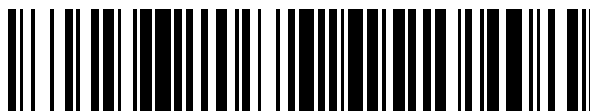


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 568 910**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.01.2012 E 12707675 (0)**

97 Fecha y número de publicación de la concesión europea: **06.04.2016 EP 2668273**

54 Título: **Reemplazo de oligonucleótidos para bibliotecas etiquetadas en dos extremos y direccionadas**

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:

05.05.2016

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:

LAZCANO GAINZA, Jesús

ES 2 568 910 T3

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

DESCRIPCIÓN

Reemplazo de oligonucleótidos para bibliotecas etiquetadas en dos extremos y direccionadas

Solicitudes relacionadas

5 Esta solicitud reivindica el beneficio de prioridad de las solicitudes provisionales estadounidenses Ser. 61/437,451, presentada el 28 de junio de 2011, y Ser. 61/506,777, presentada el 12 de julio de 2011, ambas intituladas "Fragmentación y tagmentación en dos extremos de ADN usando un transposón con reemplazamiento de oligonucleótido, llenado de brechas sin desplazamiento y ligazón".

Campo de la invención

10 La presente invención se refiere a biología molecular y más específicamente a 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

15 La preparación de muestras para secuenciación de siguiente generación puede incluir la fragmentación de ADN genómico o ADNc bicatenario (preparado a partir de ARN) en fragmentos más pequeños, seguida por la adición de secuencias de etiqueta o marca funcional ("tags" o "etiquetas") a las hebras de los fragmentos. Cuando una secuencia monocatenaria es etiquetada en ambos extremos puede utilizarse el término "bi-etiquetada" ("di-tagged"). Tales etiquetas incluyen sitios de imprimación o cebado para polimerasas de ADN para reacciones de secuenciación, sitios de restricción y dominios para captura, amplificación, detección, dirección y promotores de transcripción. Los métodos previos para generar bibliotecas de fragmentos de ADN requerían fragmentar el ADN diana mecánicamente usando un baño de ultrasonido, nebulizadores o mediante una nucleasa, y luego uniendo (por ejemplo mediante ligación) los oligonucleótidos que contienen las etiquetas a los extremos de los fragmentos.

20 Un método novedoso para usar transposones para rápidamente lograr estas etapas fue divulgado en US 2010/0120098 por Grunenwald, para generar fragmentos a partir de ADN bicatenario (por ejemplo ADNc genómico, de amplicón, viral, de fago, derivado a partir de ARN, etcétera). Los sistemas de transposones particularmente útiles incluyen el sistema de transposón Tn5 hiperactivo, descrito en U.S. 5,965,443 y U.S. 6,437,109 por Reznikoff, y el sistema de transposón Mu descrito en U.S. 6,593,113 por Tenkanen,

30 Reznikoff describió, en particular, una secuencia de extremos de transposasa de 19 bases (SEQ ID NO:3) la cual se denomina frecuentemente "ME". En algunas modalidades del método de transposón, la reacción de cadena de polimerasa (PCR) se utiliza como un paso subsiguiente para amplificación de ADN. Esto puede dar lugar a inquietudes debido al potencial de PCR para sobre- o sub-representar las cantidades relativas de una secuencia dada, dependiendo de su composición en G+C, especialmente en regiones de contenido extremo de G+C donde los márgenes de error de PCR pueden confundir la anotación y análisis de los datos.

Resumen de la invención

35 La presente invención proporciona un método para adicionar una o más etiquetas al producto bicatenario de una reacción de tagmentación (tagmentation). El método involucra proporcionar un ácido nucleico diana bicatenario y un transposoma que tiene una transposasa con dos secuencias extremas de transposón: una "hebra transferida" y una "hebra no transferida". El transposoma rompe el ácido nucleico diana en fragmentos mientras transfiere de manera covalente la hebra transferida a una primera hebra de un primer fragmento; la hebra no transferida del transposoma permanece hibridada a la hebra transferida. En una modalidad, la hebra no transferida tiene las fórmulas generales

40 BTGTYTCBTN₁₋₉ SEQ ID NO:20
NTGTMTCNTN₀₋₁₀ SEQ ID NO:21

45 donde se usa la nomenclatura IUPAC para posiciones de nucleótido degenerado, y N_{x-y} indica una secuencia que tiene un intervalo de x a y nucleótidos, inclusive. Entre los fragmentos, la hebra no transferida se retira de la hebra transferida y se reemplaza por un oligonucleótido que comprende una secuencia de etiqueta. El oligonucleótido de reemplazo se une luego a la segunda hebra del primer fragmento mediante ligación y opcionalmente mediante un paso de extensión. El resultado del método es un fragmento del ácido nucleico diana que ha sido marcado (etiquetado) con una o más secuencias de etiqueta, el cual puede ser útil para análisis subsiguiente, tal como la secuenciación.

50 La invención también proporciona un método para generar bibliotecas direccionales suministrando el ácido nucleico diana donde una hebra se modifica químicamente. Enriquecer selectivamente una hebra en el producto de la reacción de tagmentación da lugar a un fragmento que ha sido etiquetado de una manera específica para la hebra, es decir una etiqueta en el extremo 5' y una etiqueta en el extremo 3'.

También se proporcionan en la presente secuencias novedosas de extremo de transposasa que tienen las secuencias generales

MRWTGTGHWKAVGARACAV SEQ ID NO:1 y

NSHBGSHSHDDRNGAKACAN SEQ ID NO:2.

5 Estas secuencias de extremo pueden usarse en los métodos de la invención y con transposasas para reacciones de tagmentación relacionadas en general.

Breve descripción de los dibujos

10 La FIG. 1 proporciona una ilustración esquemática de una reacción de tagmentación. Los cuadrados con esquinas muescadas representan transposasas, por ejemplo transposasas Tn5 o Mu. El MEDS se refiere a un ME bicatenario ("extremo mosaico"), ejemplificado por una secuencia de extremo de transposasa Tn5, tal como SEQ ID NO:3 hibridada a la SEQ ID NO:19. Adheridas a los MEDS están las etiquetas arbitrarias, mostradas aquí como barras claras u oscuras. Conjuntamente los cuadrados, MEDS y etiquetas representan un transposoma que puede usarse para fragmentar un ácido nucleico diana. Un producto de fragmentación bicatenario se muestra como barras oscuras paralelas, con secuencias adheridas determinadas, tal como se discute más adelante. 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 etiqueta adicionales, tal como un código de barras. A y B representan secuencias adicionales que pueden adherirse por medio de PCR. Tal como se muestra, los transposomas pueden usarse para fragmentar un ácido nucleico diana para generar fragmentos bicatenarios que tienen secuencias en ambos extremos, el cual puede ser útil para codificación en barra y secueñación.

20 La Fig. 2a proporciona una geometría más detallada del producto de tagmentación de la segunda fila en la figura 1. Tal como se muestra, la reacción de tagmentación da lugar a ADN diana fragmentado donde el extremo 5' de la hebra superior está adherido covalentemente a una "hebra transferida" de 19 bases (SEQ ID NO:3). El extremo 5' de la hebra inferior está similarmente adherido a otra copia de la hebra transferida, mostrada en orientación de 3'–a–5' (SEQ ID NO:3). Sin embargo, la tagmentación (tagmentation) deja una brecha monocatenaria de 9 bases entre el extremo 3' del fragmento diana y el extremo 5' de la otra hebra (no transferida) del transposoma. Debido a que la hebra no se adhiere de manera covalente al extremo 3' del fragmento de ácido nucleico diana, se describe como "extremo no transferido" (SEQ ID NO:19), aunque el extremo no transferido permanece asociado con el fragmento mediante hibridación a la hebra transferida. La figura 2b muestra el mismo producto de tagmentación que en la figura 2a, pero usando la forma esquemática usada en las otras figuras. Para facilidad de ilustración, las etiquetas han sido omitidas de la figura 2a y la figura 2b.

30 La Fig. 3 muestra diversas versiones alternas de las hebras no transferidas. En la parte superior, el extremo transferido de ME de 19 bases (SEQ ID NO:3) con una secuencia de etiqueta adherida se muestra para referencia. Inmediatamente debajo, el extremo no transferido complementario de 19 bases se muestra como SEQ ID NO:19 en orientación 3'–a–5'. Tal como se divulga en la presente, no obstante, un extremo no transferido también puede truncarse mediante supresiones a 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), o 8 bases (SEQ ID NO:8).

40 La Fig. 4 ilustra una modalidad del método de la invención, tal como se discute con mayor detalle más adelante. Los extremos de transposón de 19–bp (hebra transferida) (SEQ ID NO:3) se muestran con etiqueta arbitraria (Arbitrary Tag) 1. Una hebra no transferida ejemplar de 14 bases (SEQ ID NO:14) se muestra hibridada a una porción de la hebra transferida. Tal como se muestra en la fila superior, se proporcionan oligonucleótidos de reemplazo (SEQ ID NO:19) (mostrados aquí adheridos a etiqueta arbitraria (Arbitrary Tag) 2). En las filas del medio y del fondo, los extremos de transposón bicatenarios de 19 bases (SEQ ID NO:3, SEQ ID NO:19) se muestran adheridos al fragmento de ácido nucleico diana. El producto mostrado en la fila inferior es un fragmento "etiquetado en dos extremos" o "bi-etiquetado" ("di-tagged") que tiene una etiqueta arbitraria 1 en el extremo 5' y una etiqueta arbitraria 2 en el extremo 3'.

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

Descripción detallada

55 La presente invención proporciona un método mejorado para preparar ADN bicatenario etiquetado en dos extremos. El método involucra (a) proporcionar un ácido nucleico diana bicatenario y un transposoma que tiene una transposasa con dos secuencias de extremo de transposón: una "hebra transferida" y una "hebra no transferida"; (b) permitir que el transposoma fragmente el ácido nucleico diana por lo cual la hebra transferida se transfiere de modo covalente a una

primera hebra de un primer fragmento y la hebra no transferida se mantiene hibridada a la hebra transferida; (c) retirar la hebra no transferida de la hebra transferida; (d) proporcionar un oligonucleótido de reemplazo que comprende una secuencia de etiqueta (tag), para hibridar a la hebra transferida; y (e) ligar el oligonucleótido de reemplazo a la segunda hebra del primer fragmento. De esta manera, el método genera un producto de tagmentación (tagmentation) que tiene una hebra transferida y un oligonucleótido de reemplazo

El ADN diana utilizado en el método puede ser cualquier ácido nucleico de interés. Los ácidos nucleicos diana pueden incluir ADN, ácido nucleico de péptido, ácido morfolino nucleico, ácido nucleico bloqueado, ácido nucleico de glicol, ácido nucleico de treosa, mezclas de los mismos e híbridos de los mismos. En una modalidad preferida, los fragmentos de ADN genómicos o las copias amplificadas de los mismos se usan como ácido nucleico diana. En otra modalidad preferida se utiliza ADN de mitocondrias o de cloroplasto.

Un ácido nucleico diana puede comprender cualquier secuencia de nucleótidos. En algunas modalidades, el ácido nucleico diana comprende secuencias de homopolímero. Un ácido nucleico diana también puede incluir secuencias de repetición. Las secuencias de repetición pueden ser 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 de repetición pueden repetirse, ya sea de manera contigua o no contigua, cualquier cantidad de veces que incluyen, por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 veces o más.

Algunas modalidades descritas en la presente pueden utilizar un único ácido nucleico diana. Otras modalidades pueden utilizar una pluralidad de ácidos nucleicos diana. En tales modalidades, una pluralidad de ácidos nucleicos diana puede incluir una pluralidad de los mismos ácidos nucleicos diana, una pluralidad de diferentes ácidos nucleicos diana donde algunos ácidos nucleicos diana son los mismos, o una pluralidad de ácidos nucleicos diana donde todos los ácidos nucleicos diana son diferentes. Las modalidades que utilizan una pluralidad de ácidos nucleicos diana pueden llevarse a cabo en formatos múltiples de modo que los reactivos se entreguen simultáneamente a los ácidos nucleicos diana, por ejemplo, en una o más cámaras o en una superficie de matriz. En algunas modalidades, la pluralidad de ácidos nucleicos diana puede incluir sustancialmente todo el genoma de un organismo particular. La pluralidad de ácidos nucleicos diana 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 modalidades particulares la porción puede tener un límite superior que es a lo sumo de aproximadamente 1%, 5%, 10%, 25%, 50%, 75%, 80%, 85%, 90%, 95%, o 99% del genoma.

Los ácidos nucleicos diana pueden obtenerse a partir de cualquier fuente. Por ejemplo, los ácidos nucleicos diana 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, pero no se limitan a, orgánulos, células, tejidos, órganos u organismos. Las células que pueden usarse como fuentes de moléculas de ácido nucleico diana pueden ser procariotas (células bacterianas, por ejemplo *Escherichia*, *Bacillus*, *Serratia*, *Salmonella*, *Staphylococcus*, *Streptococcus*, *Clostridium*, *Chlamydia*, *Neisseria*, *Treponema*, *Mycoplasma*, *Borrelia*, *Legionella*, *Pseudomonas*, *Mycobacterium*, *Helicobacter*, *Erwinia*, *Agrobacterium*, *Rhizobium*, y *Streptomyces* genera); arqueas, tales como crenarqueota, nanoarqueota o eurarqueota; o eucariotas tales como los hongos (por ejemplo, levaduras), plantas, protozoarios y otros parásitos, y animales (que incluyen insectos (por ejemplo, *Drosophila* spp.), nemátodos (por ejemplo, *Caenorhabditis elegans*), y mamíferos (por ejemplo, rata, ratón, mono, primate no humano y humano).

En algunas modalidades, puede proporcionarse el ácido nucleico diana cuando una hebra se modifica químicamente tal como con un sitio de fragmentación. Un sitio de fragmentación puede usarse para disociar la asociación física pero no la de información entre una primera secuencia de código de barras y una segunda secuencia de código de barras. La disociación debe ser por medios bioquímicos, químicos u otros. En algunas modalidades, un sitio de fragmentación puede incluir un nucleótido o secuencia de nucleótidos que pueden fragmentarse por diversos medios. Por ejemplo, un sitio de fragmentación puede ser un sustrato para una enzima, tal como una nucleasa, que disociará 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 pueden disociarse 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 otra manera puede comprender desoxirribonucleótidos y pueden disociarse con una ARNasa. Los agentes de disociación química capaces de disociar selectivamente el enlace de fosfodiéster entre un desoxirribonucleótido y un ribonucleótido incluyen iones metálicos, por ejemplo iones de metal 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)), $\text{Fe}(3)$ o $\text{Cu}(3)$, o exposición a pH elevado, por ejemplo, tratamiento con una base tal como un hidróxido de sodio. Tal como se usa en la presente, la disociación selectiva del enlace de fosfodiéster entre un desoxirribonucleótido y un ribonucleótido puede referirse al agente de disociación química que no es capaz de disociar el enlace de 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 enzima nickasa (que causa roturas (nicks)); es decir, una endonucleasa que rompe una hebra 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.

5 En otro ejemplo, un sitio de fragmentación puede incluir uno o más análogos de nucleótido que comprenden un sitio abásico y permite la disociación en el sitio de fragmentación en presencia de determinados agentes químicos tales como poliamina, N,N'- dimetiletileno-diamina (DMED) (publicación de patente estadounidense No. 2010/0022403). En una modalidad, la modificación química puede ser una conversión de citocinas a uracilos. En algunas modalidades, un sitio abásico puede crearse dentro de un ciclo de fragmentación proporcionando primero un sitio de fragmentación que comprende una desoxiuridina (U) de un ácido nucleico bicatenario. La enzima uracilo-ADN glicosilasa (UDG) puede usarse luego para retirar la base de uracilo generando un sitio abásico sobre una hebra. La hebra de polinucleótido que incluye el sitio abásico puede disociarse luego en el sitio abásico mediante tratamiento con endonucleasa (por ejemplo Endo IV endonucleasa, AP liasa, FPG glicosilasa/AP liasa, Endo VIII glicosilasa/AP liasa), calor o álcali. Los sitios abásicos también pueden generarse en análogos de nucleótido distintos de la desoxiuridina y disociarse de una manera análoga mediante tratamiento con endonucleasa, calor o álcali. For example, 8-oxo-guanina puede convertirse en un sitio abásico mediante exposición a FPG glicosilasa. Desoxiinosina puede convertirse en un sitio abásico por exposición a alca glicosilasa. Los sitios abásicos generados de esta manera pueden oxidarse luego, de manera típica mediante tratamiento con una endonucleasa adecuada (por ejemplo Endo IV, AP liasa). (Publicación de patente estadounidense No. 2011/0014657).

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

15 En algunas modalidades, un sitio de fragmentación puede incluir un residuo disociable que puede someterse a disociación fotoquímica. La disociación fotoquímica abarca cualquier método que utiliza energía de la luz para lograr disociación de ácidos nucleicos, por ejemplo una o ambas hebras de una molécula de ácido nucleico bicatenario. Un sitio para disociación fotoquímica puede proporcionarse por un residuo químico no nucleótido en un ácido nucleico, tal como fósforoamidita [4-(4,4'-dimetoxitritiloxi)butiramidometil]-1-(2-nitrofenil)-etil]-2-cianoetil-(N,N-diisopropil)-fosforamidita (Glen Research, Sterling, Va., EE.UU., Cat No. 10-4913-XX).

20 En algunas modalidades, un sitio de fragmentación puede incluir un péptido, por ejemplo, una estructura conjugada en la cual una molécula de péptido se enlaza a un ácido nucleico. La molécula de péptido puede disociarse a continuación por una enzima peptidasa de especificidad apropiada o por cualquier otro medio adecuado de disociación química o fotoquímica no enzimática. En algunas modalidades se formará un conjugado entre el péptido y ácido nucleico mediante enlace covalente de un péptido con un ácido nucleico, por ejemplo, una hebra de un ácido nucleico bicatenario. Los conjugados entre un péptido y un ácido nucleico pueden prepararse usando técnicas conocidas generalmente en el estado de arte. En una técnica de estas, los componentes de péptido y ácido nucleico de la secuencia deseada de aminoácidos y nucleótidos pueden sintetizarse por separado, por ejemplo mediante técnicas de síntesis química, automatizadas, estándar, y luego conjugarse en una solución acuosa/orgánica. Por ejemplo, el sistema OPeC™ disponible comercialmente en Glen Research se basa en la ligación nativa de un péptido funcionalizado con tioéster, N-terminal, a un oligonucleótido de 5'-cisteinilo.

25 El método de la invención puede usar cualquier transposasa que pueda aceptar una secuencia de extremo de transposasa y fragmentar un ácido nucleico diana, adherir un extremo transferido pero no un extremo no transferido. Un "transposoma" está comprendido de 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 transposón que es capaz de catalizar una reacción de transposición. La transposasa o integrasa pueden enlazarse al sitio de reconocimiento de transposasa e insertar el sitio de reconocimiento de transposasa a un ácido nucleico diana en un proceso denominado a veces "tagmentación" (tagmentation). En algunos de tales eventos de inserción, una hebra del sitio de reconocimiento de transposasa puede transformarse en el ácido nucleico diana.

30 Algunas modalidades 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 de reconocimiento de transposasa Mu que comprende secuencias de extremo R1 y R2 (Mizuuchi, K., Cell, 35: 785, 1983; Savilahti, H, et al., EMBO J., 14: 4893, 1995). Un sitio de reconocimiento de transposasa ejemplar que forma un complejo con una transposasa Tn5 hiperactiva (por ejemplo, EZ-Tn5™ Transposase, Epicentre Biotechnologies, Madison, Wisconsin).

35 Más ejemplos de sistemas de transposición que pueden usarse con determinadas modalidades proporcionadas en la presente incluyen *Staphylococcus aureus* Tn552 (Colegio et al., J. Bacteriol., 183: 2384-8, 2001; Kirby C et al., 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, N L, Science. 271: 1512, 1996; Craig, N L, Review en: Curr Top Microbiol Immunol., 204:27-48, 1996), Tn/O y IS10 (Kleckner N, et al., Curr Top Microbiol Immunol., 204:49-82, 1996), Transposasa marinera (Lampe D J, et al., EMBO J., 15: 5470-9, 1996), Tcl (Plasterk R H, Curr. Topics Microbiol.

Immunol., 204: 125–43, 1996), P Element (Gloor, G B, Methods Mol. Biol., 260: 97–114, 2004), Tn3 (Ichikawa & Ohtsubo, J Biol. Chem. 265:18829–32, 1990), secuencias de inserción bacterianas (Ohtsubo & 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 & Corces, Annu Rev Microbiol. 43:403–34, 1989). Más ejemplos incluyen IS5, Tn10, Tn903, IS911, y versiones diseñadas de enzimas de la familia transposasa (Zhang et al., (2009) PLoS Genet. 5:e1000689. Epub 2009 Oct 16; Wilson C. et al (2007) J. Microbiol. Methods 71:332–5).

El sistema de Tn5 usa transposasas Tn5 con la secuencia ME de 19 bases (SEQ ID NO:3) como el extremo transferido. Tal como se discute en el ejemplo 2, sin embargo, pueden usarse otros extremos transferidos y pueden describirse por las fórmulas genéricas

10 MRWTGTGHWKAVGARACAV SEQ ID NO:1 y
NSHBGSHSHDDRNGAKACAN SEQ ID NO:2.

Más particularmente los extremos transferidos pueden ser

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

Con base en la secuencias truncadas discutidas en el ejemplo 1 y la secuencias complementarias a SEQ ID NO:1 y SEQ ID NO:2, la presente invención también hace uso de hebras no transferidas que tienen la fórmula general:

25 BTGTYTCBTN_{1–10} SEQ ID NO:20
NTGMTMCNTN_{0–10} SEQ ID NO:21

donde el N_{0–10} indica cero a 10 nucleótidos. Como ejemplos, la hebra no transferida puede seleccionarse del grupo consistente en SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, y SEQ ID NO:18.

30 Las secuencias de extremo pueden comprender además una secuencia de etiqueta (tag) que pueden adicionarse de modo covalente a los fragmentos en el proceso del método de tagmentación (tagmentation). Tal como se usa en la presente, el término "etiqueta" significa una secuencia de nucleótidos que está adherida a otro ácido nucleico para proporcionarle al ácido nucleico alguna funcionalidad. Ejemplos de etiquetas incluyen código de barras, sitios de cebadores, etiquetas de afinidad residuos reporteros.

35 Por lo 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 de ocurrencia natural, tal como un g-código descrito en la presente. Un código de barras debe 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 modalidades, un código de barras comprende al menos aproximadamente 10, 20,
40 30, 40, 50, 60, 70 80, 90, 100 o más nucleótidos consecutivos. En algunas modalidades, al menos una porción de los códigos de barras en una población de ácidos nucleicos que comprende el código de barras es diferente. En algunas modalidades, al menos aproximadamente 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% de los códigos de barras son diferentes. En más modalidades de estas, todos los códigos de barras son diferentes. La diversidad de los diferentes códigos de barras en una población de ácidos nucleicos que comprenden código de barras puede ser
45 aleatoriamente generada o generada de modo no aleatorio.

En algunas modalidades, una secuencia de transposón comprende al menos un código de barras. En algunas modalidades, 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 estas modalidades, la primera secuencia de código de barras puede identificarse o designarse para aparejarse con la segunda secuencia de código de barras.
50 Por ejemplo, una primera secuencia de código de barras conocida puede conocerse para emparejarse con una segunda secuencia de código de barras conocida utilizando una tabla de referencia que comprende una pluralidad de una primera y una segunda secuencias de código de barra conocidas por emparejarse 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 modalidades, la primera secuencia de
55

código de barras y la segunda secuencia de código de barras son diferentes ("bicódigos"). Se entenderá que en algunas modalidades el vasto número de códigos de barras disponibles permite que cada plantilla de molécula de ácido nucleico comprenda una identificación única. La identificación única de cada molécula en una mezcla de ácidos nucleicos de plantilla puede usarse en varias aplicaciones para identificar moléculas individuales de ácido nucleico, en muestras que tienen múltiples cromosomas, genoma, células, tipos de células, estados mórbidos de células y especies, por ejemplo en secuenciación de haplotipo, discriminación de alelo parental, secuenciación metagenómica y secuenciación de muestra de un genoma.

En algunas modalidades, la etiqueta útil es un sitio cebador que puede hibridar a un cebador. La orientación de los sitios de cebador en tales modalidades puede ser tal que un cebador que hibrida al primer sitio de cebador y un cebador que hibrida al segundo sitio de cebador se encuentra en la misma orientación o en diferentes orientaciones. En una modalidad, la secuencia de cebador puede ser complementaria a un cebador usado para amplificación. En otra modalidad, la secuencia de cebador es complementaria a un cebador usado para secuenciación.

En algunas modalidades, una etiqueta puede incluir un primer sitio de cebador, un segundo sitio de cebador que tiene un sitio no amplificable dispuesto entre los mismos. El sitio no amplificable es útil para bloquear la extensión de una hebra de polinucleótido entre el primero y segundo sitio de cebador, en cuyo caso la hebra de polinucleótido hibrida a uno de los sitios de cebador. El sitio no amplificable también puede ser útil para prevenir concatémeros. Ejemplos de sitios no amplificables incluyen un análogo de nucleótido, un residuo químico no nucleótido, un aminoácido, un péptido y un polipéptido. En algunas modalidades, un sitio no amplificable comprende un análogo de nucleótido que no se empareja de modo significativo como base con A, C, G o T.

En algunas modalidades, una etiqueta puede ser una etiqueta de afinidad. Las etiquetas de afinidad pueden ser útiles para la separación en masa de los ácidos nucleicos diana hibridados a etiquetas de hibridación. Tal como se usa en la presente, el término "etiqueta de afinidad" y los equivalentes gramáticos pueden referirse a un componente de un complejo multicomponente, en cuyo caso los componentes del complejo multicomponente interactúan de modo específico entre sí o se enlazan uno a otro. Por ejemplo, una etiqueta de afinidad puede incluir biotina o His que pueden enlazar estreptavidina o níquel, respectivamente. Otros ejemplos de complejos de etiqueta de afinidad multicomponente incluyen ligandos y sus receptores, por ejemplo avidina-biotina, estreptavidina-biotina y derivados de biotina, estreptavidina o avidina incluyendo pero sin limitar a 2-iminobiotina, destiobiotina, NeutrAvidin (Molecular Probes, Eugene, Oregon), CaptAvidin (Molecular Probes), y similares; proteínas/péptidos de enlazamiento que incluyen proteína de enlazamiento de maltosa-maltosa (MBP), proteína/péptido de enlazamiento de calcio-calcio (CBP); antígeno-anticuerpo que incluye etiquetas de epítopes y sus correspondientes anticuerpos anti-epítopes; haptenos, por ejemplo dinitrofenilo y digoxigenina, y sus correspondientes anticuerpos; aptámeros y sus dianas correspondientes; etiquetas poli-His (por ejemplo, penta-His y hexa-His) y sus contrapartes de enlazamiento que incluyen los correspondientes materiales de cromatografía de afinidad de ión metálico (IMAC) y anticuerpos anti-poli-His; anticuerpos fluoróforos y anti-fluoróforos; y similares. En algunas modalidades, una etiqueta puede comprender un residuo reportero. Tal como se usa en la presente, el término "residuo reportero" y los equivalentes gramaticales pueden referirse a cualquier marca, etiqueta o grupo identificable. El experto en la materia apreciará que muchas especies diferentes de residuos reporteros pueden usarse con los métodos y composiciones descritos en la presente, ya sea individualmente o en combinación con uno o más residuos reporteros diferentes. En determinadas modalidades, un residuo reportero puede emitir una señal. Ejemplos de señales son señales fluorescentes, quimioluminiscentes, bioluminiscentes, fosforescentes, radioactivas, calorimétricas o electroquimioluminiscentes. Ejemplos de residuos reporteros incluyen fluoróforos, radioisótopos, cromógenos, enzimas, antígenos que incluyen etiquetas de epítope, nanocristales semiconductores tales como puntos cuánticos, metales pesados, tintes, grupos de fosforescencia, grupos de quimioluminiscencia, residuos de detección electroquímica, proteínas de enlazamiento, fósforo, quelatos de tierras raras, quelatos de metal de transición, tintes de infrarrojo cercano, etiquetas de electroquimioluminiscencia y residuos reporteros compatibles con espectrómetro de masas, tales como etiquetas de masa, etiquetas de carga e isótopos. Más residuos reporteros que pueden usarse con los métodos y composiciones descritos en la presente incluyen etiquetas espectrales tales como tintes fluorescentes (por ejemplo, isotiosocianato de fluoresceína, rojo de Texas, rodamina, y similares), radioetiquetas (por ejemplo, ³H, ¹²⁵I, ³⁵S, ¹⁴C, ³²P, ³³P, etc.), enzimas (por ejemplo, peroxidasa de rábano picante, fosfatasa alcalina, etc.), etiquetas calorimétricas espectrales tales como esferas de oro coloidal, o de vidrio o de plástico coloreadas (por ejemplo, poliestireno, polipropileno, látex, etc.); etiquetas magnéticas, eléctricas, térmicas; y etiquetas de masa. Los residuos reporteros también pueden incluir enzimas (peroxidasa de rábano picante, etc.) y partículas magnéticas. Más residuos reporteros incluyen cromóforos, fósforo y residuos fluorescentes, por ejemplo rojo de Texas, dixogenina, biotina, 1- y 2-aminonaftalina, p,p'-diaminoestilbenos, pirenos, sales cuaternarias de fenantridina, 9-aminoacridinas, p,p'-diaminobenzofenona-íminas, antracenos, oxacarbocianina, merocianina, 3-aminoequilenina, perileno, bis-benzoxazol, bis-p-oxazolil benceno, 1,2-benzofenazina, retinol, sales de bis-3-aminopiridinio, helebrigenina, tetraciclina, esterofenol, bencimidazolilfenilamina, 2-oxo-3-cromeno, indol, xanteno, 7-hidroxicumarina, fenoxazina, salicilato, estrofantidina, porfirinas, triarilmetanos y flavina. Compuestos fluorescentes individuales que tienen funcionalidades para enlazarse a un elemento detectado deseablemente en un aparato o ensayo, suministrados en la presente, o que pueden modificarse para incorporar tales funcionalidades incluyen, por ejemplo, cloruro de dansilo; fluoresceínas tales como 3,6-dihidroxi-9-fenilxantidrol; rodaminaisotiocianato; 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-aminoaftaleno-6-sulfonato; bromuro de etidio; estebrina; auromina-

0,2-(9'-antroil)palmitato; dansil-fosfatidiletanolamina; N,N'-dioctadecil oxacarbocianina; N,N'-dihexil oxacarbocianina; merocianina, 4-(3'-pirenil)estearato; d-3-aminodesoxi-equilenina; 12-(9'-antroil)estearato; 2-metilantraceno; 9-vinilantraceno; 2,2'(vinileno-p-fenileno) bisbenzoxazol; p-bis(2- metil-5-fenil-oxazolil)benceno; 6-dimetilamino-1,2-benzofenazina; retinol; bis(3'-aminopiridinio), 1,10-decil-diioduro; sulfonaftilhidrazona de helibrienina; clorotetraciclina; N-(7-dimetilamino-4-metil-2-oxo-3-cromenil)maleimida; N-(p-(2benzimidazolil)-fenil)maleimida; N-(4-fluorantil)maleimida; ácido bis(homovanílico); resazarina; 4-cloro7-nitro-2,1,3-benzooxadiazol; merocianina 540; resorufina; rosa bengal; 2,4-difenil-3(2H)-furanona, complejos fluorescentes de lantánidos que incluyen aquellos de europio y terbio, fluoresceína, rodamina, tetrametilrodamina, eosina, eritrosina, cumarina, metil-cumarinas, puntos cuánticos (también denominados "nanocristales": véase patente estadounidense No. 6,544,732), pireno, verde malaquita, estilbena, amarillo Lucifer Yellow, azul Cascade Blue™, rojo de Texas, tintes de Cy (Cy3, Cy5, etc.), tintes Alexa Floor®, ficoeritina, bodipy (boro-dipirrometeno), y otros descritos en la sexta edición de Molecular Probes Handbook (Manual de sondas moleculares) de Richard P. Haugly. La invención proporciona además un ácido nucleico que comprende una o dos copias de las secuencias de extremo de transposasa que pueden generarse realizando el método de la invención. Cuando se realice el método sobre una secuencia diana y se generan diferentes fragmentos, la invención proporciona una biblioteca de estos ácidos nucleicos diferentes.

La hebra no transferida con o sin grupos de protección de nucleasa y/o terminación de cadena (por ejemplo fosforotioato y/or dideoxi) se disocia luego de la hebra transferida y un oligonucleótido de reemplazo (que puede contener etiquetas de ADN adicionales, tal como se ha discutido antes, como una etiqueta de secuenciación) se empareja con la secuencia de hebra transferida complementariamente con o sin grupos protectores de nucleasas (por ejemplo, fosforotioatos). Pueden utilizarse enzimas modificadoras de ácido nucleico no desplazantes que consisten en una polimerasa de ADN y una ligasa de ADN. Las polimerasas y ligasas de ADN se utilizan para llenar y ligar la brecha entre ADN mono-etiquetada y el oligonucleótido de reemplazo, lo que da lugar a un pedazo de ADN bicatenario con una etiqueta 5' y una etiqueta 3' enlazadas de modo covalente. Por lo tanto, el método de la presente invención proporciona una novedosa manera de crear fragmentos de ADN bicatenario bi-etiquetadas, en cuyo caso la reacción de cadena de polimerasa es opcional.

La presente invención también proporciona un método mejorado para preparar una biblioteca etiquetada direccionalmente. El método comienza suministrándole al ADN una modificación específica de hebra, tal como incorporación de un nucleótido disociable sobre una hebra particular, tal como un nucleótido químicamente lábil o uno que contiene uracilo u 8-oxoguanina. Otros nucleótidos modificados útiles incluyen 8-oxoadenina, fapy-guanina, metil-fapy-guanina, fapy-adenina, aflatoxina B1-fapy-guanina, 5-hidroxi-citosina, 5-hidroxi-uracilo, y productos de adición de N-7 de anillo abierto (7-metilguanina). En una modalidad particular cada hebra puede contener una modificación diferente, por ejemplo una hebra puede contener modificaciones de uracilo y la otra hebra puede contener modificaciones de 8-oxo-guanina.

El ADN se disocia luego con un transposoma mono-etiquetado que consiste en una o más moléculas de transposasa y dos secuencias de oligonucleótido de ADN que son el ADN del extremo modificado (ME) emparejado. Una secuencia de (ME) en la hebra de ADN transferida y en la hebra no transferida de ADN puede contener una secuencia de ME de 19bp o una secuencia de ADN troncada. La hebra no transferida (con o sin grupos protectores de nucleasas y/o de terminación de cadena, por ejemplo fosforotioato y/o dideoxi, se disocian luego de la hebra transferida y un oligonucleótido de reemplazo (que puede contener secuencia de ADN adicional tal como una etiqueta de secuenciación) se empareja a la secuencia de hebra transferida complementariamente con o sin grupos protectores de nucleasa (por ejemplo fosforotioatos). Se usan enzimas modificadoras de ácido nucleico no desplazante que consisten en una polimerasa de ADN (por ejemplo, polimerasas termoestables, o polimerasas no termoestables tales como una polimerasa I de ADN o fragmento exo⁻ de Klenow) y una ligasas de ADN. Las polimerasas y las ligasas de ADN se utilizan para completar y ligar la brecha entre el ADN mono-etiquetado y el oligonucleótido de reemplazo, lo que da lugar a un pedazo de ADN bicatenario con una etiqueta 5' y una 3' adheridas de modo covalente. Como alternativa, pueden suministrarse un oligonucleótido para llenar la brecha, seguida de ligazón.

La hebra modificada o no modificada puede tratarse específicamente para enriquecer o suprimir su funcionalidad. El tratamiento puede incluir usar una enzima, tal como una glicosilasa de ADN de uracilo (UDG) también denominada uracilo N-Glicosilasa (UNG), endonucleasa apurínica/apirimidínica humana (APE I), formamidopirimidina-ADN glicosilasa (FPG) también denominada 8-oxiguanina ADN glicosilasa, endonucleasa IV y cinasa, endo III, endo VIII, hOGG1, T7 Endo I, T4 PDG y afu UDG. En otra modalidad, una hebra puede enriquecerse de modo selectivo por extensión usando una polimerasa que tiene una preferencia por nucleótidos de ocurrencia natural antes que por nucleótidos modificados químicamente. Un ejemplo de una polimerasa de este tipo es una fusión de una polimerasa del tipo Pirococcus con un dominio de enlazamiento de ADN bicatenario de *Sulfolobus solfataricus* (SSo7d). De esta manera, el método de la presente invención proporciona una manera novedosa para crear fragmentos de ADN bi-etiquetados a partir de una hebra individual predeterminada, donde no es necesaria una amplificación subsiguiente.

La invención también hace uso de un transposoma que comprende una transposasa y un ácido nucleico que contiene una o más secuencias de extremo. La invención se refiere además a un método para hacer un transposoma con las secuencias de extremo suministrando una transposasa y suministrando las secuencias de extremo transferidas y luego permitiendo que la transposasa se enlace a la secuencia de extremo transferida. Las condiciones de reacción

ejemplares se discuten en el ejemplo 1 más adelante. De esta manera, la invención proporciona un método de tagmentación que comprende los pasos de (a) proporcionar ácidos nucleicos diana, (b) proporcionar transposomas tal como se han divulgado en la presente, y (c) permitir que los transposomas fragmenten los ácidos nucleicos diana y etiqueten al menos secuencias de extremo de transposasa a los extremos de los fragmentos.

5 Definiciones

El término "comprende", tal como se usa en la presente, es sinónimo de "incluye", "contiene" o "caracterizado por" y es incluyente o de extremos abiertos y no excluye elementos o pasos del método adicionales o no declarados.

10 Tal como se usa en la presente, el término "al menos una porción" y/o los 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.

15 Tal como se usa en la presente, el término "ácido nucleico" y/o "oligonucleótido" y/o los equivalentes gramaticales del mismo pueden referirse a al menos dos monómeros de nucleótido enlazados conjuntamente. Un ácido nucleico puede generalmente contener enlaces de fósforodiéster; sin embargo, en algunas modalidades los análogos de ácido nucleico pueden tener otros tipos de ejes troncales 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 (1996)), fosfortioato (Mag, et al., *Nucleic Acids Res.*, 19:1437 (1991); y patente estadounidense no. 5,644,048), fosforoditioato (Briu, et al., *J. Am. Chem. Soc.*, 111:2321 (1989), enlaces de O-metilfosforoamidita (véase Eckstein, *Oligonucleotides y Analogues: A Practical Approach* (Oligonucleótidos y análogos: un enfoque práctico), Oxford University Press), y ejes troncales y enlaces de péptido ácido nucleico (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 (1996)).

25 Otros ácidos nucleicos análogos incluyen aquellos con ejes troncales positivos (Denpci, et al., *Proc. Natl. Acad. Sci. USA*, 92:6097 (1995)); ejes troncales no iónicos (patentes estadounidenses nos. 5,386,023; 5,637,684; 5,602,240; 5,216,141; y 4,469,863; Kiedrowski, et al., *Angew. Chem. Intl. Ed. English*, 30:423 (1991); Letsinger, et al., *J. Am. Chem. Soc.*, 110:4470 (1988); Letsinger, et al., *Nucleosides & Nucleotides*, 13:1597 (1994); Chapters 2 y 3, *ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research"*, Ed. Y. S. Sanghui y P. Dan Cook; Mesmaeker, et al., *Bioorganic & Medicinal Chem. Lett.*, 4:395 (1994); Jeffs, et al., *J. Biomolecular NMR*, 34:17 (1994); *Tetrahedron Lett.*, 37:743 (1996)) y no-ribosa (patentes estadounidenses No. 5,235,033 y No. 5,034,506, y capítulos 6 y 7, *ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research"*, Ed. Y. S. Sanghui y P. Dan Cook). Los ácidos nucleicos también pueden contener uno o más azúcares carbocíclicos (véase Jenkins, et al., *Chem. Soc. Rev.*, (1995) pp. 169-176).

35 Pueden hacerse modificaciones del eje troncal de ribosa-fosfato para facilitar la adición de residuos adicionales tales como etiquetas o para incrementar la estabilidad de tales moléculas en ciertas condiciones. Además, pueden hacerse mezclas de ácidos nucleicos y análogos de ocurrencia natural. De modo alternativo, pueden hacerse mezclas de diferentes análogos de ácido nucleico y mezclas de ácidos nucleicos y análogos de ocurrencia natural. Los ácidos nucleicos pueden ser monocatenarios o bicatenarios, según se especifique, o contener porciones de secuencia tanto bicatenaria como monocatenaria. El ácido nucleico puede ser ADN, por ejemplo genómico o ADNc, ARN o un híbrido, a partir de células individuales, células múltiples o de múltiples especies, como con muestras metagenómicas, tales como de muestras ambientales. Un ácido nucleico puede contener cualquier combinación de desoxirribosa- y ribosa-nucleótidos, y cualquier combinación de bases que incluyen uracilo, adenina, timina, citosina, guanina, inosina, xantantina, hipoxantina, isocitosina, isoguanina, y análogos de base tales como nitropirrol (incluyendo 3-nitropirrol) y nitroindol (incluyendo 5-nitroindol), etc.

45 En algunas modalidades, un ácido nucleico puede incluir al menos una base promiscua. Las bases promiscuas pueden emparejar bases con más de un tipo diferente de base. En algunas modalidades, una base promiscua puede emparejar bases 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 emparejarse con adenina, timina o citosina. Otros ejemplos incluyen hipoxantina, 5-nitroindol, 5-nitroindol acíclico, 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 pueden usarse bases promiscuas que pueden emparejarse con al menos tres, cuatro o más tipos de bases.

55 Tal como se usa en la presente, el término "análogo de nucleótido" y/o los equivalentes gramaticales del mismo pueden referirse a análogos sintéticos que tienen porciones de base de nucleótido modificadas, porciones de pentosa modificadas y/o porciones de fosfato modificadas y, en caso de polinucleótidos, enlaces internucleótidos modificados, tal como se describe en términos generales en otras fuentes (por ejemplo, Scheit, *Nucleotide Analogs*, John Wiley, New York, 1980; Englisch, *Angew. Chem. Intl. Ed. Engl.* 30:613-29, 1991; Agarwal, *Protocols for Polynucleotides y Analogs*,

Humana Press, 1994; y S. Verma y F. Eckstein, *Ann. Rev. Biochem.* 67:99–134, 1998). En términos generales, las porciones de fosfato modificadas comprenden análogos de fosfato en las que el átomo de fósforo se encuentra en el estado de oxidación +5 y uno o más de los átomos de oxígeno se reemplaza con un residuo no oxígeno, por ejemplo azufre. Análogos de fosfato ejemplares incluyen, pero no se limitan a fosforotioato, fosforoditioato, fosforoselenoato, fosforodiselenoato, fosforoanilitioato, fosforanilidato, fosforamidato, fosfatos de boro, incluyendo contra-iones asociados, por ejemplo, H⁺, NH₄⁺, Na⁺, si tales contra-iones están presentes. Las porciones de base de nucleótido modificadas ejemplares incluyen, pero no se limitan a, 5-metilcitosina (5mC); 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, patente estadounidense No. 5,432,272). Porciones modificadas ejemplares de pentosa 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 cuales la posición 2' o 3' es hidrógeno, hidroxilo, alcoxi (por ejemplo, metoxi, etoxi, aliloxi, isopropoxi, butoxi, isobutoxi y fenoxi), azido, amino, alquilamino, fluoro, cloro, o bromo. Enlaces internucleótidos modificados incluyen análogos de fosfato, análogos que tienen enlaces aquirales y no cargados de intersubunidad (por ejemplo, Sterchak, E. P. et al., *Organic Chem.*, 52:4202, 1987), y polímeros a base de morfolino no cargados que tienen enlaces aquirales de intersubunidad (véase, por ejemplo, patente estadounidense No. 5,034,506). Algunos análogos de enlace internucleótido incluyen morfolidato, acetal y heterociclos enlazados a poliamida. En una clase de análogos de nucleótido, conocidos como ácidos nucleicos de péptido, que incluyen ácidos nucleicos de péptido ("PNA") pseudocomplementarios, un azúcar convencional y un enlace internucleótido han sido reemplazados con un polímero de eje troncal de 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 y Applications*, Nielsen, ed., Horizon Bioscience, 2004).

Los siguientes ejemplos proporcionan modalidades ilustrativas y no limitan de ninguna manera las invenciones proporcionadas en la presente.

Ejemplos

Ejemplo 1 – Reemplazo de oligonucleótido

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

AGATGTGTATAAGAGACAG (SEQ ID NO:3) (ME)

(la "hebra transferida") y una de las siguientes secuencias de extremo (la "hebra no transferida"), mostradas 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
35	CTGTCTCTTAT.....	SEQ ID NO:11 11 bases
	CTGTCTCTTATA.....	SEQ ID NO:12 12 bases
	CTGTCTCTTATAAC.....	SEQ ID NO:13 13 bases
	CTGTCTCTTATAACA.....	SEQ ID NO:14 14 bases
	CTGTCTCTTATAACAC....	SEQ ID NO:15 15 bases
40	CTGTCTCTTATAACACA...	SEQ ID NO:16 16 bases
	CTGTCTCTTATAACACAT.	SEQ ID NO:17 17 bases
	CTGTCTCTTATAACACATC.	SEQ ID NO:18 18 bases
	CTGTCTCTTATAACACATCT	SEQ ID NO:19 19 bases (ME)

donde la secuencias de extremo contenían una secuencia de etiqueta tal como se representa en la figura 3. Los transposomas se prepararon en una mezcla de reacción de 2 µl de secuencia de extremo (25 µM), fosforilada; 2 µl de transposasa Tn5 a 10 U/µl, en un volumen final de 50 µl que tiene concentraciones finales de 33 mM de Tris-acetato, pH 7.8, 10 mM de acetato de magnesio y 66 mM de acetato de potasio. De modo alternativo, las concentraciones finales en el volumen de reacción pueden ser de 10 mM de tris-acetato, pH 7.6, 5 mM de cloruro de magnesio, y opcionalmente 10% (v/v) de dimetilformamida. Después de mezclar, la reacción se incubó por 1 hora a 37°C. La reacción se detuvo con 10 µl de solución de detención: 15% de sacarosa, 66 mM de EDTA, 20 mM de Tris, pH 8, 0.1% de SDS, 0.9% de Orange G (Sigma 0-7252) y Proteinasa K a 100 µg /mL. Después de adicionar la solución de detención la mezcla se calentó a 50 °C por 10 minutos.

Las reacciones de tagmentación se llevaron a cabo con los transposomas para comparar la capacidad de transposomas de haber truncado secuencias de extremo no transferidas en comparación con el extremo no transferido de 19 bases que sirve como control positivo y no un extremo no transferido como un control negativo. Con base en análisis de electroforesis en gel, el ADN diana en una muestra se fragmentó eficientemente por los transposomas que tenían

secuencias de extremo de 18 bases a 12 bases, con fragmentación menos eficiente con secuencias de extremo de 11 bases 9 bases. La secuencia de extremo de 8 bases demostró alguna fragmentación pero no de manera eficiente.

- 5 La secuencia de extremo de 14 bases se seleccionó para más experimentos, tal como se ilustra en la figura 4. Después de tagmentación con transposones que tenían la hebra transferida de 19 bases y la hebra no transferida de 14 bases, el producto se mezcló con oligonucleótidos de reemplazo (SEQ ID NO:19 con una etiqueta de secuenciación como la "etiqueta arbitraria 2") a 45°C por 1 minuto y luego 37 °C por 30 minutos para retirar la hebra no transferida y reemplazarla con el oligonucleótido de reemplazo. Las brechas remanentes de 9 bases se llenaron usando polimerasa *Tth*, fragmento grande (sin desplazamiento de hebra), seguido de ligación usando ligasas de ADN de *E. coli*, de acuerdo con las condiciones recomendadas por el fabricante.
- 10 Se demostró reemplazo mediante adición de una secuencia adicional de 100 bases, tal como se resolvió en un Agilent BioAnalyzer 2100 usando un chip de alta sensibilidad de ADN. Las bibliotecas genómicas de *Rhodobacter*, *E. coli*, y *Staphilococcus* se analizaron debido a sus contenidos divergentes de GC de 70, 50, y 33% respectivamente. Estas bibliotecas se crearon utilizando estrategias de reemplazo de oligonucleótido descritas aquí. La composición de GC a través de las primeras 30 bases no mostró sesgos de inserción adicional con base en el contenido de GC del genoma de hospedero. Los análisis de la composición de GC a través de las primeras 30 bases no demostraron sesgos de inserción alterada con base en el contenido de GC del genoma de hospedero. Los datos mostraron que hubo una profundidad incrementada de cubrimiento cuando el contenido de GC estuvo por encima de 60% y cubrimiento consistente a intervalos de contenido más bajo de GC.

Ejemplo 2 – novedosas secuencias de extremo de transposasa

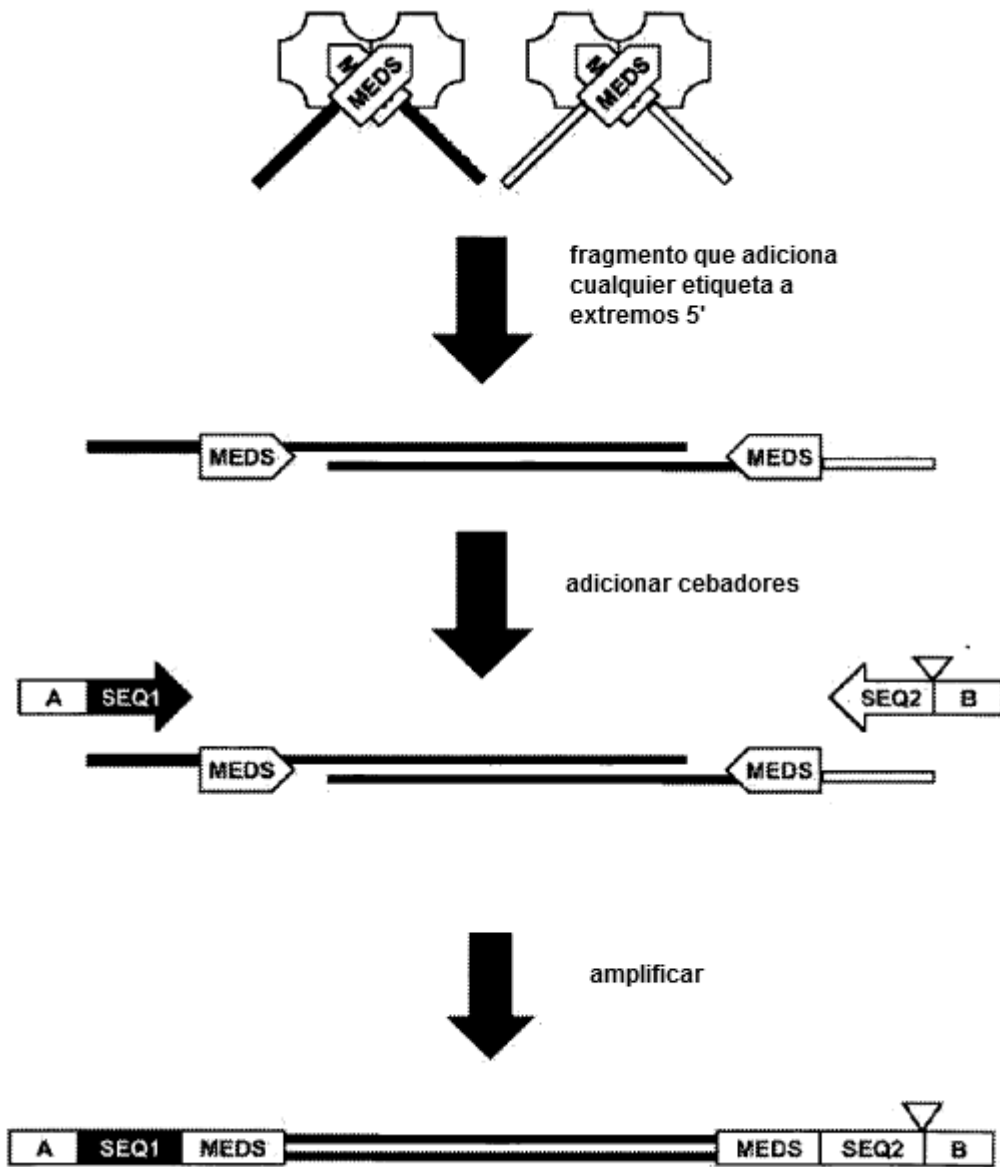
- 20 Se prepararon varios millones de secuencias de extremo de transposasa utilizando una biblioteca parcialmente aleatoria con base en la secuencia de ME de 19 bases (SEQ ID NO:3). De esta biblioteca se identificó una cantidad de extremos activos de transposón utilizando ADN genómico de *E. Coli* MG1655 como el ácido nucleico diana y fueron secuenciados en el instrumento de secuenciación Genome Analyzer_{II} (Illumina, Inc.).

- 25 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 corta
 TGACGCGGGTAAGAGACAA SEQ ID NO:22 Malta 1
 GGATGCGATGAGGAGACAA SEQ ID NO:23 Malta 6
 30 ACATGACCAAGAGAGACAG SEQ ID NO:24 Malta 8
 AGCGGTGAATAAGAGACAA SEQ ID NO:25 Malta 10
 AGCGGTGAATAAGAGACAG SEQ ID NO:26 Malta 11
 ACATGAGTATAAGAGACAA SEQ ID NO:27 Malta 12

- 35 Las secuencias de extremo alternas se confirmaron por secuenciación capilar. Un mutante designado 11G:C(G1) (SEQ ID NO:5) tuvo una temperatura de fusión significativamente alta (58°C en comparación con 44°C) y demostró actividad de transposición comparable in vitro como la secuencia ME, tal como se demostró por dilución en paralelo con transposones con la secuencia ME.

Reivindicaciones

1. Un método para adicionar una etiqueta al producto bicatenario de una reacción de tagmentación que comprende los pasos de:
- 5 (a) proporcionar un ácido nucleico bicatenario diana y un transposoma que tiene una transposasa con dos secuencias extremo de transposón: una hebra transferida y una hebra no transferida;
- (b) permitir que el transposoma fragmente el ácido nucleico diana, por lo cual la hebra transferida se transfiere de modo covalente a una primera hebra de un primer fragmento y la hebra no transferida permanece hibridada a la hebra transferida;
- (c) retirar la hebra no transferida de la hebra transferida;
- 10 (d) proporcionar un oligonucleótido de reemplazo que comprende una secuencia de etiqueta para hibridar a la hebra transferida;
- y
- (e) ligar el oligonucleótido de reemplazo a la segunda hebra del primer fragmento; generando de esta manera un producto de tagmentación que tiene una hebra transferida y un oligonucleótido de reemplazo.
- 15 2. El método de la reivindicación 1, en el cual una hebra del ácido nucleico diana está químicamente modificada.
3. El método de la reivindicación 2, en el cual la modificación química es una conversión de citocinas uracilos.
4. El método de la reivindicación 1, en el cual el transposoma comprende una transposasa Tn5 o Mu.
5. El método de la reivindicación 1, en el cual la hebra transferida comprende la ME:
- 20 MRWTGTGHWKAVGARACAV (SEQ ID NO: 1); o
NSHBGSHSHDDRNGAKACAN (SEQ ID NO: 2);
pero excluyendo AGATGTGTATAAGAGACAG (SEQ ID NO:3).
6. El método de la reivindicación 1 o de la reivindicación 5, en el cual la hebra transferida comprende además una secuencia de etiqueta.
7. El método de la reivindicación 1, en el cual la hebra transferida comprende:
- 25 BTGTYTCBTN₁₋₁₀ (SEQ ID No. 20) o
NTGTMTCNTN₀₋₁₀ (SEQ ID No. 21)
8. El método de la reivindicación 1, en el cual la hebra no transferida se selecciona del grupo consistente en: SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17 y SEQ ID NO: 18.
- 30 9. El método de la reivindicación 1, en el cual el paso (e) comprende además un paso de extensión.
10. El método de la reivindicación 2 o la reivindicación 3, que comprende además el paso de:
- (f) enriquecer para una hebra deseada.
11. El método de la reivindicación 10, en el cual el paso (f) comprende tratar el primer fragmento con un enzima que disocia preferencialmente una hebra no deseada del primer fragmento.
- 35 12. El método de la reivindicación 10 o la reivindicación 11, en el cual el paso (f) se realiza usando una polimerasa dependiente de plantilla que es sensible a modificaciones químicas.



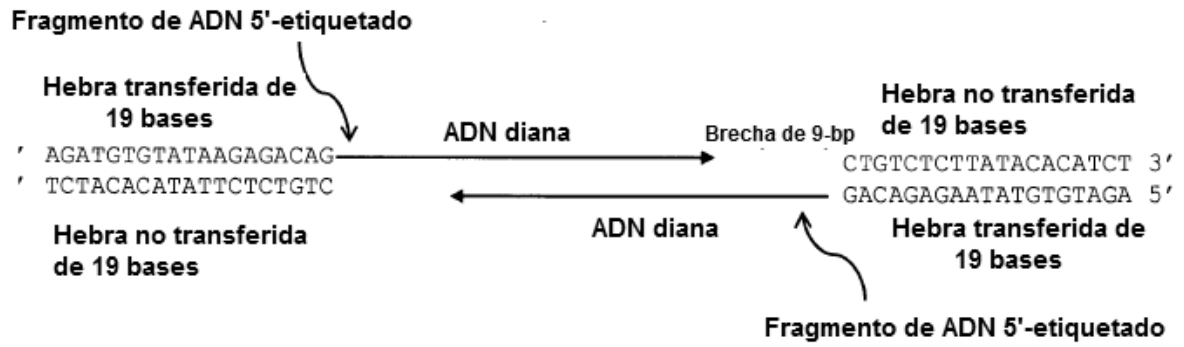


Fig. 2a



Fig. 2b

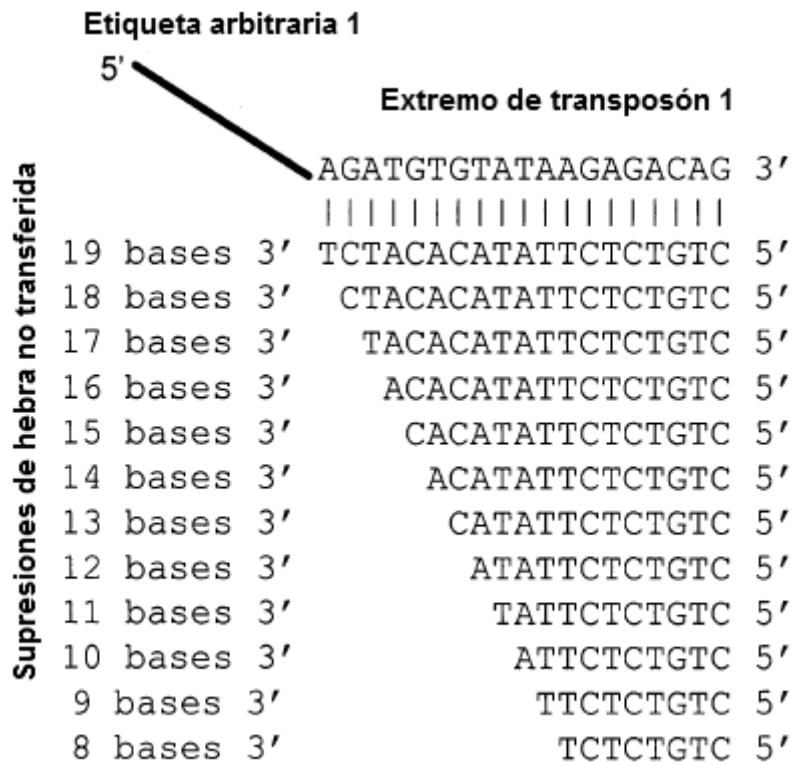
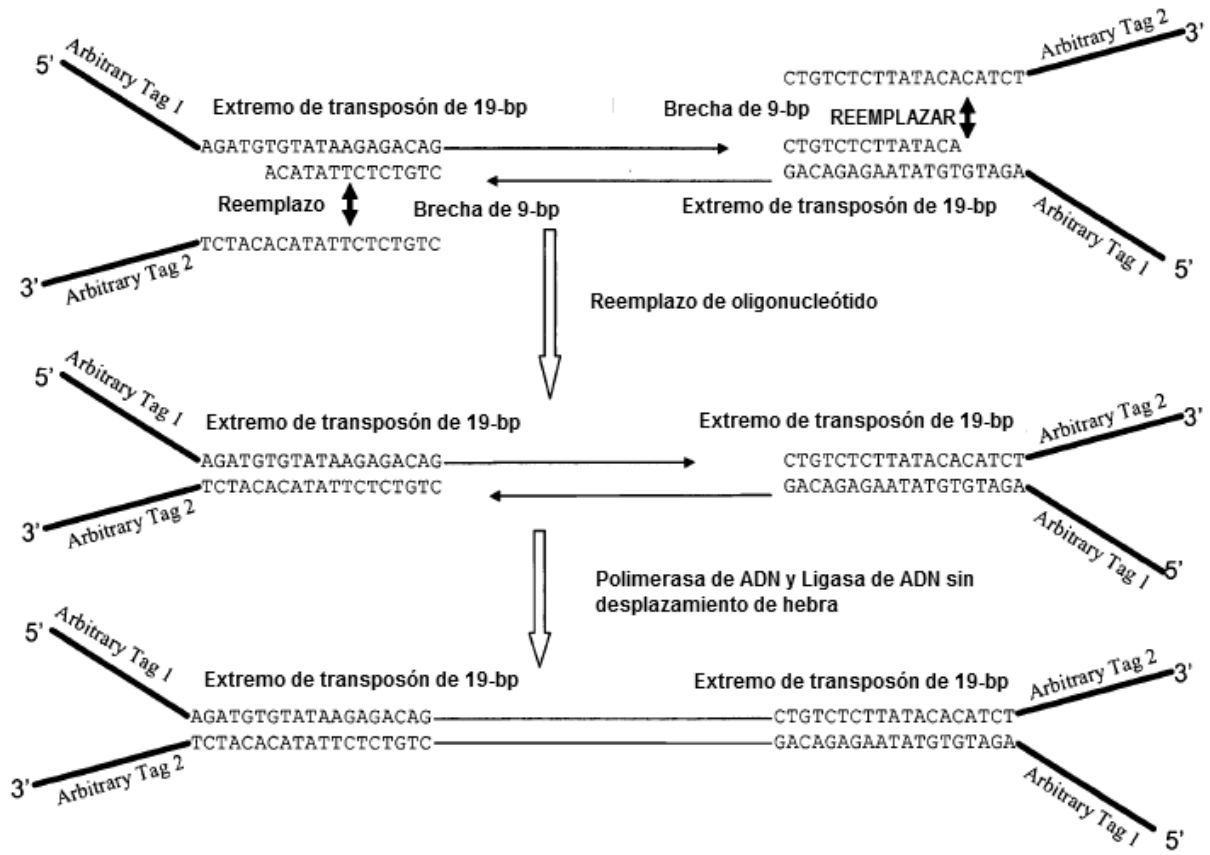


FIG. 3



Arbitrary Rag = Etiqueta arbitraria

Fig. 4

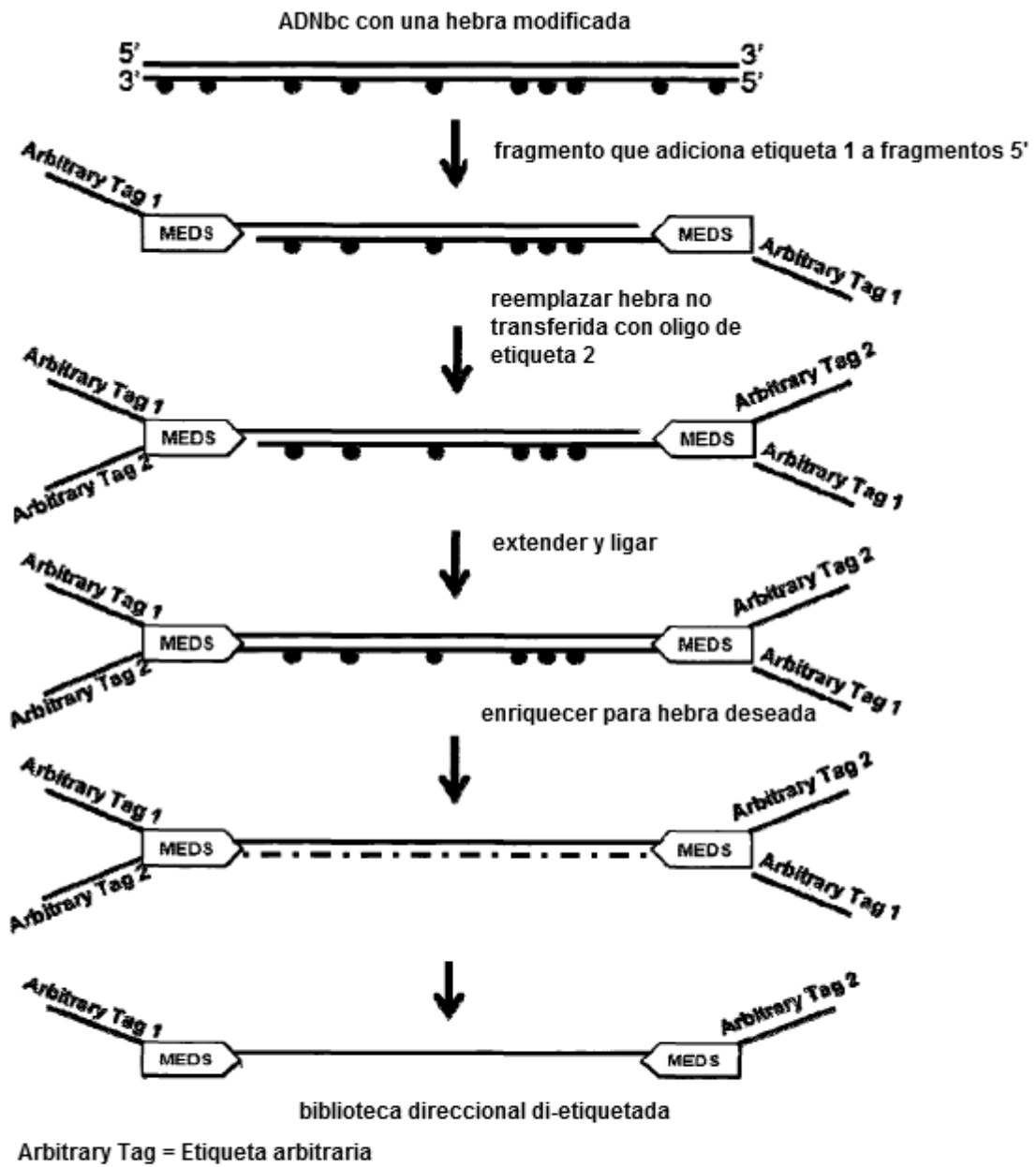


Fig. 5