(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 762 866**

(51) Int. CI.:

C12N 15/10 (2006.01)
C12Q 1/68 (2008.01)
C12N 15/11 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **30.01.2012 E 16155567 (7)**

(97) Fecha y número de publicación de la concesión europea: **27.11.2019 EP 3037536**

(54) Título: **Reemplazo de nucleótidos por bibliotecas doblemente etiquetadas y direccionales**

(30) Prioridad:

28.01.2011 US 201161437451 P
12.07.2011 US 201161506777 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
26.05.2020

(73) Titular/es:

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

(72) Inventor/es:

GORYSHIN, IGOR;
BAAS, BRADLEY;
VAIDYANATHAN, RAMESH y
MAFFITT, MARK

(74) Agente/Representante:

SÁEZ MAESO, Ana

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

Reemplazo de nucleótidos por bibliotecas doblemente etiquetadas y direccionales

Campo técnico de la invención

La presente invención se refiere a la biología molecular, y más específicamente a los métodos de uso de transposasas para fragmentar y etiquetar ácidos nucleicos, que pueden ser útiles como bibliotecas de ADN para secuenciación.

Antecedentes de la invención

La preparación de muestras para la secuenciación de próxima generación puede implicar la fragmentación de ADN genómico o ADNc bicatenario (preparado a partir de ARN) en fragmentos más pequeños, seguido de la adición de secuencias de etiquetas funcionales ("etiquetas") a las cadenas de los fragmentos. Cuando se etiqueta una secuencia monocatenaria en ambos extremos, se puede usar el término "doblemente etiquetado". Dichas etiquetas incluyen sitios de cebado para ADN polimerasas para reacciones de secuenciación, sitios de restricción y dominios para promotores de captura, amplificación, detección, direccionamiento y transcripción. Los métodos anteriores para generar bibliotecas de fragmentos de ADN requerían fragmentar el ADN objetivo mecánicamente usando un sonicador, nebulizador o mediante una nucleasa, y luego unir (por ejemplo, mediante ligación) los oligonucleótidos que contienen las etiquetas a los extremos de los fragmentos.

Un nuevo método para usar transposones para lograr rápidamente estas etapas se divulgó en el documento US 2010/0120098 por Grunenwald para generar fragmentos de cualquier ADN bicatenario (por ejemplo, genómico, amplicón, viral, fago, ADNc derivado de ARN, etc.). Los sistemas de transposón particularmente útiles incluyen el sistema del transposón Tn5 hiperactivo descrito en los documentos US 5.965.443 y US 6.437.109 por Reznikoff, y el sistema del transposón Mu en el documento US 6.593.113 por Tenkanen. Reznikoff, en particular, describió una secuencia final de transposasa de 19 bases (SEQ ID NO: 3) que se denomina frecuentemente como "ME". En algunas realizaciones del método de transposón, la reacción en cadena de la polimerasa (PCR) se usa como una etapa posterior para la amplificación de ADN. Esto puede generar inquietudes debido al potencial de PCR de representar en exceso o insuficientemente las cantidades relativas de una secuencia dada, dependiendo de su composición de G + C, especialmente en regiones de contenido extremo de G + C donde el sesgo de PCR puede confundir la anotación y el análisis de los datos.

Sumario de la invención

La presente invención describe un método para agregar una o más etiquetas al producto bicatenario de una reacción de etiquetado. El método implica proporcionar un ácido nucleico objetivo bicatenario y un transposoma que tiene una transposasa con dos secuencias finales de transposón: una "cadena transferida" y una "cadena no transferida". El transposoma rompe el ácido nucleico objetivo en fragmentos mientras transfiere covalentemente la cadena transferida a una primera cadena del fragmento; la cadena no transferida del transposoma permanece hibridada con la cadena transferida. En una realización, la cadena no transferida tiene las fórmulas generales

$BTGTYTCBTN_{1-9}$	SEQ ID NO: 20
$NTGTMTCNTN_{0-10}$	SEQ ID NO: 21

en las se usa la nomenclatura IUPAC para las posiciones de nucleótidos degenerados, y N_{x-y} indica una secuencia que tiene un rango de x a y nucleótidos, inclusive. Entre los fragmentos, la cadena no transferida se elimina de la cadena transferida y se reemplaza por un oligo que comprende una secuencia de etiqueta. El oligo de reemplazo se une luego a la segunda cadena del fragmento mediante ligación y opcionalmente mediante una etapa de extensión. El resultado del método es un fragmento del ácido nucleico objetivo que se ha etiquetado con una o más secuencias de etiquetas, que puede ser útil para análisis posteriores, como la secuenciación.

La divulgación también proporciona un método para generar bibliotecas direccionales proporcionando el ácido nucleico objetivo en el que una cadena se modifica químicamente. El enriquecimiento selectivo de una cadena en el producto de la reacción de etiquetado da como resultado un fragmento que ha sido etiquetado de una manera específica de la cadena, es decir, una etiqueta en el extremo 5' y una etiqueta en el extremo 3'.

Las secuencias finales novedosas de transposasa también se proporcionan en el presente documento, que tienen las secuencias generales

MRWT**G**TGHWKAV**GA**R**ACA**V	SEQ ID NO: 1 y
NSHB**G**HSHDDRN**GA**K**ACA**N	SEQ ID NO: 2

Estas secuencias finales se pueden usar en los métodos de la invención, y con transposasas para reacciones de etiquetado relacionadas en general.

Breve descripción de los dibujos

La Fig. 1 proporciona una ilustración esquemática de una reacción de etiquetado. Los cuadrados con esquinas grabadas representan transposasas, por ejemplo transposasas Tn5 o Mu. El MEDS se refiere a un ME de doble cadena ("extremo de mosaico"), ejemplificado por una secuencia final de transposasa de Tn5, tal como la SEQ ID NO: 3 hibridada con la SEQ ID NO: 19. Unidas a los MEDS están las etiquetas arbitrarias, que se muestran en el presente documento como barras claras u oscuras. Juntos, los cuadrados, MEDS y etiquetas representan un transposoma, que puede usarse para fragmentar un ácido nucleico objetivo. Un producto de fragmentación bicatenario se muestra como barras oscuras paralelas, con ciertas secuencias adjuntas, como se discute a continuación. La SEQ1 y la SEQ2 se refieren a secuencias complementarias a las etiquetas arbitrarias, que pueden ser parte de cebadores utilizados para PCR, por ejemplo. El triángulo invertido representa un punto de inserción opcional para secuencias de etiquetas adicionales, como un código de barras. A y B representan secuencias adicionales que pueden unirse mediante PCR. Como se muestra, los transposomas se pueden usar para fragmentar un ácido nucleico objetivo para generar fragmentos bicatenarios que tengan secuencias en ambos extremos, lo que puede ser útil para la codificación y secuenciación de barras.

La FIG. 2a proporciona una geometría más detallada del producto de etiquetado de la segunda fila de la Fig. 1. Como se muestra, la reacción de etiquetado da como resultado un ADN objetivo fragmentado en el que el extremo 5' de la cadena superior está unido covalentemente a una "cadena transferida" de 19 bases (SEQ ID NO: 3). El extremo 5' de la cadena inferior está unido de manera similar a otra copia de la cadena transferida, mostrada en orientación 3' a 5' (SEQ ID NO: 3). Sin embargo, el etiquetado deja un espacio de cadena sencilla de 9 bases entre el extremo 3' del fragmento objetivo y el extremo 5' de la otra cadena (no transferido) del transposoma. Debido a que esta cadena no se une covalentemente a los extremos 3' del fragmento de ácido nucleico objetivo, se describe como un "extremo no transferido" (SEQ ID NO: 19), aunque el extremo no transferido permanece asociado con el fragmento por hibridación con la cadena transferida. La Fig. 2b muestra el mismo producto de etiquetado que en la Fig. 2a, pero usando la forma esquemática utilizada en las otras figuras. Para facilitar la ilustración, se han omitido las etiquetas de la Fig. 2a y la Fig. 2b.

La Fig. 3 muestra varias versiones alternativas de las cadenas no transferidas. En la parte superior, se muestra como referencia el extremo transferido ME de 19 bases (SEQ ID NO: 3) con una secuencia de etiqueta unida. Inmediatamente a continuación, el extremo no transferido complementario de 19 bases se muestra como la SEQ ID NO: 19 en orientación 3' a 5'. Sin embargo, como se describe en el presente documento, un extremo no transferido también puede truncarse mediante supresiones en versiones con 18 bases (SEQ ID NO: 18), 17 bases (SEQ ID NO: 17), 16 bases (SEQ ID NO: 16), 15 bases (SEQ ID NO: 15), 14 bases (SEQ ID NO: 14), 13 bases (SEQ ID NO: 13), 12 bases (SEQ ID NO: 12), 11 bases (SEQ ID NO: 11), 10 bases (SEQ ID NO: 10), 9 bases (SEQ ID NO: 9) u 8 bases (SEQ ID NO: 8).

La Fig. 4 ilustra una realización del método de la invención, como se discute con mayor detalle a continuación. Los extremos del transposón de 19 pb (cadena transferida) (SEQ ID NO: 3) se muestran con la etiqueta arbitraria 1 unida. Un ejemplo de cadena no transferida de 14 bases (SEQ ID NO: 14) se muestra hibridada a una porción de la cadena transferida. Como se muestra en la fila superior, se proporcionan oligonucleótidos de reemplazo (SEQ ID NO: 19) (mostrados en el presente documento unidos a la etiqueta arbitraria 2). En las filas media e inferior, los extremos del transposón de 19 bases de doble cadena (SEQ ID NO: 3, SEQ ID NO: 19) se muestran unidos al fragmento de ácido nucleico objetivo. El producto que se muestra en la fila inferior es un fragmento "doblemente etiquetado" que tiene la etiqueta arbitraria 1 en el extremo 5' y la etiqueta arbitraria 2 en el extremo 3'.

La Fig. 5 ilustra una realización particular para generar un producto de fragmento direccional en el que los círculos pequeños en los que los ADNbc representan modificaciones químicas de la cadena inferior. En este diagrama, la cadena modificada (inferior) se considera la cadena no deseada. El resultado, como se muestra, es un producto que conserva la cadena (superior) deseada, la etiqueta arbitraria 1 en el extremo 5' y la etiqueta arbitraria 2 en el extremo 3'.

Descripción detallada

La presente descripción proporciona un método mejorado para preparar ADNbc marcado con doble etiqueta. El método implica (a) proporcionar un ácido nucleico objetivo bicatenario y un transposoma que tiene una transposasa con dos secuencias finales de transposón: una "cadena transferida" y una "cadena no transferida"; (b) permitir que el transposoma fragmente el ácido nucleico objetivo, por lo que la cadena transferida se transfiere covalentemente a una primera cadena del fragmento, y la cadena no transferida permanece hibridada con la cadena transferida; (c) eliminar la cadena no transferida de la cadena transferida; (d) proporcionar un oligo de reemplazo que comprende una secuencia de etiqueta, para hibridarse con la cadena transferida; y (e) ligar el oligo de reemplazo a la segunda cadena del fragmento. Por lo tanto, el método genera un producto de etiquetado que tiene una cadena transferida y un oligo de reemplazo.

El ADN objetivo utilizado en el método puede ser cualquier ácido nucleico de interés. Los ácidos nucleicos objetivo pueden incluir ADN, ácido nucleico peptídico, ácido nucleico morfolino, ácido nucleico bloqueado, ácido nucleico de glicol, ácido nucleico de treosa, mezclas de los mismos e híbridos de los mismos. En una realización preferida, los fragmentos de ADN genómico o copias amplificadas de los mismos se usan como el ácido nucleico objetivo. En otra realización preferida, se usa ADN mitocondrial o de cloroplasto.

Un ácido nucleico objetivo puede comprender cualquier secuencia de nucleótidos. En algunas realizaciones, el ácido nucleico objetivo comprende secuencias de homopolímero. Un ácido nucleico objetivo también puede incluir secuencias repetidas. Las secuencias repetidas pueden tener cualquiera de una variedad de longitudes que incluyen, por ejemplo, 2, 5, 10, 20, 30, 40, 50, 100, 250, 500, 1000 nucleótidos o más. Las secuencias repetidas pueden repetirse, ya sea de manera contigua o no contigua, en cualquiera de una variedad de veces, incluyendo, por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 veces o más.

Algunas realizaciones descritas en el presente documento puede utilizarse un solo ácido nucleico objetivo. Otras realizaciones pueden utilizar una pluralidad de ácidos nucleicos objetivo. En tales realizaciones, una pluralidad de ácidos nucleicos objetivo puede incluir una pluralidad de los mismos ácidos nucleicos objetivo, una pluralidad de ácidos nucleicos objetivo diferentes en la que algunos ácidos nucleicos objetivo son iguales, o una pluralidad de ácidos nucleicos objetivo en la que todos los ácidos nucleicos objetivo son diferentes. Las realizaciones que utilizan una pluralidad de ácidos nucleicos objetivo pueden llevarse a cabo en formatos multiplex para que los reactivos se suministren simultáneamente a los ácidos nucleicos objetivo, por ejemplo, en una o más cámaras o en una superficie de una matriz. En algunas realizaciones, la pluralidad de ácidos nucleicos objetivo puede incluir sustancialmente todo el genoma de un organismo particular. La pluralidad de ácidos nucleicos objetivo puede incluir al menos una porción del genoma de un organismo particular que incluye, por ejemplo, al menos aproximadamente 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95% o 99% del genoma. En realizaciones particulares, la porción puede tener un límite superior que es a lo sumo aproximadamente el 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95% o 99% del genoma

Los ácidos nucleicos objetivo se pueden obtener de cualquier fuente. Por ejemplo, los ácidos nucleicos objetivo pueden prepararse a partir de moléculas de ácido nucleico obtenidas de un solo organismo o de poblaciones de moléculas de ácido nucleico obtenidas de fuentes naturales que incluyen uno o más organismos. Las fuentes de moléculas de ácido nucleico incluyen, entre otras, orgánulos, células, tejidos, órganos u organismos. Las células que pueden usarse como fuentes de moléculas de ácido nucleico objetivo pueden ser procariotas (células bacterianas, por ejemplo, de los géneros *Escherichia, Bacillus, Serratia, Salmonella, Staphylococcus, Streptococcus, Clostridium, Chlamydia, Neisseria, Treponema, Mycoplasma, Borrelia, Legionella, Pseudomonas Mycobacterium, Helicobacter, Erwinia, Agrobacterium, Rhizobium y Streptomyces*); arqueas, tales como crenarchaeota, nanoarchaeota o euryarchaeotia; o eucariotas tales como hongos (por ejemplo, levaduras), plantas, protozoos y otros parásitos, y animales (incluidos insectos (por ejemplo, *Drosophila* spp.), nematodos (por ejemplo, *Caenorhabditis elegans*) y mamíferos (por ejemplo, ratas, ratón, mono, primate no humano y humano).

En algunas realizaciones, el ácido nucleico objetivo puede proporcionarse cuando una cadena se modifica químicamente, tal como con un sitio de fragmentación. Se puede usar un sitio de fragmentación para escindir la asociación física, pero no la informativa entre una primera secuencia de código de barras y una segunda secuencia de código de barras. La escisión puede ser por medios bioquímicos, químicos u otros. En algunas realizaciones, un sitio de fragmentación puede incluir un nucleótido o una secuencia de nucleótidos que puede fragmentarse por diversos medios. Por ejemplo, un sitio de fragmentación puede ser un sustrato para una enzima, tal como una nucleasa, que escindirá la asociación física entre una primera secuencia de código de barras y una segunda secuencia de código de barras. Por ejemplo, el sitio de fragmentación comprende un sitio de endonucleasa de restricción y puede escindirse con una endonucleasa de restricción apropiada. En otro ejemplo, un sitio de fragmentación puede comprender al menos un ribonucleótido en un ácido nucleico que de otro modo puede comprender desoxirribonucleótidos y puede escindirse con una ARNasa. Los agentes de escisión química capaces de escindir selectivamente el enlace fosfodiéster entre un desoxirribonucleótido y un ribonucleótido incluyen iones metálicos, por ejemplo, iones de metales de tierras raras (por ejemplo, La^{3+}, particularmente Tm^{3+}, Yb^{3+} o Lu^{3+} (Chen et al. Biotechniques. 2002, 32: 518-520; Komiyama et al. Chem. Commun. 1999, 1443-1451)), Fe(3) o Cu(3), o exposición a pH elevado, por ejemplo, tratamiento con una base tal como hidróxido de sodio. Como se usa en el presente documento, la escisión selectiva del enlace fosfodiéster entre un desoxirribonucleótido y un ribonucleótido puede referirse al agente de escisión química que no es capaz de escindir el enlace fosfodiéster entre dos desoxirribonucleótidos en las mismas condiciones.

En otro ejemplo, el sitio de fragmentación puede comprender una o más secuencias de reconocimiento para una nickasa, es decir, una endonucleasa de corte que rompe una cadena de un ácido nucleico bicatenario. Por lo tanto, el sitio de fragmentación puede comprender una primera secuencia de reconocimiento de nickasa, una segunda secuencia de reconocimiento de nickasa. El sitio de corte para cada secuencia de reconocimiento puede ser el mismo sitio o un sitio diferente.

En otro ejemplo, un sitio de fragmentación puede incluir uno o más análogos de nucleótidos que comprenden un sitio abásico y permite la escisión en el sitio de fragmentación en presencia de ciertos agentes químicos, tales como poliamina, N,N'-dimetiletilendiamina (DMED) (Publicación de patente de los Estados Unidos No. 2010/0022403). En una realización, la modificación química puede ser una conversión de citosinas a uracilos. En algunas realizaciones, se puede crear un sitio abásico dentro de un sitio de fragmentación proporcionando primero un sitio de fragmentación que comprende una desoxiuridina (U) de un ácido nucleico bicatenario. La enzima uracilo ADN glicosilasa (UDG) se puede usar para eliminar la base de uracilo, generando un sitio abásico en una cadena. La cadena polinucleotídica que incluye el sitio abásico se puede escindir luego en el sitio abásico mediante tratamiento con endonucleasa (por

ejemplo, endonucleasa Endo IV, AP liasa, FPG glicosilasa/AP liasa, Endo VIII glicosilasa/AP liasa), calor o álcali. Los sitios abásicos también pueden generarse en análogos de nucleótidos distintos de la desoxiuridina y escindirse de manera análoga mediante tratamiento con endonucleasa, calor o álcali. Por ejemplo, la 8-oxo-guanina se puede convertir en un sitio abásico por exposición a la FPG glicosilasa. La desoxiinosina puede convertirse en un sitio abásico por exposición a la AlkA glicosilasa. Los sitios abásicos generados de este modo se pueden escindir, típicamente mediante tratamiento con una endonucleasa adecuada (por ejemplo, Endo IV, AP liasa). (publicación de patente de los Estados Unidos No. 2011/0014657).

En otro ejemplo, un sitio de fragmentación puede incluir un enlace diol que permite la escisión por tratamiento con peryodato (por ejemplo, peryodato de sodio). En otro ejemplo, un sitio de fragmentación puede incluir un grupo disulfuro que permite la escisión con un agente reductor químico, por ejemplo, clorhidrato de tris(2-carboxietil)-fosfato (TCEP).

En algunas realizaciones, un sitio de fragmentación puede incluir una fracción escindible que puede someterse a escisión fotoquímica. La escisión fotoquímica abarca cualquier método que utilice energía luminosa para lograr la escisión de ácidos nucleicos, por ejemplo, una o ambas cadenas de una molécula de ácido nucleico bicatenario. Un sitio para la escisión fotoquímica puede proporcionarse mediante una fracción química no nucleotídica en un ácido nucleico, tal como fosforamidita [4-(4,4'-dimetoxitritiloxi)butiramidometil)-1-(2-nitrofenil)-etil]-2-cianoetil-(N,N-diisopropil)-fosforamidita) (Glen Research, Sterling, Virginia, EE. UU., Cat No. 10-4913-XX).

En algunas realizaciones, un sitio de fragmentación puede incluir un péptido, por ejemplo, una estructura conjugada en la que una molécula peptídica está unida a un ácido nucleico. La molécula peptídica puede segmentarse posteriormente mediante una enzima peptidasa de la especificidad apropiada, o cualquier otro medio adecuado de escisión química o fotoquímica no enzimática. En algunas realizaciones, se formará un conjugado entre péptido y ácido nucleico uniendo covalentemente un péptido a un ácido nucleico, por ejemplo, una cadena de un ácido nucleico bicatenario. Los conjugados entre un péptido y un ácido nucleico pueden prepararse usando técnicas generalmente conocidas en el arte. En una de tales técnicas, los componentes peptídicos y de ácido nucleico de la secuencia de aminoácidos y nucleótidos deseada se pueden sintetizar por separado, por ejemplo, mediante técnicas estándar de síntesis química automatizadas, y luego conjugarse en solución acuosa/orgánica. A modo de ejemplo, el sistema OPeC^{MR} disponible comercialmente a través de Glen Research se basa en la ligación nativa de un péptido funcionalizado con tioéster en el terminal N a un oligonucleótido 5'-cisteinilo.

El método de la invención puede usar cualquier transposasa que pueda aceptar una secuencia final de transposasa y fragmentar un ácido nucleico objetivo, uniendo un extremo transferido, pero no un extremo no transferido. Un "transposoma" está compuesto por al menos una enzima transposasa y un sitio de reconocimiento de transposasa. En algunos de estos sistemas, denominados "transposomas", la transposasa puede formar un complejo funcional con un sitio de reconocimiento de transposones que es capaz de catalizar una reacción de transposición. La transposasa o integrasa puede unirse al sitio de reconocimiento de transposasa e insertar el sitio de reconocimiento de transposasa en un ácido nucleico objetivo en un proceso que a veces se denomina "etiquetado". En algunos de estos eventos de inserción, una cadena del sitio de reconocimiento de transposasa puede transferirse al ácido nucleico objetivo.

Algunas realizaciones pueden incluir el uso de una transposasa Tn5 hiperactiva y un sitio de reconocimiento de transposasa de tipo Tn5 (Goryshin y Reznikoff, J. Biol. Chem., 273: 7367 (1998)), o transposasa MuA y un sitio reconocimiento de transposasa Mu que comprende secuencias finales R1 y R2 (Mizuuchi, K., Cell, 35: 785, 1983; Savilahti, H, et al., EMBO J., 14: 4893, 1995). Un ejemplo de sitio de reconocimiento de transposasa que forma un complejo con una transposasa Tn5 hiperactiva (por ejemplo, transposasa EZ-Tn5^{MR}, Epicenter Biotechnologies, Madison, Wisconsin).

Más ejemplos de sistemas de transposición que pueden usarse con ciertas realizaciones proporcionadas en el presente documento incluyen Tn552 de *Staphylococcus aureus* (Colegio et al., J. Bacteriol., 183: 2384-8, 2001; Kirby C y col., Mol. Microbiol. , 43: 173-86, 2002), Tyl (Devine & Boeke, Nucleic Acids Res., 22: 3765-72, 1994 y la publicación internacional WO 95/23875), Transposón Tn7 (Craig, NL, Science. 271: 1512, 1996; Craig, NL, Revisado en: Curr Top Microbiol Immunol., 204: 27-48, 1996), Tn/O e IS 10 (Kleckner N, et al., Curr Top Microbiol Immunol., 204: 49-82, 1996), transposasa Mariner (Lampe DJ, et al., EMBO J., 15: 5470-9, 1996), Tc1 (Plasterk RH, Curr. Topics Microbiol. Immunol., 204: 125-43, 1996), Elemento P (Gloor, GB, Methods Mol. Biol., 260: 97-114, 2004), Tn3 (Ichikawa y Ohtsubo, J Biol. Chem. 265: 18829-32, 1990), secuencias de inserción bacteriana (Ohtsubo y Sekine, Curr. Top. Microbiol. Immunol. 204: 1-26, 1996), retrovirus (Brown, et al., Proc Natl Acad Sci USA, 86: 2525-9, 1989), y retrotransposón de levadura (Boeke y Corces, Annu Rev Microbiol. 43: 403-34, 1989). Más ejemplos incluyen IS5, Tn10, Tn903, IS911 y versiones modificadas de enzimas de la familia de la transposasa (Zhang et al., (2009) PLoS Genet. 5: e1000689. Epub 16 de octubre de 2009; Wilson C. et al., (2007) J. Microbiol Métodos 71: 332-5).

El sistema Tn5 usa transposasas Tn5 con la secuencia ME de 19 bases (SEQ ID NO: 3) como el extremo transferido. Sin embargo, como se discutió en el Ejemplo 2, se pueden usar otros extremos transferidos y se pueden describir mediante las fórmulas genéricas

MRWT**G**TGHWKAV**GA**R**ACA**V SEQ ID NO: 1 y

NSHB**G**HSHDDRN**GA**K**ACA**N SEQ ID NO: 2.

Más particularmente, los extremos transferidos pueden ser

CGTTGTGTGGACGAGACAC	SEQ ID NO:	4 11G: C (C1)
CGTTGTGTGGACGAGACAG	SEQ ID NO: 5	11G: C (G1)
AGATGTGCATATGATACAG	SEQ ID NO: 6	Diff1 (G1)
AG.TGT .... AAGAGACAT	SEQ ID NO: 7	Shorty
TGACGCGGGTAAGAGACAA	SEQ ID NO: 22	Malt 1
GGATGCGATGAGGAGACAA	SEQ ID NO: 23	Malt 6
ACATGACCAAGAGAGACAG	SEQ ID NO: 24	Malt 8
AGCGGTGAATAAGAGACAA	SEQ ID NO: 25	Malt 10
AGCGGTGAATAAGAGACAG	SEQ ID NO: 26	Malt 11, o
ACATGAGTATAAGAGACAA	SEQ ID NO: 27	Malt 12.

Con base en las secuencias truncadas discutidas en el Ejemplo 1 y las secuencias complementarias a la SEQ ID NO: 1 y la SEQ ID NO: 2, la presente descripción también proporciona cadenas no transferidas que tienen la fórmula general:

$BTGTYTCBTN_{1-10}$	SEQ ID NO: 20
$NTGTMTCNTN_{0-10}$	SEQ ID NO: 21

en las que el N_{0-10} indica cero a diez nucleótidos. Como ejemplos, la cadena no transferida se puede seleccionar del grupo que consiste en la SEQ ID NO: 9, la SEQ ID NO: 10, la SEQ ID NO: 11, la SEQ ID NO: 12, la SEQ ID NO: 13, la SEQ ID NO: 14, la SEQ ID NO: 15, la SEQ ID NO: 16, la SEQ ID NO: 17 y la SEQ ID NO: 18.

Las secuencias finales pueden comprender además una secuencia de etiqueta, que puede añadirse covalentemente a los fragmentos en el proceso del método de etiquetado. Como se usa en el presente documento, el término "etiqueta" significa una secuencia de nucleótidos que está unida a otro ácido nucleico para proporcionar al ácido nucleico alguna función. Los ejemplos de etiquetas incluyen códigos de barras, sitios de cebado, etiquetas de afinidad y fracciones informadoras.

En general, un código de barras puede incluir una o más secuencias de nucleótidos que pueden usarse para identificar uno o más ácidos nucleicos particulares. El código de barras puede ser una secuencia artificial, o puede ser una secuencia natural, tal como un código g, que se describe en el presente documento. Un código de barras puede comprender al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más nucleótidos consecutivos. En algunas realizaciones, un código de barras comprende al menos aproximadamente 10, 20, 30, 40, 50, 60, 70 80, 90, 100 o más nucleótidos consecutivos. En algunas realizaciones, al menos una porción de los códigos de barras en una población de ácidos nucleicos que comprende códigos de barras es diferente. En algunas realizaciones, al menos aproximadamente 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% de los códigos de barras son diferentes. En más de tales realizaciones, todos los códigos de barras son diferentes. La diversidad de diferentes códigos de barras en una población de ácidos nucleicos que comprende códigos de barras se puede generar aleatoriamente o no aleatoriamente.

En algunas realizaciones, una secuencia de transposón comprende al menos un código de barras. En algunas realizaciones, una secuencia de transposón comprende un código de barras que comprende una primera secuencia de código de barras y una segunda secuencia de código de barras. En algunas de tales realizaciones, la primera secuencia de código de barras puede identificarse o designarse para ser emparejada con la segunda secuencia de código de barras. Por ejemplo, se puede saber que una primera secuencia de código de barras conocida se empareja con una segunda secuencia de código de barras conocida usando una tabla de referencia que comprende una pluralidad de una primera y una segunda secuencias de código de barras que se sabe que están emparejadas entre sí.

En otro ejemplo, la primera secuencia de código de barras puede comprender la misma secuencia que la segunda secuencia de código de barras. En otro ejemplo, la primera secuencia de código de barras puede comprender el complemento inverso de la segunda secuencia de código de barras. En algunas realizaciones, la primera secuencia de código de barras y la segunda secuencia de código de barras son diferentes ("bicódigos"). Se entenderá que en algunas realizaciones, la gran cantidad de códigos de barras disponibles permite que cada molécula de ácido nucleico de plantilla comprenda una identificación única. La identificación única de cada molécula en una mezcla de ácidos nucleicos plantilla puede usarse en varias aplicaciones para identificar moléculas de ácido nucleico individuales, en muestras que tienen múltiples cromosomas, genomas, células, tipos de células, estados de enfermedad celular y especies, por ejemplo, en la secuenciación de haplotipos, discriminación de alelos parentales, secuenciación metagenómica y secuenciación de muestras de un genoma.

En algunas realizaciones, la etiqueta útil es un sitio cebador que puede hibridarse con un cebador. La orientación de

los sitios de cebador en tales realizaciones puede ser tal que un cebador que hibrida con el primer sitio de cebador y un cebador que hibrida con el segundo sitio de cebador están en la misma orientación, o en diferentes orientaciones. En una realización, la secuencia del cebador puede ser complementaria a un cebador usado para la amplificación. En otra realización, la secuencia del cebador es complementaria a un cebador usado para la secuenciación.

En algunas realizaciones, una etiqueta puede incluir un primer sitio cebador, un segundo sitio cebador que tiene un sitio no amplificable dispuesto entre ellos. El sitio no amplificable es útil para bloquear la extensión de una cadena polinucleotídica entre el primer y el segundo sitio cebador, en el que la cadena polinucleotídica se hibrida con uno de los sitios cebadores. El sitio no amplificable también puede ser útil para prevenir concatámeros. Los ejemplos de sitios no amplificables incluyen un análogo de nucleótido, una fracción química no nucleotídica, un aminoácido, un péptido y un polipéptido. En algunas realizaciones, un sitio no amplificable comprende un análogo de nucleótido que no se empareja las bases significativamente con A, C, G o T.

En algunas realizaciones, una etiqueta puede ser una etiqueta de afinidad. Las etiquetas de afinidad pueden ser útiles para la separación masiva de ácidos nucleicos objetivo hibridados con etiquetas de hibridación. Como se usa en el presente documento, el término "etiqueta de afinidad" y equivalentes gramaticales pueden referirse a un componente de un complejo de múltiples componentes, en el que los componentes del complejo de múltiples componentes interactúan específicamente o se unen entre sí. Por ejemplo, una etiqueta de afinidad puede incluir biotina o His que puede unirse a estreptavidina o níquel, respectivamente. Otros ejemplos de complejos de etiqueta de afinidad de múltiples componentes incluyen ligandos y sus receptores, por ejemplo, avidina-biotina, estreptavidina-biotina y derivados de biotina, estreptavidina o avidina, que incluyen, pero no se limitan a, 2-iminobiotina, destiobiotina, NeutrAvidin (Molecular Probes, Eugene, Oregón), CaptAvidin (Molecular Probes) y similares; proteínas/péptidos de unión, que incluyen proteína de unión a maltosa-maltosa (MBP), proteína/péptido de unión a calcio-calcio (CBP); antígeno-anticuerpo, incluyendo etiquetas de epítopo, y sus correspondientes anticuerpos anti-epítopo; haptenos, por ejemplo, dinitrofenilo y digoxigenina, y sus anticuerpos correspondientes; aptámeros y sus objetivos correspondientes; etiquetas de poli-His (por ejemplo, penta-His y hexa-His) y sus compañeros de unión, incluidos los correspondientes materiales de cromatografía de afinidad de iones metálicos inmovilizados (IMAC) y anticuerpos anti-poli-His; fluoróforos y anticuerpos antifluoróforo; y similares. En algunas realizaciones, una etiqueta puede comprender una fracción informadora. Como se usa en este documento, el término "fracción informadora" y equivalentes gramaticales pueden referirse a cualquier etiqueta, marcador o grupo identificable. El experto en la materia apreciará que se pueden usar muchas especies diferentes de fracciones informadoras con los métodos y composiciones descritos en el presente documento, individualmente o en combinación con una o más fracciones informadoras diferentes. En ciertas realizaciones, una fracción informadora puede emitir una señal. Ejemplos de señales fluorescentes, son una señal quimioluminiscente, bioluminiscente, fosforescente, radiactiva, calorimétrica o electroquimioluminiscente. Los ejemplos fracciones informadoras incluyen fluoróforos, radioisótopos, cromógenos, enzimas, antígenos que incluyen etiquetas de epítopos, nanocristales semiconductores tales como puntos cuánticos, metales pesados, colorantes, grupos de fosforescencia, grupos quimioluminiscentes, fracciones de detección electroquímica, proteínas de unión, fósforos, quelatos de tierras raras, quelatos de metales de transición, tintes de infrarrojo cercano, etiquetas de electroquimioluminiscencia y fracciones informadoras compatibles con espectrómetro de masas, tales como etiquetas de masa, etiquetas de carga e isótopos. Más fracciones informadoras que pueden usarse con los métodos y composiciones descritos en el presente documento incluyen marcadores espectrales tales como colorantes fluorescentes (por ejemplo, isotiocianato de fluoresceína, rojo Texas, rodamina y similares), radiomarcadores (por ejemplo, ^{3}H, ^{125}I , ^{35}S, ^{14}C, ^{32}P, ^{33}P, etc.), enzimas (por ejemplo, peroxidasa de rábano picante, fosfatasa alcalina, etc.) etiquetas calorimétricas espectrales tales como oro coloidal o vidrio coloreado o perlas plásticas (por ejemplo, de poliestireno, polipropileno, látex, etc.); etiquetas magnéticas, eléctricas, térmicas; y etiquetas de masa. Las fracciones informadoras también pueden incluir enzimas (peroxidasa de rábano picante, etc.) y partículas magnéticas. Otras fracciones informadoras incluyen cromóforos, fósforos y fracciones fluorescentes, por ejemplo, rojo Texas, digoxigenina, biotina, 1-aminonaftaleno y 2-aminonaftaleno, p,p'-diaminoestilbenos, pirenos, sales de fenantridina cuaternarias, 9-aminoacridinas, p,p'-diamino benzofenona, iminas, antracenos, oxacarbocianina, merocianina, 3-aminoequilenina, perileno, bis-benzoxazol, bis-p-oxazolilbenceno, 1,2-benzofenazina, retinol, sales de bis-3-aminopiridinio, helebrigenina, tetraciclina, esterofenol, bencimidazolilfenilamina, 2--oxo-3-cromeno, indol, xanteno, 7-hidroxicumarina, fenoxazina, calicilato, estrofantidina, porfirinas, triarilmetanos y flavina. Los compuestos fluorescentes individuales que tienen funciones para unirse a un elemento deseablemente detectado en un aparato o ensayo proporcionado en el presente documento, o que pueden modificarse para incorporar tales funciones incluyen, por ejemplo, cloruro de dansilo; fluoresceínas tales como 3,6-dihidroxi-9-fenilxantidrol; isotiocianato de rodamina; N-fenil-1-amino-8-sulfonatonaftaleno; N-fenil-2-amino-6-sulfonatonaftaleno; ácido 4-acetamido-4-isotiocianato-estilbeno-2,2'-disulfónico; ácido pireno-3-sulfónico; 2-toluidinonaftaleno-6-sulfonato; N-fenil-N-metil-2-aminonaftaleno-6-sulfonato; bromuro de etidio; estebrina; auromina-0,2-(9'-antroil)palmitato; dansil fosfatidiletanolamina; N,N'-dioctadeciloxacarbocianina; N,N'-dihexiloxacarbocianina; merocianina, 4-(3'-pirenil)estearato; d-3-aminodesoxi-equilenina; 12-(9'-antroil)estearato; 2-metilantraceno; 9-vinilantraceno; 2,2'(vinilen-p-fenilen) bisbenzoxazol; p-bis(2-metil-5-fenil-oxazolil)benceno; 6-dimetilamino-1,2-benzofenazina; retinol; bis(3'-aminopiridinio) yoduro de 1,10-decandiilo; sulfonaftilhidrazona de helibrienina; clorotetraciclina; N-(7-dimetilamino-4-metil-2-oxo-3-cromenil)maleimida; N-(p-(2-bencimidazolil)-fenil)maleimida; N-(4-fluorantil)maleimida; bis(ácido homovanílico); resazarina; 4-cloro-7-nitro-2,1,3-benzooxadiazol; merocianina 540; resorufina; rosa de bengala; 2,4-difenil-3-(2H)-furanona, complejos de lantánidos fluorescentes, incluidos los de Europio y Terbio, fluoresceína, rodamina, tetrametilrodamina, eosina, eritrosina, cumarina, metil-cumarinas, puntos cuánticos (también denominados

"nanocristales" : véase la patente de los Estados Unidos No. 6.544.732), pireno, verde de malaquita, estilbeno, amarillo Lucifer, Cascade Blue^{MR}, Rojo Texas, tintes Cy (Cy3, Cy5, etc.), tintes Alexa Fluor®, ficoeritrina, bodipy y otros descritos en la sexta edición del Manual de Sondas Moleculares de Richard P. Haugland. La divulgación proporciona además un ácido nucleico que comprende una o dos copias de las secuencias finales de transposasa, que pueden generarse realizando el método de la invención. Cuando el método se realiza en una secuencia objetivo y se generan diferentes fragmentos, la invención proporciona una biblioteca de estos diferentes ácidos nucleicos.

La cadena no transferida con o sin grupos protectores de nucleasas y/o grupos de terminación de cadena (por ejemplo, fosforotioato y/o didesoxi) se disocia luego de la cadena transferida y un oligo de reemplazo (que puede contener etiquetas de ADN adicionales, como se discutió anteriormente, tal como una etiqueta de secuenciación) se hibrida con la secuencia de cadena transferida complementaria con o sin grupos protectores de nucleasa (por ejemplo, fosforotioatos). Se pueden usar enzimas modificadoras de ácido nucleico no desplazables, que consisten en una ADN polimerasa y una ADN ligasa. Las ADN polimerasas y la ligasa se utilizan para rellenar y ligar la brecha entre el ADN monoetiquetado y el oligonucleótido de reemplazo, lo que da como resultado un fragmento de ADNbc con una etiqueta 5' y una 3' unidas covalentemente. Por lo tanto, el método de la presente invención proporciona una forma novedosa de crear fragmentos de ADNbc con dos etiquetas, en el que la reacción en cadena de la polimerasa es opcional.

La presente descripción también proporciona un método mejorado para preparar una biblioteca etiquetada direccionalmente. El método comienza proporcionando ADN con una modificación específica de la cadena, tal como la incorporación de un nucleótido escindible en una cadena particular, tal como un nucleótido químicamente lábil o uno que contenga uracilo u 8-oxoguanina. Otros nucleótidos modificados útiles incluyen 8-oxoadenina, fapy-guanina, metil-fapy-guanina, fapy-adenina, aflatoxina B_1-fapy-guanina, 5-hidroxi-citosina, 5-hidroxi-uracilo y aductos de N-7 guanina de anillo abierto (7-metilguanina). En una realización particular, cada cadena puede contener una modificación diferente, por ejemplo, una cadena puede contener modificaciones de uracilo y la otra cadena puede contener modificaciones de 8-oxo-guanina.

El ADN luego se escinde con un transposoma monoetiquetado que consiste en una o más moléculas de transposasa y dos secuencias oligonucleotídicas de ADN que son el ADN del extremo hibridado modificado (ME). Una secuencia (ME) en la cadena de ADN transferida y una cadena de ADN no transferida que puede contener una secuencia de ME de 19 pb o una secuencia de ADN truncada. La cadena no transferida (con o sin grupos protectores de nucleasa y/o grupos de terminación de cadena, por ejemplo, fosforotioato y/o didesoxi) se disocian luego de la cadena transferida y un oligo de reemplazo (que puede contener una secuencia de ADN adicional, tal como una etiqueta de secuenciación) se hibrida con la secuencia de cadena transferida complementaria con o sin grupos protectores de nucleasa (por ejemplo, fosforotioatos). Se usan enzimas modificadoras de ácido nucleico no desplazables que consisten en ADN polimerasa (por ejemplo, polimerasas termoestables o polimerasas no termoestables tales como ADN polimerasa I o fragmento Klenow exo-) y una ADN ligasa. Las ADN polimerasas y la ligasa se utilizan para rellenar y ligar el espacio entre el ADN monoetiquetado y el oligonucleótido de reemplazo, lo que da como resultado un fragmento de ADNbc con una etiqueta 5' y 3' unida de forma covalente. Alternativamente, se puede proporcionar un oligonucleótido para llenar el espacio, seguido de la ligación.

La cadena modificada o no modificada puede tratarse específicamente para enriquecer o suprimir su función. El tratamiento puede incluir el uso de una enzima, tal como uracilo ADN glicosilasa (UDG), también conocida como uracilo N-glicosilasa (UNG), endonucleasa humana apurínica/apirimidínica (APE I), formamidopirimidina-ADN glicosilasa (FPG), también conocida como 8-oxicuanina ADN glicosilasa, endonucleasa IV y quinasa, Endo III, Endo VIII, hOGG1, T7 Endo I, T4 PDG y afu UDG. En otra realización, una cadena puede enriquecerse selectivamente por extensión usando una polimerasa que tiene preferencia por los nucleótidos naturales, en lugar de los nucleótidos modificados químicamente. Un ejemplo de tal polimerasa es una fusión de una polimerasa similar a *Pyrococcus* a un dominio de unión a ADNbc de *Sulfolobus solfataricus* (SSo7d). Por lo tanto, el método de la presente divulgación proporciona una forma novedosa de crear fragmentos de ADN con dos etiquetas a partir de una única cadena predeterminada, en la que no es necesaria una amplificación posterior.

La descripción también proporciona un transposoma que comprende una transposasa y un ácido nucleico que contiene una o más secuencias finales. La divulgación proporciona además un método para elaborar un transposoma con las secuencias finales proporcionando una transposasa y proporcionando las secuencias finales transferidas, y luego permitiendo que la transposasa se una a la secuencia final transferida. Los ejemplos de condiciones de reacción se discuten en el Ejemplo 1 a continuación. Por lo tanto, la divulgación proporciona un método de etiquetado que comprende las etapas de (a) proporcionar ácidos nucleicos objetivo (b) proporcionar los transposomas y (c) permitir que los transposomas fragmenten los ácidos nucleicos objetivo y etiqueten al menos las secuencias finales de transposasa en los extremos de los fragmentos .

Definiciones

El término "que comprende" como se usa en el presente documento es sinónimo con "que incluye", "que contiene" o "caracterizado por", y es inclusivo o abierto y no excluye elementos adicionales no mencionados o etapas del método.

Como se usa en el presente documento, el término "al menos una porción" y/o equivalentes gramaticales del mismo

pueden referirse a cualquier fracción de una cantidad total. Por ejemplo, "al menos una porción" puede referirse a al menos aproximadamente 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, 99,9 % o 100% de una cantidad total.

Como se usa en el presente documento, el término "ácido nucleico" y/u "oligonucleótido" y/o equivalentes gramaticales del mismo puede referirse a al menos dos monómeros de nucleótidos unidos entre sí. Un ácido nucleico generalmente puede contener enlaces fosfodiéster; sin embargo, en algunas realizaciones, los análogos de ácido nucleico pueden tener otros tipos de cadenas principales, que comprenden, por ejemplo, fosforamida (Beaucage, et al., Tetrahedron, 49: 1925 (1993); Letsinger, J. Org. Chem., 35: 3800 (1970); Sprinzl, et al., Eur. J. Biochem., 81: 579 (1977); Letsinger, et al., Nucl. Acids Res., 14: 3487 (1986); Sawai, et al., Chem Lett., 805 (1984), Letsinger, et al., J. Am. Chem. Soc., 110: 4470 (1988); y Pauwels, et al., Chemica Scripta, 26: 141 (1986)), fosforotioato. (Mag, et al., Nucleic Acids Res., 19: 1437 (1991); y la patente de los Estados Unidos No. 5.644.048), fosforoditioato (Briu, et al., J. Am. Chem. Soc., 111: 2321 (1989), Enlaces O-metilfosforoamidita (véase Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), y esqueletos de ácido nucleico peptídico y enlaces (véase Egholm, J. Am. Chem. Soc., 114: 1895 (1992); Meier, et al., Chem. Int. Ed. Engl., 31: 1008 (1992); Nielsen, Nature, 365: 566 (1993); Carlsson, et al., Nature, 380: 207 (19 96)).

Otros ácidos nucleicos análogos incluyen aquellos con cadenas principales positivas (Denpcy, et al., Proc. Natl. Acad. Sci. USA, 92: 6097 (1995)); cadenas principales no iónicas (patentes de los Estados Unidos Nos. 5,386,023; 5,637,684; 5,602,240; 5,216,141; y 4,469,863; Kiedrowshi, et al., Angew. Chem. Intl. Ed. English, 30: 423 (1991); Letsinger, et al., J. Am. Chem. Soc., 110: 4470 (1988); Letsinger, et al., Nucleosides & Nucleotides, 13: 1597 (1994); Capítulos 2 y 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. YS Sanghui y P. Dan Cook; Mesmaeker, et al., Bioorganic & Medicinal Chem. Lett., 4: 395 (1994); Jeffs, y col., J. Biomolecular NMR, 34:17 (1994); Tetrahedron Lett., 37: 743 (1996)) y sin ribosa (patente de los Estados Unidos No. 5.235.033 y No. 5.034.506, y Capítulos 6 y 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. YS Sanghui y P. Dan Coo). Los ácidos nucleicos también pueden contener uno o más azúcares carbocíclicos (véase Jenkins, et al., Chem. Soc. Rev., (1995) páginas 169 -176).

Las modificaciones de la cadena principal de fosfato de ribosa se pueden hacer para facilitar la adición de fracciones adicionales tales como marcadores, o para aumentar la estabilidad de tales moléculas en ciertas condiciones. Además, se pueden hacer mezclas de ácidos nucleicos y análogos naturales. Alternativamente, se pueden hacer mezclas de diferentes análogos de ácido nucleico y mezclas de ácidos nucleicos y análogos naturales. Los ácidos nucleicos pueden ser monocatenarios o bicatenarios, según se especifique, o pueden contener porciones de secuencias tanto bicatenarias como monocatenarias. El ácido nucleico puede ser ADN, por ejemplo, genómico o ADNc, ARN o un híbrido, de células individuales, células múltiples o de especies múltiples, como con muestras metagenómicas, tales como muestras ambientales. Un ácido nucleico puede contener cualquier combinación de desoxirribonucleótidos y ribonucleótidos, y cualquier combinación de bases, incluidos uracilo, adenina, timina, citosina, guanina, inosina, xantanina, hipoxantanina, isocitosina, isoguanina y análogos de bases como nitropirrol (incluyendo 3-nitropirrol) y nitroindol (incluido 5-nitroindol), etc.

En algunas realizaciones, un ácido nucleico puede incluir al menos una base promiscua. Las bases promiscuas pueden emparejarse con más de un tipo diferente de base. En algunas realizaciones, una base promiscua puede emparejarse con al menos dos tipos diferentes de bases y no más de tres tipos diferentes de bases. Un ejemplo de una base promiscua incluye inosina que puede combinarse con adenina, timina o citosina. Otros ejemplos incluyen hipoxantina, 5-nitroindol, 5-nitroindol acílico, 4-nitropirazol, 4-nitroimidazol y 3-nitropirrol (Loakes et al., Nucleic Acid Res. 22: 4039 (1994); Van Aerschot et al., Nucleic Acid Res. 23: 4363 (1995); Nichols et al., Nature 369: 492 (1994); Bergstrom et al., Nucleic Acid Res. 25: 1935 (1997); Loakes et al., Nucleic Acid Res. 23: 2361 (1995); Loakes et al., J. Mol. Biol. 270: 426 (1997); y Fotin et al., Nucleic Acid Res. 26: 1515 (1998)). También se pueden usar bases promiscuas que pueden emparejarse con al menos tres, cuatro o más tipos de bases.

Como se usa en el presente documento, el término "análogo de nucleótido" y/o equivalentes gramaticales del mismo pueden referirse a análogos sintéticos que tienen porciones de base de nucleótidos modificadas, porciones de pentosa modificadas y/o porciones de fosfato modificadas, y, en el caso de polinucleótidos, enlaces internucleotídicos modificados, como se describe generalmente en otra parte (por ejemplo, Scheit, Nucleotide Analogs, John Wiley, Nueva York, 1980; Englisch, Angew. Chem. Int. Ed. Engl. 30: 613-29, 1991; Agarwal, Protocols for Polynucleotides and Analogs, Humana Press, 1994 y S. Verma y F. Eckstein, Ann. Rev. Biochem. 67: 99-134, 1998). Generalmente, las porciones de fosfato modificadas comprenden análogos de fosfato en los que el átomo de fósforo está en el estado de oxidación +5 y uno o más de los átomos de oxígeno se reemplazan con una fracción diferente de oxígeno, por ejemplo, azufre. Los ejemplos de análogos de fosfato incluyen, entre otros, fosforotioato, fosforoditioato, fosforoselenoato, fosforodiselenoato, fosforoanilotioato, fosforoanilidato, fosforamidato, boronofosfatos, incluidos los contraiones asociados, por ejemplo, H^+, NH_4^+, Na^+, si tales contra iones están presentes. Los ejemplos de porciones de bases nucleotídicas modificadas incluyen, pero no se limitan a, 5-metilcitosina (5 mC); análogos de C-5-propinilo, que incluyen pero no se limitan a, C-5 propinil-C y C-5 propinil-U; 2,6-diaminopurina, también conocida como 2-amino adenina o 2-amino-dA); hipoxantina, pseudouridina, 2-tiopirimidina, isocitosina (isoC), 5-metil isoC e isoguanina (isoG; véase, por ejemplo, la patente de Estados Unidos número 5.432.272). Los ejemplos de porciones de pentosa modificadas incluyen, pero no se limitan a, análogos de ácido nucleico bloqueado (LNA) que incluyen, sin limitación,

Bz-A-LNA, 5-Me-Bz-C-LNA, dmf-G-LNA y T-LNA (véase, por ejemplo, The Glen Report, 16 (2): 5, 2003; Koshkin et al., Tetrahedron 54: 3607-30, 1998), y modificaciones 2' o 3' en las que la posición 2' o 3' es hidrógeno, hidroxi, alcoxi (por ejemplo, metoxi, etoxi, aliloxi, isopropoxi, butoxi, isobutoxi y fenoxi), azido, amino, alquilamino, fluoro, cloro o bromo. Los enlaces internucleotídicos modificados incluyen análogos de fosfato, análogos que tienen enlaces entre subunidades aquirales y no cargadas (por ejemplo, Sterchak, EP et al., Organic Chem., 52: 4202, 1987), y polímeros basados en morfolino sin carga que tienen enlaces entre subunidades aquirales (véase, por ejemplo, la patente de Estados Unidos No. 5.034.506). Algunos análogos de enlaces internucleotídicos incluyen heterociclos unidos a morfolidato, acetal y poliamida. En una clase de análogos de nucleótidos, conocidos como ácidos nucleicos peptídicos, incluidos los ácidos nucleicos peptídicos seudocomplementarios ("PNA"), un enlace convencional de azúcar e internucleótido se ha reemplazado con un polímero de cadena principal 2-aminoetilglicina amida (véase, por ejemplo, Nielsen et al. , Science, 254: 1497-1500, 1991; Egholm et al., J. Am. Chem. Soc., 114: 1895-1897 1992; Demidov et al., Proc. Natl. Acad. Sci. 99: 5953-58 , 2002; Peptide Nucleic Acids: Protocols and Applications, Nielsen, ed., Horizon Bioscience, 2004).

Los siguientes ejemplos proporcionan realizaciones ilustrativas y de ninguna manera limitan las invenciones proporcionadas en este documento.

Ejemplos

Ejemplo 1: reemplazo de oligonucleótidos

Se preparó una serie de transposomas Tn5 hiperactivos, cada uno con una secuencia final de transposón de 19 bases:

AGATGTGTATAAGAGACAG (SEQ ID NO: 3) (ME)

(la "cadena transferida") y una de las siguientes secuencias finales (la "cadena no transferida"), mostrada en orientación 5' a 3':

CTGTCTCT	SEQ ID NO: 8	8 bases
CTGTCTCTT	SEQ ID NO: 9	9 bases
CTGTCTCTTA	SEQ ID NO: 10	10 bases
CTGTCTCTTAT	SEQ ID NO: 11	11 bases
CTGTCTCTTATA	SEQ ID NO: 12	12 bases
CTGTCTCTTATAC	SEQ ID NO: 13	13 bases
CTGTCTCTTATACA	SEQ ID NO: 14	14 bases
CTGTCTCTTATACAC	SEQ ID NO: 15	15 bases
CTGTCTCTTATACACA ...	SEQ ID NO: 16	16 bases
CTGTCTCTTATACACAT ..	SEQ ID NO: 17	17 bases
CTGTCTCTTATACACATC.	SEQ ID NO: 18	18 bases
CTGTCTCTTATACACATCT	SEQ ID NO: 19	19 bases (ME)

en las que las secuencias finales contenían una secuencia de etiqueta como se representa en la Figura 3. Los transposomas se prepararon en una mezcla de reacción de 2 µL de secuencia final (25 µM), fosforilada; 2 µL de transposasa Tn5 a razón de 10 U/µL, en un volumen final de 50 µL con concentraciones finales de Tris-acetato 33 mM, pH 7,8, acetato de magnesio 10 mM y acetato de potasio 66 mM. Alternativamente, las concentraciones finales en el volumen de reacción pueden ser Tris-acetato 10 mM, pH 7,6, cloruro de magnesio 5 mM y opcionalmente 10% (v/v) de dimetilformamida. Después de mezclar, la reacción se incubó durante 1 hora a 37 °C. La reacción se detuvo con 10 µL de solución de detención: sacarosa al 15%, EDTA 66 mM, Tris 20 mM pH 8, SDS al 0.1%, Orange G al 0.9% (Sigma O-7252) y Proteinasa K a razón de 100 µg/mL. Después de la adición de la solución de parada, la mezcla se calentó a 50 °C durante 10 minutos.

Se realizaron reacciones de etiquetado con los transposomas para comparar la capacidad de los transposomas que tienen secuencias finales truncadas no transferidas en comparación con el extremo no transferido de 19 bases, que sirve como control positivo y ningún extremo no transferido como control negativo. Con base en el análisis de electroforesis en gel, el ADN objetivo en una muestra fue fragmentado eficientemente por los transposomas que tienen secuencias finales de 18 bases a 12 bases, con una fragmentación menos eficiente con secuencias finales de 11 bases a 9 bases. La secuencia final de 8 bases demostró cierta fragmentación, pero no de manera eficiente.

La secuencia final de 14 bases se seleccionó para experimentos adicionales como se ilustra en la Figura 4. Después del etiquetado con transposones que tienen la cadena transferida de 19 bases y la cadena no transferida de 14 bases, el producto se mezcló con oligonucleótido de reemplazo (SEQ ID NO: 19 con una etiqueta de secuenciación como la "Etiqueta arbitraria 2") a 45 °C durante 1 minuto y luego a 37 °C durante 30 minutos para eliminar el oligonucleótido no transferido y reemplazarlo con el oligonucleótido de reemplazo. Los espacios restantes de 9 bases se rellenaron con polimerasa Tth, fragmento grande (sin desplazamiento de cadena), seguido de ligación usando ADN ligasa de *E. coli*, de acuerdo con las condiciones recomendadas por el fabricante.

El reemplazo se demostró mediante la adición de una secuencia adicional de 100 bases tal como se resolvió en un Agilent BioAnalyzer 2100 usando un chip de ADN de alta sensibilidad. Se analizaron las bibliotecas genómicas de *Rhodobacter*, *E. coli* y *Staphylococcus* debido a sus contenidos de GC divergentes de 70, 50 y 33%, respectivamente. Estas bibliotecas se crearon utilizando las estrategias de reemplazo de oligonucleótidos descritas en el presente documento. La composición de GC en las primeras 30 bases no mostró sesgo de inserción adicional basado en el contenido de GC del genoma huésped. El análisis de la composición de GC en las primeras 30 bases no demostró sesgo de inserción alterado en función del contenido de GC del genoma huésped. Los datos mostraron que hubo una mayor profundidad de cobertura donde el contenido de GC es superior al 60% y una cobertura constante en intervalos de contenido de GC más bajos.

Ejemplo 2 - nuevas secuencias finales de transposasa

Se prepararon varios millones de secuencias finales de transposasa usando una biblioteca parcialmente aleatorizada con base en la secuencia de ME de 19 bases (SEQ ID NO: 3). A partir de esta biblioteca, se identificaron varios extremos activos del transposón utilizando ADN genómico de *E. coli* MG1655 como ácido nucleico objetivo y se secuenciaron en un instrumento de secuenciación Genome Analyzer$_{II}$ (Illumina, Inc.).

CGTTGTGTGGACGAGACAC	SEQ ID NO: 4	11G: C (C1)
CGTTGTGTGGACGAGACAG	SEQ ID NO: 5	11G: C (G1)
AGATGTGCATATGATACAG	SEQ ID NO: 6	Diff1 (G1)
AG.TGT .... AAGAGACAT	SEQ ID NO: 7	Shorty
TGACGCGGGTAAGAGACAA	SEQ ID NO: 22	Malt 1
GGATGCGATGAGGAGACAA	SEQ ID NO: 23	Malt 6
ACATGACCAAGAGAGACAG	SEQ ID NO: 24	Malt 8
AGCGGTGAATAAGAGACAA	SEQ ID NO: 25	Malt 10
AGCGGTGAATAAGAGACAG	SEQ ID NO: 26	Malt 11
ACATGAGTATAAGAGACAA	SEQ ID NO: 27	Malt 12

Las secuencias finales alternativas se confirmaron mediante secuenciación capilar. Un mutante designado 11G: C(G1) (SEQ ID NO: 5) tenía una temperatura de fusión significativamente alta (58 °C en comparación con 44 °C) y demostró una actividad de transposición comparable *in vitro* como la secuencia de ME, como lo demuestra la dilución en paralelo con transposones con la secuencia de ME.

Listado de secuencias

<110 Illumina, Inc.

<120 REEMPLAZO DE NUCLEÓTIDOS POR BIBLIOTECAS DOBLEMENTE ETIQUETADAS Y DIRECCIONALES

<130 TXH/71337EP2

<140 PCT/US 12/023139
<141 2012-01-30

<150 US 61/506.777
<151 2011-07-12

<150 US 61/437.451
<151 2011-01-28

<160 27

<170 PatentIn versión 3.5

<210 1
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético
<400 1
mrwtgtghwk avgaracav 19

<210 2
<211 19
<212 ADN

<213 Secuencia Artificial

<220
<223 constructo sintético

<220
<221 característica miscelánea
<222 (1)..(1)
<223 n en posición 1 puede ser A, C, T, o G

<220
<221 característica miscelánea
<222 (12)..(12)
<223 n en posición 12 puede ser A, C, T, o G

<220
<221 característica miscelánea
<222 (19)..(19)
<223 n en posición 19 puede ser A, C, T, o G

<400 2
nshbghshdd rngakacan 19

<210 3
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 3
agatgtgtat aagagacag 19

<210 4
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 4
cgttgtgtgg acgagacac 19

<210 5
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 5
cgttgtgtgg acgagacag 19

<210 6
<211 19
<212 ADN
<213 Secuencia Artificial
<220
<223 constructo sintético

<400 6
agatgtgcat atgatacag 19

<210 7
<211 14
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 7
agtgtaagag acat 14

<210 8
<211 8
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 8
ctgtctct 8

<210 9
<211 9
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 9
ctgtctctt 9

<210 10
<211 10
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 10
ctgtctctta 10

<210 11
<211 11
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 11
ctgtctctta t 11

<210 12
<211 12
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 12
ctgtctctta ta 12

<210 13
<211 13
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 13
ctgtctctta tac 13

<210 14
<211 14
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 14
ctgtctctta taca 14

<210 15
<211 15
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 15
ctgtctctta tacac 15

<210 16
<211 16
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 16
ctgtctctta tacaca 16

<210 17
<211 17
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 17
ctgtctctta tacacat 17

<210 18
<211 18
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 18
ctgtctctta tacacatc 18

<210 19
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 19
ctgtctctta tacacatct 19

<210 20
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<220
<221 característica miscelánea
<222 (10)..(19)
<223 n en las posiciones 10-19 puede ser A, C, T, o G y hasta 9 de ellas pueden estar ausentes

<400 20
btgtytcbtn nnnnnnnnn 19

<210 21
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<220
<221 característica miscelánea
<222 (1)..(1)
<223 n en posición 1 puede ser A, C, T, o G

<220
<221 característica miscelánea
<222 (8)..(8)
<223 n en la posición 8 puede ser A, C, T, o G

<220
<221 característica miscelánea
<222 (10)..(19)
<223 n en la posición 8 puede ser A, C, T, o G hasta 9 de ellas pueden estar ausentes

<400 21
ntgtmtcntn nnnnnnnnn 19

<210 22
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético
<400 22
tgacgcgggt aagagacaa 19

<210 23
<211 19
<212 ADN

<213 Secuencia Artificial

<220
<223 constructo sintético

<400 23
ggatgcgatg aggagacaa 19

<210 24
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 24
acatgaccaa gagagacag 19

<210 25
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 25
agcggtgaat aagagacaa 19

<210 26
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 26
agcggtgaat aagagacag 19

<210 27
<211 19
<212 ADN
<213 Secuencia Artificial

<220
<223 constructo sintético

<400 27
acatgagtat aagagacaa 19

REIVINDICACIONES

1. Un producto de ácido nucleico al menos parcialmente bicatenario de una reacción de etiquetado que comprende:

(a) un fragmento de un ácido nucleico objetivo bicatenario;
(b) una o dos copias de una secuencia final de transposasa que es:

MRWTGTGHWKAVGARACAV (SEQ ID NO: 1) o
NSHBGHSHDDRNGAKACAN (SEQ ID NO: 2),
pero excluyendo AGATGTGTATAAGAGACAG (SEQ ID NO: 3),

como cadenas transferidas, en las que cada secuencia final de transposasa está unida covalentemente a un extremo 5' del fragmento del ácido nucleico objetivo bicatenario;
(c) una primera secuencia de etiqueta en el extremo 5' de al menos una de las secuencias finales del transposón unidas covalentemente; y
(d) un oligonucleótido hibridado con al menos una de las secuencias finales transferidas y que comprende una segunda secuencia de etiqueta ligada en el extremo 3' del oligonucleótido, en el que la segunda secuencia de etiqueta no es complementaria a la primera secuencia de etiqueta.

2. El ácido nucleico de la reivindicación 1, en el que un oligonucleótido se hibrida con cada cadena transferida y una secuencia de etiqueta se liga en el extremo 3' de cada oligonucleótido.

3. El ácido nucleico de la reivindicación 1, en el que la secuencia final de la transposasa se selecciona del grupo que consiste en:

CGTTGTGTGGACGAGACAC (SEQ ID NO: 4),
CGTTGTGTGGACGAGACAG (SEQ ID NO: 5),
AGATGTGCATATGATACAG (SEQ ID NO: 6),
TGACGCGGGTAAGAGACAA (SEQ ID NO: 22),
GGATGCGATGAGGAGACAA (SEQ ID NO: 23),
ACATGACCAAGAGAGACAG (SEQ ID NO: 24),
AGCGGTGAATAAGAGACAA (SEQ ID NO: 25),
AGCGGTGAATAAGAGACAG (SEQ ID NO: 26), y
ACATGAGTATAAGAGACAA (SEQ ID NO: 27).

4. El ácido nucleico de la reivindicación 1, en el que la secuencia final de la transposasa es:

CGTTGTGTGGACGAGACAG (SEQ ID NO: 5).

5. El ácido nucleico de una cualquiera de las reivindicaciones 1-4, en el que la secuencia final de transposasa tiene una secuencia de etiqueta ligada en su extremo 5'.

6. El ácido nucleico de una cualquiera de las reivindicaciones 1-5, en el que cada secuencia de etiqueta se selecciona de un código de barras, un sitio cebador, una etiqueta de afinidad y una fracción informadora.

7. El ácido nucleico de la reivindicación 5, en el que el ácido nucleico es bicatenario.

8. Una biblioteca de diferentes ácidos nucleicos de cualquiera de las reivindicaciones 1-7.

9. Un método de etiquetado que comprende las etapas de:

(a) proporcionar ácidos nucleicos objetivo bicatenarios;
(b) proporcionar transposomas que comprenden una transposasa y un ácido nucleico que comprende una o dos copias de las secuencias finales de transposasa como las cadenas transferidas:

MRWTGTGHWKAVGARACAV (SEQ ID NO: 1) o
NSHBGHSHDDRNGAKACAN (SEQ ID NO: 2),
pero excluyendo AGATGTGTATAAGAGACAG (SEQ ID NO: 3);

(c) permitir que los transposomas fragmenten los ácidos nucleicos objetivo y etiqueten al menos las cadenas transferidas a los extremos de los fragmentos; y
(d) hibridar oligonucleótidos con las cadenas transferidas, cada oligonucleótido tiene una secuencia de etiqueta ligada en su extremo 3';

generando así una biblioteca de fragmentos etiquetados.

10. El método de etiquetado de la reivindicación 9, en el que las secuencias finales de transposasa se seleccionan del grupo que consiste en:

CGTTGTGTGGACGAGACAC	(SEQ ID NO: 4),
CGTTGTGTGGACGAGACAG	(SEQ ID NO: 5),
AGATGTGCATATGATACAG	(SEQ ID NO: 6),
TGACGCGGGTAAGAGACAA	(SEQ ID NO: 22),
GGATGCGATGAGGAGACAA	(SEQ ID NO: 23),
ACATGACCAAGAGAGACAG	(SEQ ID NO: 24),
AGCGGTGAATAAGAGACAA	(SEQ ID NO: 25),
AGCGGTGAATAAGAGACAG	(SEQ ID NO: 26), y
ACATGAGTATAAGAGACAA	(SEQ ID NO: 27).

11. El método de etiquetado de la reivindicación 9 o la reivindicación 10, en el que cada secuencia final de transposasa tiene una secuencia de etiqueta ligada en su extremo 5'.

12. El método de etiquetado de una cualquiera de las reivindicaciones 9-11, en el que cada secuencia de etiqueta se selecciona de un código de barras, un sitio de cebador, una etiqueta de afinidad y una fracción informadora.

13. El método de etiquetado de la reivindicación 11, en el que los fragmentos etiquetados son bicatenarios.

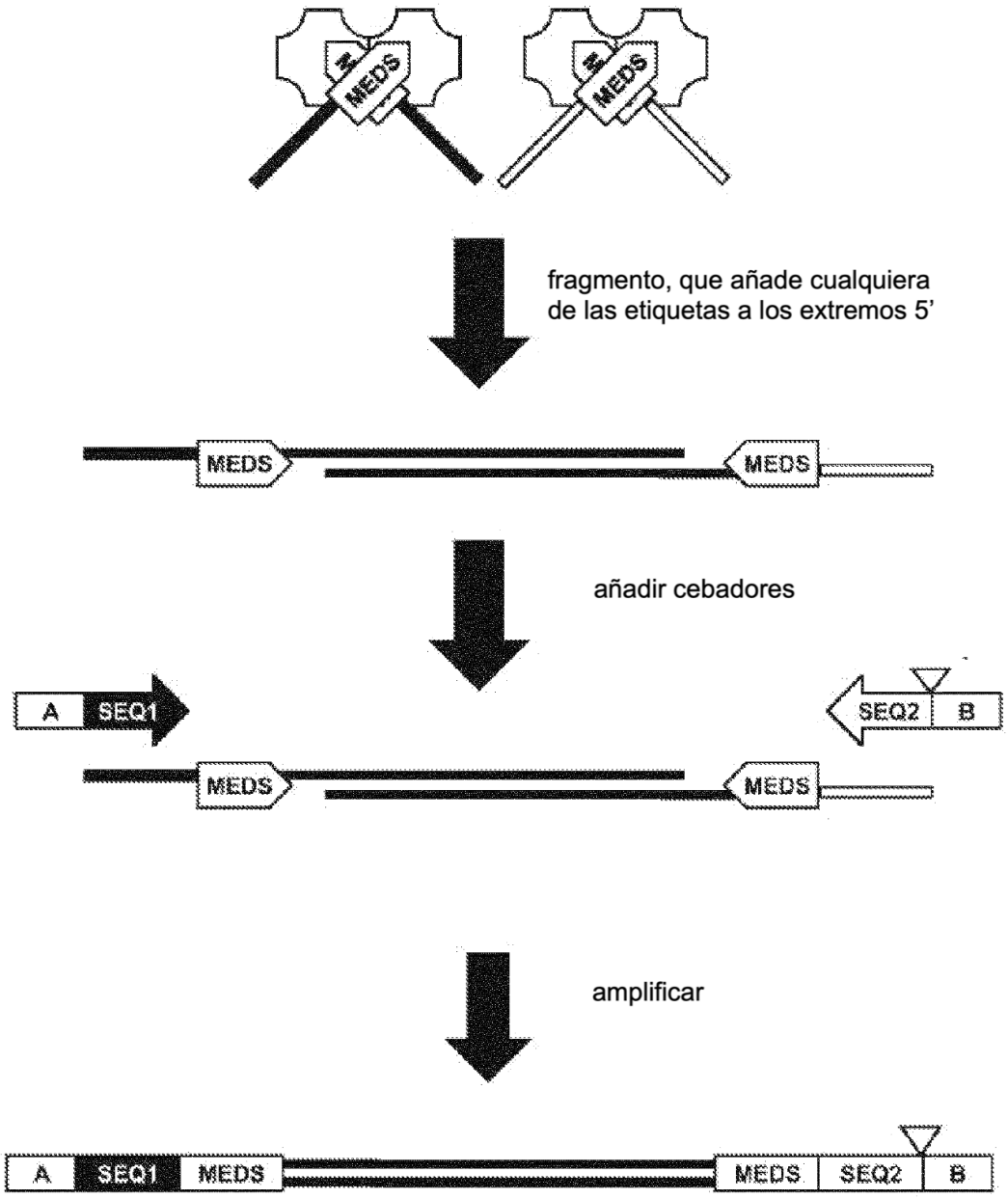
MEDS
MEDS
fragmento, que añade cualquiera
de las etiquetas a los extremos 5'
MEDS
MEDS
añadir cebadores
A
SEQ1
SEQ2
B
MEDS
MEDS
amplificar
A
SEQ1
MEDS
MEDS
SEQ2
B

Fig. 1

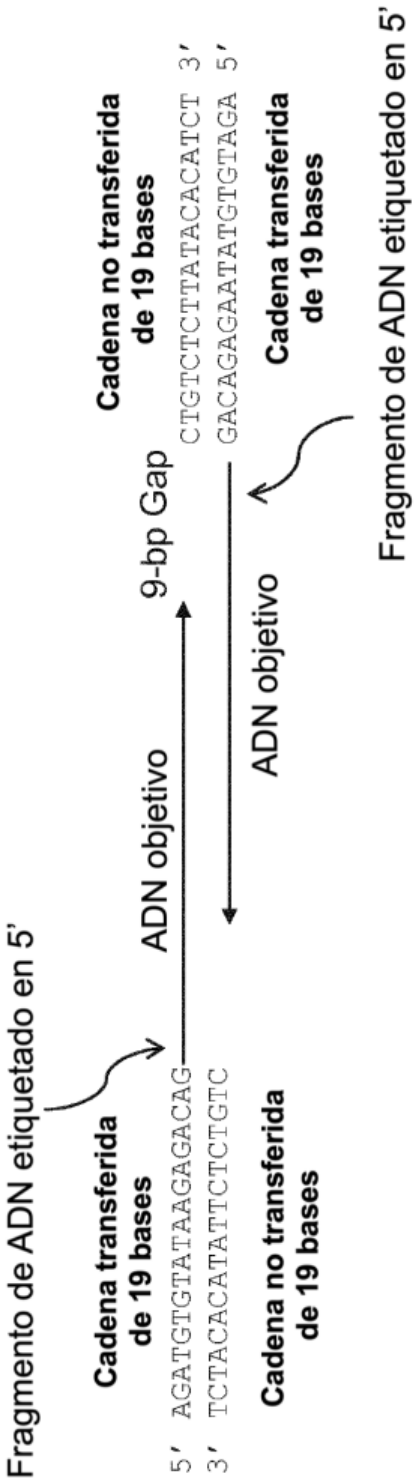
Fragmento de ADN etiquetado en 5'
Cadena transferida de 19 bases
5' AGATGTGTATAAGAGACAG
3' TCTACACATATTCTCTGTC
Cadena no transferida de 19 bases
ADN objetivo
9-bp Gap
ADN objetivo
Cadena no transferida de 19 bases
CTGTCTCTTATACACATCT 3'
GACAGAGAATATGTGTAGA 5'
Cadena transferida de 19 bases
Fragmento de ADN etiquetado en 5'

Fig. 2a

MEDS
MEDS

Fig. 2b

Supresiones de cadena no transferida
Extremo 1 del transposón
Etiqueta Arbitraria 1
5'
AGATGTGTATAAGAGACAG 3'
19 bases 3' TCTACACATATTCTCTGTC 5'
18 bases 3' CTACACATATTCTCTGTC 5'
17 bases 3' TACACATATTCTCTGTC 5'
16 bases 3' ACACATATTCTCTGTC 5'
15 bases 3' CACATATTCTCTGTC 5'
14 bases 3' ACATATTCTCTGTC 5'
13 bases 3' CATATTCTCTGTC 5'
12 bases 3' ATATTCTCTGTC 5'
11 bases 3' TATTCTCTGTC 5'
10 bases 3' ATTCTCTGTC 5'
9 bases 3' TTCTCTGTC 5'
8 bases 3' TCTCTGTC 5'

Fig. 3

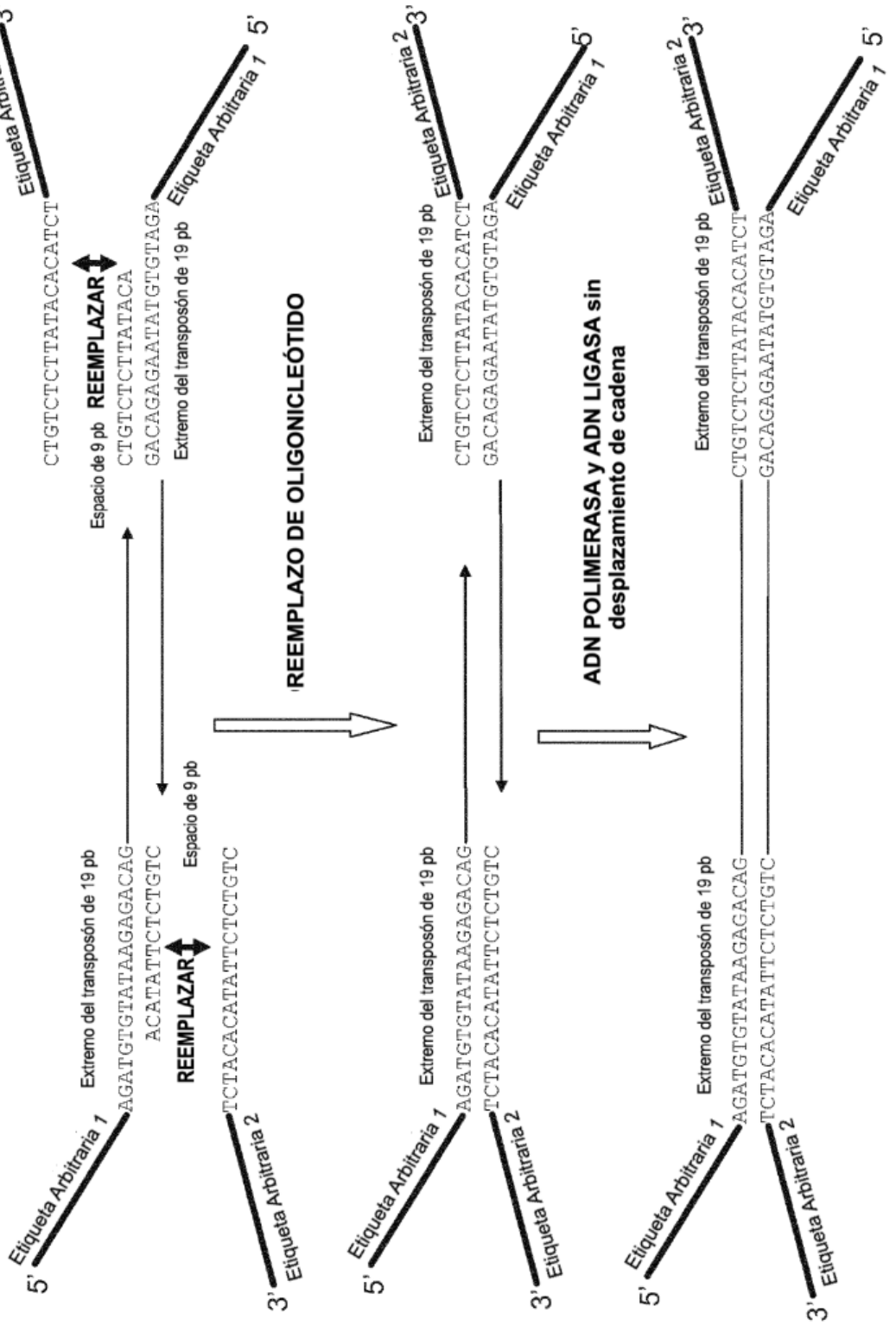
Etiqueta Arbitraria 2
3'
CTGTCTCTTATACACATCT
Espacio de 9 pb
REEMPLAZAR
Extremo del transposón de 19 pb
AGATGTGTATAAGAGACAG
ACATATTCTCTGTC
REEMPLAZAR
Espacio de 9 pb
TCTACACATATTCTCTGTC
5'
Etiqueta Arbitraria 1
3'
Etiqueta Arbitraria 2
CTGTCTCTTATACA
GACAGAGAATATGTGTAGA
Extremo del transposón de 19 pb
Etiqueta Arbitraria 1
5'
REEMPLAZO DE OLIGONICLEÓTIDO
Extremo del transposón de 19 pb
Etiqueta Arbitraria 2
3'
CTGTCTCTTATACACATCT
GACAGAGAATATGTGTAGA
Etiqueta Arbitraria 1
5'
Extremo del transposón de 19 pb
Etiqueta Arbitraria 1
5'
AGATGTGTATAAGAGACAG
TCTACACATATTCTCTGTC
Etiqueta Arbitraria 2
3'
ADN POLIMERASA y ADN LIGASA sin desplazamiento de cadena
Extremo del transposón de 19 pb
Etiqueta Arbitraria 2
3'
CTGTCTCTTATACACATCT
GACAGAGAATATGTGTAGA
Etiqueta Arbitraria 1
5'
Extremo del transposón de 19 pb
Etiqueta Arbitraria 1
5'
AGATGTGTATAAGAGACAG
TCTACACATATTCTCTGTC
Etiqueta Arbitraria 2
3'

Fig. 4

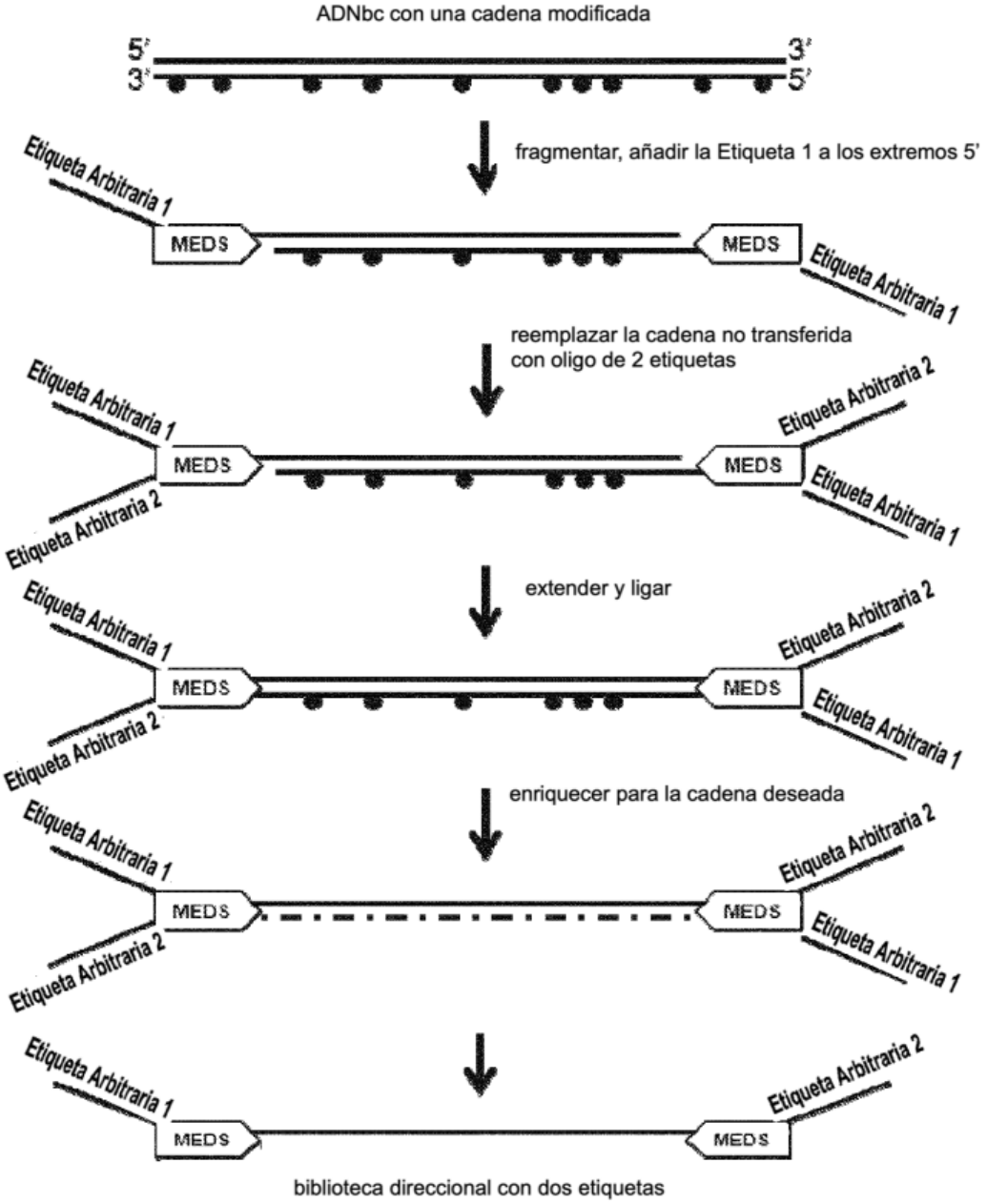
ADNbc con una cadena modificada
5'
3'
3'
5'
fragmentar, añadir la Etiqueta 1 a los extremos 5'
Etiqueta Arbitraria 1
MEDS
MEDS
Etiqueta Arbitraria 1
reemplazar la cadena no transferida
con oligo de 2 etiquetas
Etiqueta Arbitraria 2
Etiqueta Arbitraria 2
extender y ligar
enriquecer para la cadena deseada
biblioteca direccional con dos etiquetas

Fig. 5