

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 697 974**

51 Int. Cl.:

C12Q 1/6855 (2008.01)

C12Q 1/6869 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.06.2011 PCT/NL2011/050411**

87 Fecha y número de publicación internacional: **15.12.2011 WO11155833**

96 Fecha de presentación y número de la solicitud europea: **08.06.2011 E 11738836 (3)**

97 Fecha y número de publicación de la concesión europea: **29.08.2018 EP 2580351**

54 Título: **Códigos de barras de secuencia combinatoria para selección de alto rendimiento**

30 Prioridad:

06.12.2010 US 457005 P
09.06.2010 US 352910 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
30.01.2019

73 Titular/es:

KEYGENE N.V. (100.0%)
P.O. Box 216
6700 AE Wageningen, NL

72 Inventor/es:

VAN EIJK, MICHAEL JOSEPHUS THERESIA y
VAN DER POEL, HENRICUS JOHANNES ADAM

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

ES 2 697 974 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

Códigos de barras de secuencia combinatoria para selección de alto rendimiento

5 Campo de la invención

[0001] La presente invención se refiere al campo de biología molecular, en particular a la preparación de ADN de muestra para métodos de secuenciación. Más específicamente, la presente invención se refiere al uso de identificadores de secuencia de nucleótidos para secuenciación de alto rendimiento.

10

Antecedentes de la invención

[0002] La alta demanda de secuenciación de bajo coste ha llevado al desarrollo de tecnologías de secuenciación de alto rendimiento. En tales tecnologías, millones de secuencias se producen en paralelo. Por ejemplo, 454 Life Sciences, ahora Roche Applied Sciences, desarrolló una tecnología de secuenciación de alto rendimiento de un ADN de muestra que implica los pasos de fragmentación de ADN, ligamiento de adaptadores a los fragmentos de ADN, captura de fragmentos de ADN únicos con un reborde recubierto con cebadores, amplificando cada fragmento de ADN en un reborde dentro de unas gotas de agua en aceite (emulsión PCR), y posteriormente cargar cada reborde en un pocillo de picolitro y secuenciar cada fragmento de ADN amplificado con pirosecuenciación. En general, las tecnologías de secuenciación de alto rendimiento implican el ligamiento de adaptadores a fragmentos de ADN, donde estos adaptadores pueden comprender sitios de unión a cebador utilizados para captura, amplificación y/o secuenciación de los fragmentos de ADN. Debido al hecho de que se pueden producir un gran número de secuencias, frecuentemente se combinan muestras de distintos orígenes en un único recorrido de secuenciación de alto rendimiento. Para determinar el origen de cada muestra de una agrupación de muestras, las aplicaciones de secuenciación de alto rendimiento actuales dependen del uso de identificadores de secuencia de nucleótidos. El término identificador de secuencias de nucleótidos (NSI, (código de barras basado en secuencia o índice de secuencia son términos que son intercambiables y tienen el mismo significado. Un identificador de secuencia de nucleótidos es una secuencia de nucleótidos particular que se usa como un identificador. Un identificador de secuencia de nucleótidos se incluye en el adaptador debajo del sitio de unión del cebador de manera que cuando se secuencia del sitio de unión de cebador, se determina la secuencia de nucleótidos de la secuencia de identificador. Se ligan adaptadores diferentes que comprenden identificadores de secuencia de nucleótidos diferentes a muestras diferentes, después de lo cual las muestras se pueden agrupar. Cuando se determinan las secuencias de las muestras agrupadas, el identificador de secuencia de nucleótidos se secuencia junto con parte de la secuencia del fragmento al que el adaptador se liga. La presencia o ausencia del identificador de secuencia de nucleótidos determina así la presencia o ausencia de una muestra de ADN en la agrupación. La secuencia de la secuencia interna que está secuenciada junto con el identificador de secuencia de nucleótidos habilita además a asignar esa secuencia a una muestra particular donde esta se originó, puesto el identificador de secuencias de nucleótidos sirve para identificar la muestra origen de ADN.

[0003] Por ejemplo, en el sistema de secuencia de alto rendimiento desarrollado por Roche, se utilizan secuencias identificadoras multiplexadas (MIDs) Genome Sequencer FLX. Las MIDs son secuencias 10-mer que se incorporan en los adaptadores para asignar lecturas de secuencia a muestras individuales. Se usan actualmente más de 100 MIDs diferentes (454 Life Science Corp (2009) Technical Bulletin N° 005-2009). Están disponibles identificadores de secuencias de nucleótidos similares para otros sistemas de secuenciación.

45

[0004] Se describen por ejemplo métodos, donde se incorporan identificadores de secuencia de nucleótidos en el extremo 5' de un cebador, por Rigola et al. PLoS ONE. 2009; 4(3): e4761 y en WO 2007/037678. Típicamente, los identificadores de secuencias de nucleótidos no tienen complementariedad significativa con la secuencia objetivo. Un cebador por tanto comprende en el extremo 5' una sección que comprende un identificador de secuencia de nucleótidos y en el extremo 3' la secuencia que es complementaria a la secuencia objetiva. Cuando una muestra se amplifica con un par de cebadores de los cuales un cebador comprende un identificador de secuencia de nucleótidos, el amplicón incluirá el identificador de secuencia de nucleótidos. Cuando las muestras son posteriormente agrupadas y sometidas a métodos de secuenciación de alto rendimiento, el identificador de secuencia de nucleótidos servirá para identificar el origen del amplicón secuenciado. Por lo tanto, el origen del amplicón se determina mediante la determinación del identificador de secuencia de nucleótidos. Al mismo tiempo, puede rastrearse el origen de la secuencia interna que ha sido amplificada y está secuenciada también junto con el identificador de secuencia de nucleótidos a partir de las muestras de las que se origina.

50

[0005] En ambos escenarios, un adaptador o cebador que comprende un identificador de secuencias de nucleótidos, el concepto es el mismo, es decir, determinar el origen de muestra de secuencias producidas utilizando plataformas de secuenciación de alto rendimiento a partir de una pluralidad de muestras de ADN que han sido multiplexadas, por ejemplo, combinadas o agrupadas, en algún lugar en el proceso de preparación de la muestra.

60

65 Resumen de la invención

[0006] La capacidad de las de tecnologías de secuenciación de alto rendimiento ha aumentado un orden de magnitud por un periodo de dos años desde su introducción. Con la secuenciación de alto rendimiento permitiendo el multiplexado de un mayor número en aumento de muestras, el número de adaptadores o cebadores únicos requeridos para identificar el origen de las muestras está también en aumento. Aunque el uso de 100 cebadores diferentes o adaptadores puede ya ser desafiante, cuando el número aumente a 1000, puede convertirse en un obstáculo. Por lo tanto es deseable que el número de cebadores y/o adaptadores que tienen que usarse se pueda reducir porque esto puede simplificar la preparación de la muestra, puede reducir la carga de trabajo, puede optimizar el rendimiento técnico y puede reducir costes. La presente invención tal y como se define por las reivindicaciones 1-11 habilita la reducción del número de diferentes cebadores y/o adaptadores requeridos. El número podría reducirse por el uso de los llamados "códigos de barras de división". Los códigos de barras de división según la invención tal y como se define por las reivindicaciones 1-11 son identificadores de secuencia de nucleótidos que están presentes en al menos dos adaptadores y/o cebadores. Un ADN de muestra (o una combinación de ADNs de muestra) se prepara utilizando por ejemplo un par de cebadores y/o un par de adaptadores, cada cebador o adaptador del par comprendiendo un identificador de secuencia de nucleótidos. El amplicón o fragmento de ADN ligado a un adaptador producido comprende al menos dos identificadores de secuencias de nucleótidos. Para cada muestra diferente puede utilizarse una combinación única de identificadores de nucleótidos. La combinación de identificadores de secuencias de nucleótidos, también indicada como el código de barras de división, sirve como identificador.

20 Breves descripciones de los dibujos

[0007]

25 Figura 1. Método para preparar un amplicón a partir de una muestra ADN con dos identificadores de secuencias de nucleótidos. Un ADN de muestra se proporciona (1) comprendiendo una secuencia interna (SI) flanqueada por dos sitios de unión de cebador (P1 y P2), al igual que un par de cebadores de amplificación (2) comprendiendo secuencias complementarias a un sitio de unión de cebador en el extremo 3' e identificadores de secuencia de nucleótidos (NSI1 y NSI2) 5' del mismo. El ADN de muestra se amplifica con los cebadores de amplificación (3), dando como resultado un amplicón con dos identificadores de secuencias de nucleótidos en cada lado. (5' indica el extremo 5' de una cadena de nucleótidos, el extremo 3' no tiene anotación).

35 Figura 2. Método para preparar un fragmento de ADN ligado a un adaptador de un ADN de muestra con dos identificadores de secuencias de nucleótidos. Se proporciona un ADN de muestra (1), que se fragmenta proporcionando fragmentos de ADN (2), se proporciona un par de adaptadores comprendiendo un primer y segundo NSI (NSI1 y NSI2) (3) los cuales se ligan a ambos extremos de un fragmento de ADN, dando como resultado un fragmento de ADN ligado a un adaptador (4). (5' indica el extremo 5' de una cadena de nucleótidos).

40 Figura 3. Método para la preparación de un fragmento de ADN ligado a un adaptador amplificado de un ADN de muestra con 2, 3 o 4 identificadores de secuencias de nucleótidos. Se proporciona un ADN de muestra (1), que se fragmenta proporcionando fragmentos de ADN (2), un par de adaptadores se proporcionan de los cuales al menos uno comprende un NSI (NS1 y opcionalmente (NS2)) y ambos comprenden un sitio de unión de cebador (P1 y P2), que están ligados con un fragmento de ADN (3), es decir secuencia interna (IS), dando como resultado un fragmento de ADN ligado a un adaptador (4) comprendiendo las secuencias de unión de cebador en ambos extremos de los fragmentos de ADN ligados a un adaptador. Se proporciona un par de cebadores de amplificación (5) comprendiendo cada uno en el extremo 3' una secuencia complementaria a un sitio de unión de cebador de secuencia, y al menos un cebador de amplificación comprende un identificador de secuencias de nucleótidos en el extremo 5' (NSI3 u opcionalmente (NSI4)). El fragmento de ADN ligado a un adaptador se amplifica con el par de cebadores de amplificación (6). El resultado (7) es un fragmento de ADN ligado a un adaptador amplificado que comprende al menos dos NSIs. Los al menos dos NSIs pueden estar flanqueando IS, y/o al menos dos NSIs pueden estar en el mismo lado que la IS. (5' indica el extremo 5' de una cadena de nucleótido; los paréntesis indican que la inclusión del segundo y/o cuarto NSI puede ser opcional en este método).

60 Figura 4. Método para preparar un amplicón ligado a un adaptador de una muestra ADN con dos, tres o cuatro identificadores de secuencia de nucleótidos. Se proporciona (1) una muestra de ADN que comprende una secuencia interna (IS) flanqueada por dos sitios de unión de cebador (P1 y P2), al igual que un par de cebadores de amplificación (2) comprendiendo en el extremo 3' secuencias complementarias a un sitio de unión de cebador y al menos uno de los cebadores que comprende un identificador de secuencia de nucleótidos (NSI1 y opcional (NSI2)) en el extremo 5'. La muestra de ADN se amplifica con los cebadores de amplificación (3), dando como resultado un amplicón con al menos un identificador de secuencias de nucleótidos (4). Se proporcionan un par de adaptadores que comprenden un tercer y un opcional cuarto nucleótido NSI (NSI3 y opcional (NSI4)) que se ligan a cualquier extremo del amplicón proporcionando así un amplicón ligado a un adaptador (6). Los al menos dos NSIs pueden

estar flanqueando la IS, y/o al menos dos NSIs pueden estar en el mismo lado que la IS. (5' indica el extremo 5' de una cadena de nucleótidos; los paréntesis indican que la inclusión del segundo y/o cuarto NSI pueden ser opcionales en este método).

5 Figura 5. Método para determinar las dos secuencias de identificadores de secuencias de nucleótidos de una muestra de fragmento de ADN ligado a un adaptador. Se proporciona un ADN de muestra (1), que está fragmentado proporcionando fragmentos de ADN (2), se proporciona (3) un par de adaptadores comprendiendo un primer y segundo NSI (NSI1 y NSI2) que se ligan a cualquier extremo de un fragmento de ADN, dando como resultado un fragmento de ADN ligado a un adaptador (4). Los adaptadores comprenden cada uno un sitio de unión de cebador de secuenciación (SEQ1 y SEQ2), y opcionalmente cada uno comprende un sitio de unión de cebador de amplificación ((P1) y (P2)). El orden de los sitios presentes en un adaptador es: (P) SEC - NSI, por ejemplo (P1)-SEQ1-NSI1. El lado del adaptador que se liga al fragmento de ADN es el lado que comprende el NSI. El fragmento de ADN ligado a un adaptador puede opcionalmente amplificarse con cebadores dirigidos a los sitios de unión de cebador (4). Cada cadena del adaptador de ADN ligado a un adaptador puede servir como un modelo para una reacción de secuenciación. Un modelo de cadena usada se representa de la siguiente manera: 3'-(P1)-SEQ1-NSI1-IS NSI2(P2)-SEQ2-5', para que un cebador de secuenciación se usa contra SEQ1. Se proporcionan cebadores de secuenciación de manera que de cada modelo de SEQ1 o SEQ2 se determinen la(s) secuencia(s) NSI. Las secuencias se pueden determinar separadamente. Las secuencias se pueden determinar consecutivamente, por ejemplo, tal como en la secuenciación de extremo emparejado. (5' indica el extremo 5' de una cadena de nucleótidos).

25 Figura 6. Método para determinar las dos secuencias de identificadores de secuencia de nucleótidos de un fragmento de ADN ligado a un adaptador: lectura única etiquetado doble. Se proporciona un ADN de muestra (1), que está fragmentado proporcionando fragmentos de ADN (2), se proporciona un par de adaptadores que comprenden un primer y un segundo NSI (NSI1 y NSI2) que se ligan a cualquier extremo de un fragmento de ADN (3), dando como resultado un fragmento de ADN ligado a un adaptador (4). Los adaptadores comprenden cada uno un sitio de unión a cebador de secuenciación (SEQ1 o SEQ2), y opcionalmente cada uno comprende un sitio de unión a cebador de amplificación ((P1) o (P2)). La orden de los sitios presente en los dos adaptadores es (P1)-SEQ1-NSI1 y SEQ2-NSI2-(P2). Los adaptadores se ligan al fragmento de ADN de manera que los sitios (P1) y (P2) son los sitios externos del fragmento de ADN ligado a un adaptador (4), que puede opcionalmente ser amplificado con cebadores dirigidos a él (5). Una cadena del fragmento de ADN ligado a un adaptador puede servir como un modelo para secuenciación utilizando los sitios de unión a cebador de secuenciación SEQ1 y SEQ2 en dos reacciones de secuenciación diferentes, es decir utilizando los cebadores de secuenciación diferentes correspondientes. La cadena de modelo utilizada se representa de la siguiente manera: 3'-(P1)-SEQ1-NSI1-IS-SEQ2-NSI2-(P2)-5'. (5' indica el extremo 5' de una cadena de nucleótido).

40 Figura 7. Método para determinar la secuencia de dos identificadores de secuencia de nucleótidos de un amplicón de un ADN de muestra. Se proporciona una muestra de ADN (1) comprendiendo una secuencia interna (IS) flanqueada por dos sitios de unión de cebador (P1 y P2), al igual que un par de cebadores de amplificación (2) comprendiendo las secuencias en el extremo 3' complementarias a un sitio de unión de cebador y sitios de unión de cebador de secuenciación (SEQ) en los extremos 5'. Entremedias, en los cebadores de amplificación, se localizan identificadores de secuencias de nucleótidos. El ADN de muestra se amplifica con los cebadores de amplificación (3), dando como resultado un amplicón con dos identificadores de secuencias de nucleótidos en cada lado, con los dos SEQs (SEQ1 y SEQ2) en los extremos externos del amplicón. Cada cadena del amplicón puede servir como un modelo para una reacción de secuenciación. Un modelo de cadena utilizada se representa de la siguiente manera: 3-SEQ1-NSI1-P1-IS-P2-NSI2-SEQ2-5', para la cual se usa un cebador de secuenciación contra SEQ1. Se proporcionan cebadores de secuenciación de manera que de cada modelo se determina(n) la(s) secuencia(s) NSI. Las secuencias se pueden determinar separadamente. Las secuencias se pueden determinar consecutivamente, por ejemplo, tal como en la secuenciación del extremo emparejado. (5' indica el extremo 5' de una cadena de nucleótidos).

55 Figura 8. Método para determinar la secuencia de dos identificadores de secuencias de nucleótidos de un amplicón de una muestra de ADN: única lectura etiquetado doble. Se proporciona una muestra ADN (1) comprendiendo una secuencia interna (IS) flanqueada por dos sitios de unión de cebador (P1 y P2), al igual que un par de cebadores de amplificación (2) comprendiendo secuencias complementarias a un sitio de unión de cebador en el extremo 3' (C1 o C2). Los cebadores comprenden además NSIs y sitios de unión de cebador de secuenciación, los cebadores diferentes comprendiendo los sitios/secuencias diferentes que se representan de la siguiente manera: 5'-SEQ1-NSI1-C1 y C2-SEQ2-NSI2-5'. El ADN de muestra se amplifica con los cebadores de amplificación (3), dando como resultado un amplicón con un identificador de secuencia de nucleótidos a cada lado, con SEQ1 en un extremo externo del amplicón y en el otro extremo externo NSI2 (4). Una cadena del amplicón puede servir como un modelo para secuenciación utilizando los sitios de unión de cebador de secuenciación SEQ1 y SEQ2 en dos reacciones de secuenciación diferentes, es decir, utilizando los cebadores de secuenciación diferentes

correspondientes. El modelo de cadena utilizado se representa de la siguiente manera: 3'-SEQ1-NSI1-P1-IS-P2-SEQ2-NSI2-5'. (5' indica el extremo 5' de una cadena de nucleótido).

5
10
15
20

Figura 9. Método para determinar la secuencia de cuatro identificadores de secuencia de nucleótidos de un amplicón de una muestra de ADN: lectura única etiquetado doble. Se proporciona una muestra de ADN (1) comprendiendo una secuencia interna (IS) flanqueada por dos sitios de unión a cebador (P1 y P2), al igual que un par de cebadores de amplificación (2) que comprende secuencias complementarias a un sitio de unión de cebador en el extremo 3' (C1 o C2). Los cebadores comprenden además NSIs y uno de los cebadores comprende un sitio de unión a cebador de secuenciación. Los cebadores diferentes que comprenden las secciones diferentes están representados de la siguiente manera: 5'-NSI1-C1 y C2-SEQ2-NSI2-5'. El ADN de muestra se amplifica con los cebadores de amplificación (3), dando como resultado un amplicón con dos identificadores de secuencia de nucleótidos en los extremos externos del amplicón. A continuación, se proporciona un par de adaptadores (4). Un adaptador comprende un sitio de unión a cebador de secuenciación (SEQ1) y un NSI (NSI3), el otro cebador comprende un NSI (NSI4). Los adaptadores se ligan a cualquier extremo del amplicón, dando como resultado un amplicón ligado a un adaptador, donde la sección SEQ1 está sobre el extremo externo del amplicón ligado a un adaptador (5), y SEQ1 y SEQ2 están flanqueando la IS. Una cadena del fragmento de ADN ligado a un adaptador puede servir como un modelo para secuenciación utilizando los sitios de unión a cebador de secuenciación SEQ1 y SEQ2 en dos reacciones de secuenciación diferentes, es decir utilizando los cebadores de secuenciación diferentes correspondientes. El modelo de cadena utilizado se representa de la siguiente manera: 3'-SEQ1-NSI3-NSI1-IS -SEQ2-NSI2-NSI4-5'. (5' indica el extremo 5' de una cadena de nucleótidos).

25
30
35
40

Figura 10. Método para determinar la secuencia de cuatro identificadores de secuencia de nucleótidos de un amplicón de un ADN de muestra. Se proporciona un ADN de muestra (1) que comprende una secuencia interna (IS) flanqueada por dos sitios de unión a cebador (P1 y P2), al igual que un par de cebadores de amplificación (2) que comprende secuencias complementarias a un sitio de unión a cebador en el extremo 3' (C1 o C2). Los cebadores comprenden además NSIs. Los diferentes cebadores comprendiendo las diferentes secciones que se representan de la siguiente manera: 5'- NSI1-C1 y C2 -NSI2-5'. El ADN de muestra se amplifica con los cebadores de amplificación (3), dando como resultado un amplicón con los dos identificadores de secuencia de nucleótidos en los extremos externos del amplicón. A continuación, se proporcionan un par de adaptadores (4). Cada adaptador comprende un sitio de unión a cebador de secuenciación (SEQ1 o SEQ2) y un NSI (NSI3 o NSI4). Los adaptadores se ligan a cualquier extremo del amplicón, dando como resultado un amplicón ligado a un adaptador, donde las secciones SEQ1 y SEQ2 están en los extremos externos del amplicón ligado a un adaptador (5). Cada cadena del amplicón ligado a un adaptador puede servir como un modelo para una reacción de secuenciación. Se proporcionan cebadores de secuenciación de manera que de cada modelo se determine(n) la(s) secuencia(s) NSI. Las secuencias se pueden determinar separadamente. Las secuencias se pueden determinar consecutivamente, por ejemplo, tal como en la secuenciación de extremo emparejado. Uno de los modelos de cadena utilizados se representa de la siguiente manera: 3'-SEQ1-NSI3-NSI1-P1-IS-P2- NSI2-NSI4-SEQ2-5', para el que se usa un cebador de secuenciación contra SEQ1. (5' indica el extremo 5' de una cadena de nucleótidos).

45

Figura 11. Amplificación con el par de cebadores. (UT1, cola universal 1; BC1, parte 1 de código de barras, UT2 cola universal 2; BCP2, parte 2 de código de barras)

50

A. La cola universal 1 puede ser el sitio de cebador de secuencia 1 (flecha negra grande) y la cola universal 2 puede ser el sitio de cebador de secuencia 2 (flecha punteada), por ejemplo, P5 y P7 en caso de la secuenciación de extremo emparejado Illumina GA.

B. La cola universal 1 puede ser el sitio de cebador de secuencia 1 (flecha negra grande) y la cola universal 2 puede ser el sitio de cebador de secuencia 2 (flecha discontinua), por ejemplo, P5 y P7 en caso de la secuenciación Illumina GA con dos eventos de cebador de la misma cadena.

55
60

Figura 12. Ligamiento de un par de adaptadores codificados por barras. (P5, P5 + sitio de cebador de secuencia; BC1, parte de código de barras 1; BC2, parte de código de barras 2; P7, B, adaptador de extremo romo). A. Ligamiento de adaptadores codificados por barras EcoRI y MseI para ADN digerido EcoRI/MseI, donde la combinación de partes 1 de código de barras (EcoRI lateral) y 2 (MseI lateral) definen la muestra únicamente. B. Ligamiento de EcoRI y adaptadores de extremo romo codificados por barras para mostrar que fue primero digerido con EcoRI, seguido de ligamiento a un adaptadores codificados por barras EcoRI (código de barras 1), seguidos de fragmentación de los fragmentos ligados a adaptador, pulido del extremo opcional y seguido de ligamiento a un adaptador del extremo romo (parte de código de barras 2), donde la combinación de partes de código de barras 1 (lado EcoRI) y parte de código de barras 2 (lado de extremo romo) definen la muestra únicamente.

[0008] En la siguiente descripción y ejemplos se utilizan varios términos. Para proporcionar una comprensión clara y consistente de la especificación y reivindicaciones, con el alcance de tales términos, se proporcionan las siguientes definiciones. A menos que se defina aquí lo contrario, todos los términos técnicos y científicos utilizados tienen el mismo significado como se entiende comúnmente por una de las habilidades en la materia a la que esta invención pertenece.

[0009] Serán evidentes al trabajador experto los métodos de realización de las técnicas convencionales utilizados en los métodos descritos. La práctica de técnicas convencionales en la biología molecular, bioquímica, química computacional, cultivo celular, DNA recombinante, bioinformática, genómicos secuenciación y campos relacionados son bien conocidos a los expertos en la técnica y se analizan, por ejemplo, en las siguientes referencias de bibliografía: Sambrook et al., Molecular Cloning. A Laboratory Manual, segunda edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989; Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1987 y actualizaciones periódicas; y los métodos en serie en Enzymology, Academic Press, San Diego.

[0010] Como se utiliza en este caso, las formas singulares "un" y "el" incluyen referentes plurales a menos que el contexto claramente dictamine lo contrario. Por ejemplo, el aislamiento del ADN incluye el aislamiento de una pluralidad de moléculas de ADN (por ejemplo 10's, 100's, 1000's, 10's de millares, 100's de millares, millones, o más moléculas).

[0011] Una "secuencia de nucleótidos" según la presente invención puede incluir cualquier polímero u oligómero de nucleótidos tal como pirimidina y bases de purina, preferiblemente citosina, timina, y uracilo, y adenina y guanina, respectivamente y combinaciones de los mismos (ver Alberto L. Lehninger, Principles of Biochemistry, en 793-800 (Worth Pub. 1982). La presente invención contempla cualquier desoxirribonucleótido, ribonucleótida o componente de ácido nucleico péptido, y cualquier variante química de lo mismo, tal como metilado, hidroximetilado o formas glicosiladas de estas bases, y similares. Los polímeros u oligómeros pueden ser heterogéneos u homogéneos en la composición, y se pueden aislar de fuentes de origen natural o pueden ser producidos artificialmente o sintéticamente. Además, una secuencia de nucleótidos puede ser de ADN o ARN, o una mezcla de lo mismo, y puede existir permanente o transicionalmente en la forma monocatenaria o bicatenaria, incluyendo homodúplex, heterodúplex, y estados híbridos.

[0012] Un "ADN de muestra" según la invención es una muestra derivada de un organismo y que comprende ADN. Un "ADN de muestra" puede comprender células de un organismo comprendiendo ADN, pero también ADN aislado de las células de un organismo. Siempre y cuando "ADN de muestra" comprenda ADN que se pueda usar en el método de la invención tal ADN de muestra se puede utilizar en la invención. Los organismos de donde se puedan obtener muestras de ADN son por ejemplo plantas, mamíferos, hongos, y microorganismos. El ADN de muestra también puede comprender identificativos de secuencia expresada o ADNc, donde ARN como se expresa en células de organismos se convierte en el ADN bicatenario a través de transcripción inversa. El ADN de muestra también puede comprender ADNs de muestra agrupados obtenidos de distintos sitios de un organismo, y/o de diferentes organismos. Los ADNs de muestra agrupados se pueden agrupar por ejemplo en un esquema de fusión 3-D de manera que se pueda determinar el origen de cada muestra que se comprende en un ADN de muestra (por ejemplo, como se describe en WO2007/037678).

[0013] "Fragmentando el ADN" incluye cualquier técnica que, cuando se aplica a una muestra ADN, produce fragmentos de ADN. La sonicación, el corte y/o la restricción enzimática son técnicas bien conocidas en la técnica, pero también pueden estar previstas otras técnicas.

[0014] Una "endonucleasa de restricción" o "enzima de restricción" es una enzima que reconoce una secuencia de nucleótidos específica (sitio de reconocimiento), por ejemplo en una molécula de ADN bicatenario, y disociará ambas cadenas de la molécula de ADN cerca de cada sitio de reconocimiento, dejando un extremo romo o un extremo saliente 3' o 5'. La secuencia de nucleótidos específica que se reconoce puede determinar la frecuencia de quebradura, por ejemplo, una secuencia de nucleótidos de 6 nucleótidos ocurre de media cada 4096 nucleótidos, mientras que una secuencia de nucleótidos de 4 nucleótidos ocurre mucho más frecuentemente, de media cada 256 nucleótidos. Las enzimas de restricción tipo I cortan en un sitio que difiere, y a cierta distancia (a mínimo 1000 pares de bases) de su sitio de reconocimiento. El sitio de reconocimiento es asimétrico y está compuesto de dos partes, una que contiene 3-4 nucleótidos, y otra que contiene 4-5 nucleótidos, separados por un separador de aproximadamente 6-8 nucleótidos. Las enzimas de restricción tipo II tienen sitios de reconocimiento que son normalmente enteros y palindrómicos y 4-8 nucleótidos de longitud. Estas reconocen y disocian el ADN en el mismo sitio. El tipo II corta en el exterior de su secuencia de reconocimiento y el tipo IIB corta el ADN en ambos lados de su sitio de reconocimiento para recortar el sitio de reconocimiento. Las enzimas de restricción tipo III (por ejemplo, EcoP15) reconocen dos secuencias no-palindrómicas separadas que se orientan inversamente. Estas cortan ADN aproximadamente 20-30 pares de bases después del sitio de reconocimiento. Las enzimas de restricción tipo IV cortan el ADN metilado.

[0015] "Pulir" incluye cualquier técnica utilizada para hacer secuencias de nucleótidos bicatenarias que pueden tener salientes 3' o 5' de punta roma. Por ejemplo, en el caso de que un ADN de muestra se fragmente utilizando

la sonificación o enzimas que suponen extremos escalonados (salientes). Polimerasa I de ADN, el fragmento grande (Klenow) puede utilizarse para llenar los salientes 5' (también llamados extremos ahuecados 3') y digerir salientes 3' o la nucleasa Mung Bean pueden utilizarse para digerir los salientes 3' o 5'.

5 [0016] "Ligamiento" según la invención implica la unión de secuencias de nucleótidos bicatenarias separadas. Las Moléculas de ADN bicatenarias pueden acabar en forma redondeada, o pueden tener preponderancias compatibles (preponderancias pegajosas) de manera que las preponderancias pueden hibridar entre sí. La unión de los fragmentos de ADN puede ser enzimática, con una enzima de ligasa, ligasa de ADN. No obstante, un
10 ligamiento no enzimático también se puede utilizar, mientras que se unen los fragmentos de ADN, es decir formando un enlace covalente. Típicamente un enlace de fosfodiéster entre el hidroxilo y grupo de fosfato de las cadenas separadas se forma en una reacción de ligamiento. Las secuencias de nucleótidos bicatenarias pueden tener que ser fosforiladas antes del ligamiento.

15 [0017] "Cebadores de amplificación" se refieren a secuencias de nucleótidos monocatenarias que pueden cebar la síntesis de ADN. La polimerasa de ADN no puede sintetizar ADN de nuevo sin cebadores. Un cebador de amplificación hibrida con el ADN, es decir se forman pares de bases. Nucleótidos que pueden formar pares de bases, que son complementarios unos a otros, son por ejemplo citosina y guanina, timina y adenina, adenina y uracilo, guanina y uracilo. La complementariedad entre el cebador de amplificación y la cadena de ADN existente no debe ser 100%, es decir no todas bases de un cebador necesitan ser par de base de la cadena de ADN
20 existente. La secuencia de la cadena de ADN existente, por ejemplo, el ADN de muestra o un fragmento de ADN ligado a un adaptador, a lo que un cebador de amplificación (parcialmente) hibrida se refiere frecuentemente como sitio de unión de cebador (PBS). Del extremo 3' de un cebador hibridado con la cadena de ADN existente, se incorporan nucleótidos utilizando la cadena existente como un modelo (síntesis de ADN dirigido por modelo). También puede referirse a las moléculas de oligonucleótido sintético que se usan en una reacción de
25 amplificación como "cebadores". Las secuencias de nucleótidos sintetizadas recientemente en la reacción de amplificación pueden designarse como una secuencia interna. En el caso de que se realice una reacción por PCR, la secuencia interna es típicamente la secuencia en medio de los dos sitios de unión de cebador. Según la invención, un cebador se puede utilizar en un paso de amplificación para introducir secuencias adicionales al ADN. Esto se puede conseguir proporcionando cebadores con secuencias adicionales tal como un identificador, un adaptador de secuenciación o un ligando de captura tal como una fracción de biotina. Se pueden introducir
30 modificaciones proporcionándolas en el extremo 5' del cebador, encima de la parte del cebador la cual consigue cebar la síntesis de ADN.

35 [0018] "Amplificación" o "amplificando" se refiere a una reacción de amplificación de polinucleótidos, es decir, una población de polinucleótidos que se replican a partir de una o varias secuencias de inicio. Amplificación puede referirse a una variedad de reacciones de amplificación, incluyendo pero no limitada a una reacción de cadena de polimerasa (PCR), reacciones de polimerasa lineal, amplificación de base de secuencia de ácido nucleico, amplificación de círculo rodante y reacciones similares. Típicamente, los cebadores de amplificación se usan para la amplificación, siendo un amplicón el resultado de la reacción de amplificación.

40 [0019] "Cebadores secuenciadores" se refiere a secuencias de nucleótidos monocatenarias que pueden cebar la síntesis de ADN y se utilizan para secuenciar el ADN. Un cebador de amplificación también se puede utilizar como un cebador de secuenciación. Un cebador de secuenciación se puede utilizar como un cebador de amplificación. Polimerasa de ADN no puede sintetizar ADN de nuevo sin cebadores. Un cebador de
45 secuenciación hibrida el ADN, es decir se forman pares de bases. Nucleótidos que pueden formar pares de bases, que son complementarios uno a otro, son por ejemplo citosina y guanina, timina y adenina, adenina y uracilo, guanina y uracilo. La complementariedad entre el cebador de amplificación y la cadena de ADN existente no debe ser 100%, es decir, no todas las bases de un cebador necesitan ser par base de la cadena de ADN existente. La secuencia de la cadena de ADN existente, por ejemplo el ADN de muestra o un fragmento de ADN
50 ligado a un adaptador, a lo que un cebador de secuenciación (parcialmente) hibrida se refiere frecuentemente como sitio de unión de cebador de secuenciación (SEQ). Del extremo 3' de un cebador de secuenciación hibridado con la cadena de ADN existente, se incorporan nucleótidos utilizando la cadena existente como un modelo (síntesis de ADN dirigido de modelo). La incorporación de un nucleótido particular (A, T, C, o G) se puede detectar durante la síntesis, por ejemplo, en la pirosecuenciación o cuando se utilizan nucleótidos
55 marcados fluorescentemente. Alternativamente, se puede usar un método de terminación de cadena, por ejemplo secuenciación de Sanger o secuenciación de terminación de tinte. En cualquier caso, se pueden contemplar estos y otros métodos, siempre y cuando el orden de los nucleótidos de un modelo de ADN se pueda determinar por sintetización ADN con un cebador de secuenciación y detectando nucleótidos incorporados y/o fragmentos sintetizados.

60 [0020] "Secuenciando" se refiere a determinar a la orden de nucleótidos (secuencias de base) en una muestra de ácido nucleico, por ejemplo ADN o ARN. Muchas técnicas están disponibles tal como la secuenciación de Sanger y tecnologías de secuenciación de alto rendimiento (HTS). La secuenciación de Sanger puede implicar
65 secuenciación mediante la detección a través de electroforesis (capilar), donde pueden analizarse hasta 384 capilares en secuencia de una sola vez. La secuenciación de alto rendimiento implica la secuenciación paralela de millares o millones de secuencias o más a la vez. HTS se puede definir como secuenciación de próxima

generación, es decir técnicas basadas en pirosecuenciación de fase sólida o como secuenciación de próxima generación basada en secuenciación en tiempo real de nucleótido único (SMRT). Están disponibles tecnologías HTS tales como las ofrecidas por Roche, Illumina y Applied Biosystems (Life Technologies). Además se describen tecnologías de secuenciación de alto rendimiento y/o disponibles por Helicos, Pacific Biosciences, Complete Genomics, Ion Torrent Systems, Oxford Nanopore Technologies, Nabsys, ZS Genetics, GnuBio. Cada una de estas tecnologías de secuenciación tienen su propia forma de preparar muestras antes del paso de secuenciación real. Estos pasos se pueden incluir en el método de secuenciación de alto rendimiento. En casos determinados, se pueden integrar pasos particulares para el paso de secuenciación en el protocolo de preparación de la muestra antes del paso de secuenciación real por cuestiones de eficiencia o economía. Por ejemplo, los adaptadores que se ligan a fragmentos pueden contener secciones que se pueden utilizar en los pasos de secuenciación posterior (denominados adaptadores secuenciadores). O bien los cebadores que se utilizan para amplificar un subconjunto de fragmentos antes de la secuenciación pueden contener partes dentro de su secuencia que introducen secciones que pueden utilizarse más tarde en el paso de secuenciación, por ejemplo, por introducción a través de un paso de amplificación o un adaptador de secuenciación o una fracción de captura en un amplicón que se puede utilizar en un paso de secuenciación posterior. Dependiendo también de la tecnología de secuenciación utilizada, pueden omitirse los pasos de amplificación.

[0021] Un "adaptador" es una molécula de ADN bicatenario corto con un número limitado de pares de bases, por ejemplo aproximadamente 10 a aproximadamente 100 pares de bases de longitud, que están diseñados de manera que se pueden ligar a los extremos de fragmentos de ADN o amplicones. Los adaptadores están generalmente compuestos por dos oligonucleótidos sintéticos que tienen secuencias de nucleótidos que son al menos parcialmente complementarias entre sí. Un adaptador puede tener extremos romos o tener extremos escalonados, o un extremo romo y un extremo escalonado. Un extremo escalonado es un saliente 3' o 5'. Cuando se mezclan los dos oligonucleótidos sintéticos en la solución bajo condiciones apropiadas, se reconocerán entre sí formando una estructura bicatenaria. Después del reconocimiento, un extremo de la molécula adaptadora se puede diseñar de manera que sea compatible con el extremo de un fragmento de restricción y se pueda ligar a ellos; el otro extremo del adaptador se puede diseñar de modo que este no pueda ser ligado, pero no tiene por qué ser el caso, por ejemplo cuando un adaptador va a ser ligado en medio de fragmentos de ADN. En determinados casos los adaptadores se pueden ligar a fragmentos para proveer un punto de partida para manipulación posterior del fragmento ligado a adaptador, por ejemplo para amplificación o secuenciación. En este último caso, los denominados adaptadores de secuenciación se pueden ligar a los fragmentos.

Descripción detallada de la invención

[0022] En un primer aspecto de la invención, tal y como se define por las reivindicaciones 1-11, se proporciona un uso de una combinación de al menos dos identificadores de secuencia de nucleótidos en la preparación de una muestra de ADN para secuenciación de alto rendimiento. Por consiguiente, se proporciona un método que comprende un paso donde una combinación de al menos dos identificadores de secuencia de nucleótidos se utiliza en la preparación de un ADN de muestra para secuenciación de alto rendimiento. Con la preparación de un ADN de muestra según esto se entiende que un ADN de muestra se prepara de manera que al menos dos NSIs se incluyen en el ADN de muestra, es decir al menos dos NSIs se incluyen por ejemplo en un amplicón y/o un fragmento de ADN ligado a un adaptador o amplicón de lo mismo. Al menos dos NSIs se incluyen así en una secuencia de nucleótidos del ADN de muestra de manera que una única molécula polinucleótida comprende al menos dos NSIs. La combinación de los NSIs sirve como identificadores únicos (código de barras de división) para el ADN de muestra.

[0023] Desde una perspectiva de diseño, no hay limitación práctica del número de identificadores de secuencia de nucleótidos que se pueden usar. Por ejemplo, un nucleótido puede ya servir como un identificador de secuencia de nucleótidos. Por lo tanto, se pueden diseñar 4 identificadores de secuencia de nucleótidos diferentes: A, G, C o T. Las secuencias que flanquean tal identificador de nucleótido único pueden servir como guía de la identificación del NSI. Mediante el aumento del tamaño de un código de barras, el número de posibilidades aumentan. Tres bases de ADN permiten 64 secuencias posibles 3-mer (4^3), 256 posibles 4-mer ($=4^4$), 1024 posibles 5-mer ($=4^5$), y 4096 posibles 7-mer ($=4^6$) etc. En la práctica, puede no obstante preferirse la selección de un subconjunto de estas secuencias para evitar el uso en el mismo experimento de identificadores de secuencia de nucleótidos que difieren por solo una base (tal como por ejemplo GATC y GATT en el caso de 4-mer), como esto podría suponer un atribución incorrecta en caso de una base amplificación- o error de secuenciación. De forma similar, se puede preferir evitar el uso de identificadores de secuencia de nucleótidos que tienen dos bases consecutivas idénticas (por ejemplo AATGC teniendo dos consecutivas A's en caso de un 5-mer) porque determinadas plataformas NGS tienen índices de error más altos para las llamadas secuencias "homopoliméricas". A pesar de tales criterios de selección, en general no hay escasez de identificadores de secuencias de nucleótidos adecuados ya que el incremento de su longitud con una base crea un pliegue cuaternario partiendo de número más alto de donde seleccionar.

[0024] Así, cuando por ejemplo en un método de secuenciación de alto rendimiento de una muestra preparada de ADN se determinan las dos secuencias de NSI, la combinación de los dos NSIs determina el origen de la muestra preparada de ADN. De esta manera, el número de NSIs, y así por ejemplo el número de diferentes

cebadores y/o adaptadores que tiene que usarse se pueden reducir considerablemente. Por ejemplo, para 100 muestras, actualmente 100 NSIs se utilizan, por ejemplo, se combinan 100 cebadores diferentes en adelante comprendiendo el NSIs con un cebador inverso. Según la invención, mediante el uso de un código de barras de división, bastarían 10 NSIs, y así se pueden usar 10 cebadores directos diferentes con 10 cebadores inversos diferentes de los cuales se pueden hacer 100 combinaciones únicas. Así, el número total de cebadores que tiene que usarse se reduce considerablemente, reduciéndose el número de 101 cebadores a 20 cebadores. Por lo tanto, reduciendo la complejidad de la preparación de la muestra flujo de trabajo, aumentando la probabilidad de la representación equivalente de muestras, reduciendo la carga de trabajo y la capacidad de memoria requerida y reduciendo los costes experimentales.

[0025] En otra forma de realización, el uso se proporciona de una combinación de al menos dos NSIs en la preparación de un ADN de muestra para secuenciación de alto rendimiento, donde en la secuenciación de alto rendimiento se utilizan una pluralidad de ADN de muestra preparada, donde cada preparación de un ADN de muestra comprende una combinación única de al menos dos NSIs donde un primer NSI se selecciona de un grupo de NSIs y un segundo NSI se selecciona del grupo de NSIs.

[0026] El grupo de NSIs utilizado comprende todos los NSIs. Para cada muestra de ADN, se selecciona un identificador de secuencia de nucleótidos del grupo. Esto significa que para un ADN de muestra para al menos dos NSIs, el mismo NSI se puede seleccionar en la combinación de NSIs. Además, para un ADN de muestra para al menos dos NSIs, se pueden seleccionar diferentes NSIs en la combinación de NSIs. Siempre y cuando la combinación de NSIs sea única para cada muestra ADN, se puede utilizar tal combinación. El grupo de NSIs también puede comprender al menos dos subgrupos de NSIs, donde cada primer y segundo NSI se puede seleccionar de un subgrupo diferente. Además del primer y segundo NSIs, se pueden utilizar otros NSIs seleccionados del grupo de NSIs. Un grupo de NSIs puede comprender al menos 4, 10, 100 o 1000 NSIs.

[0027] Se entiende que donde se dispone en toda la divulgación de un grupo de NSIs, el grupo puede estar virtualmente dispuesto, es decir no físico. El grupo puede por ejemplo estar dispuesto *in silico* y/o *in physico*. Por ejemplo, un grupo de NSIs se puede disponer como una lista de secuencias. Los NSIs se pueden seleccionar de esta lista y utilizarse para diseñar *in silico* un cebador y/o adaptador. Se entiende así que los pasos posteriores de proporcionar un grupo de NSIs y proporcionar un adaptador pueden comprender proporcionar un grupo de NSIs *in silico*, seleccionar de este grupo un NSI *in silico*, diseñar un adaptador *in silico*, y a continuación proporcionar *in physico* el adaptador comprendiendo el NSI. Alternativamente, el NSIs también se puede proporcionar *in physico* y utilizarse directamente, por ejemplo, en un escenario donde se utiliza un adaptador consistente en un identificador de secuencia de nucleótidos. O por ejemplo, los NSIs se pueden disponer *in physico* y enlazarse (por ejemplo ligados o de otra manera) a otras cadenas de nucleótido de manera que se genere un adaptador y/o cebador de amplificación, proporcionando así un adaptador y/o cebador de amplificación que incluye un NSI.

[0028] Las razones de la invención es que se crea un gran número de combinaciones únicas mediante el uso de al menos dos identificadores de secuencia de nucleótidos diferentes que se incorporan en cada muestra y aprovechando la potencia de multiplicación para reducir costes iniciales de reactivo. Esto se muestra en la tabla 1 para varias situaciones matemáticamente óptimas para dos NSIs.

Tabla 1. Combinaciones óptimas de NSIs con dos NSIs.

# de combinaciones únicas (A)	# de diferentes NSIs (B)	# de diferentes NSIs (C)	# de diferentes NSIs, B+C (D)	# de códigos de barras guardados (A-D)
100	10	10	20	80
1,024	32	32	64	960
4,096	64	64	128	3,968
16,384	128	128	256	16,128

Como se ha dicho, el concepto no está limitado a los ejemplos mostrados en la tabla 2 sino que también se pueden considerar otras combinaciones. Por ejemplo, se pueden seleccionar combinaciones menos óptimas. Combinaciones que pueden implicar más de dos NSIs se pueden seleccionar también. Por ejemplo, con 10 NSIs, y combinando cuatro NSIs, $10 \times 10 \times 10 \times 10 = 10,000$ combinaciones únicas son posibles. En el diseño y/o estrategia de combinación utilizados, también se toman en cuenta consideraciones prácticas, por ejemplo, consideraciones prácticas relacionadas con la preparación del ADN de muestra.

[0029] La muestra preparada de ADN es un ADN de muestra que ha sido sometida a tratamiento por el que al menos dos identificadores de secuencia de nucleótidos se incluyen en el ADN, es decir, al menos dos NSIs se incluyen en una molécula de ADN comprendiendo los dos NSIs y una secuencia de ADN del ADN de muestra. La molécula de ADN puede ser una molécula de ADN bicatenario o una única molécula de ADN monocatenario.

Se entiende que la muestra preparada de ADN puede comprender una pluralidad de diferentes moléculas de ADN, comprendiendo cada molécula de ADN la combinación única de NSIs de manera que cada molécula de ADN se puede asignar a una muestra ADN de donde se origina. En el método de secuenciación de alto rendimiento, de cada una de las moléculas de ADN de la pluralidad de moléculas de ADN se pueden determinar una secuencia con secuenciación de alto rendimiento.

[0030] En una forma de realización, la muestra preparada de ADN comprende un amplicón. Por ejemplo al menos dos NSIs se incluyen en cebadores utilizados para preparar un amplicón. El amplicón comprende así al menos dos NSIs de al menos dos cebadores diferentes utilizados. Los amplicones se pueden preparar por ejemplo en una reacción por PCR. Un amplicón también puede obtenerse a partir de un PCR anidado, es decir preparando un primer amplicón con un primer conjunto de cebadores en una primera reacción por PCR seguido de una segunda reacción por PCR con un segundo conjunto de cebadores, estos cebadores son diferentes desde el primer conjunto de cebadores y cuyos cebadores amplifican el amplicón de la primera reacción por PCR. Un PCR anidado puede por ejemplo realizarse para amplificar secuencias de ADN que tienen una concentración muy baja, o puede por ejemplo servir para proporcionar un identificador adicional o identificadores adicionales.

[0031] En una forma de realización, la muestra preparada de ADN comprende un fragmento de ADN ligado a un adaptador, el ADN de muestra se fragmenta y adaptadores se ligan a los fragmentos de ADN. Al menos dos adaptadores se ligan a un fragmento de ADN, y cada uno de al menos dos adaptadores comprende un NSI.

[0032] En una forma de realización, la muestra preparada de ADN comprende un amplicón y un fragmento ligado a un adaptador.

[0033] En una forma de realización, la muestra preparada de ADN comprende un fragmento ligado a un adaptador amplificado y/o un amplicón ligado a un adaptador. Por ejemplo, un ADN de muestra puede someterse posteriormente a fragmentación, ligamiento a un adaptador, y amplificación. Por el contrario, un ADN de muestra puede por ejemplo posteriormente someterse a amplificación, fragmentación, y ligamiento a un adaptador. En esta forma de realización, al menos dos de todos los cebadores de amplificación y/o adaptadores utilizados en la preparación de la muestra ADN comprenden un NSI.

[0034] En una forma de realización, el método según la divulgación comprende:

- a) proporcionar adaptadores y/o cebadores de amplificación, donde al menos un primer adaptador o cebador de amplificación comprendiendo un primer identificador de secuencia de nucleótidos seleccionado del grupo de identificadores de secuencia de nucleótidos y un segundo adaptador o cebador de amplificación se proporciona comprendiendo un segundo identificador de secuencia de nucleótidos seleccionado del grupo de identificadores de secuencia de nucleótidos, y donde opcionalmente, se proporcionan otro adaptador o cebadores de amplificación comprendiendo otro identificador de secuencia de nucleótidos seleccionado del grupo de identificadores de secuencia de nucleótidos;
- b) proporcionar una pluralidad de muestras de ADN;
- c) realizar un ligamiento y/o reacciones de amplificación en las muestras de ADN utilizando los adaptadores y/o cebadores de amplificación, para proporcionar un ADN de muestra ligada y/o amplificada comprendiendo el primer, segundo y opcionalmente otros identificadores de secuencia de nucleótidos;
- d) determinar al menos la secuencia del primer, segundo y otros identificadores de secuencia de nucleótidos utilizando secuenciación de alto rendimiento;
- e) determinar el origen de muestra de las muestras de ADN ligado y/o amplificado.

[0035] En tanto en cuanto en la preparación de un ADN de muestra se usen al menos dos NSIs los cuales acaban en una molécula de ADN única comprendiendo los dos NSIs y una secuencia de ADN del ADN de muestra, tal método de preparación de la muestra se puede contemplar según la divulgación. Así, se pueden combinar pasos de ligamiento a un adaptador diferentes separados y/o pasos de amplificación separados para incorporar al menos dos NSIs en una muestra preparada de ADN. La preparación de un ADN de muestra también puede comprender pasos que no implican la adición de NSIs.

[0036] Se entiende que en la preparación del ADN de muestra, se pueden realizar pasos en el paso de secuenciación posterior también se pueden incluir y viceversa. Se pueden incluir también pasos requeridos para el paso de secuenciación posterior en la preparación del ADN de muestra. Por ejemplo, en una reacción de secuenciación, se pueden utilizar cebadores de secuenciación para enlazar con sitios de unión a cebador de secuenciación presentes en un modelo. Por lo tanto, los adaptadores y/o cebadores de amplificación utilizados en la preparación de un ADN de muestra pueden comprender sitios de unión a cebador de secuenciación por añadidura. Alternativamente, se puede añadir secuenciación de sitios de unión al cebador en el método de secuenciación de alto rendimiento proporcionando estas secuencias adicionales en una fase posterior.

[0037] Además de la secuenciación de al menos dos NSIs en el método de secuenciación de alto rendimiento, también se pueden secuenciar secuencias del ADN de muestra. Tales secuencias también se pueden referir

como secuencias internas, ya que son secuencias las cuales son capturadas y/o amplificadas con el método de preparación de ADN de muestra y pueden representar secuencias desconocidas de ADNs de muestra que son de interés. Estas secuencias internas también se pueden ordenar junto con los NSIs, y así el origen de la muestra que comprende estas secuencias internas se puede determinar por la combinación de los NSIs. De esta manera, por ejemplo, los polimorfismos se pueden detectar entre muestras diferentes, tales como pequeños polimorfismos de nucleótido, deleciones, inserciones, etc., por comparación de las secuencias internas y/o comparación de las secuencias internas con una secuencia de referencia. Además, estas secuencias internas también pueden ayudar a la atribución de diferentes lecturas de una muestra preparada de ADN, por ejemplo, en el escenario como se describe a continuación donde diferentes lecturas de secuenciación se utilizan para construir un cóntigo.

[0038] La secuenciación de alto rendimiento implica la secuenciación paralela de millares o millones o más de secuencias a la vez. Sea cual sea el método de secuenciación de alto rendimiento utilizado, los NSIs de los ADNs de muestra preparada son determinados de manera que cada combinación de NSIs se puedan asignar a una muestra preparada de ADN. Por ejemplo, se pueden construir secuencias contiguas alineando y combinando diferentes lecturas de secuencias por las que pueden ser acoplados al menos dos NSIs, y así se pueden asignar a una molécula de ADN única comprendiendo al menos dos NSIs. También, dos reacciones de secuenciación se pueden realizar en las cadenas complementarias de una muestra preparada de ADN que se prefiere en caso de que las secuencias internas del fragmento se requieran de ambos extremos, cuando por ejemplo la secuencia interna es relativamente grande. Cuando las dos lecturas de secuenciación se pueden asignar a la muestra preparada de ADN, comprendiendo al menos dos NSIs y una secuencia de ADN del ADN de muestra, pueden contemplarse tales métodos de secuenciación de alto rendimiento. También, se pueden determinar al menos dos NSIs de una única lectura de secuenciación en un método de secuenciación de alto rendimiento.

[0039] Por ejemplo, en un denominado método de secuenciación de extremo emparejado, en una primera reacción de secuenciación de una primera secuencia se puede determinar incluyendo un NSI que utiliza una de las cadenas como modelo. Después de la primera reacción de secuenciación, una cadena complementaria se puede generar de la cadena primeramente utilizada como un modelo. Una segunda reacción de secuenciación puede realizarse posteriormente utilizando esta cadena recién generada como un segundo modelo. Así, se utilizan en este método dos cadenas de modelo de ADN. Por ejemplo, la estructura de una primera cadena de modelo de secuenciación puede ser - sitio 1 de unión a cebador de secuencia 3' - NSI 1 - secuencia interna - un complemento inverso NSI 2 - sitio de unión a cebador de secuencia de complemento inverso 2-5'. Después de la primera reacción de secuenciación, se puede generar un complemento inverso de la primera cadena de modelo de secuenciación, el cual puede utilizarse posteriormente en una segunda reacción de secuenciación. La segunda reacción de secuenciación tiene por tanto el siguiente modelo: - sitio 2 de unión a cebador de secuencia 3' - NSI 2 - secuencia interna de complemento inverso - complemento inverso NSI 1 - sitio 1-5' de unión a cebador de secuencia de complemento inverso. Debido a que ambas lecturas de secuencia colocalizan (por ejemplo en el mismo depósito, el mismo reborde), las dos lecturas de secuencia que comprenden NSI 1 y NSI 2 se pueden asignar a la misma muestra preparada de ADN y utilizarse para identificar la muestra preparada de ADN. Tales escenarios de preparación de la muestra y secuenciación posterior se representan en las figuras 5, 7 y 10.

[0040] En una forma de realización, en el método de secuenciación de alto rendimiento, se utiliza un único modelo de ADN de la muestra preparada de ADN. Con un único modelo de ADN según la invención tal y como se define por las reivindicaciones 1-11 se entiende una única molécula de ADN monocatenaria la cual comprende al menos dos identificadores de secuencia de nucleótidos. Se entiende que un único modelo de ADN de una muestra preparada de ADN puede comprender una pluralidad de moléculas de modelo de ADN único, que comprende por ejemplo diferentes secuencias internas derivadas de un ADN de muestra donde cada secuencia interna diferente se enlaza con la combinación única de NSIs. Cuando por ejemplo, la muestra preparada de ADN es un amplicón (o una pluralidad de amplicones), el amplicón comprende dos cadenas de ADN. En esta forma de realización, solo una de las cadenas de un amplicón se usa en la reacción de secuenciación para determinar las secuencias NSI. De esta manera, el origen de una muestra preparada de ADN se puede determinar sin requerir la construcción de un cóntigo y/o sin requerir una secuencia derivada de otro modelo de ADN. Los NSIs pueden flanquear la secuencia interna de la muestra preparada de ADN. Los NSIs también pueden estar en un lado de (una secuencia interna de) una muestra preparada de ADN. En estos escenarios, una única molécula de modelo de ADN de una muestra preparada de ADN puede tener la siguiente estructura: 3'- sitio de unión a cebador de secuencia - NSI 1 - secuencia interna - NSI 2 -5' o 3'- sitio de unión a cebador de secuencia - NSI 1 - NSI 2 - secuencia interna -5'. La muestra preparada de ADN puede comprender secuencias adicionales. El orden del sitio de unión al cebador de secuencia, los NSIs, y la secuencia interna es el objetivo en esta representación estructural. El sitio de unión al cebador de secuencia se puede incorporar durante la preparación de un ADN de muestra y/o se puede incorporar en el método de secuenciación de alto rendimiento. La longitud y/o calidad de secuencias generadas en la secuenciación de alto rendimiento pueden ser limitadas. Por otro lado, la longitud de secuenciación se puede restringir de manera que no puedan determinarse ambos NSIs que flanquean la secuencia interna. Puede ser ventajoso tener el NSIs en un lado de la secuencia interna, es decir la parte del modelo de ADN único que está secuenciado primero, de manera que ambas secuencias se pueden determinar en una única lectura. Los NSIs también pueden estar en ambos lados

de la secuencia interna.

[0041] En una forma de realización, el modelo de ADN único utilizado puede comprender dos sitios de unión a cebador de secuenciación, donde cada sitio de unión de cebador de secuenciación está localizado 3' de un identificador de secuencia de nucleótidos diferente. En general, un único modelo de ADN puede comprender una primera sección y una segunda sección, con una secuencia interna derivada de la muestra ADN entremedias. La primera sección comprende un sitio de unión de cebador de secuenciación con 5' localizado en un NSI y en otros NSIs opcionales, y la segunda sección comprende un segundo sitio de unión a cebador de secuenciación con 5' localizado en el mismo NSI y otros NSIs opcionales. En este escenario, un único modelo de ADN de una muestra preparada de ADN puede tener la siguiente estructura: sitio de unión a cebador de secuencia 3' 1 - NSI 1 - secuencia interna - sitio de unión a cebador de secuencia 1 NSI 2 -5'. La muestra preparada de ADN puede comprender secuencias adicionales. El orden de los sitios de unión a cebador de secuencia, los NSIs, y la secuencia interna es el objetivo en esta representación estructural. Así, el sitio de unión a cebador de secuenciación se puede localizar directamente 3' de unas secuencias pero también pueden estar presentes identificadores de secuencia de nucleótidos adicionales entre un sitio de unión de cebador de secuenciación y un identificador de secuencia de nucleótidos. En este escenario, de un único modelo, se pueden realizar consecutivamente dos reacciones de secuenciación diferentes en el método de secuenciación de alto rendimiento. Una reacción de secuenciación determinará uno (o más) NSI, y una segunda reacción de secuenciación un segundo NSI (o más). Las dos reacciones de secuenciación de esta forma de realización realizadas en un método de secuenciación de alto rendimiento utilizando el mismo modelo se pueden referir como secuenciación "lectura única etiquetado doble" de ahora en adelante. Tales escenarios de lectura única etiquetado doble se representan en las Figuras 6, 8 y 9.

[0042] Durante la preparación de la muestra y/o durante el método de secuenciación de alto rendimiento utilizado, las muestras diferentes o parte de las muestras diferentes se pueden agrupar de manera que pasos que se pueden realizar simultáneamente se puedan realizar simultáneamente. El origen de muestra puede todavía determinarse como ninguna de las muestras o parte de diferentes muestras se agrupan por lo que ya no se puede determinar el origen de la muestra. Por ejemplo, en un escenario donde en la preparación de un ADN de muestra, los NSIs se agregan en diferentes pasos, puede ser beneficioso agrupar al menos parte de las muestras bajo la preparación después de tal paso. Por ejemplo, en un escenario donde se usan 6 NSIs durante 36 muestras, cada ADN de muestra se somete antes a un paso que añade uno de 6 diferente NSIs (A-F). Se pueden agrupar muestras que comprenden identificadores únicos (A1, B1, C1, D1, E1, F1) y cada grupo puede ahora sufrir la adición de uno de los 6 identificadores, cada grupo de seis con un segundo identificador único (A2, B2, C2, D2, E2, o F2), por lo que A1 y A2 etc. pueden ser o no ser idénticos. Finalmente, una vez todos los ADNs de muestra se han preparado, los ADN de muestra preparados se pueden agrupar (parcialmente o como conjunto) en cualquier forma posible, debido a que se han incorporado ahora al menos dos identificadores únicos.

[0043] Son posibles diferentes métodos para identificar el origen de muestra de una muestra preparada de ADN que comprende una combinación única de al menos dos NSIs.

[0044] En una forma de realización de la invención, se proporciona un método para identificar el origen de la muestra de amplicones de una pluralidad de ADNs de muestra que incluyen las etapas de:

- a) proporcionar una pluralidad de ADNs de muestra;
- b) proporcionar un grupo de identificadores de secuencia de nucleótidos;
- c) proporcionar los primeros cebadores de amplificación, cada primer cebador comprendiendo un primer identificador de secuencia de nucleótidos seleccionado del grupo de identificadores de secuencia de nucleótidos;
- d) proporcionar los segundos cebadores de amplificación, cada segundo cebador comprendiendo un segundo identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- e) amplificar cada ADN de muestra con un par único de un primer y segundo cebador de amplificación para dar amplicones;
- f) opcionalmente, agrupar al menos parte de los amplicones;
- g) determinar la secuencia de la primera secuencia de identificador y la segunda secuencia de identificador de los amplicones utilizando secuenciación de alto rendimiento;
- h) determinar el origen de muestra de los amplicones.

Un esquema de la preparación de la muestra de este método de esta forma de realización se muestra en la figura 1. En este método, dos NSIs se incluyen en un primer y segundo cebador de amplificación. El amplicón comprende los dos NSIs. Los cebadores de amplificación se pueden diseñar para amplificar una secuencia interna particular que es de interés. Mediante la secuenciación de al menos parte de la secuencia interna y los dos NSIs del amplicón, cada secuencia interna secuenciada (parcial) se puede asignar a un ADN de muestra donde se origina. Alternativamente, los cebadores de amplificación se pueden diseñar de manera que sean selectivos hacia un sitio de unión de cebador particular. Mediante la determinación de solo la secuencia de los

dos NSIs se determina la presencia o ausencia de un amplicón para una muestra particular de ADN. Los cebadores de amplificación utilizados pueden tener extremos fosforilados 5' que son adecuados para el ligamiento de adaptadores que se pueden utilizar en un método de secuenciación de alto rendimiento posterior. Alternativamente, los amplicones se pueden fosforilar si fuera necesario.

5

[0045] En una forma de realización, se proporciona un método para identificar el origen de la muestra de fragmentos de ADN ligados a un adaptador de una pluralidad de ADNs de muestra que incluye las etapas de:

10

- a) proporcionar una pluralidad de ADNs de muestra;
- b) proporcionar un grupo de identificadores de secuencias de nucleótidos;
- c) proporcionar los primeros adaptadores, cada primer adaptador comprendiendo un primer identificador de secuencia de nucleótidos seleccionado del grupo de identificadores de secuencia de nucleótidos;
- d) proporcionar los segundos adaptadores, cada segundo adaptador comprendiendo un segundo identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- e) fragmentar cada ADN de muestra;
- f) ligar un par único de un primer y segundo adaptador para cada muestra fragmentada de ADN para dar fragmentos de ADN ligados a un adaptador;
- g) opcionalmente, agrupar al menos parte de los fragmentos de ADN ligados a un adaptador;
- h) determinar la secuencia de la primera secuencia de identificador y la segunda secuencia de identificador de los fragmentos de ADN ligados a un adaptador utilizando secuenciación de alto rendimiento;
- i) determinar el origen de muestra de los fragmentos de ADN ligados a un adaptador.

15

20

25

Un esquema que muestra la preparación de un método de esta forma de realización se muestra en la figura 2. En este método, dos adaptadores se ligan a un fragmento de ADN. Como se muestra en la figura 2, los adaptadores se pueden ligar a cualquier sitio de un fragmento. Esto es en particular adecuado para métodos de secuenciación de alto rendimiento donde se utilizan tales estrategias de ligamiento a un adaptador. Son posibles muchas estrategias para enlazar dos adaptadores diferentes a fragmentos. Por ejemplo, primero el ADN se puede fragmentar con dos enzimas de restricción con dos sitios de reconocimiento diferente. Esto produce fragmentos de ADN que tienen extremos que son el resultado de una enzima de restricción, pero también en fragmentos que tienen extremos que son el resultado de las dos enzimas de restricción. Cuando dos adaptadores diferentes se diseñan pudiéndose ligar a los extremos de restricción específicos de cada extremo, se pueden formar fragmentos ligados a un adaptador que comprenden los dos adaptadores diferentes. Además, se forman fragmentos ligados a un adaptador que comprenden dos del mismo adaptador. Alternativamente, el ADN puede por ejemplo fragmentarse con una única enzima de restricción seguida del ligamiento de adaptadores compatibles al mismo. A continuación, el fragmento ligado a un adaptador se puede fragmentar otra vez pero ahora mediante por ejemplo sonificación. Los extremos del fragmento se pulen a continuación y se ligan adaptadores a los extremos romos. El resultado es una mezcla de fragmentos ligados a un adaptador incluyendo fragmentos ligados a un adaptador comprendiendo el adaptador compatible de enzima de restricción y el adaptador compatible de extremo romo. En ambos escenarios, se forman fragmentos ligados a un adaptador que pueden comprender los dos NSIs. Solo se puede determinar el origen de la muestra de los fragmentos ligados a un adaptador que comprenden los dos adaptadores diferentes, puesto que es una combinación de los dos NSIs que se requieren para determinar el origen de muestra.

30

35

40

45

[0046] En una forma de realización, se proporciona un método para identificar el origen de muestra de los amplicones ligados a un adaptador comprendiendo las etapas de:

50

- a) proporcionar una pluralidad de ADNs de muestra;
- b) proporcionar un grupo de identificadores de secuencias de nucleótidos;
- c) proporcionar los primeros cebadores de amplificación;
- d) provisión de los segundos cebadores de amplificación;
- e) amplificar un ADN de muestra con un par de un primer y segundo cebador de amplificación para dar amplicones;
- f) opcionalmente, agrupar al menos parte de los amplicones de los ADNs de muestra cada uno amplificado con un par de cebadores diferentes;
- g) opcionalmente, fragmentar los amplicones;
- h) proporcionar los primeros adaptadores, cada primer adaptador comprendiendo un primer identificador de secuencia de nucleótidos seleccionado del grupo de identificadores de secuencia de nucleótidos;
- i) proporcionar segundos adaptadores, cada segundo adaptador comprendiendo un segundo identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- j) opcionalmente, proporcionar otros adaptadores, cada adaptador comprendiendo otros identificadores de secuencia de nucleótidos seleccionados del grupo de identificadores de secuencia de nucleótidos;
- k) ligar el primer adaptador a los amplicones (fragmentados);
- l) opcionalmente, agrupar al menos parte de los amplicones ligados a un adaptador del paso k);
- m) repetir el paso de ligamiento con el segundo y los otros adaptadores, donde cada paso de ligamiento

55

60

65

está seguido de, opcionalmente, la agrupación de al menos parte de los amplicones ligados a adaptador obtenidos;

- n) determinar la secuencia del primer, segundo, y las otras secuencias opcionales de identificador de los amplicones ligados a un adaptador obtenidos en el paso m) utilizando secuenciación de alto rendimiento;
- o) determinar el origen de la muestra de los amplicones ligados a un adaptador.

En esta forma de realización, cada ADN de muestra está sujeto al menos a una reacción de amplificación de PCR. Se pueden realizar reacciones PCR múltiples en cada muestra, por ejemplo, para secuencias de diferente objetivo. Estos amplicones diferentes de cada muestra pueden agruparse opcionalmente. Los NSIs en esta forma de realización se pueden añadir en pasos separados y después de cada paso, al menos parte de los amplicones ligados a un adaptador se pueden agrupar.

[0047] En otras formas de realización, se combinan el ligamiento a adaptadores y la amplificación, donde el origen de la muestra de una muestra preparada de ADN se puede determinar mediante la determinación de 2-4 NSIs de una muestra preparada de ADN.

[0048] En una forma de realización, se proporciona un método para identificar el origen de muestra de fragmentos de ADN ligados a un adaptador amplificados de una pluralidad de ADNs de muestra que incluyen las etapas de:

- a) proporcionar una pluralidad de ADNs de muestra;
- b) proporcionar un grupo de identificadores de secuencia de nucleótidos;
- c) proporcionar los primeros adaptadores, cada primer adaptador comprendiendo un primer identificador de secuencia de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- d) proporcionar segundos adaptadores, cada segundo adaptador comprendiendo opcionalmente un segundo identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencia de nucleótidos;
- e) fragmentar cada ADN de muestra;
- f) ligar al menos un primer adaptador y opcionalmente un segundo adaptador a la muestra fragmentada de ADN para dar fragmentos de ADN ligado a un adaptador;
- g) opcionalmente, agrupar al menos parte de los fragmentos de ADN ligado a un adaptador;
- h) proporcionar los primeros cebadores de amplificación, cada primer cebador comprendiendo un tercer identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- i) opcionalmente, proporcionar segundos cebadores de amplificación, cada segundo cebador opcionalmente comprendiendo un cuarto identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- j) amplificar los fragmentos de ADN ligados a un adaptador con un primer cebador de amplificación y con opcionalmente un segundo cebador de amplificación, donde la combinación de un primero y un segundo opcional, tercero y cuarto opcional NSI es única para cada muestra, para dar fragmentos de ADN ligados a un adaptador amplificados;
- k) opcionalmente, fusionar al menos parte de los fragmentos de ADN ligados a un adaptador amplificados;
- l) determinar la secuencia de la primera, segunda opcional, tercera y cuarta opcional secuencia de identificador de los fragmentos de ADN ligado a un adaptador amplificado utilizando secuenciación de alto rendimiento;
- m) determinar el origen de la muestra de los fragmentos de ADN ligado a un adaptador amplificado.

Un esquema de la preparación de la muestra de un método de esta forma de realización se muestra en la figura 3. En este escenario, los fragmentos ligados a un adaptador se pueden preparar como se ha descrito anteriormente. Alternativamente, solo un adaptador puede haber sido ligado a los fragmentos. El(Los) adaptador(es) puede(n) comprender además de un NSI, un sitio de unión de cebador que se puede utilizar en una amplificación posterior y preferiblemente un sitio de unión de cebador de secuenciación. En el escenario donde el mismo adaptador se utiliza en cada lado de un fragmento, el mismo cebador de amplificación se puede utilizar para amplificar el fragmento ligado a un adaptador. El sitio de unión de cebador también puede incluir (parte del) NSI. Alternativamente, (más) cebadores selectivos se pueden utilizar requiriendo (diferentes) nucleótidos adicionales complementarios a la secuencia interna del fragmento de ADN más allá de la secuencia de sitio de reconocimiento de restricción de la secuencia interna. El concepto de cebadores selectivos se describe bien por ejemplo en WO2006/137733, y el método de esta forma de realización puede implicar la preparación de ADN de muestra que usa tales cebadores selectivos. Por ejemplo, un cebador se diseña de manera que sea complementario a (parte de) la secuencia de adaptador y la secuencia interna del ADN de muestra a la que el adaptador se liga que comprende el sitio de reconocimiento de restricción, y un nucleótido adicional. El nucleótido adicional es el nucleótido selectivo que hace selectivo el cebador. De media, para uno de cuatro fragmentos de restricción ligados a un adaptador, el cebador selectivo puede enlazar y tener el extremo 3' extendido. El concepto de cebadores selectivos es bien conocido de AFLP (EP 534858) como un método de reducción de complejidad.

[0049] En cualquier caso, el resultado final es un fragmento ligado a un adaptador amplificado, que puede comprender al menos dos NSIs, o incluso 3 o 4 NSIs. Incluir más NSIs puede ser ventajoso ya que esto puede reducir el número de NSIs incluso más. Por ejemplo, donde de 10,000 muestras, se requieren 100 NSIs en una una combinación de dos NSIs (100x100), se requieren 10 NSIs en una una combinación de cuatro NSIs (10x10x10x10).

[0050] En una forma de realización, se proporciona un método para identificar el origen de muestra de amplicones ligados a un adaptador comprendiendo las etapas de:

- a) proporcionar una pluralidad de ADN de muestra;
- b) proporcionar un grupo de identificadores de secuencias de nucleótidos;
- c) proporcionar los primeros cebadores de amplificación, cada primer cebador comprendiendo un primer identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- d) proporcionar los segundos de cebadores de amplificación, cada segundo cebador opcionalmente comprendiendo un segundo identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- e) amplificar cada ADN de muestra con un par de un primer y segundo cebador de amplificación para dar amplicones;
- f) opcionalmente, agrupar al menos parte de los amplicones de ADN de muestra, cada uno amplificado con un par de cebadores diferente;
- g) opcionalmente, fragmentar los amplicones;
- h) proporcionar los primeros adaptadores, cada primer adaptador comprendiendo un tercer identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- i) opcionalmente, proporcionar segundos adaptadores, cada segundo adaptador comprendiendo opcionalmente un cuarto identificador de secuencias de nucleótidos seleccionado del grupo de identificadores de secuencias de nucleótidos;
- j) ligar al menos un primer adaptador y opcionalmente un segundo adaptador a los amplicones (fragmentados), donde la combinación de un primer, segundo opcional, tercero y cuarto opcional identificador de secuencias de nucleótidos es único para cada muestra, para dar amplicones ligados a un adaptador;
- k) opcionalmente, agrupar al menos parte de los amplicones ligados a un adaptador;
- l) determinar la secuencia de la primera, segunda opcional, tercera y cuarta opcional secuencias de identificador de los amplicones ligados a un adaptador utilizando secuenciación de alto rendimiento;
- m) determinar el origen de la muestra de los amplicones ligados a un adaptador.

Un esquema de la preparación de la muestra de un método de esta forma de realización se muestra en la figura 4. En este escenario, un ADN de muestra se puede someter a amplificación como se ha descrito anteriormente, donde ahora al menos uno de los cebadores comprende un NSI. El amplicón se puede utilizar directamente (con o sin pulido), y un adaptador se puede ligar al amplicón. En este escenario, el amplicón ligado a un adaptador comprenderá en cualquier extremo el mismo adaptador. Este amplicón ligado a un adaptador se puede someter a más fragmentación y ligamiento a un adaptador a un segundo adaptador para obtener amplicones ligados a un adaptador que tienen dos adaptadores diferentes. Alternativamente, un amplicón se puede someter a un paso de fragmentación que produce fragmentos con dos extremos diferentes que son compatibles con dos adaptadores diferentes. En cualquier caso, se forma un amplicón ligado a un adaptador comprendiendo 2-4 NSIs.

[0051] En una forma de realización, en un método según cualquiera de los métodos como se han descrito anteriormente, en el paso de determinar la secuencia de las secuencias de identificador utilizando secuenciación de alto rendimiento, la secuencia de las secuencias de identificador se determina de un único modelo de ADN de la muestra preparada de ADN. Se entiende que la muestra preparada de ADN en los métodos anteriores comprende amplicones, fragmentos ligados a un adaptador, amplicones ligados a un adaptador y/o fragmentos ligados a un adaptador amplificados.

[0052] En una forma de realización, en el modelo único de ADN, el primero, segundo y otras secuencias opcionales de identificador son al menos 3' o 5' de una secuencia interna. De la muestra preparada de ADN, donde las secuencias de identificador de secuencia se determinan de un modelo único de ADN de la muestra preparada de ADN, la combinación de secuencias de identificador requerida para únicamente identificar el origen de muestra es al menos 3' o 5' de la secuencia interna. Un modelo de secuenciación del ADN puede prepararse así, teniendo las secuencias de identificador flanqueadas por el sitio de unión al cebador de secuencia y la secuencia interna. Por ejemplo, el extremo 3' de tal modelo de secuenciación del ADN se puede representar por el siguiente esquema: '3-SEQ1-NSI4-NSI3-NSI2-NSI1-IS - (etc.). De esta manera, cuando la secuencia del modelo se determina, todos los identificadores de secuencia se determinan en primer lugar. Tales modelos de ADN se pueden generar solo mediante la adición de identificadores de secuencia en un extremo de la secuencia interna. Por ejemplo, mediante la adición de adaptadores a solo un extremo de un fragmento de ADN, y/o utilizando adaptadores asimétricos, o mediante el uso de conjuntos de cebadores de amplificación donde solo un cebador comprende un identificador de secuencia. Alternativamente, se pueden añadir adaptadores y/o

cebadores de amplificación para ambos extremos 5' y 3' del fragmento de ADN, de manera que se localizan todos los identificadores de secuencia en ambos extremos. Tal modelo de ADN se puede representar por el siguiente esquema '3-SEQ1-NSI4-NSI3-NSI2-NS11-IS-NSI1-NSI2-NSI3-NSI4-5'. Son también posibles las combinaciones de estas estrategias diferentes, teniendo en uno o más pasos separados un NSI añadido a ambos extremos de la IS y a solo un extremo. Tal modelo de ADN se puede representar por el siguiente esquema '3-SEQ1-NSI2-NSI1-IS-NSI1-5'. Mientras se utilice una estrategia donde se genera un modelo de ADN en el cual una combinación única de identificadores de secuencia esté flanqueada por el sitio de unión de cebador de secuencia y la secuencia interna, tal modelo de ADN bastará. Así, pueden utilizarse combinaciones diferentes de cebadores de amplificación y/o adaptadores con y sin identificadores de secuencia de nucleótidos. Por ejemplo, como se muestra en los ejemplos 3 y 4, cuando NSI2 y NSI4, que son opcionales, corresponderían a NSI1 y NS3, se generan modelos de ADN adecuados (es decir necesitaría todavía de un sitio de unión de cebador de secuencia, que puede añadirse en una preparación de la muestra a través de secuenciación de alto rendimiento). De forma similar, las secuencias NSI2 y NSI4 pueden no estar incluidas. También, el NSI2 puede no estar incluido, y el NSI4 corresponde a NS3, o NS12 puede corresponder a NSI1 y NSI4 no está incluido. En este escenario, solo un extremo del modelo de ADN puede comprender todos los NSIs. En otra forma de realización, cuando la secuencia de las secuencias de identificador se determina de un único modelo de ADN de la muestra preparada de ADN, el modelo de ADN único puede comprender dos sitios de unión al cebador de secuenciación, donde cada sitio de unión al cebador de secuenciación está localizado 3' de un identificador de secuencia de nucleótidos diferente, y donde en el método de secuenciación de alto rendimiento se realizan dos reacciones de secuenciación diferentes con dos cebadores de secuenciación de los dos sitios de unión a cebadores de secuenciación del modelo de ADN único. Los dos sitios de unión al cebador de secuenciación diferentes y que corresponden a 5' NSI o NSIs pueden estar flanqueando una secuencia interna de una muestra preparada de ADN. Una muestra preparada de ADN puede ser un amplicón, un fragmento ligado a un adaptador, amplicones ligados a un adaptador y/o un fragmento ligado a un adaptador amplificado como se ha descrito anteriormente.

[0053] En una forma de realización, el modelo de ADN único comprende dos sitios de unión al cebador de secuenciación, donde al menos un sitio de unión de cebador de secuenciación está localizado 3' de al menos dos identificadores de secuencias de nucleótidos, y donde se realizan dos reacciones de secuenciación diferentes en el método de secuenciación de alto rendimiento con dos cebadores de secuenciación de los dos sitios de unión al cebador de secuenciación del modelo de ADN único.

[0054] En una forma de realización, cuando la secuencia de las secuencias del identificador se determina de un único modelo de ADN de la muestra preparada de ADN, el modelo de ADN único puede comprender dos sitios de unión al cebador de secuenciación, donde cada sitio de unión al cebador de secuenciación está localizado 3' de uno o varios identificadores de secuencia de nucleótidos, y donde se realizan dos reacciones de secuenciación diferente en el método de secuenciación de alto rendimiento con dos cebadores de secuenciación de los dos sitios de unión al cebador de secuenciación del modelo de ADN único. Los dos sitios de unión al cebador de secuenciación diferentes y que corresponden a 5' NSI o NSIs pueden estar flanqueando una secuencia interna de una muestra preparada de ADN. Cada uno de uno o varios identificadores de secuencias de nucleótidos pueden ser el mismo. Tal configuración se puede representar de la siguiente manera: '-3'-SEQ1-NS1-NS2-IS-SEQ2-NS1-NS2-5'. De esta forma, mediante la secuenciación de un único modelo ambas combinaciones únicas de identificadores únicos se secuencian dos veces. Tal muestra preparada de ADN puede ser un amplicón, un fragmento ligado a un adaptador, amplicones ligados a un adaptador y/o un fragmento ligado a un adaptador amplificado como se ha descrito anteriormente.

[0055] En una forma de realización, cuando la secuencia de las secuencias de identificador se determina de un único modelo de ADN de la muestra preparada de ADN, el modelo de ADN único puede comprender dos sitios de unión al cebador de secuenciación, donde un sitio de unión al cebador de secuenciación está localizado 3' de dos o más identificadores de secuencias de nucleótidos, y donde el otro sitio de unión de cebador de secuenciación se puede localizar adyacente a la secuencia interna. En el método de secuenciación de alto rendimiento se realizan dos reacciones de secuenciación diferente con dos cebadores de secuenciación de los dos sitios de unión al cebador de secuenciación del modelo de ADN único. Los dos sitios de unión al cebador de secuenciación diferentes y correspondiendo pueden estar flanqueando una secuencia interna de una muestra preparada de ADN. Tal configuración se puede representar esquemáticamente de la siguiente manera: '-3'-SEQ1-IS-SEQ2-NS1-NS2-5'. De esta manera, mediante la secuenciación de un único modelo en un recorrido de secuenciación se puede determinar la secuencia interna y en el otro recorrido de secuenciación se puede determinar la combinación única de identificadores de secuencia.

60 Ejemplos

Ejemplo 1

[0056] A continuación se ejemplifican dos aplicaciones diferentes para la preparación de la muestra con códigos de barras de división,

- 1) amplificación de PCR con dos cebadores codificados por barras;
- 2) ligamiento de un adaptador a una muestra digerida con dos enzimas de restricción o una muestra digerida con una única enzima seguida de ligamiento a un adaptador 1, seguido de fragmentación y reducción de los extremos fragmentados, seguido del ligamiento a un adaptador 2.

5

Amplificación de PCR

[0057] Se proporciona una descripción de los elementos funcionales de un par de cebadores que contienen códigos de barras de división (código de barras 1 y código de barras 2), que se determinan bien por secuenciación de extremo emparejado (A) o bien por secuenciación de la misma cadena con dos eventos de cebado (B). Se proporciona una vista esquemática en la figura 11. Se observa en la figura 11A que la cola universal 1 (flecha grande) puede ser el sitio de cebador de secuencia 1 (es decir el sitio de cebador que se utiliza en la secuenciación) y la cola universal 2 (flecha rayada) puede ser el sitio de cebador de secuencia 1 (es decir el sitio de cebador que se utiliza en la secuenciación), ejemplos son los cebadores P5 y P7 respectivamente como se utilizan en la secuenciación Illumina GA de extremo pareado. En la figura 11B la cola universal 1 (flecha grande) puede ser el sitio de cebador de secuencia 1 (es decir el sitio de cebador que se utiliza en la secuenciación) y cola universal 2 (flecha rayada) puede ser el sitio de cebador de secuencia 2 (es decir el sitio de cebador que se utiliza en la secuenciación), ejemplos son los cebadores P5 y P7 respectivamente como se utilizan en Illumina GA con dos eventos de cebador de la misma cadena. El concepto se puede utilizar en cualquier método implicando amplificación con un par de cebadores. Ejemplos del mismo son secuenciación de amplicón (por ejemplo, detección de mutaciones, polimorfismos naturales), genotipificación multiplexada SNP implicando cebadores PCR tales como cebadores KASP, cebadores escorpiones etc.

Ligamiento a un adaptador

25

[0058] Se combinaron diferentes muestras grupales BAC en un experimento de mapeo físico de genoma entero 480. Esto requeriría 480 adaptadores diferentes codificados por barras EcoRI. Esta cantidad de adaptadores ha sido evitada utilizando 80 adaptadores EcoRI con códigos de barras 5 nt en combinación con adaptadores 6 Msel con códigos de barras 3 nt, que en la combinación genera 480 códigos de barras 8-mer que son únicos. En este caso, la secuenciación fue realizada mediante la realización de dos eventos de cebadores de secuencia, como se describe en la amplificación de PCR anterior y la figura 11B. En la figura 12, se describe el perfil general para el uso de dos adaptadores codificados por barras en el contexto de secuenciación Illumina GA (utilizando regiones de cebador de secuencia de amplificación P5 y P7). La parte A describe la preparación de la muestra utilizando dos enzimas de restricción para asimilar el ADN, mientras la parte B describe la preparación de la muestra utilizando una combinación de una enzima de restricción y ligamiento a un adaptador de extremo romo. El concepto se puede usar en cualquier método que implique el ligamiento de dos adaptadores, tal como secuenciación de fragmento de restricción, AFLP, RAD, WGP, secuenciación de genoma entero, secuenciación de extremo emparejado, secuenciación de representaciones reducidas etc.

35

REIVINDICACIONES

1. Método para identificar el origen de muestra de ADN de muestra ligado a un adaptador y/o amplificado comprendiendo las etapas de:

5

a) proporcionar un primer y un segundo identificador de secuencias de nucleótidos, donde cada identificador de secuencia de nucleótidos es una secuencia de nucleótidos de al menos un nucleótido;

10

b) proporcionar adaptadores y/o cebadores de amplificación, donde al menos un primer adaptador o cebador de amplificación comprende el primer identificador de secuencia de nucleótidos y un segundo adaptador o cebador de amplificación comprende el segundo identificador de secuencia de nucleótidos, y donde opcionalmente se proporcionan otros adaptadores o cebadores de amplificación que comprenden otro identificador de secuencias de nucleótidos;

c) proporcionar una pluralidad de ADNs de muestra;

15

d) realizar reacciones de ligamiento y/o de amplificación en los ADNs de muestra utilizando los adaptadores y/o cebadores de amplificación, para proporcionar modelos de ADN amplificado y/o ligado a un adaptador preparados para secuenciación de alto rendimiento, donde un modelo monocatenario preparado de ADN comprende, además de una secuencia capturada y/o amplificada a partir del ADN de muestra, una combinación única del primer y el segundo identificador de secuencia de nucleótidos y un primer y un segundo sitio de unión al cebador de secuenciación, y donde el primer sitio de unión al cebador de secuenciación está localizado 3' del primer identificador de secuencia de nucleótidos y el segundo sitio de unión al cebador de secuenciación está localizado 3' del segundo identificador de secuencia de nucleótidos;

20

e) determinar al menos la secuencia del primer y el segundo identificador de secuencia de nucleótidos, donde la primera secuencia de identificador de secuencias de nucleótidos se determina en una primera reacción de secuenciación de alto rendimiento utilizando un primer cebador de secuenciación que hibrida con el primer sitio de unión de cebador de secuenciación y la segunda secuencia de identificador de secuencia de nucleótidos se determina en una segunda reacción de secuenciación de alto rendimiento utilizando un segundo cebador de secuenciación que hibrida con el segundo sitio de unión de cebador de secuenciación y

25

30

f) determinar, basándose en la combinación única del primer y segundo identificador de secuencia de nucleótidos, el origen de muestra de los ADNs de muestra amplificados y/o ligados a un adaptador.

2. Método según la reivindicación 1 para identificar el origen de muestra de ADN de muestras amplificadas, donde:

35

- en el paso b), se proporcionan los primeros y segundos cebadores de amplificación, cada primer cebador comprendiendo un primer identificador de secuencia de nucleótidos y cada segundo cebador comprendiendo un segundo identificador de secuencia de nucleótidos;

40

- en el paso d), cada ADN de muestra se amplifica con un par único de un primer y segundo cebador de amplificación para proporcionar modelos de ADN preparado amplificado; y,
- el método comprende opcionalmente además un paso de agrupación de al menos parte de los modelos de ADN preparados amplificados después del paso d) y antes del paso e).

3. Método según la reivindicación 1, para identificar el origen de muestra del ADN de muestra ligado a un adaptador, donde:

45

- en el paso b), se proporcionan los primeros y segundos adaptadores, cada primer adaptador comprendiendo un primer identificador de secuencia de nucleótidos y cada segundo adaptador comprendiendo un segundo identificador de secuencia de nucleótidos;

50

- el método comprende además un paso de fragmentación de cada ADN de muestra después del paso c) y antes del paso d);

- en el paso d), un par único de un primer y un segundo adaptador se liga a cada muestra fragmentada de ADN para proporcionar modelos de ADN ligados a un adaptador para secuenciación de alto rendimiento; y

55

- el método comprende opcionalmente además un paso de agrupar al menos parte de los modelos de ADN ligado a un adaptador obtenidos en el paso d) antes del paso e).

4. Método según la reivindicación 1 para identificar el origen de muestra del ADN de muestra amplificado ligado a un adaptador, donde:

60

- en el paso b) se proporcionan el primer, segundo y opcionalmente otros adaptadores, y primeros y segundos cebadores de amplificación, donde cada primer adaptador comprende un primer identificador de secuencia de nucleótidos, donde cada segundo adaptador comprende un segundo identificador de secuencia de nucleótidos, y cada adaptador adicional comprende otro identificador de secuencia de nucleótidos;

65

- el método comprende además el paso de amplificación de cada ADN de muestra con un par de un

primer y segundo cebador de amplificación para proporcionar amplicones antes del paso d);
 - el método comprende además opcionalmente un paso de fragmentación después de dicho paso de amplificación y antes del paso d);
 - en el paso d) el primer, segundo y otros adaptadores opcionales se ligan a los amplicones opcionalmente fragmentados para proporcionar modelos de ADN amplificados ligados a un adaptador para secuenciación de alto rendimiento; y
 - el método comprende además opcionalmente un paso de agrupación dentro del paso d) después del ligamiento del primer adaptador y antes del ligamiento del segundo adaptador.

5

10 5. Método según la reivindicación 1 para identificar el origen de muestra de un ADN de muestra amplificado y ligado a un adaptador donde:

15

- en el paso b) se proporcionan los primeros adaptadores, los primeros cebadores de amplificación y opcionalmente segundos adaptadores y/o segundos cebadores de amplificación, donde cada primer adaptador comprende un primer identificador de secuencia de nucleótidos, y cada primer cebador comprendiendo un segundo identificador de secuencia de nucleótidos, cada segundo adaptador opcional comprendiendo un tercer identificador de secuencia de nucleótidos y cada segundo cebador de amplificación opcional comprende un cuarto identificador de secuencia de nucleótidos;

20

- el método comprende además un paso de fragmentación de cada muestra ADN antes del paso d)
 - en el paso d) los fragmentos de ADN se ligan con al menos un primer adaptador, y opcionalmente con un segundo adaptador, para proporcionar fragmentos de ADN ligado a un adaptador, donde el fragmento de ADN ligado a un adaptador se amplifican con un primer cebador de amplificación y un segundo cebador de amplificación para proporcionar modelos de ADN adaptador ligado a un adaptador, amplificado para secuenciación de alto rendimiento; y opcionalmente

25

- el método comprende además un paso de agrupación de al menos parte de los modelos de ADN amplificado ligado a un adaptador antes del paso e).

30

6. Método según la reivindicación 1 para identificar el origen de muestra del ADN de muestra amplificado ligado a un adaptador, donde:

35

- en el paso b) se proporcionan los primeros y segundos cebadores de amplificación y el primer y el segundo adaptador opcional, donde cada primer cebador comprende un primer identificador de secuencias de nucleótidos, cada primer adaptador comprende un segundo identificador de secuencias de nucleótidos, cada segundo cebador comprende opcionalmente un tercer identificador de secuencias de nucleótidos, y cada segundo adaptador comprende un cuarto identificador de secuencias de nucleótidos;

40

- en el paso d) cada muestra de ADN se amplifica con un par de un primer y segundo cebador de amplificación para proporcionar amplicones, donde los amplicones se fragmentan opcionalmente, y donde los amplicones opcionalmente fragmentados se ligan a al menos un primer adaptador y opcionalmente un segundo adaptador para obtener modelos de ADN amplificado ligado a un adaptador para secuenciación de alto rendimiento; y
 - el método opcionalmente comprende además el agrupamiento de al menos parte de los modelos de ADN amplificado y ligado a un adaptador antes del paso e).

45

7. Método según cualquiera de las reivindicaciones anteriores, donde el modelo de ADN monocatenario comprende en una dirección 3' a 5' el primer sitio de unión al cebador de secuenciación, el primer identificador de secuencia de nucleótidos, la secuencia capturada y/o amplificada del ADN de muestra, el segundo sitio de unión al cebador de secuenciación y el segundo identificador de secuencia de nucleótidos.

50

8. Método según cualquiera de las reivindicaciones anteriores, donde además del primer y segundo identificador de secuencias de nucleótidos, se utilizan otros identificadores de secuencias de nucleótidos para determinar el origen de muestra del ADN de muestra amplificado y/o ligado a un adaptador.

55

9. Método según cualquiera de las reivindicaciones anteriores donde el identificador no contiene dos bases idénticas consecutivas.

60

10. Método según cualquiera de las reivindicaciones anteriores donde dos identificadores difieren en más de un nucleótido.

60

11. Método según cualquiera de las reivindicaciones anteriores donde dos identificadores difieren en más de dos nucleótidos y no contienen dos bases idénticas consecutivas.

Figura 1

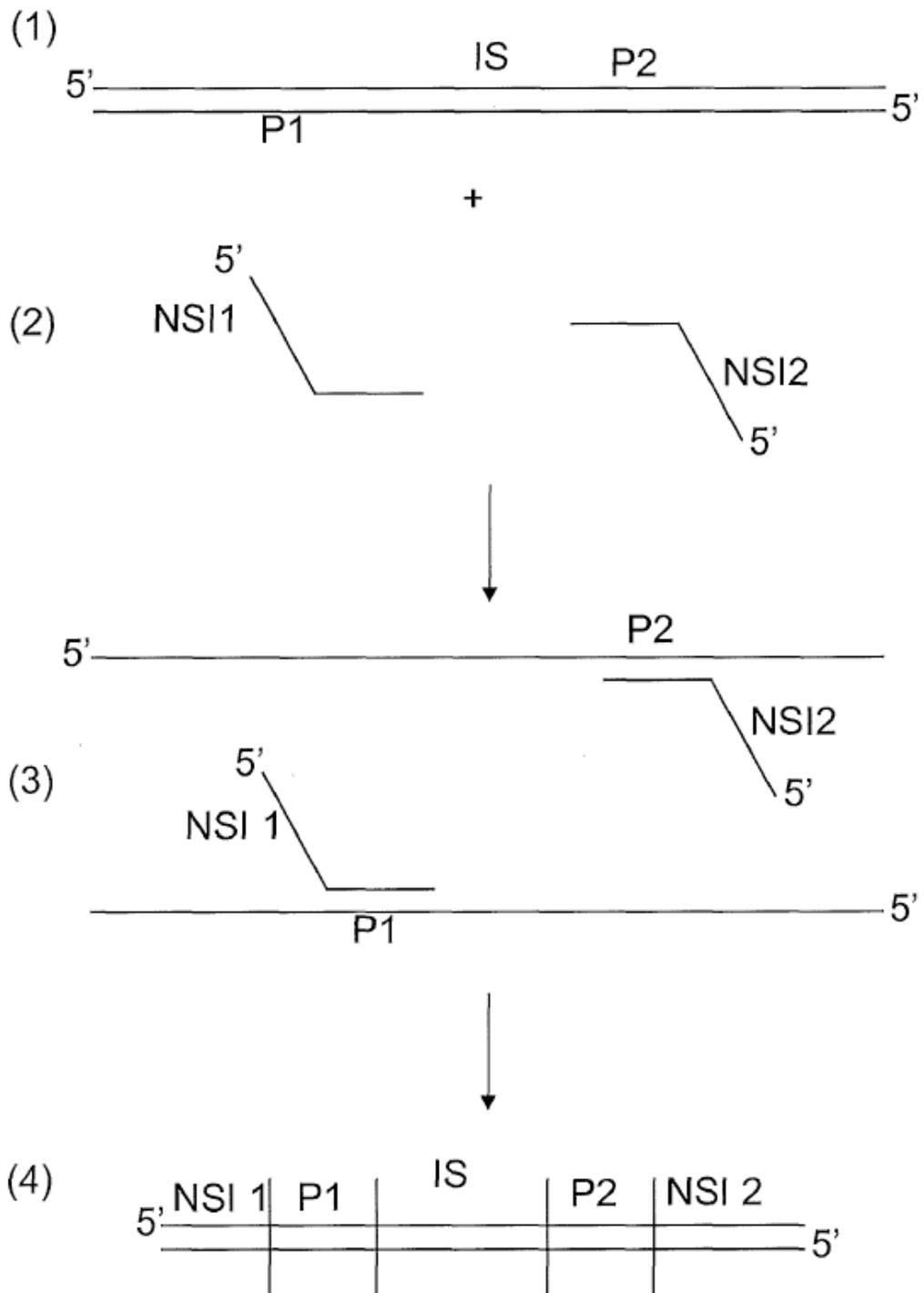


Figura 2

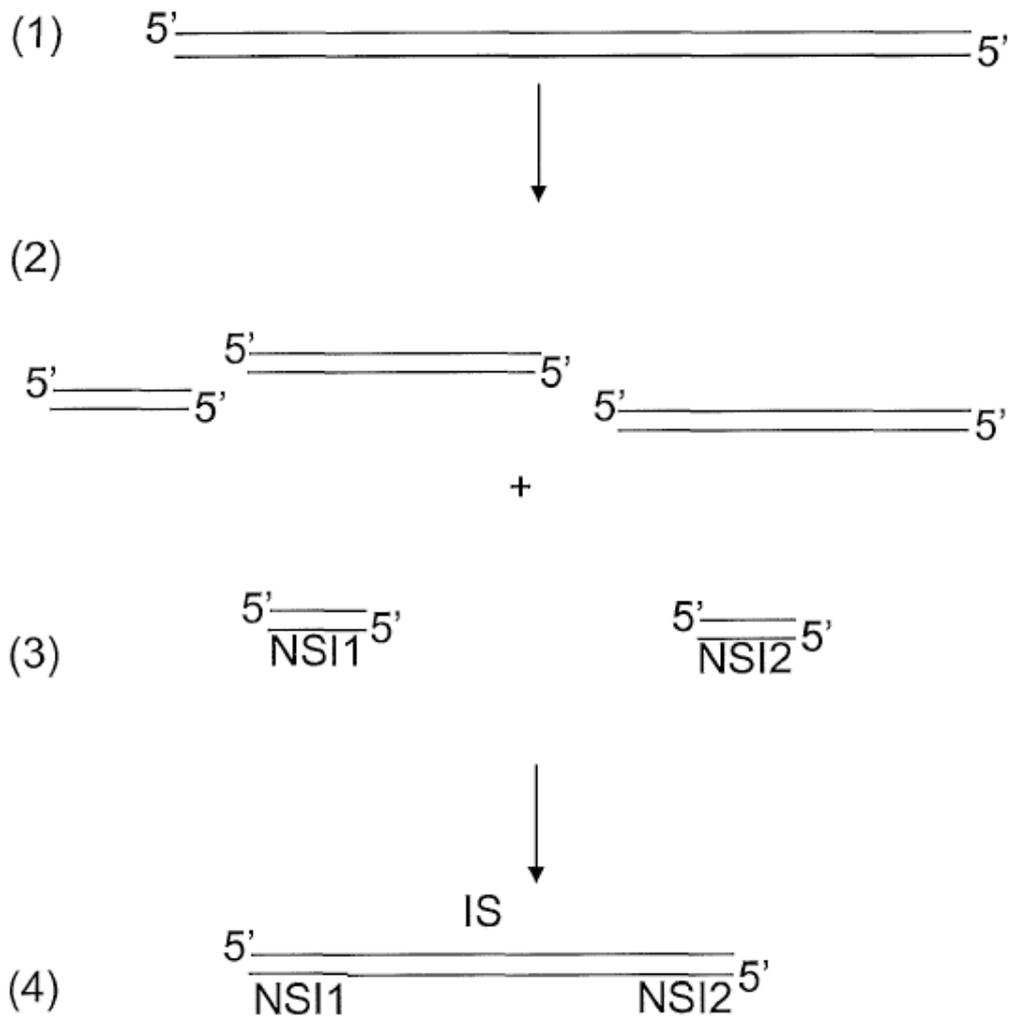


Figura 3

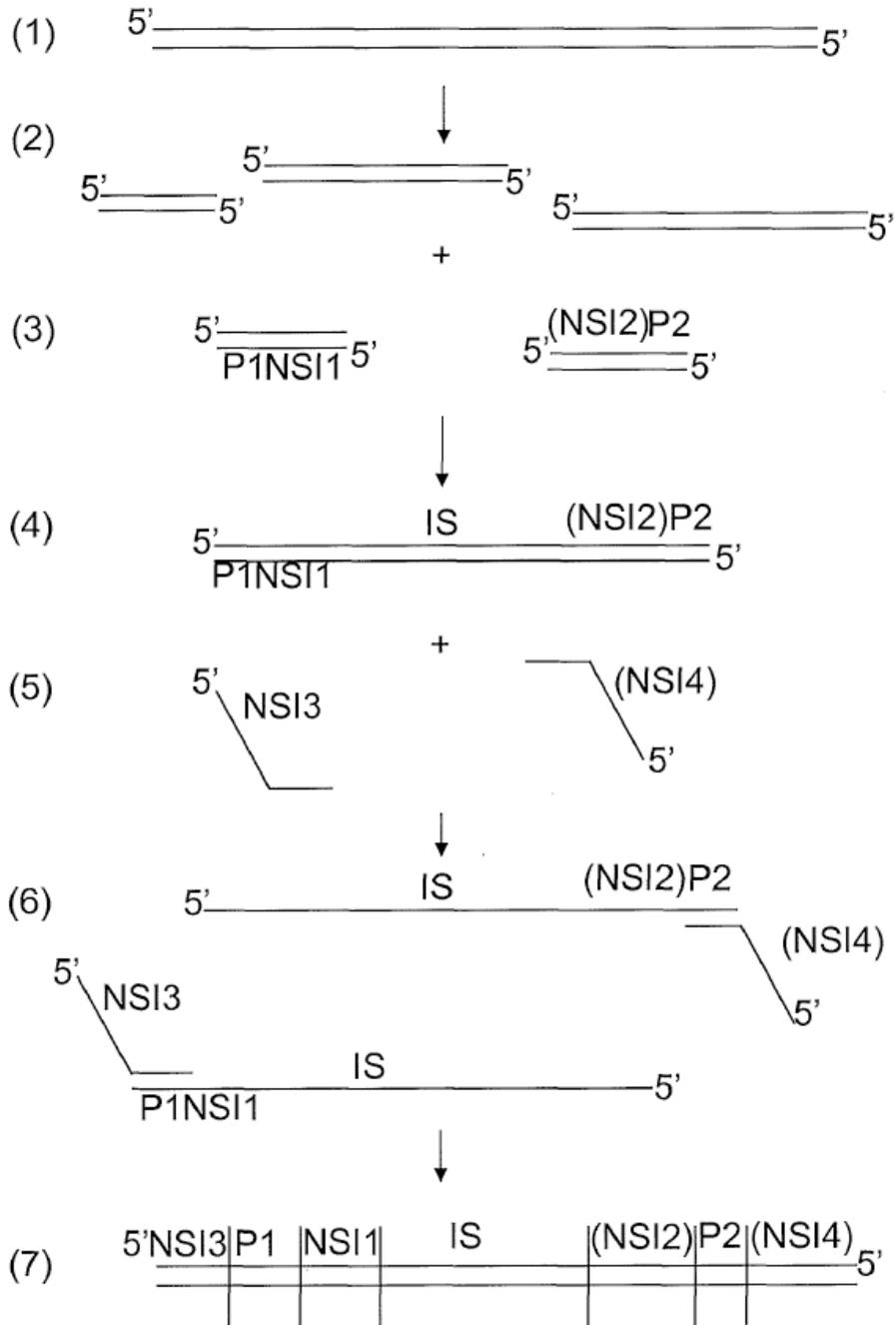


Figura 4

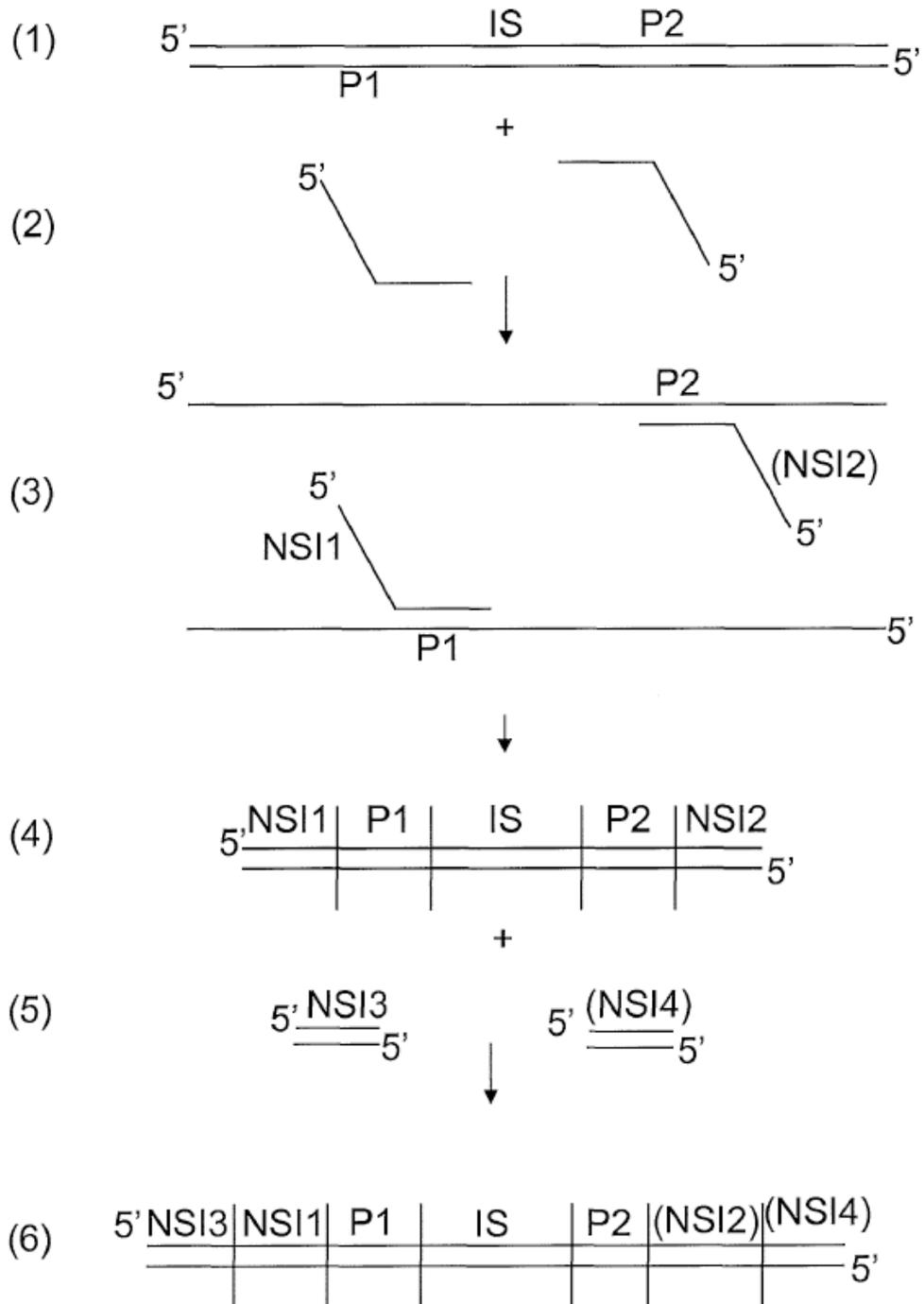


Figura 6

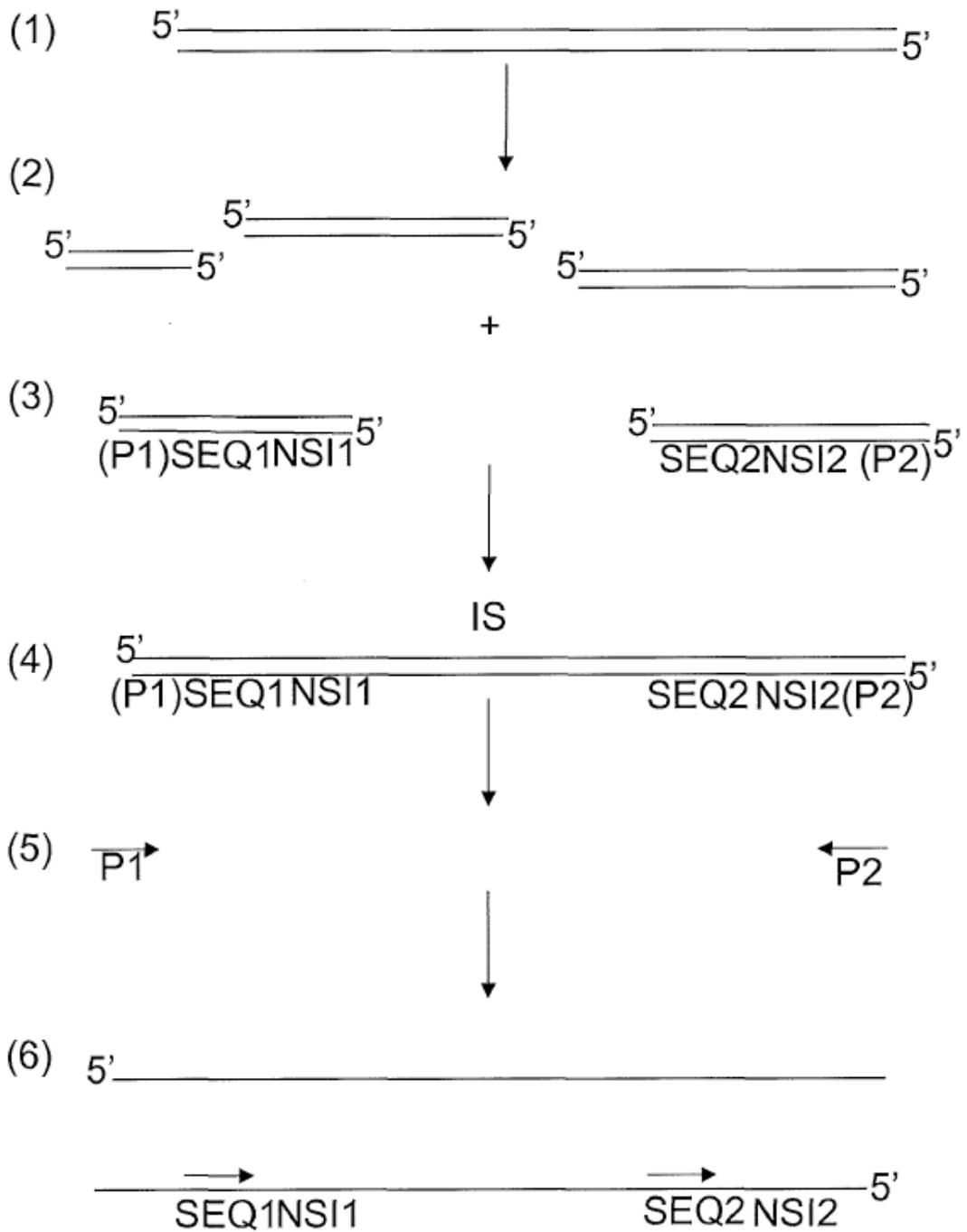


Figura 7

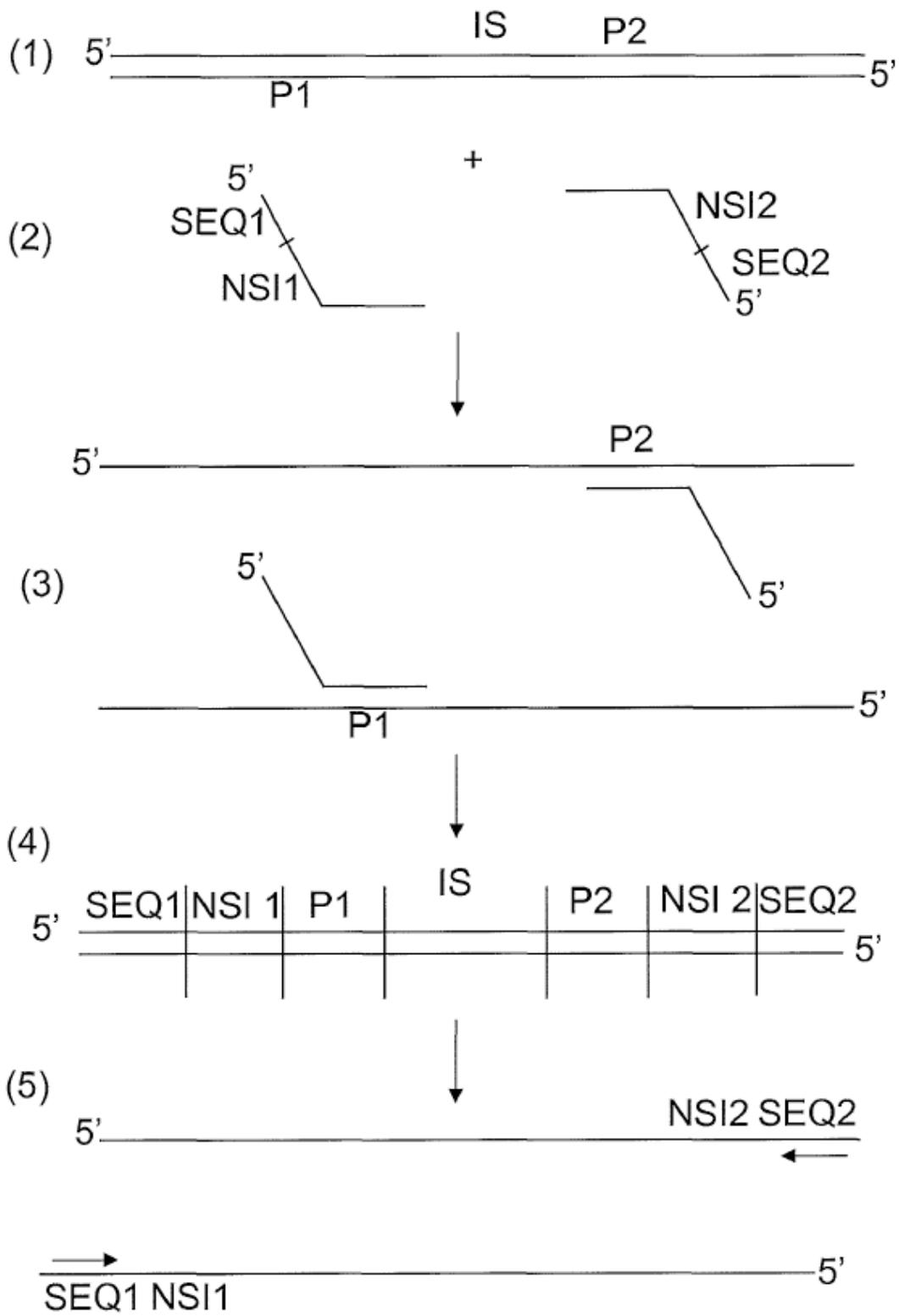


Figura 8

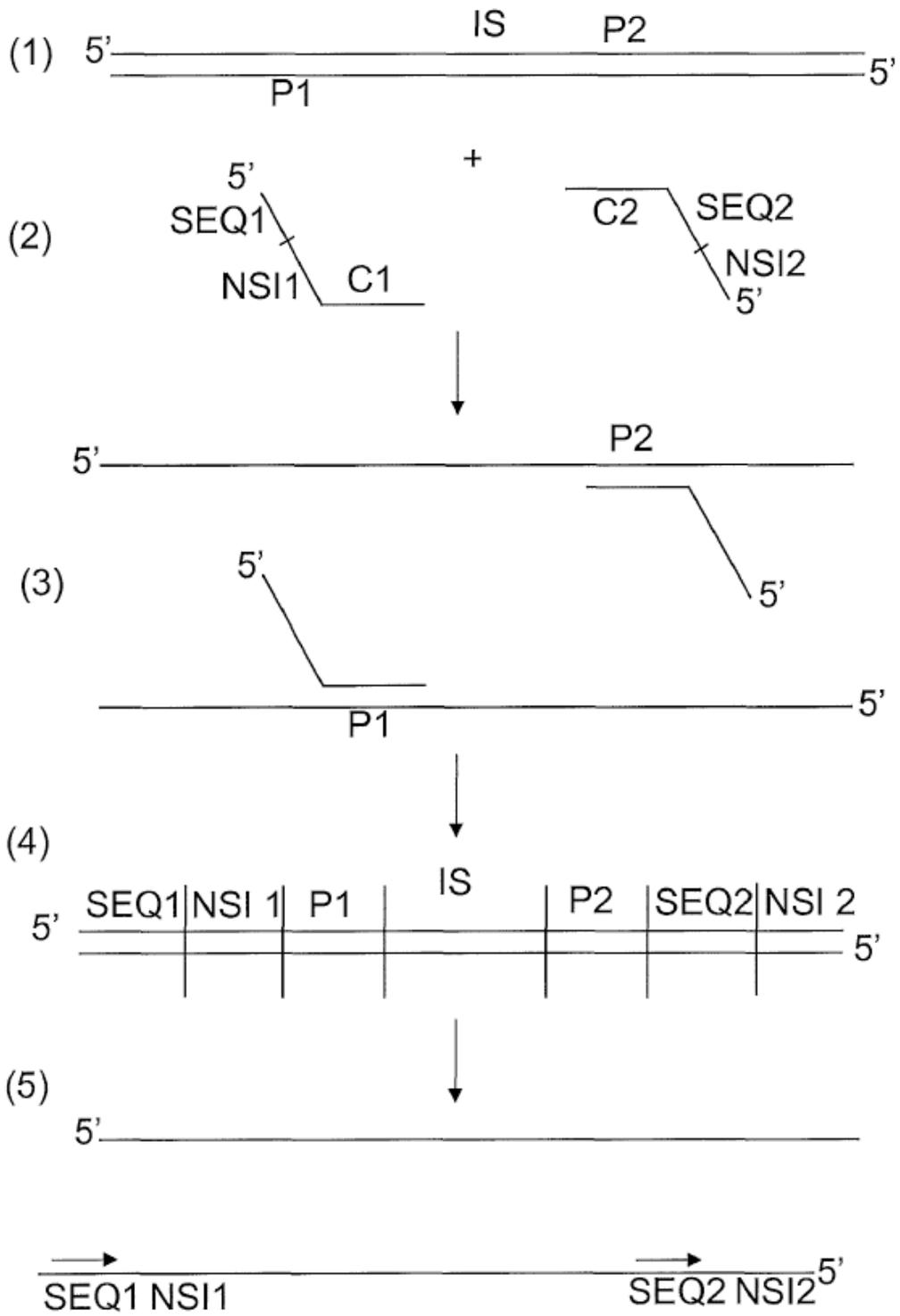


Figura 9

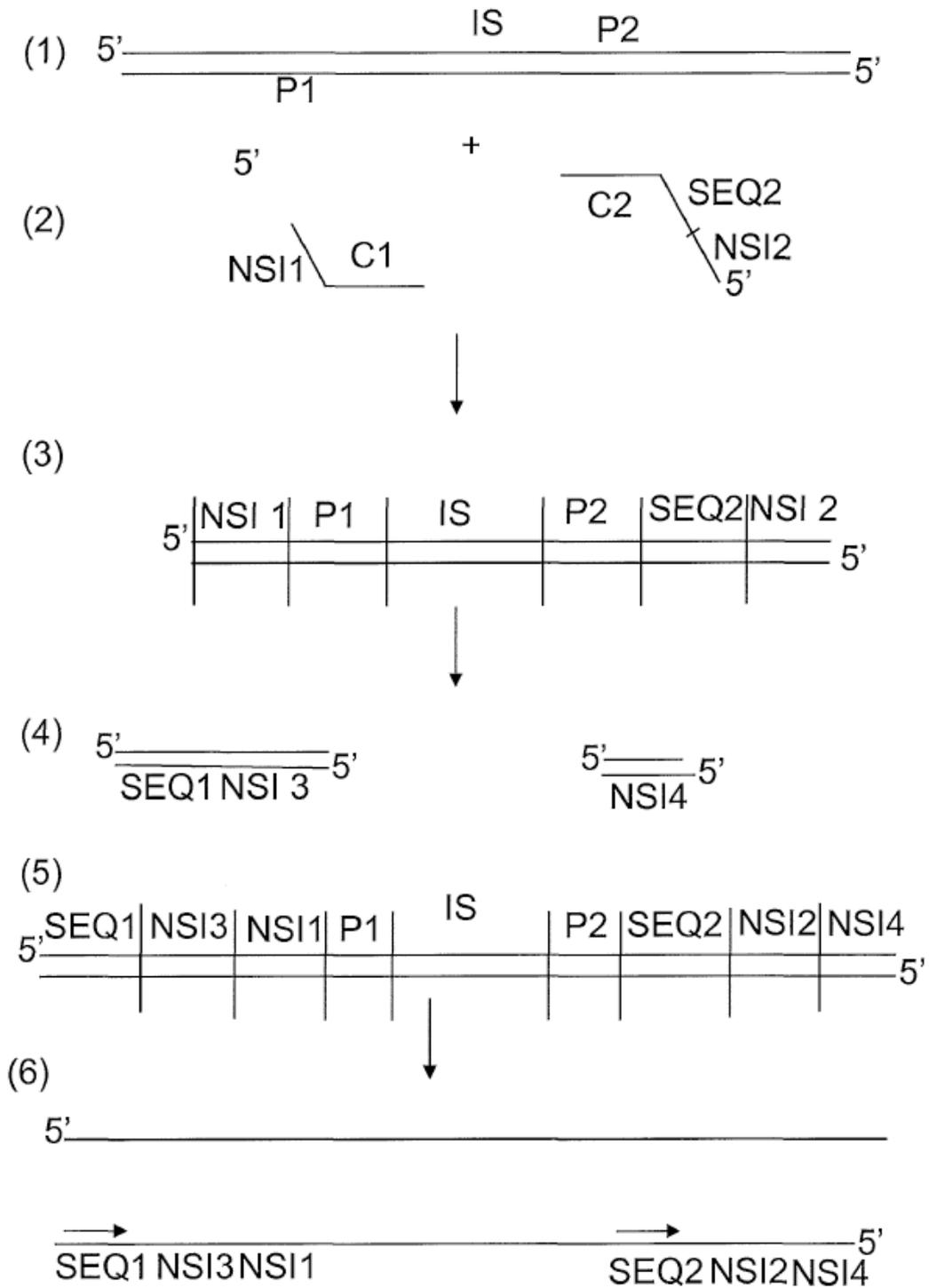


Figura 10

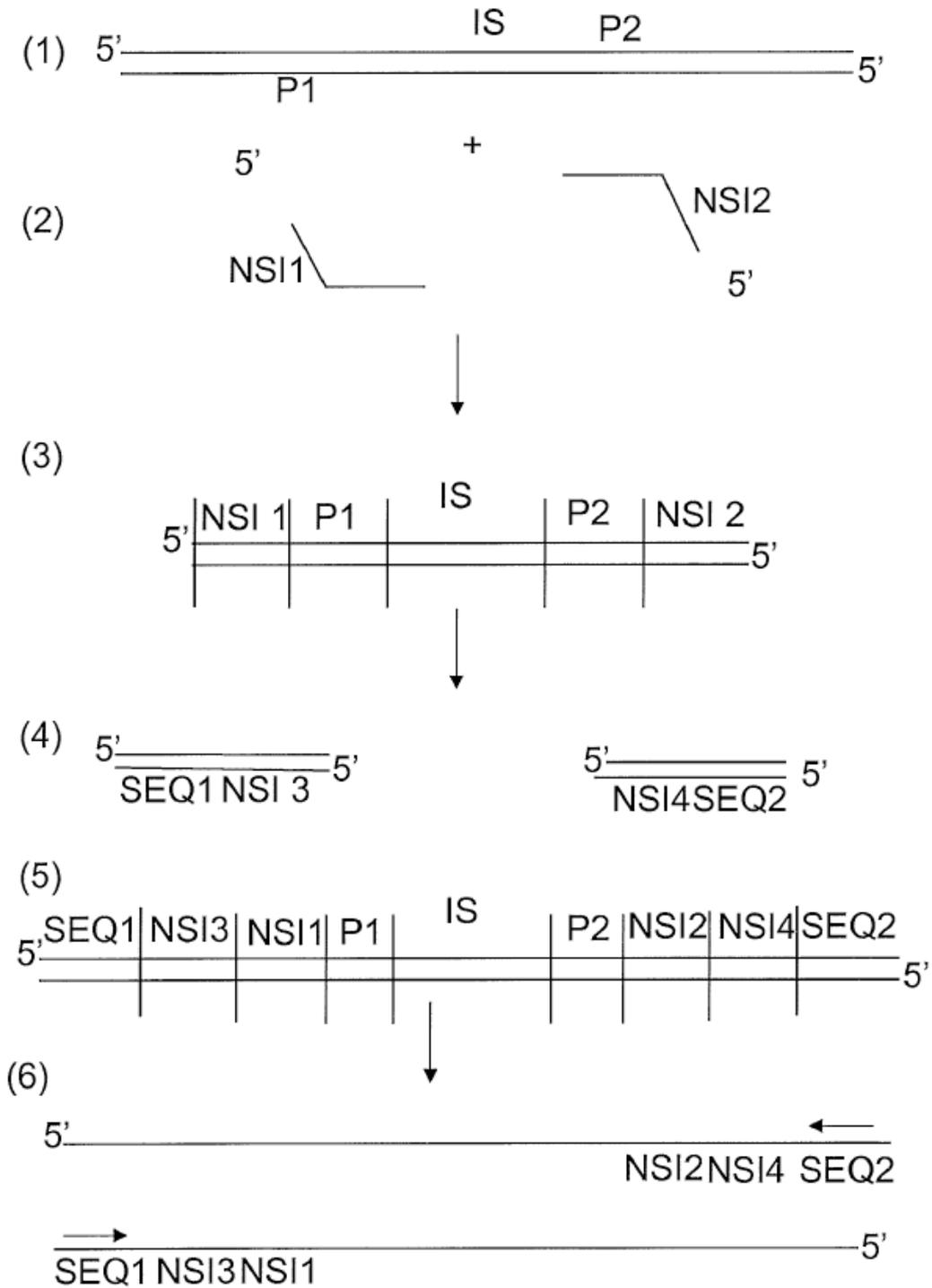
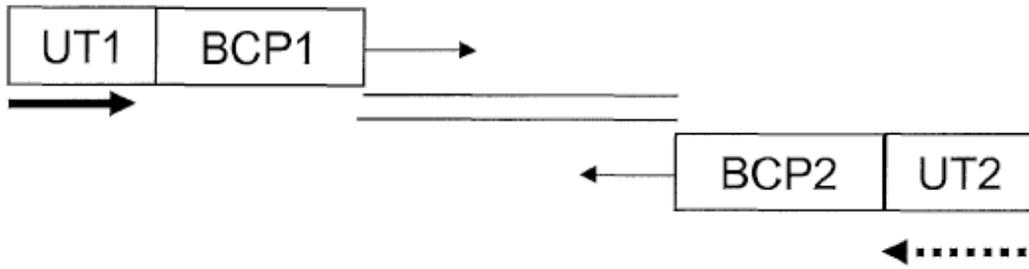


Figura 11

A



B

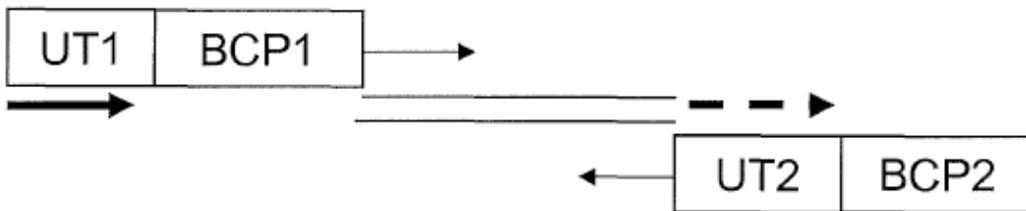
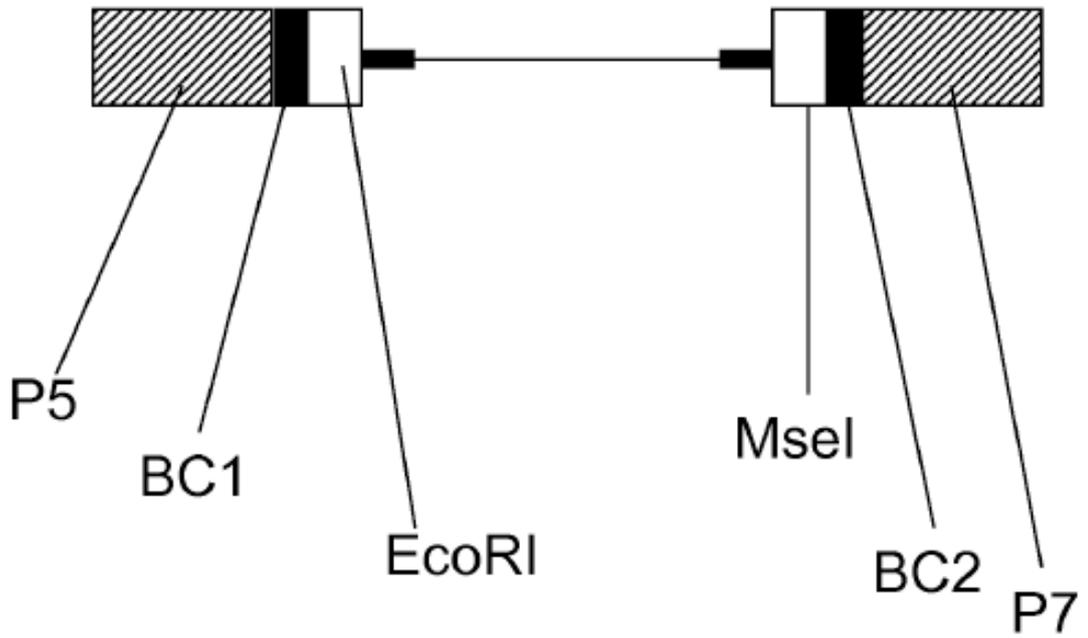


Figura 12

A



B

