras/tejidos embebidos en parafina, fijados con formalina (FFPE).

Como se ha indicado anteriormente, una de las principales desventajas del enfoque de la técnica anterior es que crea uniones artificiales en la biblioteca de WGA que no pueden identificarse fácilmente mediante ordenador. Los algoritmos de secuenciación de próxima generación que alinean lecturas cortas están limitados en su capacidad para procesar datos con uniones artificiales. En ciertas realizaciones, el presente procedimiento genera dos tipos de productos en la biblioteca. El primer tipo contiene fragmentos aleatorios de ADN flanqueados por los RE-adaptadores, que no contienen uniones artificiales. El segundo tipo contiene múltiples fragmentos aleatorios de ADN flanqueados por los RE-adaptadores. Estos fragmentos de ADN concatenados están separados por una secuencia conocida (sitio de RE con código de barras, seguido de una secuencia aleatoria de X unidades) que se puede identificar fácilmente por ordenador. Como resultado, la secuencia leída puede dividirse en sus lecturas individuales, no concatenadas.

En ciertas realizaciones, el presente procedimiento puede usarse para tomar muestras desconocidas que contienen ácido nucleico (por ejemplo, ARN o ADN) y producir moldes para WGA usando phi29 (u otras polimerasas de WGA). Los adaptadores específicos para las plataformas de secuenciación (o secuenciación por síntesis) de próxima generación (Pacific Biosciences, Illumina, Ion Torrent, etc.) pueden ligarse a los datos de la biblioteca y de secuencias obtenidos de cantidades residuales de entrada.

Las secuencias de códigos de barras se usan en ciertos cebadores del presente procedimiento. La secuencia del código de barras puede ser cualquier secuencia identificable localizada entre el sitio de la enzima de restricción y la secuencia hexamérica aleatoria. Generalmente, estas secuencias tienen aproximadamente 5-10 nucleótidos de longitud. Su propósito es distinguir los sitios para enzimas de restricción introducidos desde el cebado con los oligos aleatorios de X unidades/código de barras/enzima de restricción de los sitios para las enzimas de restricción de origen natural en el genoma diana (que produciría una unión artificial). Los códigos de barras del ADN pueden variar ampliamente en tamaño y composiciones. Las siguientes referencias proporcionan una guía para seleccionar conjuntos de códigos de barras de oligonucleótidos apropiados para realizaciones particulares: Brenner, patente de Estados Unidos n.º 5.635,400; Brenner y col., Proc. Natl. Acad. Sci. 97:1665-1670 (2000), Shoemaker y col., Nature Genetics, 14:450-456 (1996); Morris y col., publicación de patente europea 0799897A1; Wallace, patente de Estados Unidos n.º 5.981.179; y similares. En diferentes aplicaciones del procedimiento, los códigos de barras de oligonucleótidos pueden tener cada uno una longitud dentro de un intervalo de 4 a 36 nucleótidos o de 6 a 30 nucleótidos, o de 8 a 20 o de 5 a 10 nucleótidos, respectivamente.

En ciertas realizaciones, se secuencian las secuencias amplificadas que se generan. El presente procedimiento no está limitado por la técnica de secuenciación empleada. A continuación se describen ejemplos de procedimientos de secuenciación. Entre los ejemplos no limitantes ilustrativos de las técnicas de secuenciación de ácidos nucleicos se incluyen, pero no se limitan a, secuenciación de terminación de cadena (Sanger), secuenciación por terminador colorante y procedimientos de secuenciación de la próxima generación.

La secuenciación de terminación de cadena utiliza la terminación específica de secuencia de una reacción de síntesis de ADN usando sustratos de nucleótidos modificados. La extensión se inicia en un sitio específico en el ADN molde utilizando un cebador oligonucleotídico radioactivo corto, u otro marcado, complementario del molde en esa región. El cebador oligonucleotídico se extiende utilizando una ADN polimerasa, cuatro bases de desoxinucleótidos estándar y una concentración baja de un nucleótido de terminación de cadena, más habitualmente un didesoxinucleótido. Esta reacción se repite en cuatro tubos separados, turnándose cada una de las bases como didesoxinucleótido. La incorporación limitada del nucleótido de terminación de cadena por la ADN polimerasa da

como resultado una serie de fragmentos de ADN relacionados que se terminan solamente en las posiciones en las que se utiliza dicho didesoxinucleótido particular. Para cada tubo de reacción, los fragmentos se separan por tamaño mediante electroforesis en un gel de poliacrilamida plano o un tubo capilar relleno con un polímero viscoso. La secuencia se determina leyendo qué calle produce una marca visualizada desde el cebador marcado mientras se escanea desde la parte superior del gel hasta la parte inferior.

La secuenciación por terminador colorante marca, alternativamente los terminadores. La secuenciación completa se puede llevar a cabo en una sola reacción marcando cada uno de los terminadores de cadena de los didesoxinucleótidos con un colorante fluorescente separado, que brilla a una longitud de onda diferente.

Un conjunto de procedimientos conocidos como técnicas de "secuenciación de próxima generación" han surgido como alternativas a los procedimientos de secuenciación de Sanger y por terminador colorante (Voelkerding y col., *Clinical Chem.*, 55: 641–658, 2009; MacLean y col., *Nature Rev. Microbiol.*, 7: 287–296). La tecnología de secuenciación de próxima generación permite la secuenciación *de novo* de genomas completos para determinar la secuencia de ácidos nucleicos primaria de un organismo. La tecnología de secuenciación de próxima generación también proporciona resecuenciación específica (secuenciación profunda), que permite la detección de mutaciones sensibles dentro de una población de secuencias de tipo salvaje. Algunos ejemplos incluyen trabajos recientes que describen la identificación de variantes de VIH resistentes a fármacos, así como mutaciones del EGFR para determinar la respuesta a fármacos terapéuticos anti-TK. Las publicaciones que describen la secuenciación de próxima generación permiten la secuenciación simultánea de múltiples muestras durante un ciclo de secuenciación típica, incluyendo, por ejemplo: Margulies, M. y col., "Genome Sequencing in Microfabricated High-Density Picolitre Reactors", *Nature*, 437, 376–80 (2005); Mikkelsen, T. y col., "Genome-Wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells", *Nature*, 448, 553–60 (2007); McLaughlin, S. y col., "Whole-Genome Resequencing with Short Reads: Accurate Mutation Discovery with Mate Pairs and Quality Values", *ASHG Annual Meeting* (2007); Shendure J. y col., "Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome", *Science*, 309, 1728–32 (2005); Harris, T. y col., "Single-Molecule DNA Sequencing of a Viral Genome", *Science*, 320, 106–9 (2008); Simen, B. y col., "Prevalence of Low Abundance Drug Resistant Variants by Ultra Deep Sequencing in Chronically HIV-infected Antiretroviral (ARV) Naive Patients and the Impact on Virologic Outcomes", 16th International HIV Drug Resistance Workshop, Barbados (2007); Thomas, R. y col., "Sensitive Mutation Detection in Heterogeneous Cancer Specimens by Massively Parallel Picoliter Reactor Sequencing", *Nature Med.*, 12, 852–855 (2006); Mitsuya, Y. y col., "Minority Human Immunodeficiency Virus Type 1 Variants in Antiretroviral-Naive Persons with Reverse Transcriptase Codon 215 Revertant Mutations", *J. Vir.*, 82, 10747–10755 (2008); Binladen, J. y col., "The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing", *PLoS ONE*, 2, e197 (2007); y Hoffmann, C. y col., "DNA Bar Coding and Pyrosequencing to Identify Rare HIV Drug Resistance Mutations", *Nuc. Acids Res.*, 35, e91 (2007).

En comparación con la secuenciación tradicional de Sanger, la tecnología de secuenciación de próxima generación produce grandes cantidades de puntos de datos de secuenciación. Un ciclo típico puede generar fácilmente de decenas a cientos de megabases por ciclo, con una potencia salida diaria que llega al rango de gigabases. Esto se traduce en varios órdenes de magnitud mayor que una placa estándar de 96 pocillos, que puede generar varios cientos de puntos de datos en ciclo múltiplex típico. Los amplicones diana que difieren en tan poco como un nucleótido pueden distinguirse fácilmente, incluso cuando están presentes múltiples dianas de especies u organismos relacionados. Esto mejora en gran medida la capacidad de realizar un genotipado exacto. Los programas de software de alineación de secuencia de próxima generación usados para producir secuencias consenso pueden identificar fácilmente nuevas mutaciones puntuales, lo que podría dar lugar a nuevas cepas con resistencias a fármacos asociadas. El uso de la codificación de barras de cebadores también permite la multiplexación de diferentes muestras de pacientes dentro de un solo ciclo de secuenciación.

Los procedimientos de secuenciación de próxima generación (NGS) comparten la característica común de las estrategias masivas paralelas de alto rendimiento, con el objetivo de reducir los costes en comparación con los procedimientos de secuenciación más antiguos. Los procedimientos de NGS pueden dividirse en líneas generales en aquellos que requieren amplificación del molde y los que no lo hacen. Los procedimientos que requieren amplificación incluyen la pirosecuenciación comercializada por Roche como las plataformas tecnológicas 454 (por ejemplo, GS 20 y GS FLX), la plataforma Solexa comercializada por Illumina y la plataforma de Supported Oligonucleotide Ligation and Detection (SOLiD) comercializada por Applied Biosystems. Los enfoques de no amplificación, también conocidos como secuenciación de una sola molécula, se ilustran mediante la plataforma HeliScope comercializada por Helicos BioSciences, Ion Torrent y plataformas emergentes comercializadas por VisiGen y Pacific Biosciences, respectivamente.

En la pirosecuenciación (Voelkerding y col., *Clinical Chem.*, 55: 641–658, 2009; MacLean y col., *Nature Rev. Microbiol.*, 7: 287–296; patente de Estados Unidos n.º 6.210,891; patente de Estados Unidos n.º 6.258,568), el ADN molde se fragmenta, se reparan los extremos, se liga a los adaptadores y se amplifica clonalmente *in situ* capturando moléculas molde sencillas con perlas que llevan oligonucleótidos complementarios a los adaptadores. Cada perla portadora de un solo tipo de molde se compartimenta en una microvesícula de agua en aceite y el molde se amplifica clonalmente utilizando una técnica denominada PCR en emulsión. La emulsión se rompe después de la amplificación y las perlas se depositan en pocillos individuales de una placa de picotitulación que funciona como una

celda de flujo durante las reacciones de secuenciación. La introducción iterativa ordenada de cada uno de los cuatro reactivos de dNTP se produce en la celda de flujo en presencia de enzimas de secuenciación e indicador luminiscente, tal como luciferasa. En el caso de que se añada un dNTP apropiado al extremo 3' del cebador de secuenciación, la producción resultante de ATP provoca una explosión de luminiscencia dentro del pocillo, que se registra utilizando una cámara CCD. Es posible conseguir longitudes de lectura mayores o iguales a 400 bases y se pueden conseguir  $1 \times 10^6$  lecturas de secuencias, dando como resultado hasta 500 millones de pares de bases (Mb) de secuencia.

En la plataforma Solexa/Illumina (Voelkerding y col., *Clinical Chem.*, 55: 641–658, 2009; MacLean y col., *Nature Rev. Microbiol.*, 7: 287–296; patente de Estados Unidos n.º 6.833.246; patente de Estados Unidos n.º 7.115.400; patente de Estados Unidos n.º 6.969.488), los datos de secuenciación se producen en forma de lecturas de longitud más corta. En este procedimiento, el ADN fragmentado monocatenario se repara en los extremos para generar extremos romos fosforilados en 3', seguido de la adición mediada por el fragmento Klenow de una única base A al extremo 3' de los fragmentos. La adición de A facilita la adición de oligonucleótidos adaptadores sobresalientes de T, que se utilizan posteriormente para capturar las moléculas adaptadoras del molde sobre la superficie de una celda de flujo que está tachonada con anclajes de oligonucleótidos. El anclaje se utiliza como cebador de PCR, pero debido a la longitud del molde y su proximidad a otros oligonucleótidos de anclaje cercanos, la extensión por PCR da como resultado el "arqueamiento" de la molécula para hibridar con un oligonucleótido de anclaje adyacente para formar una estructura de puente sobre la superficie de la celda de flujo. Estos bucles de ADN son desnaturalizados y escindidos. A continuación, las hebras delanteras se secuencian con terminadores colorantes reversibles. La secuencia de los nucleótidos incorporados se determina mediante detección de fluorescencia posterior a la incorporación, con cada fluorescencia y bloque eliminados antes del siguiente ciclo de adición de dNTP. La longitud de la lectura de la secuencia varía desde 36 nucleótidos a más de 50 nucleótidos, con una producción total superior a mil millones de pares de nucleótidos por análisis.

La secuenciación de moléculas de ácido nucleico usando tecnología SOLiD (Voelkerding y col., *Clinical Chem.*, 55: 641–658, 2009; MacLean y col., *Nature Rev. Microbiol.*, 7: 287–296; patente de Estados Unidos n.º 5.912.148; patente de Estados Unidos n.º 6.130.073) también implica la fragmentación del molde la unión a los adaptadores de oligonucleótidos, la unión a perlas y la amplificación clonal por PCR en emulsión. Después de esto, las perlas portadoras de molde se inmovilizan sobre una superficie derivatizada de una celda de flujo de vidrio y se hibrida un cebador complementario al oligonucleótido adaptador. Sin embargo, en lugar de utilizar este cebador para la extensión en 3', en lugar de ello se utiliza para proporcionar un grupo fosfato en 5' para la unión sondas identificadoras que contienen dos bases específicas de sonda seguidas de 6 bases degeneradas y uno de cuatro marcadores fluorescentes. En el sistema SOLiD, las sondas identificadoras tienen 16 posibles combinaciones de las dos bases en el extremo 3' de cada sonda y uno de cuatro fluorescencias en el extremo 5'. El color fluorado y, por lo tanto, la identidad de cada sonda corresponde a esquemas de codificación de espacio de color especificados. A varias rondas (normalmente 7) de hibridación de la sonda, unión y detección de fluorescencia les sigue desnaturalización y, después, una segunda ronda de secuenciación usando un cebador que está compensado por una base con respecto al cebador inicial. De esta manera, la secuencia molde puede reconstruirse por ordenador y las bases del molde se interrogan dos veces, lo que da como resultado una mayor precisión. La longitud de lectura de la secuencia es de 35 nucleótidos y la producción total supera los 4 mil millones de bases por ciclo de secuenciación.

En ciertas realizaciones, se usa secuenciación en nanoporos (véase, por ejemplo, Astier y col., *J Am Chem Soc.* 2006 Feb. 8; 128(5):1705–10). La teoría detrás de la secuenciación en nanoporos tiene que ver con lo que se produce cuando el nanoporo se sumerge en un fluido conductor y se aplica un potencial (voltaje) a su través: en estas condiciones se puede observar una ligera corriente eléctrica debido a la conducción de iones a través del nanoporo y la cantidad de corriente es extremadamente sensible al tamaño del nanoporo. Si las moléculas de ADN pasan (o parte de la molécula de ADN pasa) a través del nanoporo, se puede crear un cambio en la magnitud de la corriente a través del nanoporo, lo que permite determinar las secuencias de la molécula de ADN.

HeliScope de Helicos BioSciences (Voelkerding y col., *Clinical Chem.*, 55: 641–658, 2009; MacLean y col., *Nature Rev. Microbiol.*, 7: 287–296; patente de Estados Unidos n.º 7.169.560; patente de Estados Unidos n.º 7.282.337; patente de Estados Unidos n.º 7.482.120; patente de Estados Unidos n.º 7.501.245; patente de Estados Unidos n.º 6.818.395; patente de Estados Unidos n.º 6.911.345; patente de Estados Unidos n.º 7.501.245) es la primera plataforma de secuenciación de una sola molécula comercializada. Este procedimiento no requiere amplificación clonal. El ADN molde se fragmenta y poliadenila en el extremo 3', portando la adenosina final un marcador fluorescente. Los fragmentos del molde poliadenilados desnaturalizados se ligan a oligonucleótidos poli(dT) sobre la superficie de una celda de flujo. Las localizaciones físicas iniciales de las moléculas molde capturadas son grabadas por una cámara CCD y, después, se escinde el marcador y se elimina mediante lavado. La secuenciación se consigue mediante la adición de polimerasa y la adición en serie de reactivos de dNTP marcados con fluorescencia. Los acontecimientos de incorporación producen una señal de fluorescencia correspondiente al dNTP y la señal es capturada por una cámara CCD antes de cada ronda de adición de dNTP. La longitud de la lectura de la secuencia varía desde 25 nucleótidos a 50 nucleótidos, con una producción total superior a mil millones de pares de nucleótidos por análisis.



Otros procedimientos emergentes de secuenciación de moléculas individuales incluyen la secuenciación en tiempo real mediante la síntesis utilizando una plataforma VisiGen (Voelkerding y col., *Clinical Chem.*, 55: 641–658, 2009; patente de Estados Unidos n.º 7.329,492; solicitud de patente de Estados Unidos n.º de serie 11/671.956; solicitud de patente de Estados Unidos n.º de serie 11/781.166) en la que el molde de ADN cebado inmovilizado se somete a extensión de cadena usando una polimerasa modificada por fluorescencia y moléculas aceptoras fluorescentes, lo que da como resultado una transferencia de energía de resonancia de fluorescencia detectable (FRET) después de la adición de nucleótidos.

Otro sistema de secuenciación de moléculas individuales en tiempo real desarrollado por Pacific Biosciences (Voelkerding y col., *Clinical Chem.*, 55: 641–658, 2009; MacLean y col., *Nature Rev. Microbiol.*, 7: 287–296; patente de Estados Unidos n.º 7.170.050; patente de Estados Unidos n.º 7.302,146; patente de Estados Unidos n.º 7.313.308; patente de Estados Unidos n.º 7.476.503) utiliza pocillos de reacción de 50-100 nm de diámetro y que abarca un volumen de reacción de aproximadamente 20 zeptolitros ( $10 \times 10^{-21}$  l). Las reacciones de secuenciación se llevan a cabo usando un molde inmovilizado, ADN polimerasa de phi29 modificada y concentraciones locales altas de dNTP marcados con fluorescencia. Las concentraciones locales altas y las condiciones de reacción continuas permiten capturar acontecimientos de incorporación en tiempo real mediante la detección de la señal fluorescente mediante excitación por láser, una guía de ondas óptica y una cámara CCD.

La secuenciación Ion torrent es un procedimiento de secuenciación de ADN basado en la detección de iones de hidrógeno que se liberan durante la polimerización de ADN. Este es un procedimiento de "secuenciación mediante síntesis", durante el cual se construye una cadena complementaria basada en la secuencia de un soporte del molde. Un micropocillo que contiene una cadena de ADN molde para secuenciar se inunda con una sola especie de desoxirribonucleótido (dNTP). Si el dNTP introducido es complementario al nucleótido molde principal, se incorpora en la cadena complementaria en crecimiento. Esto provoca la liberación de un ion de hidrógeno que desencadena un sensor de iones hipersensible, lo que indica que se ha producido una reacción. Si hay repeticiones de homopolímero presentes en la secuencia molde, se incorporarán múltiples moléculas de dNTP en un único ciclo. Esto conduce a un número correspondiente de hidrógenos liberados y una señal electrónica proporcionalmente más alta.

Diversas modificaciones y variaciones de los procedimientos y realizaciones descritos en la presente invención serán evidentes para los expertos en la materia. Aunque la presente invención se ha descrito en relación con realizaciones preferidas específicas, debe entenderse que la invención, tal como se reivindica, no debe limitarse indebidamente a dichas realizaciones específicas.

**REIVINDICACIONES**

1. Un procedimiento para generar ácido nucleico amplificado a partir de ARN que comprende:

- 5 a) exponer una secuencia molde de ARN a un conjunto de cebadores en condiciones de transcripción inversa, de tal manera que se genere una población mixta de las primeras cadenas de ADNc, en la que dicho conjunto de cebadores comprende cebadores individuales, comprendiendo cada uno: i) un sitio de secuencia de restricción en 5', (ii) una secuencia hexamérica aleatoria en 3' y (iii) una secuencia de código de barras localizada entre el sitio de secuencia de restricción en 5' y la secuencia hexamérica aleatoria en 3', en la que dicho sitio de secuencia de restricción en cada uno de dichos cebadores individuales es idéntico y en la que dicho conjunto de cebadores comprende todas o casi todas las posibles secuencias hexaméricas aleatorias y en la que cada uno de dichas primeras cadenas de ADNc tiene uno de dichos cebadores individuales en su extremo 5';
- 10 b) exponer dicha población mixta de las primeras cadenas de ADNc a dicho conjunto de cebadores en condiciones de polimerización, de tal manera que se genera una población mixta de moléculas de ADNc de doble cadena,
- 15 c) digerir dicha población mixta de moléculas de ADNc de doble cadena con una enzima de restricción específica de dicho sitio de secuencia de restricción en 5',
- 20 d) tratar dicha población mixta de moléculas de ADNc de doble cadena con un agente de unión, de tal manera que las moléculas individuales de ADNc de doble cadena se ligan entre sí para formar una población mixta de concatémeros;
- 25 e) exponer dicha población mixta de concatémeros a cebadores aleatorios en condiciones de amplificación del genoma completo de tal manera que se genere ácido nucleico amplificado;
- f) digerir, al menos parcialmente, dicho ácido nucleico amplificado con la misma enzima de restricción específica de dicho sitio de secuencia de restricción en 5', generando de este modo una pluralidad de secuencias digeridas;
- 30 y
- g) ligar las secuencias adaptadoras de la secuenciación a los extremos de dicha pluralidad de secuencias digeridas para generar una población mixta de moldes de secuenciación ligados al adaptador, en los que dichas secuencias adaptadoras de la secuenciación contienen un sitio de enzima de restricción que es idéntico al sitio de secuencia de restricción en 5' en el conjunto de cebadores.

2. El procedimiento de la reivindicación 1, en el que dichas secuencias adaptadoras de la secuenciación son secuencias en horquilla.

35 3. El procedimiento de la reivindicación 1, en el que dicha pluralidad de secuencias digeridas comprende secuencias digeridas individuales que contienen, cada una, las secuencias de bases de solo una de dichas moléculas de ADNc de doble cadena de dicha población mixta de moléculas de ADNc de doble hebra.

40 4. El procedimiento de la reivindicación 1, en el que dicha pluralidad de secuencias digeridas comprende secuencias digeridas individuales que contienen, cada una, las secuencias de bases de solo una de dichas moléculas de ADNc de doble cadena de dicha población mixta de moléculas de ADNc de doble hebra, en el que dichas secuencias de base están separadas entre sí por dichas secuencias de código de barras.

45 5. El procedimiento de la reivindicación 4, que comprende además secuenciar al menos una de dichas secuencias digeridas individuales para generar información de secuencia electrónica y procesar dicha información de secuencia electrónica con un algoritmo de alineación en el que dichas secuencias de códigos de barras se usan para identificar uniones artificiales entre dichas secuencias de bases de dos o más de dichas moléculas de ADNc de doble cadena.

50 6. El procedimiento de la reivindicación 1, en el que dicha secuencia de código de barras en cada uno de dichos cebadores individuales es idéntica.

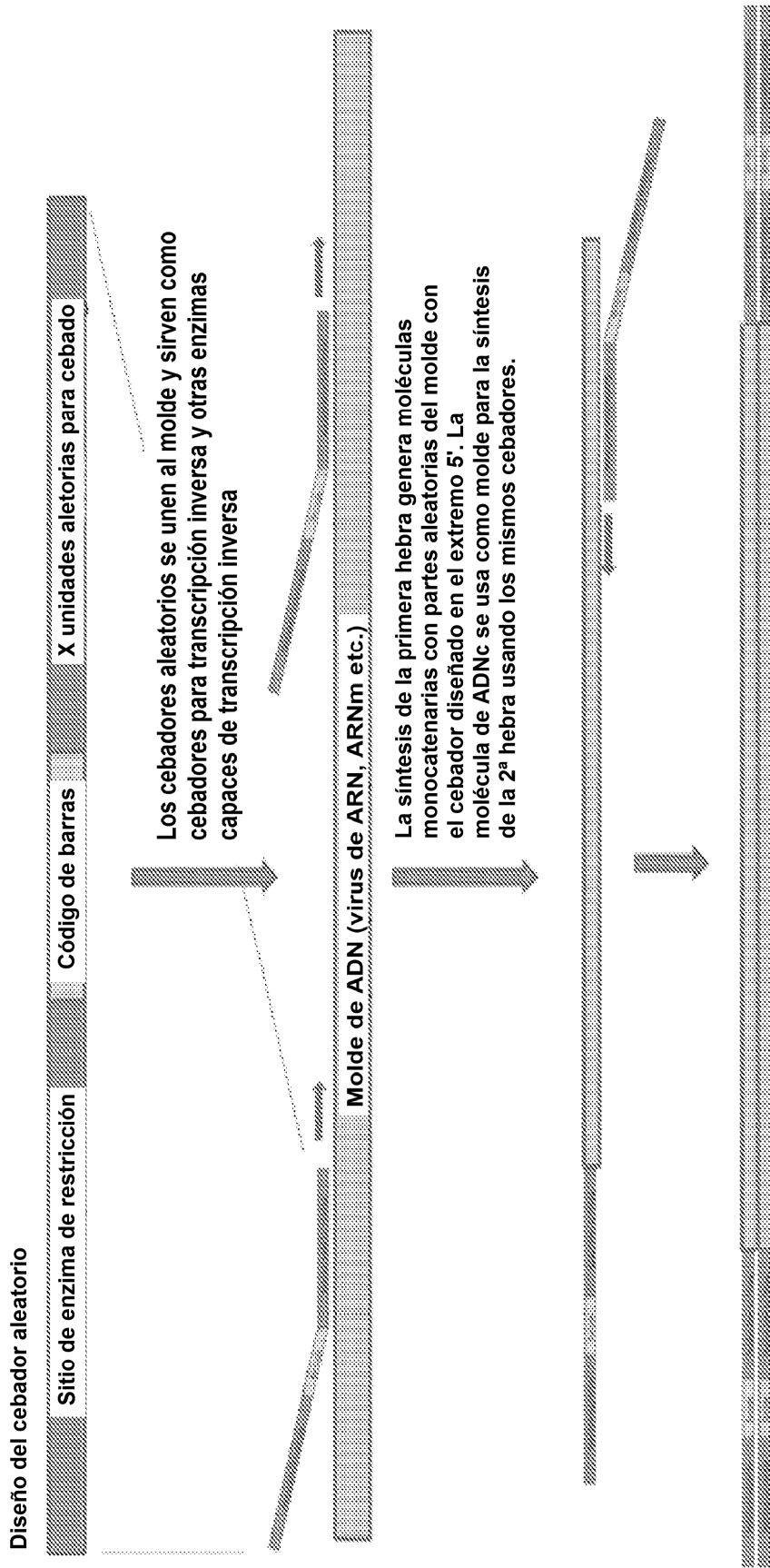
7. El procedimiento de la reivindicación 1, en el que dicho molde de ARN tiene una longitud inferior a 2.000 bases.

55 8. El procedimiento de la reivindicación 1, en el que dicha población mixta de concatémeros comprende concatémeros individuales que tienen aproximadamente 2.000 bases de longitud o más.

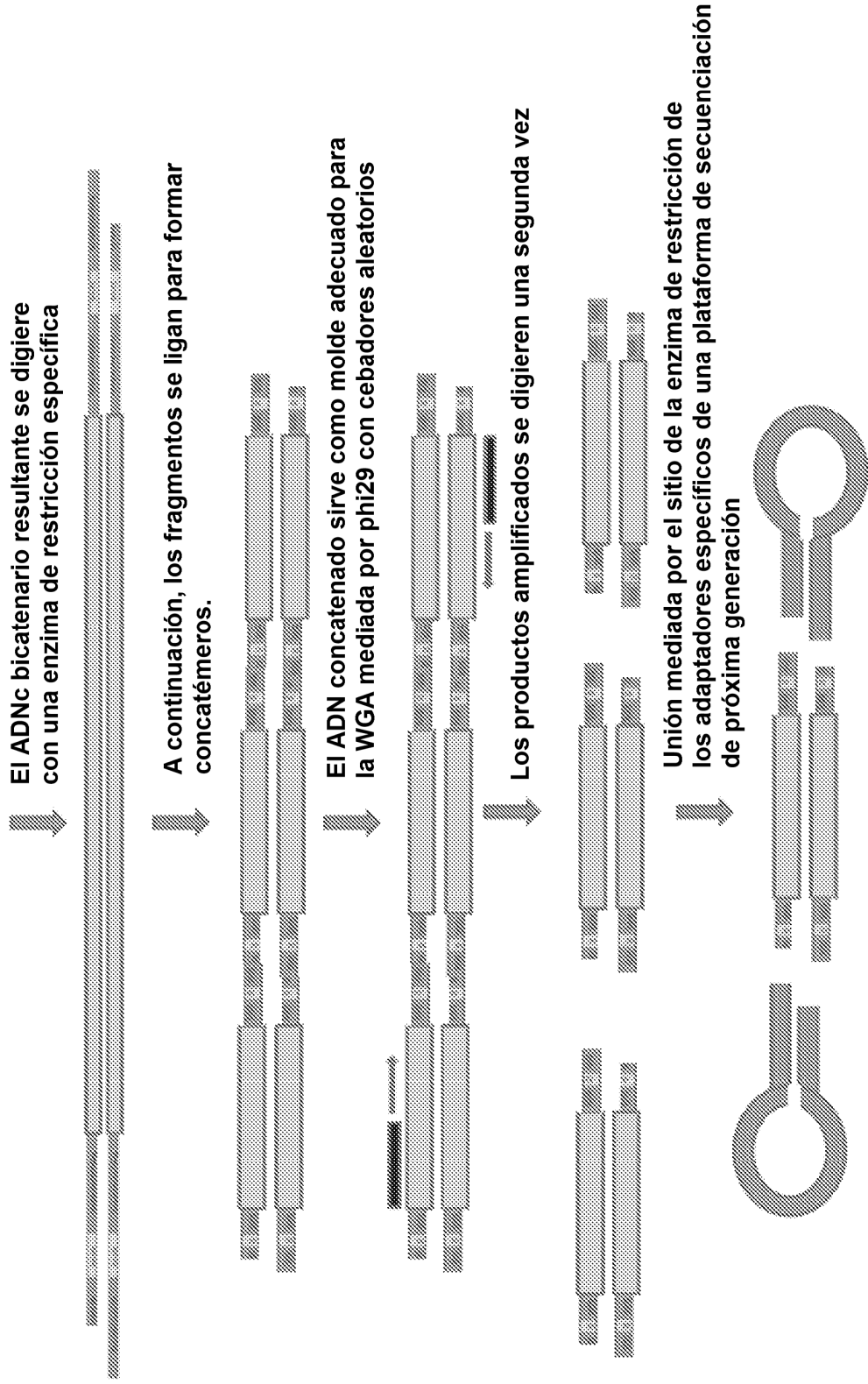
60

65

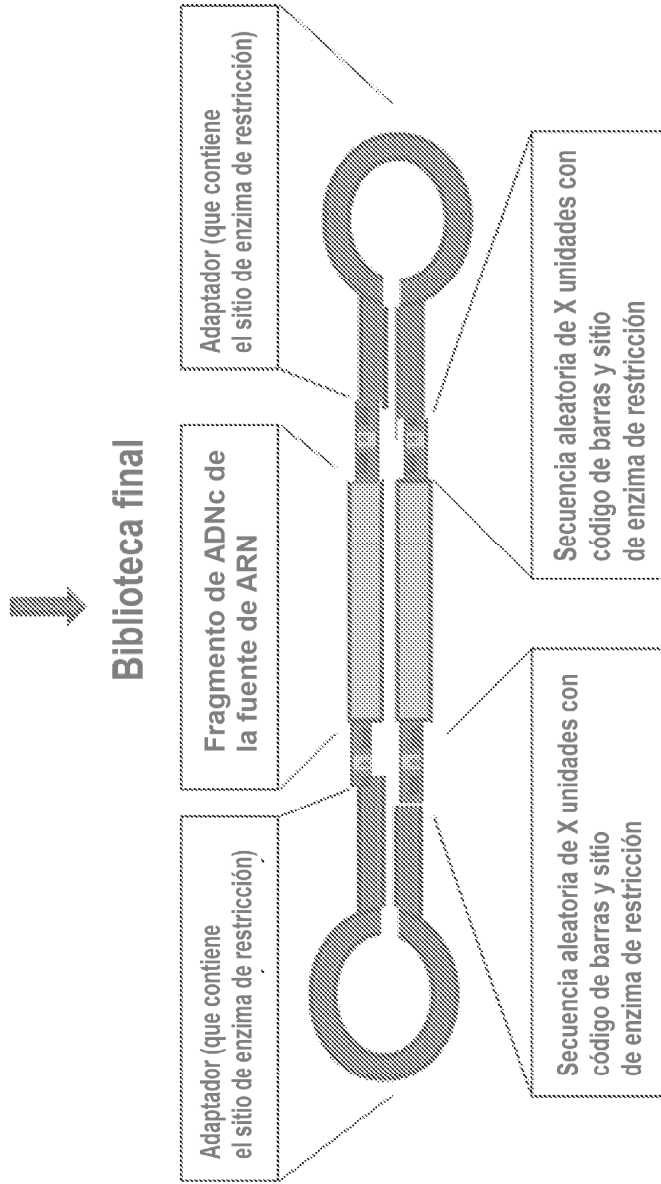
**FIGURA 1A**



**FIGURA 1B**



**FIGURA 1C**



Otros productos: construcciones de la biblioteca con múltiples insertos de ADNc. Se pueden identificar por ordenador debido a las presencias del código de barras y el sitio de enzima de restricción.

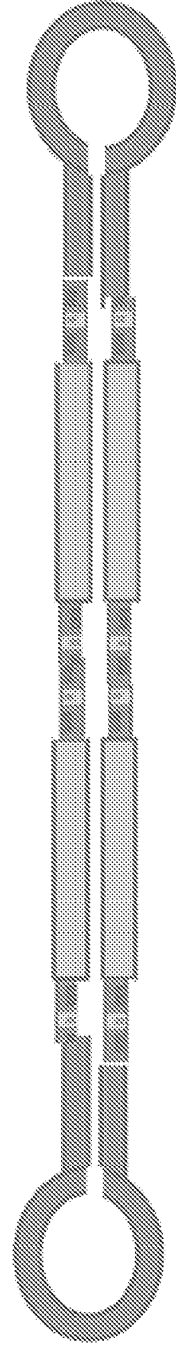
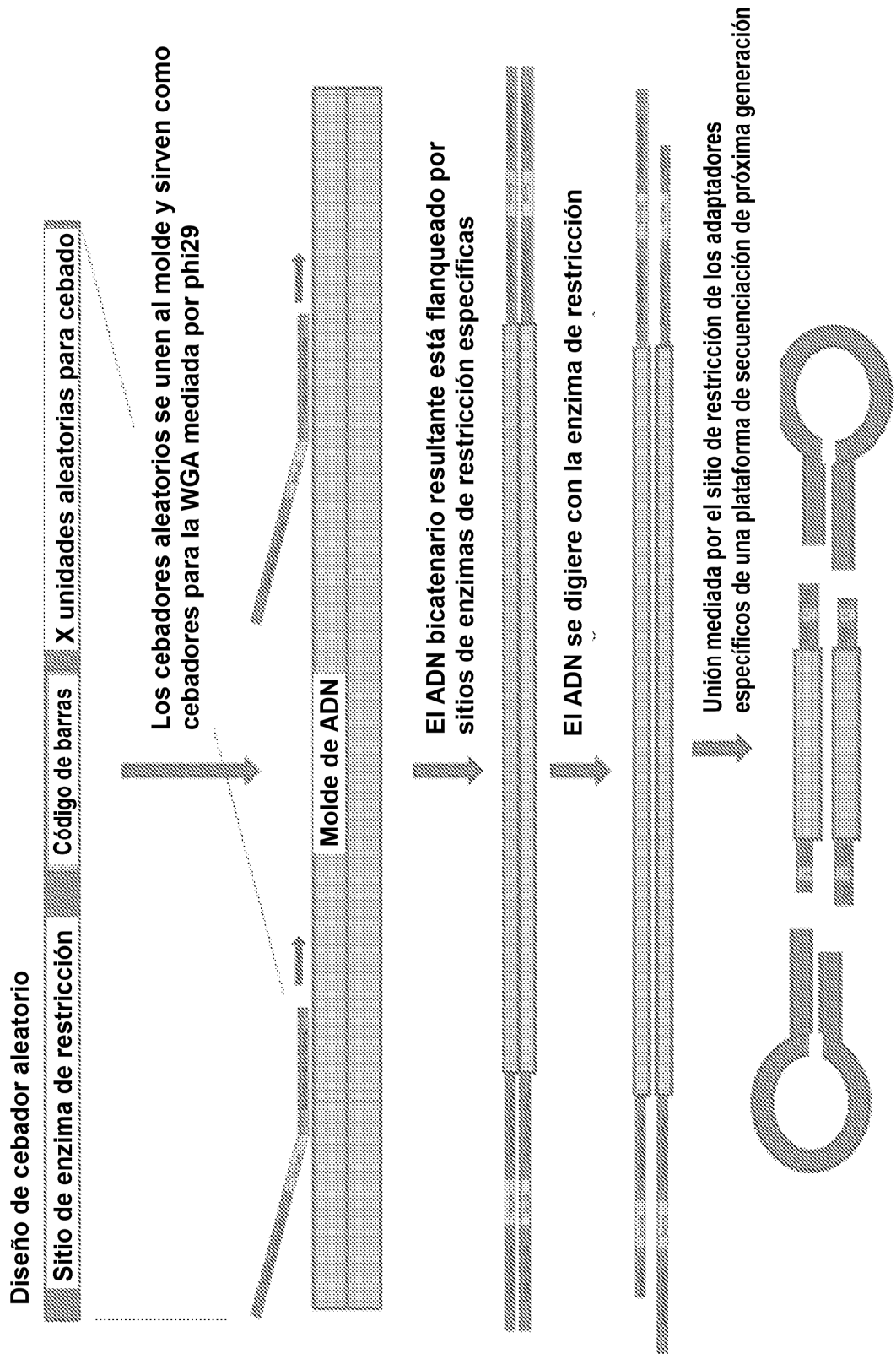
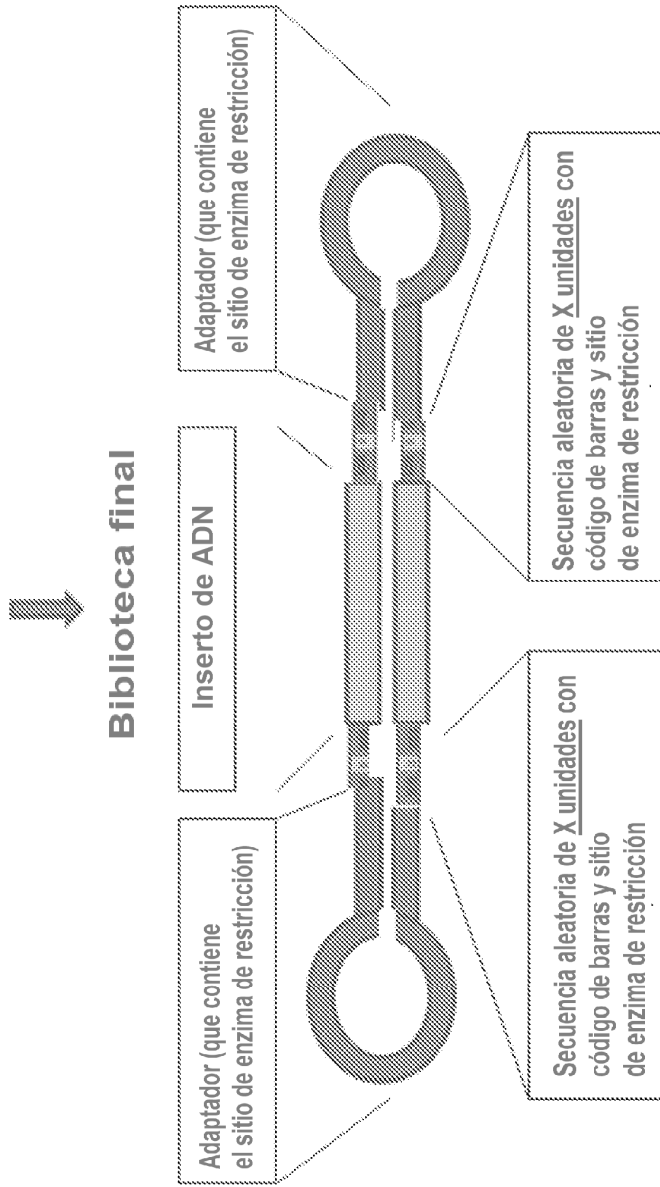


FIGURA 2A



**FIGURA 2B**



Otros productos: construcciones de la biblioteca con múltiples insertos de ADN. Se pueden identificar por ordenador debido a las presencias de código de barras y el sitio de enzima de restricción.

