

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 965 222**

51 Int. Cl.:
C12Q 1/6844 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **18.12.2019 PCT/US2019/067233**
- 87 Fecha y número de publicación internacional: **25.06.2020 WO20132103**
- 96 Fecha de presentación y número de la solicitud europea: **18.12.2019 E 19842471 (5)**
- 97 Fecha y número de publicación de la concesión europea: **18.10.2023 EP 3899037**

54 Título: **Métodos para mejorar la prioridad de clonalidad de agrupamientos de polinucleótidos**

30 Prioridad:
19.12.2018 US 201862782279 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
11.04.2024

73 Titular/es:
**ILLUMINA, INC. (100.0%)
5200 Illumina Way
San Diego, CA 92122, US**

72 Inventor/es:
**FISHER, JEFFREY S. y
GUO, MINGHAO**

74 Agente/Representante:
DEL VALLE VALIENTE, Sonia

ES 2 965 222 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

Métodos para mejorar la prioridad de clonalidad de agrupamientos de polinucleótidos

5 **Campo**

La presente memoria se refiere, entre otras cosas, al uso de la amplificación de exclusión de ácidos nucleicos diana para generar agrupamientos de secuenciación de amplicones; y más particularmente al aumento del número de agrupamientos que son monoclonales.

10

Antecedentes

Las mejoras en la técnica de secuenciación de próxima generación (NGS, del inglés *next-generation sequencing*) han aumentado considerablemente la velocidad de secuenciación y la producción de datos, lo que ha dado como resultado una productividad masiva de muestras de las plataformas de secuenciación actuales. Hace aproximadamente 10 años, el Analizador del genoma Illumina era capaz de generar hasta 1 *gigabyte* de datos de secuencia por ciclo. Hoy día, la Serie de sistemas NovaSeq™ de Illumina es capaz de generar hasta 2 *terabytes* de datos en dos días, lo que representa un aumento de capacidad de más de 2000x.

15

20

Un aspecto para lograr esta mayor capacidad es la generación de agrupamientos (*clusters*). La generación de agrupamientos puede incluir la producción de una biblioteca donde los miembros de la biblioteca incluyan una secuencia universal presente en cada extremo. La biblioteca se carga en una celda de flujo y los miembros individuales de la biblioteca se capturan en un “césped” de oligonucleótidos unidos a la superficie complementarios a la secuencia universal. A continuación, cada miembro se amplifica en agrupamientos clonales distintos mediante amplificación en puente. Cuando se completa la generación de agrupamientos, un agrupamiento individual puede incluir aproximadamente 1000 copias de un solo miembro de la biblioteca, y la biblioteca está lista para la secuenciación.

25

El documento WO 2016/075204 describe un método para producir una matriz que comprende amplicones monoclonales con fines de secuenciación.

30

Un método de amplificación en puente es la amplificación de exclusión (ExAmp), también conocida como amplificación de exclusión cinética. Este método es una reacción de amplificación facilitada por recombinasa que utiliza una matriz con patrones y condiciones isotérmicas para amplificar la biblioteca, dando como resultado una amplificación más rápida y el uso de menos reactivos para generar agrupamientos clonales en los pocillos de una matriz. Los métodos de ExAmp han demostrado ser muy útiles para la generación de agrupamientos clonales; sin embargo, las condiciones que dan como resultado más pocillos ocupados también causan la producción de más pocillos policlonales.

35

Resumen de la solicitud

40

La técnica de secuenciación de próxima generación (NGS) se basa en la secuenciación altamente paralela de poblaciones monoclonales de amplicones que se produjeron a partir de un único ácido nucleico diana. La secuenciación de poblaciones monoclonales de amplicones produce relaciones señal-ruido mucho más altas, mayor intensidad y mayor porcentaje de agrupamientos que pasan el filtro, todo lo cual contribuye a una mayor producción y calidad de datos.

45

Los métodos de amplificación de exclusión permiten la amplificación de un único nucleico diana por pocillo en una celda de flujo con patrones y la producción de una población monoclonal de amplicones en un pocillo. Típicamente, la velocidad de amplificación del primer ácido nucleico diana capturado dentro de un pocillo es más rápida en relación con velocidades mucho más lentas de transporte y captura del ácido nucleico diana en el pocillo. El primer ácido nucleico diana capturado en un pocillo se puede amplificar rápidamente y llenar todo el pocillo, evitando la captura de ácidos nucleicos diana adicionales en el mismo pocillo. Alternativamente, si un segundo ácido nucleico diana se une al mismo pocillo después del primero, la rápida amplificación del primero frecuentemente llena de forma suficiente del pocillo para dar como resultado una señal que pasa el filtro. El uso de amplificación de exclusión también puede dar como resultado distribuciones de super-Poisson de pocillos monoclonales, es decir, la fracción de pocillos en una matriz que son monoclonales puede superar la fracción predicha por la distribución de Poisson.

50

55

Es muy deseable aumentar las distribuciones de super-Poisson de agrupamientos útiles porque más pocillos monoclonales dan como resultado una mayor producción de datos; sin embargo, la siembra de ácidos nucleicos diana en pocillos generalmente sigue una distribución espacial de Poisson, donde la compensación por más pocillos ocupados son más pocillos policlonales. Un método para obtener distribuciones de super-Poisson más altas es hacer que la siembra se produzca rápidamente, seguida de un retraso entre los ácidos nucleicos diana sembrados. El retraso, denominado “retraso cinético” porque se cree que surge a través de cinéticas de reacciones bioquímica, le da a un ácido nucleico diana sembrado un inicio más temprano que las otras dianas sembradas.

60

65

La amplificación de exclusión trabaja mediante el uso de una recombinasa para facilitar la invasión de los cebadores (p. ej., cebadores unidos a un pocillo) en el ADN bicatenario (p. ej., un ácido nucleico diana) cuando encuentra un emparejamiento de secuencias. Para maximizar la eficiencia de la amplificación, es una práctica estándar para la amplificación de exclusión

utilizar una identidad completa entre los cebadores de invasión y las secuencias adaptadoras. Los inventores han identificado una manera de codificar un retraso cinético en ácidos nucleicos diana sembrados ajustando el grado de homología entre los adaptadores de los ácidos nucleicos diana y los cebadores unidos a los pocillos. Al reducir la homología promedio entre los cebadores de invasión y las secuencias adaptadoras, se produjo una sorprendente mejora de la velocidad de la llamada monoclonalidad de los pocillos, aunque se redujo la velocidad de amplificación promedio. Generalmente, a medida que se introducían más emparejamientos erróneos, la eficiencia de amplificación disminuía. Inesperadamente, cuando se usaron mezclas de secuencias adaptadoras que tenían eficiencias de amplificación tanto más altas como más bajas, las mezclas no se desempeñaron como un promedio del desempeño de los componentes individuales - a medio camino entre las eficiencias alta y baja - sino que superaron a todas las secuencias adaptadoras de tipo único en tanto la intensidad como los agrupamientos que pasaban el filtro.

Definiciones

Debe entenderse que los términos utilizados en la presente memoria adoptarán su significado habitual en la técnica correspondiente, a menos que se especifique lo contrario. A continuación se exponen varios términos utilizados en la presente memoria, y sus significados.

Como se utiliza en la presente memoria, cuando el término "amplicón" se utiliza en referencia a un ácido nucleico, significa el producto de copiar el ácido nucleico, en donde el producto tiene una secuencia de nucleótidos que es igual o complementaria a al menos una parte de la secuencia de nucleótidos del ácido nucleico. Un amplicón puede producirse por cualquiera de una variedad de métodos de amplificación que utilicen el ácido nucleico, p. ej. un ácido nucleico diana o un amplicón del mismo, tal como un molde, incluyendo, por ejemplo, extensión por polimerasa, reacción en cadena de la polimerasa (PCR), amplificación de círculo rodante (RCA), extensión por ligamiento o reacción en cadena de ligamiento. Un amplicón puede ser una molécula de ácido nucleico que tenga una sola copia de una secuencia de nucleótidos particular (p. ej., un producto de extensión por polimerasa) o múltiples copias de la secuencia de nucleótidos (p. ej., un producto concatamérico de RCA). Un primer amplicón de un ácido nucleico diana es normalmente una copia complementaria. Los amplicones posteriores son copias que se crean, después de la generación del primer amplicón, a partir del ácido nucleico diana o del primer amplicón. Un amplicón posterior puede tener una secuencia que sea sustancialmente complementaria a la del ácido nucleico diana o sustancialmente idéntica a la del ácido nucleico diana.

Como se utiliza en la presente memoria, la expresión "sitio de amplificación" se refiere a un sitio en o sobre una matriz donde se pueden generar uno o más amplicones. Un sitio de amplificación puede configurarse además para contener, sujetar o unir al menos un amplicón que se genera en el sitio.

Como se utiliza en la presente memoria, el término "matriz" se refiere a una población de sitios que pueden diferenciarse entre sí según la ubicación relativa. Las diferentes moléculas que se encuentran en diferentes sitios de una matriz se pueden diferenciar entre sí según las ubicaciones de los sitios en la matriz. Un sitio individual de una matriz puede incluir una o más moléculas de un tipo particular. Por ejemplo, un sitio puede incluir una sola molécula de ácido nucleico diana que tenga una secuencia particular o un sitio puede incluir varias moléculas de ácido nucleico que tengan la misma secuencia (y/o secuencia complementaria, de la misma). Los sitios de una matriz pueden ser diferentes casillas situadas en el mismo sustrato. Como casillas ilustrativas se incluyen, sin limitación, pocillos en un sustrato, perlas (u otras partículas) en o sobre un sustrato, proyecciones desde un sustrato, crestas sobre un sustrato o canales en un sustrato. Los sitios de una matriz pueden ser diferentes sustratos, cada uno con una molécula diferente. Se pueden identificar diferentes moléculas unidas a diferentes sustratos según las ubicaciones de los sustratos en una superficie a la que están asociados los sustratos o según las ubicaciones de los sustratos en un líquido o gel. Como matrices ilustrativas en que se ubican distintos sustratos en una superficie se incluyen, sin limitación, las que tienen perlas en pocillos.

Como se utiliza en la presente memoria, cuando el término "capacidad" se utiliza en referencia a un sitio y a un material de ácido nucleico, significa la cantidad máxima de material de ácido nucleico, p. ej., amplicones procedentes de un ácido nucleico diana, que puede ocupar el sitio. Por ejemplo, el término puede referirse al número total de moléculas de ácido nucleico que pueden ocupar el sitio en una condición particular. También pueden utilizarse otras medidas incluyendo, por ejemplo, la masa total de material de ácido nucleico o el número total de copias de una secuencia de nucleótidos particular que puede ocupar el sitio en una condición particular. Normalmente, la capacidad de un sitio para un ácido nucleico diana será sustancialmente equivalente a la capacidad del sitio para amplicones del ácido nucleico diana.

Como se utiliza en la presente memoria, la expresión "agente de captura" se refiere a un material, producto químico, molécula o resto del mismo, que es capaz de acoplarse, retener o unirse a una molécula diana (p. ej., un ácido nucleico diana). Los agentes de captura ilustrativos incluyen, sin limitarse a, un ácido nucleico de captura que es complementario de al menos una parte de un ácido nucleico diana modificado (p. ej., una secuencia de unión de captura universal), un miembro de un par de unión receptor-ligando (p. ej. avidina, estreptavidina, biotina, lectina, carbohidrato, proteína de unión a ácido nucleico, epítipo, anticuerpo, etc.) capaz de unirse a un ácido nucleico diana modificado (o resto de unión unido al mismo), o un reactivo químico capaz de formar un enlace covalente con un ácido nucleico diana modificado (o resto de unión unido al mismo). En una realización, un agente de captura es un ácido nucleico. También se puede utilizar un agente de captura de ácido nucleico como un cebador de amplificación.

Los términos “P5” y “P7” pueden utilizarse para referirse a un agente de captura de ácido nucleico. Los términos “P5” (P5 prima) y “P7” (P7 prima) se refieren a los complementos de P5 y P7, respectivamente. Se entenderá que en los métodos presentados en la presente memoria puede utilizarse cualquier agente de captura de ácido nucleico adecuado, y que el uso de P5 y P7 son sólo realizaciones ilustrativas. Los usos de agentes de captura de ácido nucleico tales como P5 y P7 en células de flujo son conocidos en la técnica, como se ilustra mediante las descripciones de WO 2007/010251, WO 2006/064199, WO 2005/065814, WO 2015/106941, WO 1998/044151 y WO 2000/018957. Un experto en la técnica reconocerá que un agente de captura de ácido nucleico también puede actuar como un cebador de amplificación. Por ejemplo, cualquier agente de captura de ácido nucleico adecuado puede actuar como cebador de amplificación directo, ya sea inmovilizado o en solución, y puede ser útil en los métodos presentados en la presente memoria para la hibridación a una secuencia (p. ej., una secuencia de unión de captura universal) y la amplificación de una secuencia. Por ejemplo, cualquier agente de captura de ácido nucleico adecuado puede actuar como cebador de amplificación inverso, ya sea inmovilizado o en solución, y puede ser útil en los métodos presentados en la presente memoria para la hibridación a una secuencia (p. ej., una secuencia de unión de captura universal) y la amplificación de una secuencia. En vista del conocimiento general disponible y las enseñanzas de la presente memoria, un experto en la técnica comprenderá cómo diseñar y usar secuencias que sean adecuadas para la captura y amplificación de ácidos nucleicos diana como se presentan en la presente memoria.

Como se utiliza en la presente memoria, la expresión “secuencia universal” se refiere a una región de secuencia que es común a dos o más ácidos nucleicos diana, donde las moléculas también tienen regiones de secuencia que difieren entre sí. Una secuencia universal que está presente en diferentes miembros de una colección de moléculas puede permitir la captura de múltiples ácidos nucleicos diferentes utilizando una población de ácidos nucleicos de captura que son complementarios a una parte de la secuencia universal, p. ej., una secuencia de unión de captura universal. Los ejemplos no limitativos de secuencias de unión de captura universales incluyen secuencias que son idénticas o complementarias a los cebadores P5 y P7. Otros ejemplos no limitativos de secuencias de unión de captura universales descritas en detalle en la presente memoria incluyen secuencias con identidad reducida (p. ej., uno o más emparejamientos erróneos) o complementariedad reducida con los cebadores P5 y P7, y/o tienen una longitud menor que los cebadores P5 y P7. Similarmente, una secuencia universal presente en distintos miembros de una colección de moléculas puede permitir la replicación o amplificación de múltiples ácidos nucleicos diferentes utilizando una población de cebadores universales que sean complementarios a una parte de la secuencia de universal, p. ej., un sitio de unión de cebadores universal. Las moléculas de ácido nucleico diana se pueden modificar para unir adaptadores universales (también denominados en la presente memoria adaptadores), por ejemplo, en uno o ambos extremos de las diferentes secuencias diana, como se describe en la presente memoria.

Como se utiliza en la presente memoria, el término “adaptador” y sus derivados, p. ej., adaptador universal, se refiere generalmente a cualquier oligonucleótido lineal que pueda ligarse a un ácido nucleico diana. En algunas realizaciones, el adaptador es sustancialmente no complementario al extremo 3' o al extremo 5' de cualquier secuencia diana presente en una muestra. En algunas realizaciones, las longitudes adecuadas del adaptador están en el intervalo de aproximadamente 10-100 nucleótidos, aproximadamente 12-60 nucleótidos y aproximadamente 15-50 nucleótidos de longitud. En general, el adaptador puede incluir cualquier combinación de nucleótidos y/o ácidos nucleicos. En algunos aspectos, el adaptador puede incluir uno o más grupos escindibles en una o más ubicaciones. En otro aspecto, el adaptador puede incluir una secuencia que sea sustancialmente idéntica, o sustancialmente complementaria, a al menos una parte de un cebador, por ejemplo, un ácido nucleico de captura. En algunas realizaciones, el adaptador puede incluir un código de barras, también denominado índice o etiqueta, para ayudar en la corrección de errores, identificación o secuenciación posteriores. Los términos “adaptador” y “adaptador” se utilizan indistintamente.

Como se define en la presente memoria, “muestra” y sus derivados se utilizan en su sentido más amplio e incluye cualquier muestra de ensayo, cultivo y similares que se sospeche que incluyen un ácido nucleico diana. En algunas realizaciones, la muestra comprende ADN, ARN, APN, ANB, formas quiméricas o híbridas de ácidos nucleicos. La muestra puede incluir cualquier muestra de ensayo biológica, clínica, quirúrgica, agrícola, atmosférica o acuática que contenga uno o más ácidos nucleicos. El término también incluye cualquier muestra de ácido nucleico aislada, tal como ADN genómico, muestra de ensayo de ácido nucleico congelada en fresco o fijada con formalina incluida en parafina. También se contempla que la muestra pueda ser de un único individuo, una colección de muestras de ácido nucleico de miembros genéticamente relacionados, muestras de ácido nucleico de miembros genéticamente no relacionados, muestras de ácido nucleico (emparejadas) de un único individuo tal como una muestra de tumor y una muestra de tejido normal, o muestra de una sola fuente que contiene dos formas distintas de material genético, tal como ADN materno y fetal obtenido de un sujeto materno, o la presencia de ADN bacteriano contaminante en una muestra que contiene ADN vegetal o animal. En algunas realizaciones, la fuente de material de ácido nucleico puede incluir ácidos nucleicos obtenidos de un recién nacido, por ejemplo, como se usa típicamente para el cribado de recién nacidos.

Como se utiliza en la presente memoria, la expresión “población clonal” se refiere a una población de ácidos nucleicos que es homogénea con respecto a una secuencia de nucleótidos particular. La secuencia homogénea tiene típicamente una longitud de al menos 10 nucleótidos, pero puede ser incluso más larga incluyendo, por ejemplo, una longitud de al menos 50, al menos 100, al menos 250, al menos 500 o al menos 1000 nucleótidos. Una población clonal puede proceder de un solo ácido nucleico diana. Normalmente, todos los ácidos nucleicos de una población clonal tendrán la misma secuencia de nucleótidos. Se entenderá que, en una población clonal, puede producirse un pequeño número de mutaciones (p. ej., debidas a artefactos de amplificación) sin apartarse de la clonalidad. También se entenderá que una pequeña cantidad de

ácidos nucleicos diana diferentes (por ejemplo, debido a un ácido nucleico diana que no fue amplificado o se amplificó en un grado limitado) puede estar presente en una población clonal sin apartarse de la clonalidad.

Como se utiliza en la presente memoria, el término “diferente”, cuando se usa en referencia a ácidos nucleicos, significa que los ácidos nucleicos tienen secuencias de nucleótidos que no son iguales entre sí. Dos o más ácidos nucleicos pueden tener secuencias de nucleótidos que son diferentes en toda su longitud. Alternativamente, dos o más ácidos nucleicos pueden tener secuencias de nucleótidos que son diferentes en una porción sustancial de su longitud. Por ejemplo, dos o más ácidos nucleicos pueden tener porciones de secuencias de nucleótidos diana que son diferentes entre sí y al mismo tiempo tener una región de secuencia universal que son iguales entre sí.

Como se utiliza en la presente memoria, la expresión “acceso fluídico”, cuando se usa en referencia a una molécula en un fluido y un sitio en contacto con el fluido, se refiere a la capacidad de la molécula de moverse en o a través del fluido para entrar en contacto o entrar en el sitio. La expresión también puede referirse a la capacidad de la molécula para separarse o salir del sitio para entrar en la solución. El acceso fluídico puede tener lugar cuando no hay barreras que impidan que la molécula entre en el sitio, entre en contacto con el sitio, se separe del sitio y/o salga del sitio. Sin embargo, se entiende que el acceso fluídico existe incluso si la difusión se retarda, reduce o altera, siempre que no se impida absolutamente el acceso.

Como se utiliza en la presente memoria, el término “bicatenario”, cuando se usa en referencia a una molécula de ácido nucleico, significa que sustancialmente todos los nucleótidos en la molécula de ácido nucleico están unidos por enlaces de hidrógeno a un nucleótido complementario. Un ácido nucleico parcialmente bicatenario puede tener al menos el 10 %, al menos el 25 %, al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 90 % o al menos el 95 % de sus nucleótidos con un enlace de hidrógeno a un nucleótido complementario.

Como se utiliza en la presente memoria, cuando el término “cada uno” se utiliza en referencia a una colección de artículos, pretende identificar un artículo individual en la colección, pero no se refiere necesariamente a cada artículo de la colección, a menos que el contexto indique otra cosa con claridad.

Como se utiliza en la presente memoria, la expresión “volumen excluido” se refiere al volumen del espacio ocupado por una molécula particular con exclusión de otras moléculas similares.

Como se utiliza en la presente memoria, la expresión “región intersticial” se refiere a un área en un sustrato o en una superficie que separa otras áreas del sustrato o superficie. Por ejemplo, una región intersticial puede separar una característica de una matriz de otra característica de la matriz. Las dos regiones que están separadas entre sí pueden ser discretas, sin contacto entre sí. En otro ejemplo, una región intersticial puede separar una primera parte de una característica de una segunda parte de una característica. La separación proporcionada por una región intersticial puede ser una separación parcial o total. Las regiones intersticiales tendrán típicamente un material de superficie que difiere del material de superficie de las casillas de la superficie. Por ejemplo, las casillas de una matriz pueden tener una cantidad o concentración de agentes de captura superior a la cantidad o concentración presente en las regiones intersticiales. En algunas realizaciones, los agentes de captura pueden no estar presentes en las regiones intersticiales.

Como se utiliza en la presente memoria, el término “polimerasa” pretende ser concordante con su uso en la técnica e incluye, por ejemplo, una enzima que produce una réplica complementaria de una molécula de ácido nucleico utilizando el ácido nucleico como cadena molde. De forma típica, las polimerasas de ADN se unen a la hebra molde y luego se mueven hacia abajo de la cadena molde añadiendo secuencialmente nucleótidos al grupo hidroxilo libre del extremo 3' de una cadena en crecimiento de ácido nucleico. Las ADN polimerasas sintetizan de forma típica moléculas de ADN complementarias a partir de patrones de ADN y las ARN polimerasas sintetizan de forma típica moléculas de ARN a partir de patrones de ADN (transcripción). Las polimerasas pueden utilizar una cadena corta de ARN o ADN, llamada cebador, para iniciar el crecimiento de la cadena. Algunas polimerasas pueden desplazar la cadena, secuencia arriba del sitio donde están añadiendo bases a una cadena. Se dice que dichas polimerasas desplazan la cadena, lo que significa que tienen una actividad que elimina una cadena complementaria de una cadena molde que lee la polimerasa. Entre las polimerasas ilustrativas con actividad de desplazamiento de cadena se incluyen, sin limitarse a, el fragmento grande de la Bsu (*Bacillus subtilis*), la Bst (*Bacillus stearothermophilus*) polimerasa, la exo-Klenow polimerasa o la exopolimerasa T7 con calidad de secuenciación. Algunas polimerasas degradan la cadena delante de ellas, reemplazándola efectivamente por la cadena en crecimiento por detrás (actividad exonucleasa 5'). Algunas polimerasas tienen una actividad que degrada la cadena detrás de ellas (actividad exonucleasa 3'). Se han modificado algunas polimerasas útiles, ya sea por mutación o de otro modo, para reducir o eliminar la actividad exonucleasa 3' y/o 5'.

Como se utiliza en la presente memoria, la expresión “ácido nucleico” pretende ser concordante con su uso en la técnica e incluye los ácidos nucleicos naturales y análogos funcionales de los mismos. Los análogos funcionales particularmente útiles pueden hibridarse a un ácido nucleico de una manera específica de secuencia o pueden utilizarse como un molde para la replicación de una secuencia de nucleótidos particular. Los ácidos nucleicos de origen natural generalmente tienen una cadena principal que contiene enlaces fosfodiéster. Una estructura análoga puede tener un enlace de cadena principal alternativo que incluye cualquiera de una variedad de los conocidos en la técnica. Los ácidos nucleicos naturales suelen tener un azúcar de desoxirribosa (por ejemplo, en el ácido desoxirribonucleico (ADN)) o un azúcar de ribosa (por ejemplo, en el ácido ribonucleico (ARN)). Un ácido nucleico puede contener cualquiera de una variedad de análogos de estos restos de azúcar que se conocen en la técnica. Los ácidos nucleicos pueden incluir bases naturales o no naturales. En este sentido,

un ácido desoxirribonucleico natural puede tener una o más bases seleccionadas de adenina, timina, citosina o guanina y un ácido ribonucleico puede tener una o más bases seleccionadas de uracilo, adenina, citosina o guanina. En la técnica se conocen bases no naturales y útiles que pueden incluirse en un ácido nucleico. El término “diana”, cuando se utiliza en referencia a un ácido nucleico, pretende ser un identificador semántico para el ácido nucleico en el contexto de un método o composición expuesto en la presente memoria y no limita necesariamente la estructura o función del ácido nucleico más allá de lo que se indica explícitamente. Un ácido nucleico diana que tiene una secuencia universal en cada extremo, por ejemplo un adaptador universal en cada extremo, puede denominarse ácido nucleico diana modificado.

Como se utiliza en la presente memoria, las expresiones “proteína de carga recombinasa” y “recombinasa” se usan indistintamente y pretenden ser concordantes con su uso en la técnica e incluyen, por ejemplo, la proteína RecA, la proteína T4 UvsX, la proteína UvsX del bacteriófago RB69, cualquier proteína homóloga o complejo proteico de cualquier filo, o variantes funcionales de los mismos. Los homólogos eucarióticos de RecA generalmente se llaman Rad51 en alusión al primer miembro de este grupo identificado. Se pueden usar otras recombinasas no homólogas en lugar de RecA, por ejemplo, RecT o RecO.

Como se utiliza en la presente memoria, la expresión “proteína de unión monocatenaria”, también denominada “proteína SSB” o “SSB”, pretende referirse a cualquier proteína que tenga la función de unirse a un ácido nucleico monocatenario, por ejemplo, para impedir la hibridación prematura, para proteger el ácido nucleico monocatenario de la digestión con nucleasas, para eliminar la estructura secundaria del ácido nucleico o para facilitar la replicación del ácido nucleico. La expresión pretende incluir, pero no se limita a, proteínas identificadas formalmente como proteínas de unión monocatenaria por el Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (NC-IUBMB). Las proteínas de unión monocatenaria ilustrativas incluyen, pero no se limita a, SSB de *E. coli*, T4 gp32, SSB de T7 gen 2.5, SSB de fago phi 29, proteína gp32 del bacteriófago RB69, cualquier proteína homóloga o complejo proteico de cualquier filo, y variantes funcionales de las mismas.

Como se utiliza en la presente memoria, la expresión “proteína accesoria” pretende referirse a cualquier proteína que tenga la función de interactuar con una recombinasa y una proteína de unión monocatenaria para ayudar en la producción de la nucleación de un filamento de UvsX en un ADNmc. Las expresiones “proteína accesoria”, “proteína accesoria de recombinasa” y “proteína auxiliar de recombinasa” se usan indistintamente. Las proteínas accesorias ilustrativas incluyen, pero no se limitan a, UvsY de T4, proteína UvsY del bacteriófago RB69, RecO de *E. coli*, RecR de *E. coli*, cualquier proteína homóloga o complejo proteico de cualquier filo, y variantes funcionales de las mismas.

Como se utiliza en la presente memoria, el término “transporte” se refiere al movimiento de una molécula a través de un líquido. El término puede incluir transporte pasivo, como el movimiento de moléculas a lo largo de su gradiente de concentración (p. ej., difusión pasiva). El término también puede incluir transporte activo mediante el cual las moléculas pueden moverse a lo largo de su gradiente de concentración o en contra de su gradiente de concentración. Por tanto, el transporte puede incluir la aplicación de energía para mover una o más moléculas en una dirección deseada o en una ubicación deseada, tal como un sitio de amplificación.

Como se utiliza en la presente memoria, el término “velocidad”, cuando se utiliza en referencia al transporte, amplificación, captura u otros procesos químicos, pretende ser concordante con su significado en cinética química y cinética bioquímica. Las velocidades de dos procesos se pueden comparar con respecto a velocidades máximas (p. ej., en saturación), velocidades previas al estado estacionario (p. ej., antes del equilibrio), constantes de velocidad cinética u otras medidas conocidas en la técnica. En realizaciones particulares, se puede determinar una velocidad para un proceso particular con respecto al tiempo total para completar el proceso. Por ejemplo, se puede determinar una velocidad de amplificación con respecto al tiempo necesario para que se complete la amplificación. Sin embargo, no es necesario determinar una velocidad para un proceso particular con respecto al tiempo total para completar el proceso.

El término “y/o” significa uno o todos los elementos enumerados o una combinación de cualquiera de dos o más de los elementos enumerados.

Las palabras “preferido(a)” y “preferiblemente” se refieren a realizaciones de la invención que pueden proporcionar determinados beneficios, en determinadas circunstancias. Sin embargo, en las mismas circunstancias o en otras, también pueden preferirse otras realizaciones. Además, la mención de una o más realizaciones preferidas no implica que otras realizaciones no sean útiles y no pretende excluir otras realizaciones del alcance de la invención.

Los términos “comprende” y sus variaciones, no tienen un significado limitativo cuando estos términos aparecen en la descripción y en las reivindicaciones.

Se entiende que siempre que en la presente memoria se describan realizaciones con las frases “incluye”, “que incluye” o “incluyendo” y similares, también se proporcionan realizaciones análogas descritas en términos de “que consiste en” y/o “que consiste esencialmente en”.

A menos que se especifique lo contrario, “un”, “uno”, “una”, “el”, “la” y “al menos un, uno, una”, se usan indistintamente y significan uno(a) o más de uno(a).

Las condiciones que son “adecuadas” para que tenga lugar un suceso, tal como la hibridación de dos secuencias de ácidos nucleicos, o condiciones “adecuadas”, son condiciones que no impiden que tales sucesos tengan lugar. Por lo tanto, estas condiciones permiten, potencian, facilitan y/o favorecen el evento.

5 Como se utiliza en la presente memoria, “proporcionar” en el contexto de una composición, un artículo o un ácido nucleico significa fabricar la composición, el artículo o el ácido nucleico, adquirir la composición, el artículo o el ácido nucleico, u obtener de otro modo el compuesto, la composición, el artículo o el ácido nucleico.

10 También en la presente memoria, las enumeraciones de intervalos numéricos mediante extremos incluyen todos los números comprendidos dentro de ese intervalo (por ejemplo, 1 a 5 incluye 1, 1,5, 2, 2,75, 3, 3,80, 4, 5, etc.).

15 A lo largo de la presente memoria descriptiva, la referencia a “una realización”, a “determinadas realizaciones” o a “algunas realizaciones” etc., significa que, en al menos una realización de la descripción, se incluye un rasgo, una configuración, una composición o una característica particular descrita en relación con la realización. Por tanto, la aparición de dichas expresiones en varios lugares a lo largo de esta memoria descriptiva, no se refieren necesariamente a la misma realización de la descripción. Además, los rasgos, configuraciones, composiciones o características particulares pueden combinarse de cualquier manera adecuada en una o más realizaciones.

20 En cualquier método desvelado en la presente memoria que incluya distintos pasos, estas pueden realizarse en cualquier orden factible. Y, según sea apropiado, cualquier combinación de dos o más pasos puede realizarse simultáneamente.

25 El resumen anterior de la presente invención no está destinado a describir cada realización descrita o cada implementación de la presente invención. La descripción siguiente ilustra más particularmente las realizaciones ilustrativas. En varios lugares de la solicitud, se proporciona orientación a través de listados de ejemplos, ejemplos que se pueden utilizar en diversas combinaciones. En cada caso, el listado citado sirve solo como un grupo representativo y no debe interpretarse como un listado excluyente.

Breve descripción de las figuras

30 La siguiente descripción detallada de las realizaciones ilustrativas de la presente memoria puede entenderse mejor cuando se lee junto con los siguientes dibujos.

35 Las FIG. 1A y 1B son un esquema de un ejemplo ilustrativo o una primera y una segunda secuencia de captura unidas a un pocillo de una matriz (FIG. 1A), y un esquema de un ejemplo ilustrativo de un ácido nucleico diana que tiene un adaptador universal unido a cada extremo (FIG. 1B).

La FIG. 2 es un esquema de un ejemplo ilustrativo de un primer ácido nucleico de captura unido a un pocillo de una matriz y una cadena individual hibridada de un ácido nucleico diana.

40 La FIG. 3A muestra el filtro de paso de densidad de adaptadores individuales y de grupos de adaptadores. k/mm², miles por milímetro cuadrado. Los carriles se refieren a los carriles de la celda de flujo que se muestran en la Tabla 2 de los Ejemplos. La FIG. 3B muestra la relación de las lecturas finales asociadas con los adaptadores mutantes individuales.

45 Las FIG. 4A y 4B muestran esquemas de ejemplos ilustrativos de invasión y duplicación de cadenas (“APR” se refiere a amplificación por recombinasa y polimerasa). En la FIG. 4A, una recombinasa facilita la invasión de cebadores P7 libres en los moldes bicatenarios que contienen secuencias homólogas (es decir, que se emparejan con los extremos de P7). No se precisa una homología perfecta (en este caso se muestra mediante dos emparejamientos erróneos deliberados introducidos en P7), pero la velocidad de invasión y amplificación disminuirá mediante los adaptadores con homología reducida (en este caso representado por una flecha más pequeña para las cadenas mutantes). En la FIG. 4B, la invasión mediada por la recombinasa desde cualquiera de los extremos se produce con un cebador del “césped” no mutado y corrige eficazmente las mutaciones de las cadenas hijas, transformándolas nuevamente en adaptadores perfectos. Sin embargo, dado que la homología entre la cadena original y la cadena del “césped” se ha reducido, el retraso hasta que se produce la primera copia es proporcional al número y grado de mutaciones.

55 Las FIG. 5A y 5B muestran los efectos de bibliotecas de adaptadores cortos y mutantes sobre las velocidades de amplificación. En la FIG. 5A, una copia satisfactoria transforma cada molde en uno perfecto. Sin embargo, la constante de tiempo para esa transición depende del grado de no homología a superar (las velocidades más altas se indican con flechas más gruesas). En la FIG. 5B, las velocidades de amplificación más lentas en las bibliotecas de adaptadores cortos y mutantes están indicadas por los desplazamientos hacia la derecha en las curvas de amplificación en tiempo real.

60 La FIG. 6 ilustra la competencia entre diferentes moldes por el dominio clonal en un parche individual. Los moldes sembrados se muestran con su tendencia de amplificación (es decir, retraso cinético); 1 = más rápido, 6 = más lento. No son necesarias ni siquiera deseables relaciones molares iguales de los moldes. Se prefieren números más altos de los moldes más rápidos. Sin embargo, incluso el molde más lento (6) puede poblar un parche con un agrupamiento monoclonal si no tiene competencia en el parche.

65

Los dibujos esquemáticos no son necesariamente a escala. Los mismos números que se utilizan en las figuras se refieren a los mismos componentes, pasos y similares. Sin embargo, se entenderá que el uso de un número para referirse a un componente en una figura dada no pretende estar limitado al componente en otra figura etiquetada con el mismo número. Además, el uso de números distintos para referirse a componentes no pretende indicar que los distintos componentes numerados no puedan ser iguales o similares a otros componentes numerados.

Descripción detallada

En la presente memoria se proporcionan composiciones y métodos relacionados con el aumento de la producción de agrupamientos monoclonales que pueden usarse en secuenciación.

La presente memoria proporciona métodos para amplificar ácidos nucleicos y métodos para determinar secuencias de ácidos nucleicos. En una realización, un método incluye proporcionar un reactivo de amplificación que incluye (i) una matriz de sitios de amplificación y (ii) una solución que tiene una pluralidad de ácidos nucleicos diana diferentes. Los sitios de amplificación incluyen al menos dos poblaciones de ácidos nucleicos de captura. Una población, una primera población, incluye una primera secuencia de captura y la segunda población incluye una segunda secuencia de captura. Los diferentes ácidos nucleicos diana incluyen en el extremo 3' una primera secuencia de unión de captura universal. En una realización, los ácidos nucleicos diana son bicatenarios. La primera secuencia de unión de captura universal tiene menos afinidad por la primera secuencia de captura que una primera secuencia de unión de captura universal que tiene el 100 % de complementariedad con la primera secuencia de captura. Por ejemplo, como se muestra en la FIG. 1A, un ácido nucleico **100** de una primera población de ácidos nucleicos de captura incluye una primera secuencia **110** de captura, donde el ácido nucleico **100** está unido a la superficie de un sitio **120** de amplificación. En la FIG. 1B se muestra un ácido nucleico diana bicatenario **130** que incluye un adaptador universal **140** en cada extremo, y una primera secuencia **150** de unión de captura universal en el extremo 3' de cada adaptador universal **140**.

Opcionalmente, los diferentes ácidos nucleicos diana también incluyen en el extremo 5' una segunda secuencia de unión de captura universal. El complemento de la segunda secuencia de unión de captura universal tiene menos afinidad por la segunda secuencia de captura que una segunda secuencia de unión de captura universal que tiene un complemento con el 100 % de complementariedad con la segunda secuencia de captura. Por ejemplo, como se muestra en la FIG. 1A, un ácido nucleico **160** de una segunda población de ácidos nucleicos de captura incluye una primera secuencia **170** de captura, donde el ácido nucleico **160** está unido a la superficie de un sitio **120** de amplificación. En la FIG. 1B se muestra un ácido nucleico diana bicatenario **130** que incluye un adaptador universal **140** en cada extremo, y una segunda secuencia de unión de captura universal **180** en el extremo 5' de cada adaptador universal **140**.

El método incluye además hacer reaccionar el reactivo de amplificación para producir una pluralidad de sitios de amplificación que tienen cada uno una población clonal de amplicones de un ácido nucleico diana individual de la solución. La reacción incluye transportar los diferentes ácidos nucleicos diana a los sitios de amplificación y amplificar los ácidos nucleicos diana en los sitios de amplificación. Por ejemplo, como se muestra en la FIG. 2, un ácido nucleico **200** de una primera población de ácidos nucleicos de captura incluye una primera secuencia **210** de captura, donde el ácido nucleico **200** está unido a la superficie de un sitio **220** de amplificación. Una cadena de un ácido nucleico diana **230** que incluye una primera secuencia **250** de unión de captura universal en el extremo 3' de la cadena individual híbrida con la primera secuencia **210** de captura del ácido nucleico **200**. La primera secuencia **250** de unión de captura universal incluye una 'X' para indicar la presencia de un emparejamiento erróneo entre la primera secuencia **250** de unión de captura universal y la primera secuencia **210** de captura. Esta puede a continuación experimentar una amplificación de agrupamiento, por ejemplo mediante amplificación de puente, para dar como resultado la generación de un agrupamiento.

Matrices

Una matriz de sitios de amplificación utilizada en un método expuesto en la presente memoria puede estar presente como uno o más sustratos. Los tipos ilustrativos de materiales de sustrato que se pueden usar para una matriz incluyen vidrio, vidrio modificado, vidrio funcionalizado, vidrios inorgánicos, microesferas (p. ej., partículas inertes y/o magnéticas), plásticos, polisacáridos, nailon, nitrocelulosa, cerámica, resinas, sílice, materiales a base sílice, carbono, metales, una fibra óptica o haces de fibras ópticas, polímeros y placas de multipocillo (p. ej., de microtitulación). Los plásticos ilustrativos incluyen acrílicos, poliestireno, copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos y TeflonTM. Los materiales a base de sílice ilustrativos incluyen silicio y diversas formas de silicio modificado.

En realizaciones particulares, un sustrato puede estar dentro o ser parte de un recipiente tal como un pocillo, tubo, canal, cubeta, placa de Petri, frasco o similar. Un recipiente particularmente útil es una celda de flujo, por ejemplo, como se describe en la patente estadounidense N.º 8.241.573 o Bentley y col., Nature 456:53-59 (2008). Las celdas de flujo ilustrativas son las que son comercializados por Illumina, Inc. (San Diego, California). Otro recipiente particularmente útil es un pocillo en una placa de multipocillo o en una placa de microtitulación.

En algunas realizaciones, los sitios de una matriz se pueden configurar como casillas en una superficie. Las casillas pueden estar presentes en cualquiera de una variedad de formatos deseados. Por ejemplo, los sitios pueden ser pocillos, huecos, canales, crestas, regiones elevadas, clavijas, postes o similares. Como se indicó anteriormente, los sitios pueden contener perlas. Sin embargo, en realizaciones particulares los sitios no necesitan contener una perla o partícula. Los sitios ilustrativos

incluyen pocillos que están presentes en sustratos utilizados para plataformas de secuenciación comerciales vendidas por 454 LifeSciences (una subsidiaria de Roche, Basilea, Suiza) o Ion Torrent (una subsidiaria de Life Technologies, Carlsbad California). Otros sustratos que tienen pocillos incluyen, por ejemplo, fibra óptica grabada y otros sustratos descritos en la patente estadounidense N.º 6.266.459; patente estadounidense N.º 6.355.431; patente estadounidense N.º 6.770.441; patente estadounidense N.º 6.859.570; patente estadounidense N.º 6.210.891; patente estadounidense N.º 6.258.568; patente estadounidense N.º 6.274.320; patente estadounidense N.º 8.262.900; patente estadounidense N.º 7.948.015; publicación de patente estadounidense N.º 2010/0137143; patente estadounidense N.º 8.349.167 o la publicación PCT N.º WO 00/63437. En varios casos, los sustratos se ilustran en estas referencias para aplicaciones que utilizan perlas en los pocillos. Los sustratos que contienen pocillos se pueden usar con o sin perlas en los métodos o composiciones de la presente memoria. En algunas realizaciones, los pocillos de un sustrato pueden incluir material de gel (con o sin perlas) como se expone en la patente estadounidense N.º 9.512.422.

Los sitios de una matriz pueden ser casillas metálicas sobre una superficie no metálica tal como vidrio, plástico u otros materiales ilustrados anteriormente. Se puede depositar una capa metálica sobre una superficie usando métodos conocidos en la técnica tales como grabado con plasma en húmedo, grabado con plasma en seco, deposición de capas atómicas, grabado con haz iónico, deposición química de vapor, pulverización catódica en vacío o similares. Se puede utilizar cualquiera de una variedad de instrumentos comerciales según sea apropiado, incluido, por ejemplo, los sistemas FlexAL®, Opal®, Ionfab 300Plus® u Optofab 3000® (Oxford Instruments, Reino Unido). También se puede depositar una capa de metal mediante evaporación por haz de electrones o pulverización catódica como se expone en Thornton, Ann. Rev. Mater. Sci. 7:239-60 (1977). Las técnicas de deposición de capa metálica, tales como las ilustradas anteriormente, se pueden combinar con técnicas de fotolitografía para crear regiones o parches metálicos en una superficie. Se proporcionan métodos ilustrativos para combinar técnicas de deposición de capas metálicas y técnicas de fotolitografía en la patente estadounidense N.º 8.778.848 y la patente estadounidense N.º 8.895.249.

Una variedad de características puede aparecer como una cuadrícula de puntos o manchas. Las casillas pueden ubicarse en un patrón repetitivo o en un patrón irregular no repetitivo. Los patrones particularmente útiles son patrones hexagonales, patrones rectilíneos, patrones en cuadrícula, patrones que tienen simetría reflectante, patrones que tienen simetría rotacional o similares. Los patrones asimétricos también pueden resultar útiles. La inclinación puede ser el mismo entre diferentes pares de casillas vecinas más cercanas o la inclinación puede variar entre diferentes pares de entidades vecinas más cercanas. En realizaciones particulares, las casillas de una matriz pueden tener cada una un área que sea mayor que aproximadamente 100 nm², 250 nm², 500 nm², 1 μm², 2,5 μm², 5 μm², 10 μm², 100 μm² o 500 μm². Alternativamente o de forma adicional, las casillas de una matriz pueden tener cada una un área que sea menor que aproximadamente 1 mm², 500 μm², 100 μm², 25 μm², 10 μm², 5 μm², 1 μm², 500 nm² o 100 nm². De hecho, una región puede tener un tamaño que esté en un intervalo entre un límite superior e inferior seleccionado de los ilustrados anteriormente.

Para realizaciones que incluyen una matriz de casillas en una superficie, las casillas pueden ser discretas estando por regiones intersticiales. El tamaño de las casillas y/o el espaciado entre las regiones puede variar de modo que las matrices puedan ser de alta densidad, densidad media o densidad más baja. Las matrices de alta densidad se caracterizan por tener regiones separadas por menos que aproximadamente 15 μm. Las matrices de densidad media tienen regiones separadas por aproximadamente 15 a 30 μm, mientras que las matrices de baja densidad tienen regiones separadas por más que 30 μm. Una matriz útil en la descripción puede tener regiones que estén separadas por menos de 100 μm, 50 μm, 10 μm, 5 μm, 1 μm o 0,5 μm.

En realizaciones particulares, una matriz puede incluir una colección de perlas u otras partículas. Las partículas pueden suspenderse en una solución o pueden ubicarse en la superficie de un sustrato. Los ejemplos de matrices de perlas en solución son las comercializadas por Luminex (Austin, Tex.). Los ejemplos de matrices que tienen perlas ubicadas en una superficie incluyen aquellas en donde las perlas están ubicadas en pocillos tales como una matriz BeadChip (Illumina Inc., San Diego California) o sustratos usados en plataformas de secuenciación de 454 LifeSciences (una subsidiaria de Roche, Basilea, Suiza) o de Ion Torrent (una subsidiaria de Life Technologies, Carlsbad California). Otras matrices que tienen perlas ubicadas sobre una superficie se describen en la patente estadounidense N.º 6.266.459; patente estadounidense N.º 6.355.431; patente estadounidense N.º 6.770.441; patente estadounidense N.º 6.859.570; patente estadounidense N.º 6.210.891; patente estadounidense N.º 6.258.568; patente estadounidense N.º 6.274.320; US 2009/0026082 A1; US 2009/0127589 A1; US 2010/0137143 A1; US 2010/0282617 A1 o la publicación PCT N.º WO 00/63437. Varias de las referencias anteriores describen métodos para unir ácidos nucleicos diana a perlas antes de cargar las perlas en o sobre un sustrato de matriz. Sin embargo, se entenderá que las perlas se pueden fabricar para que incluyan cebadores de amplificación y las perlas se pueden usar a continuación para cargar una matriz, formando así sitios de amplificación para su uso en un método expuesto en la presente memoria. Como se expuso anteriormente en la presente memoria, los sustratos se pueden usar sin perlas. Por ejemplo, los cebadores de amplificación se pueden unir directamente a los pocillos o al material de gel en los pocillos. Por lo tanto, las referencias son ilustrativas de materiales, composiciones o aparatos que pueden modificarse para su uso en los métodos y composiciones expuestos en la presente memoria.

Los sitios de amplificación de una matriz pueden incluir una pluralidad de agentes de captura capaces de unirse a ácidos nucleicos diana. En una realización, un agente de captura incluye un ácido nucleico de captura. En condiciones típicas utilizadas para preparar matrices para secuenciación, la secuencia de nucleótidos del ácido nucleico de captura es complementaria a una secuencia de uno o más ácidos nucleicos diana. Por el contrario, la secuencia de nucleótidos del ácido nucleico de captura de la presente memoria no es completamente complementaria a una secuencia de uno o más

ácidos nucleicos diana. La secuencia de nucleótidos de los ácidos nucleicos de captura útiles en los métodos presentados en la presente memoria se describe en detalle en la presente memoria. En algunas realizaciones, el ácido nucleico de captura también puede actuar como un cebador para la amplificación del ácido nucleico diana (contenga o no también una secuencia universal). En algunas realizaciones, una población de ácido nucleico de captura incluye un cebador P5 o su complemento, y la segunda población de ácido nucleico de captura incluye un cebador P7 o su complemento.

En realizaciones particulares, se puede unir un agente de captura, tal como un ácido nucleico de captura, al sitio de amplificación. Por ejemplo, el agente de captura se puede unir a la superficie de una casilla de una matriz. La unión puede realizarse a través de una estructura intermedia tal como una perla, partícula o gel. Un ejemplo de unión de ácidos nucleicos de captura a una matriz mediante un gel se describe en la patente estadounidense N.º 8.895.249 y además ilustrada por celdas de flujo disponibles en el mercado de Illumina Inc. (San Diego, California) o descritas en el documento WO 2008/093098. Los geles ilustrativos que pueden usarse en los métodos y aparatos expuestos en la presente memoria incluyen, pero no se limita a, los que tienen una estructura coloidal, tal como agarosa; una estructura de malla polimérica, tal como gelatina; o una estructura polimérica reticulada, tal como poli(acrilamida), SFA (véase, por ejemplo, la solicitud de patente estadounidense N.º 2011/0059865 A1) o PAZAM (véase, por ejemplo, solicitud provisional de patente estadounidense de N.º de serie 61/753.833 y la patente estadounidense N.º 9.012.022). La unión mediante una perla se puede lograr como se ilustra en la descripción y la bibliografía citada expuesta anteriormente en la presente memoria.

En algunas realizaciones, las casillas en la superficie de un sustrato de matriz no son contiguas y están separadas por regiones intersticiales de la superficie. Son ventajosas las regiones intersticiales que tienen una cantidad o concentración sustancialmente menor de agentes de captura, en comparación con las casillas de la matriz. Las regiones intersticiales que carecen de agentes de captura son particularmente ventajosas. Por ejemplo, una cantidad relativamente pequeña o ausencia de restos de captura en las regiones intersticiales favorece la ubicación de los ácidos nucleicos diana y de los agrupamientos generados posteriormente en las casillas deseadas. En realizaciones particulares, las casillas pueden ser casillas cóncavas en una superficie (p. ej., pocillos) y las casillas pueden contener un material de gel. Las casillas que contienen gel pueden estar separadas entre sí por regiones intersticiales en la superficie donde el gel está sustancialmente ausente o, si está presente, el gel es sustancialmente incapaz de soportar la ubicación de ácidos nucleicos. Los métodos y composiciones para fabricar y usar sustratos que tienen casillas que contienen gel, tales como pocillos, se exponen en la patente estadounidense N.º 9.512.422.

Ácidos nucleicos diana

La solución del reactivo de amplificación utilizada en un método descrito en la presente memoria incluye ácidos nucleicos diana. Las expresiones “ácido nucleico diana”, “fragmento diana”, “fragmento de ácido nucleico diana”, “molécula diana” y “molécula de ácido nucleico diana” se usan indistintamente para referirse a moléculas de ácido nucleico que se desea secuenciar, tal como en una matriz. El ácido nucleico diana puede ser esencialmente cualquier ácido nucleico de secuencia conocida o desconocida. Puede ser, por ejemplo, un fragmento de ADN genómico o ADNc. La secuenciación puede dar lugar a la determinación de la secuencia de toda la molécula diana o de una parte de ella. Las dianas pueden proceder de una muestra primaria de ácido nucleico fragmentada al azar. En una realización, las dianas pueden procesarse en moldes adecuados para la amplificación mediante la colocación de secuencias de amplificación universales, por ejemplo, secuencias presentes en un adaptador universal, en los extremos de cada fragmento de diana.

La muestra primaria de ácido nucleico puede originarse en forma de ADN bicatenario (“ADNbc”) (por ejemplo, fragmentos de ADN genómico, productos de PCR y amplificación y similares) a partir de una muestra o puede haberse originado en forma monocatenaria a partir de una muestra, tal como ADN o ARN, y haberse convertido en forma de ADNbc. A modo de ejemplo, las moléculas de ARNm pueden copiarse en ADNc bicatenario adecuado para su uso en el método descrito en la presente memoria, utilizando técnicas estándar bien conocidas en la técnica. La secuencia precisa de las moléculas de polinucleótido de una muestra primaria de ácido nucleico generalmente no es material de la descripción, y puede ser conocida o desconocida.

En una realización, las moléculas de polinucleótido primarias de una muestra de ácido nucleico primaria son moléculas de ADN. Más particularmente, las moléculas de polinucleótido primarias representan todo el complemento genético de un organismo, y son moléculas de ADN genómico que incluyen tanto secuencias de intrones como de exones, así como secuencias reguladoras no codificantes, tales como secuencias promotoras y potenciadoras. En una realización, se pueden usar subconjuntos particulares de secuencias de polinucleótido o ADN genómico, tales como, por ejemplo, cromosomas particulares. Aún más en particularmente, se desconoce la secuencia de las moléculas de polinucleótido primarias. Todavía más particularmente, las moléculas de polinucleótido primarias son moléculas de ADN genómico humano. Los fragmentos de ADN diana pueden tratarse química o enzimáticamente antes o después de cualquier proceso de fragmentación al azar, y antes o después del ligamiento de las secuencias adaptadoras universales.

La muestra de ácido nucleico puede incluir material de elevado peso molecular tal como ADN genómico (ADNg). La muestra puede incluir material de bajo peso molecular, como moléculas de ácido nucleico obtenidas de muestras FFIP o de ADN archivadas. En otra realización, el material de bajo peso molecular incluye ADN fragmentado enzimática o mecánicamente. La muestra puede incluir ADN circulante extracelular. En algunas realizaciones, la muestra puede incluir moléculas de ácido nucleico obtenidas de biopsias, tumores, raspados, hisopados, sangre, moco, orina, plasma, semen, cabello, microdisecciones por captura con láser, extirpaciones quirúrgicas y otras muestras obtenidas clínicas o de laboratorio. En

algunas realizaciones, la muestra puede ser una muestra epidemiológica, agrícola, forense o patógena. En algunas realizaciones, la muestra puede incluir moléculas de ácido nucleico obtenidas de un animal tal como una fuente humana o de mamífero. En otra realización, la muestra puede incluir moléculas de ácido nucleico obtenidas de una fuente no de mamífero tal como una planta, bacteria, virus u hongo. En algunas realizaciones, la fuente de las moléculas de ácido nucleico puede ser una muestra o especie archivada o extinta.

Además, los métodos y composiciones descritos en la presente memoria pueden ser útiles para amplificar una muestra de ácido nucleico que tiene moléculas de ácido nucleico de baja calidad, tal como ADN genómico degradado y/o fragmentado de una muestra forense. En una realización, las muestras forenses pueden incluir ácidos nucleicos obtenidos de la escena de un crimen, ácidos nucleicos obtenidos de una base de datos de ADN de personas desaparecidas, ácidos nucleicos obtenidos de un laboratorio asociado con una investigación forense o incluir muestras forenses obtenidas por autoridades policiales y judiciales, una o más fuerzas armadas o cualquier personal similar. La muestra de ácido nucleico puede ser una muestra purificada o un lisado que contiene ADN en bruto, por ejemplo obtenido de un hisopado bucal, papel, tela u otro sustrato que puede estar impregnado con saliva, sangre u otros fluidos corporales. Como tal, en algunas realizaciones, la muestra de ácido nucleico puede comprender cantidades bajas o porciones fragmentadas de ADN, tal como ADN genómico. En algunas realizaciones, las secuencias diana pueden estar presentes en uno o más fluidos corporales que incluyen, pero no se limita a, sangre, esputo, plasma, semen, orina y suero. En algunas realizaciones, las secuencias diana se pueden obtener a partir de cabello, piel, muestras de tejido, autopsias o restos de una víctima. En algunas realizaciones, los ácidos nucleicos que incluyen una o más secuencias diana se pueden obtener de un animal muerto o un ser humano fallecido. En algunas realizaciones, las secuencias diana pueden incluir ácidos nucleicos obtenidos de ADN no humano, tal como ADN microbiano, vegetal o entomológico. En algunas realizaciones, las secuencias diana o las secuencias diana amplificadas están dirigidas a fines de identificación de seres humanos. En algunas realizaciones, la descripción se refiere generalmente a métodos para identificar características de una muestra forense. En algunas realizaciones, la descripción se refiere generalmente a métodos de identificación de seres humanos que utilizan uno o más cebadores específicos de diana descritos en la presente memoria o uno o más cebadores específicos de diana diseñados utilizando los criterios de diseño de cebadores presentados en la presente memoria. En una realización, una muestra de identificación forense o humana que contiene al menos una secuencia diana se puede amplificar usando uno cualquiera o más de los cebadores específicos de diana descritos en la presente memoria o usando los criterios para cebadores presentados en la presente memoria.

Los ejemplos adicionales no limitativos de fuentes de muestras biológicas pueden incluir organismos completos así como una muestra obtenida de un paciente. La muestra biológica se puede obtener de cualquier fluido o tejido biológico y puede estar en una diversidad de formas, incluyendo fluido y tejido líquidos, tejido sólido y formas conservadas tales como formas secas, congeladas y fijadas. La muestra puede ser de cualquier tejido, célula o fluido biológico. Dichas muestras incluyen, pero no se limita a, esputo, sangre, suero, plasma, células sanguíneas (por ejemplo, glóbulos blancos), líquido ascítico, orina, saliva, lágrimas, esputo, fluido vaginal (flujo), lavados obtenidos durante un procedimiento médico (p. ej., lavados pélvicos u otros lavados obtenidos durante una biopsia, endoscopia o cirugía), tejido, aspirado del pezón, muestras de biopsia con aguja gruesa o de biopsia con aguja fina, fluidos corporales que contienen células, ácidos nucleicos flotantes libres, líquido peritoneal y líquido pleural, o células a partir de los mismos. Las muestras biológicas también pueden incluir cortes de tejidos tales como cortes congelados o fijados tomados con fines histológicos o células microdisecionadas o partes extracelulares de los mismos. En algunas realizaciones, la muestra puede ser una muestra de sangre, tal como, por ejemplo, una muestra de sangre entera. En otro ejemplo, la muestra es una muestra de gota de sangre seca (GSS) sin procesar. En otro ejemplo más, la muestra es una muestra fijada con formalina incluida en parafina (FFIP). En otro ejemplo más, la muestra es una muestra de saliva. En otro ejemplo más, la muestra es una muestra de gota de saliva seca (GSS).

Las muestras biológicas ilustrativas de las que se pueden obtener ácidos nucleicos diana incluyen, por ejemplo, las procedentes de un eucariota, por ejemplo un mamífero, tal como un roedor, ratón, rata, conejo, cobaya, ungulado, caballo, oveja, cerdo, cabra, vaca, gato, perro, primate, primate humano o no humano; una planta, tal como *Arabidopsis thaliana*, maíz, sorgo, avena, trigo, arroz, colza o soja; un alga, tal como *Chlamydomonas reinhardtii*; un nematodo tal como *Caenorhabditis elegans*; un insecto, tal como *Drosophila melanogaster*, mosquito, mosca de la fruta, abeja melífera o araña; un pez, tal como el pez cebra; un reptil; un anfibio, tal como una rana o *Xenopus laevis*; un *Dictyostelium discoideum*; un hongo, tal como *Pneumocystis carinii*, *Takifugu rubripes*, levaduras, *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe*; o un *Plasmodium falciparum*. Los ácidos nucleicos diana también pueden obtenerse de un procariota tal como una bacteria, *Escherichia coli*, estafilococos o *Mycoplasma pneumoniae*; una arquea; un virus tal como el virus de la hepatitis C o el virus de la inmunodeficiencia humana; o un viroide. Los ácidos nucleicos diana pueden obtenerse de un cultivo o población homogénea de los organismos anteriores o, alternativamente, de una colección de varios organismos diferentes, por ejemplo, en una comunidad o ecosistema.

Fragmentación al azar se refiere a la fragmentación de una molécula de polinucleótido de una muestra de ácido nucleico primario de forma no ordenada por medios enzimáticos, químicos o mecánicos. Dichos métodos de fragmentación se conocen en la técnica y usan métodos estándar (Sambrook y Russell, Molecular Cloning, A Laboratory Manual, tercera edición). En una realización, la fragmentación se puede lograr usando un proceso al que frecuentemente se hace referencia como "tagmentación". La tagmentación utiliza un complejo de transposoma y se combina en un solo paso de fragmentación y ligamiento para añadir adaptadores universales (Gunderson y col., documento WO 2016/130704). Para mayor claridad, generar fragmentos más pequeños de un fragmento mayor de ácido nucleico mediante la amplificación específica por PCR de dichos fragmentos más pequeños no equivale a fragmentar la pieza más grande de ácido nucleico, ya que la pieza más grande de la secuencia de ácido nucleico permanece intacta (es decir, no se fragmenta por la amplificación por PCR).

Además, la fragmentación aleatoria está diseñada para producir fragmentos independientemente de la identidad de secuencia o posición de los nucleótidos que comprenden y/o rodean la ruptura. Más particularmente, la fragmentación al azar se realiza por medios mecánicos tales como la nebulización o la sonicación para producir fragmentos de unos 50 pares de bases de longitud a aproximadamente 1500 pares de bases de longitud, aún más particularmente de 50-700 pares de bases de longitud, todavía más particularmente de 50-400 pares de bases de longitud. Más particularmente, el método se utiliza para generar fragmentos más pequeños de 50 - 150 pares de bases de longitud.

La fragmentación de moléculas de polinucleótido por medios mecánicos (nebulización, sonicación y cizallamiento hidráulico, por ejemplo) da como resultado fragmentos con una mezcla heterogénea de extremos romos y de extremos protuberantes 3' y 5'. Por lo tanto, es deseable reparar los extremos del fragmento utilizando métodos o kits (como el kit de reparación de extremos del terminador de ADN Lucigen) conocidos en la técnica para generar extremos que sean óptimos para la inserción, por ejemplo, en sitios romos de vectores de clonación. En una realización particular, los extremos de los fragmentos de la población de ácidos nucleicos son extremos romos. Más particularmente, los extremos de los fragmentos tienen extremos romos y están fosforilados. El resto fosfato puede introducirse mediante tratamiento enzimático, por ejemplo, utilizando polinucleótido cinasa.

Una población de ácidos nucleicos diana, o amplicones de los mismos, puede tener una longitud de cadena promedio que se desea o que sea apropiada para una aplicación particular de los métodos o composiciones expuestos en la presente memoria. Por ejemplo, la longitud de cadena promedio puede ser menor que aproximadamente 100.000 nucleótidos, 50.000 nucleótidos, 10.000 nucleótidos, 5.000 nucleótidos, 1.000 nucleótidos, 500 nucleótidos, 100 nucleótidos o 50 nucleótidos. Alternativamente, o de forma adicional, la longitud de cadena promedio puede ser mayor que aproximadamente 10 nucleótidos, 50 nucleótidos, 100 nucleótidos, 500 nucleótidos, 1.000 nucleótidos, 5.000 nucleótidos, 10.000 nucleótidos, 50.000 nucleótidos o 100.000 nucleótidos. La longitud de cadena promedio para la población de ácidos nucleicos diana, o amplicones de los mismos, puede estar en un intervalo de entre un valor máximo y mínimo expuesto anteriormente. Se entenderá que los amplicones generados en un sitio de amplificación (o fabricados o utilizados de otro modo en la presente memoria) pueden tener una longitud de cadena promedio que está en un intervalo entre un límite superior e inferior seleccionado de los ilustrados anteriormente.

En algunos casos, una población de ácidos nucleicos diana puede producirse en condiciones o configurarse de otro modo para que tenga una longitud máxima para sus miembros. Por ejemplo, la longitud máxima para los miembros que se usan en una o más pasos de un método expuesto en la presente memoria o que están presentes en una composición particular puede ser de menos que 100.000 nucleótidos, menos que 50.000 nucleótidos, menos que 10.000 nucleótidos, menos que 5.000 nucleótidos, menos que 1.000 nucleótidos, menos que 500 nucleótidos, menos que 100 nucleótidos o menos que 50 nucleótidos. Alternativamente o de forma adicional, una población de ácidos nucleicos diana, o amplicones de los mismos, puede producirse en condiciones o configurarse de otro modo para tener una longitud mínima para sus miembros. Por ejemplo, la longitud mínima para los miembros que se usan en una o más pasos de un método expuesto en la presente memoria o que están presentes en una composición particular puede ser de más que 10 nucleótidos, más que 50 nucleótidos, más que 100 nucleótidos, más que 500 nucleótidos, más que 1.000 nucleótidos, más que 5.000 nucleótidos, más que 10.000 nucleótidos, más que 50.000 nucleótidos o más que 100.000 nucleótidos. La longitud de cadena máxima y mínima para los ácidos nucleicos diana en una población puede estar en un intervalo entre un valor máximo y mínimo expuesto anteriormente. Se entenderá que los amplicones generados en un sitio de amplificación (o fabricados o utilizados de otro modo en la presente memoria) pueden tener longitudes de cadena máximas y/o mínimas en un intervalo entre los límites superior e inferior ilustrados anteriormente.

En realizaciones particulares, los ácidos nucleicos diana tienen un tamaño con respecto al área de los sitios de amplificación, por ejemplo, para facilitar la amplificación de exclusión. Por ejemplo, el área para cada uno de los sitios de una matriz puede ser mayor que el diámetro del volumen excluido de los ácidos nucleicos diana para lograr la amplificación de exclusión. Tomando, por ejemplo, realizaciones que utilizan una matriz de casillas en una superficie, el área para cada una de las casillas puede ser mayor que el diámetro del volumen excluido de los ácidos nucleicos diana que se transportan a los sitios de amplificación. El volumen excluido de un ácido nucleico diana y su diámetro se pueden determinar, por ejemplo, a partir de la longitud del ácido nucleico diana. Los métodos para determinar el volumen excluido de ácidos nucleicos y el diámetro del volumen excluido se describen, por ejemplo, en la patente estadounidense N.º 7.785.790; Rybenkov y col., Proc. Natl. Acad. Sci. EE. UU. 90: 5307-5311 (1993); Zimmerman y col., J. Mol. Biol. 222:599-620 (1991); o Sobel y col., Biopolymers 31:1559-1564 (1991).

En una realización particular, las secuencias de fragmentos diana se preparan con nucleótidos protuberantes individuales mediante, por ejemplo, la actividad de determinados tipos de polimerasa de ADN tal como la polimerasa Taq o la polimerasa Klenow exo minus que tiene una actividad de transferasa terminal no dependiente de molde que añade un único desoxinucleótido, por ejemplo, desoxiadenosina (A) a los extremos 3' de, por ejemplo, un producto de PCR. Dichas enzimas pueden utilizarse para añadir un único nucleótido 'A' al extremo romo 3' de cada cadena de los fragmentos diana bicatenarios. Por lo tanto, podría añadirse una 'A' al extremo 3' de cada cadena de extremos reparados de los fragmentos diana bicatenarios por reacción con la polimerasa Taq o Klenow exo minus, mientras que la construcción de polinucleótido adaptador universal podría ser una construcción de T con una protuberancia 'T' compatible presente en el extremo 3' de cada región de ácido nucleico bicatenario del adaptador universal. Esta modificación del extremo también impide el autoligamiento tanto del vector como de la diana, de modo que existe una tendencia hacia la formación de las moléculas combinadas de adaptador-diana-adaptador ligadas.

En algunos casos, los ácidos nucleicos diana que se obtienen de tales fuentes se pueden amplificar antes de su uso en un método o composición de la presente memoria. Se puede utilizar cualquiera de una diversidad de técnicas de amplificación conocidas, incluyendo, pero no se limita a, reacción en cadena de la polimerasa (PCR), amplificación de círculo rodante (RCA), amplificación de desplazamiento múltiple (MDA) o la amplificación primaria al azar (RPA). Se entenderá que la amplificación de los ácidos nucleicos diana antes de su uso en un método o composición expuesta en la presente memoria es opcional. Como tales, los ácidos nucleicos diana no se amplificarán antes de su uso en algunas realizaciones de los métodos y composiciones expuestos en la presente memoria. Los ácidos nucleicos diana pueden proceder opcionalmente de bibliotecas sintéticas. Los ácidos nucleicos sintéticos pueden tener composiciones de ADN o ARN naturales o pueden ser análogos de los mismos.

Adaptadores universales

Un ácido nucleico diana usado en un método o composición descrito en la presente memoria incluye un adaptador universal unido a cada extremo. Los expertos en la técnica conocen los métodos para unir un adaptador universal a cada extremo de un ácido nucleico diana usado en un método descrito en la presente memoria. La unión puede realizarse mediante técnicas estándar de preparación de bibliotecas mediante ligación (Chesney y col. publicación de patente estadounidense N.º 2018/0305753 A1), o mediante tagmentación usando complejos de transposasa (Gunderson y col., documento WO 2016/130704).

En una realización, los ácidos nucleicos diana bicatenarios de una muestra, p. ej., una muestra fragmentada, se tratan ligando primero moléculas adaptadoras universales idénticas ('adaptadores con emparejamiento erróneo', cuyas características generales se definen a continuación y se describen además en Gormley y col., documento US 7.741.463, y Bignell y col., documento US 8.053.192) a los extremos 5' y 3' de los ácidos nucleicos diana bicatenarios (que pueden ser de secuencia conocida, parcialmente conocida o desconocida). En una realización, el adaptador universal incluye las secuencias de unión de captura universales necesarias para inmovilizar los ácidos nucleicos diana en una matriz para su posterior secuenciación. En otra realización, se usa un paso de PCR para modificar adicionalmente el adaptador universal presente en cada extremo de los ácidos nucleicos diana antes de la inmovilización y la secuenciación. Por ejemplo, se lleva a cabo una reacción de extensión de cebador inicial utilizando un sitio de unión de cebador universal en que se forman productos de extensión complementarios con ambas cadenas de cada ácido nucleico diana individual y se añade una secuencia de unión de captura universal. Los productos de extensión de cebador resultantes, y opcionalmente las copias amplificadas de los mismos, proporcionan colectivamente una biblioteca de ácidos nucleicos diana modificados que pueden inmovilizarse y a continuación secuenciarse. El término biblioteca se refiere a la colección de ácidos nucleicos diana que contienen secuencias comunes conocidas en sus extremos 3' y 5', y también puede denominarse biblioteca modificada en 3' y 5'. Los extremos 3', y opcionalmente los extremos 5', de los adaptadores universales unidos a los ácidos nucleicos diana pueden incluir una población homogénea o una población heterogénea de secuencias de unión de captura universales descritas en la presente memoria.

Los adaptadores universales utilizados en el método de la descripción se denominan adaptadores 'con emparejamiento erróneo' debido a que, como se explicará en detalle en la presente memoria, los adaptadores incluyen una región de emparejamiento erróneo de secuencia, es decir, no se forman por hibridación de cadenas de polinucleótido totalmente complementarias.

Los adaptadores con emparejamiento erróneo para su uso en la presente memoria se forman mediante hibridación de dos cadenas de polinucleótido parcialmente complementarias para proporcionar, cuando las dos cadenas están hibridadas, al menos una región bicatenaria, también denominada región de ácido nucleico bicatenario, y al menos una región no emparejada monocatenaria, también denominada región de cadenas de ácido nucleico monocatenarias no complementarias.

La 'región bicatenaria' del adaptador universal es una región bicatenaria corta, que incluye típicamente 5 o más pares de bases consecutivos, formados por hibridación de las dos cadenas de polinucleótido parcialmente complementarias. Este término se refiere a una región bicatenaria de ácido nucleico en que las dos cadenas están hibridadas y no implica ninguna conformación estructural particular. Como se utiliza en la presente memoria, el término "bicatenario", cuando se usa en referencia a una molécula de ácido nucleico, significa que sustancialmente todos los nucleótidos en la molécula de ácido nucleico están unidos por enlaces de hidrógeno a un nucleótido complementario. Un ácido nucleico parcialmente bicatenario puede tener al menos el 10 %, 25 %, 50 %, 60 %, 70 %, 80 %, 90 % o 95 % de sus nucleótidos unidos por enlaces de hidrógeno a un nucleótido complementario.

Es generalmente ventajoso que la región bicatenaria sea lo más corta posible sin pérdida de función. En este contexto, 'función' se refiere a la capacidad de la región bicatenaria para formar un dúplex estable en condiciones de reacción estándar para una reacción de ligamiento de ácidos nucleicos catalizada por enzimas, que serán bien conocidas por el lector experto (p. ej., incubación a una temperatura comprendida en el intervalo de 4 °C a 25 °C en un tampón de ligamiento apropiado para la enzima), de forma que las dos cadenas que forman el adaptador universal permanezcan parcialmente hibridadas durante el ligamiento del adaptador universal a una molécula diana. No es absolutamente necesario que la región bicatenaria sea estable en las condiciones normalmente utilizadas en los pasos de hibridación de la extensión de cebador o en las reacciones PCR.

La región bicatenaria de los adaptadores universales suele ser idéntica en todos los adaptadores universales utilizados en un ligamiento. Dado que los adaptadores universales se ligan a ambos extremos de cada molécula diana, el ácido nucleico diana modificado estará flanqueado por secuencias complementarias procedentes de la región bicatenaria de los adaptadores universales. Cuanto más larga sea la región bicatenaria y, por tanto, las secuencias complementarias procedentes de la misma en las construcciones de ácido nucleico diana modificado, mayor será la posibilidad de que la construcción de ácido nucleico diana modificado sea capaz de plegarse y emparejarse consigo misma en estas regiones de autocomplementariedad interna en las condiciones de hibridación utilizadas en la extensión de cebadores y/o la PCR. Por lo tanto, generalmente se prefiere que la región bicatenaria tenga 20 o menos, 15 o menos, o 10 o menos pares de bases de longitud para reducir este efecto. La estabilidad de la región bicatenaria puede aumentarse y, por lo tanto, su longitud verse potencialmente reducida, por la inclusión de nucleótidos no naturales que muestran emparejamiento de bases más fuerte que los pares de bases estándar de Watson-Crick.

En una realización, las dos cadenas del adaptador universal son el 100 % complementarias en la región bicatenaria. Se apreciará que pueden tolerarse uno o más emparejamientos erróneos de nucleótidos dentro de la región bicatenaria, siempre que las dos cadenas sean capaces de formar un dúplex estable en condiciones de ligadura convencionales.

Los adaptadores universales para su uso en la presente memoria generalmente incluirán una región bicatenaria que forma el extremo 'ligable' del adaptador, p. ej., el extremo que está unido a un ácido nucleico diana bicatenario en la reacción de ligamiento. El extremo ligable del adaptador universal puede ser romo o, en otras realizaciones, puede haber presentes protuberancias 5' o 3' cortas de uno o más nucleótidos para facilitar/propiciar el ligamiento. El nucleótido terminal 5' en el extremo ligable del adaptador universal está típicamente fosforilado para permitir la unión fosfodiéster a un grupo hidroxilo 3' en el polinucleótido diana.

La expresión "región no emparejada" se refiere a una región del adaptador universal, la región de cadenas de ácido nucleico monocatenario no complementarias, en donde las secuencias de las dos cadenas de polinucleótido que forman el adaptador universal presentan un grado de no complementariedad tal que las dos cadenas no son capaces de hibridar completamente entre sí en condiciones de hibridación estándar para una reacción de extensión de cebador o de PCR. La región o regiones no emparejadas pueden presentar cierto grado de hibridación en condiciones de reacción estándar para una reacción de ligamiento catalizada por enzimas, siempre que las dos cadenas vuelvan a la forma monocatenaria en condiciones de hibridación en una reacción de amplificación.

Debe entenderse que la "región no emparejada" la proporcionan diferentes porciones de las mismas dos cadenas de polinucleótidos que forman la región o regiones bicatenarias. Los emparejamientos erróneos en la construcción del adaptador pueden tomar la forma de una cadena más larga que la otra, de modo que hay una región monocatenaria en una de las cadenas, o una secuencia seleccionada de tal manera que las dos cadenas no hibridan y, por lo tanto, forman una región monocatenaria en ambas cadenas. Los emparejamientos erróneos también pueden tomar la forma de 'burbujas', en donde ambos extremos de la construcción o construcciones del adaptador universal son capaces de hibridar entre sí y formar un dúplex, pero la región central no. Las partes de la cadena o cadenas que forman la región no emparejada no están hibridadas en condiciones en que otras partes de las mismas dos cadenas hibridan para formar una o más regiones bicatenarias. Para evitar dudas, debe entenderse que una protuberancia monocatenaria o de una sola base en el extremo 3' de un dúplex de polinucleótido que posteriormente se somete a ligamiento con las secuencias diana no constituye una "región no emparejada" en el contexto de la presente memoria.

El límite inferior de la longitud de la región no emparejada típicamente estará determinado por la función, por ejemplo, la necesidad de proporcionar una secuencia adecuada para i) la unión de un cebador para extensión de cebadores, PCR y/o secuenciación (por ejemplo, unión de un cebador a un sitio de unión de cebador universal), o para ii) la unión de una secuencia de unión de captura universal a una secuencia de captura para la inmovilización de un ácido nucleico diana modificado a una superficie. Teóricamente no hay límite superior en la longitud de la región no emparejada, excepto que generalmente es ventajoso minimizar la longitud total del adaptador universal, por ejemplo, para facilitar la separación de adaptadores universales no unidos de construcciones de ácido nucleico diana modificado tras el paso de ligamiento. Por lo tanto, generalmente se prefiere que la región no emparejada tenga menos de 50, o menos de 40, o menos de 30, o menos de 25 nucleótidos consecutivos de longitud.

La región de cadenas de ácido nucleico monocatenario no complementario incluye al menos una secuencia de unión de captura universal en el extremo 3' (véase la FIG. 1B, secuencia de unión de captura universal 150). El extremo 3' de un adaptador universal incluye una primera secuencia de unión de captura universal que hibridará con una primera secuencia de captura presente en un ácido nucleico de captura. Por ejemplo, como se muestra en la FIG. 2, un ácido nucleico 200 de una primera población de ácidos nucleicos de captura incluye una primera secuencia de captura 210. Una cadena de un ácido nucleico diana 230 modificado que incluye una primera secuencia de unión de captura universal 250 en el extremo 3' de una cadena individual se muestra hibridada con la primera secuencia de captura 210. Es la interacción entre la primera secuencia de unión de captura universal 250 y la primera secuencia de captura 210 que se altera para reducir la afinidad y codificar un retraso cinético en los ácidos nucleicos diana sembrados en un pocillo. Los métodos de ExAmp estándar utilizan secuencias de unión de captura universales y secuencias de captura que son completamente complementarias en toda la longitud de la secuencia de captura. Los métodos de ExAmp descritos en la presente memoria utilizan secuencias de unión de captura universales que incluyen uno o más emparejamientos erróneos, tienen una longitud reducida o una combinación

de los mismos. El resultado del emparejamiento erróneo o emparejamientos erróneos y/o la longitud reducida es una afinidad reducida entre las dos secuencias en comparación con la afinidad de las dos secuencias de longitud completa completamente complementarias. La afinidad reducida provoca una disminución en la eficiencia de amplificación, donde la eficiencia de amplificación resultante es, generalmente, una función del número de diferencias entre la secuencia de unión de captura universal y la secuencia de captura.

Opcionalmente, el extremo 5' de un adaptador universal incluye una segunda secuencia de unión de captura universal unida a cada extremo de un ácido nucleico diana, donde la segunda secuencia de unión de captura universal hibridará con una segunda secuencia de captura presente en un ácido nucleico de captura. Por ejemplo, como se muestra en la **FIG. 1B**, la secuencia de unión de captura universal **180**. Por lo tanto, a menos que se indique lo contrario, el siguiente análisis sobre cómo se ajusta una secuencia de unión de captura universal para reducir la afinidad se aplica a las secuencias de unión de captura universales tanto 3' como 5'.

El extremo 3' de una secuencia de captura sirve como punto de inicio para la síntesis de ADN mediante una ADN polimerasa en los métodos descritos en la presente memoria. El experto reconocerá que el nucleótido en el extremo 3' de una secuencia de captura y el nucleótido correspondiente en la secuencia de unión de captura universal deben ser complementarios para conservar la capacidad de una ADN polimerasa para iniciar la síntesis de ADN.

Una secuencia de unión de captura universal puede incluir uno o más nucleótidos que no son complementarios con la secuencia de captura. En una realización, una secuencia de unión de captura universal puede incluir de 1 a 5 nucleótidos con emparejamiento erróneo (también denominados nucleótidos no complementarios), por ejemplo, al menos 1, al menos 2, al menos 3, al menos 4 o 5 nucleótidos con emparejamiento erróneo en comparación con una secuencia de captura utilizada en una reacción de amplificación descrita en la presente memoria. Un nucleótido con emparejamiento erróneo puede ser un emparejamiento erróneo oscilante o un emparejamiento erróneo verdadero.

Un emparejamiento erróneo oscilante se refiere a una posición en que los cuatro nucleótidos están representados en la población de la secuencia de unión de captura universal. Por ejemplo, si N es el nucleótido oscilante en ACTNGC, entonces la población de la secuencia de unión de captura universal incluirá ACTTGC, ACTAGC, ACTCGC y ACTGGC, y el 25 % de las secuencias de unión de captura universales en la población serán complementarias con el correspondiente nucleótido de la secuencia de captura. En una realización, una secuencia de unión de captura universal puede incluir de 1 a 5 nucleótidos oscilantes, por ejemplo, al menos 1, al menos 2, al menos 3, al menos 4 o 5 nucleótidos oscilantes en comparación con una secuencia de captura usada en una reacción de amplificación descrita en la presente memoria. En una realización, los nucleótidos oscilantes pueden ubicarse en cualquier lugar de la secuencia de unión de captura universal.

Un emparejamiento erróneo verdadero se refiere a una posición donde sólo tres de los cuatro nucleótidos están representados en una posición particular en la población de la secuencia de unión de captura universal. Por ejemplo, si G es la ubicación del verdadero nucleótido con emparejamiento erróneo en ACTTGC, entonces la población de la secuencia de unión de captura universal incluirá ACTTCC, ACTTTC y ACTTAC, y ninguna de las secuencias de unión de captura universales en la población será complementaria con el nucleótido correspondiente, una C en este ejemplo, de la secuencia de captura. En una realización, una secuencia de unión de captura universal puede incluir de 1 a 5 nucleótidos con emparejamiento erróneo, por ejemplo, al menos 1, al menos 2, al menos 3, al menos 4 o 5 nucleótidos oscilantes en comparación con una secuencia de captura usada en una reacción de amplificación descrita en la presente memoria. En una realización, los nucleótidos oscilantes pueden ubicarse en cualquier lugar de la secuencia de unión de captura universal.

El experto reconocerá que el uso de un emparejamiento erróneo oscilante o un emparejamiento erróneo verdadero proporciona un mayor control de la alteración de la afinidad de una secuencia de unión de captura universal. El uso de una secuencia de unión de captura universal con solo un nucleótido oscilante da como resultado que el 25 % de las secuencias de unión de captura universales tengan complementariedad en esa posición, mayor afinidad que el otro 75 % y una eficiencia de amplificación esperada mayor que el otro 75 %. El uso de una secuencia de unión de captura universal con solo un único nucleótido de emparejamiento erróneo verdadero da como resultado que todas las secuencias de unión de captura universales no tengan complementariedad en esa posición, afinidad reducida y una eficiencia de amplificación reducida esperada.

En otra realización, una secuencia de unión de captura universal tiene una longitud más corta que da como resultado una afinidad que es menor que la afinidad entre la secuencia de unión de captura universal de longitud completa y la secuencia de captura. Las secuencias de captura útiles en los métodos de amplificación estándar descritos en la presente memoria típicamente tienen una longitud de aproximadamente 20 a aproximadamente 30 nucleótidos, aunque pueden ser más largas o más cortas si es necesario. Una secuencia de unión de captura universal útil en los métodos descritos en la presente memoria puede tener una longitud que es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 o 12 nucleótidos más corta que la secuencia de captura utilizada en una reacción de amplificación descrita en la presente memoria. En una realización, la longitud de la secuencia de unión de captura universal se reduce mediante la eliminación de nucleótidos del extremo 3' de la primera secuencia de unión de captura universal y/o del extremo 5' de la segunda secuencia de unión de captura universal.

Una reacción de amplificación descrita en la presente memoria puede usar una población heterogénea de secuencias de unión de captura universales (p. ej., una pluralidad de ácidos nucleicos diana diferentes pueden incluir una población heterogénea de secuencias de unión de captura universales presentes en los extremos 3' y opcionalmente presentes en los

extremos 5'). En una realización, la población heterogénea incluye secuencias de unión de captura universales individuales que tienen nucleótidos con emparejamiento erróneo. En una realización, las secuencias de unión de captura universales tienen 1, 2, 3, 4 o 5 nucleótidos con emparejamiento erróneo. Los nucleótidos con emparejamiento erróneo pueden ser emparejamientos erróneos oscilantes, emparejamientos erróneos verdaderos o una combinación de los mismos.

5 En una realización, la población heterogénea incluye secuencias de unión de captura universales individuales que tienen una longitud acortada. En una realización, la población heterogénea incluye secuencias de unión de captura universales individuales que tienen una longitud que es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 o 12 nucleótidos más corta que la secuencia de captura utilizada en una reacción de amplificación descrita en la presente memoria.

10 En una realización, la población heterogénea incluye secuencias de unión de captura universales individuales que tienen una combinación de uno o más nucleótidos con emparejamiento erróneo y una longitud acortada. El número de nucleótidos con emparejamiento erróneo y el número de nucleótidos faltantes en la secuencia de unión de captura universal puede estar presentes en cualquier combinación, por ejemplo, el número de nucleótidos con emparejamiento erróneo y el número de nucleótidos faltantes son independientes.

15 La población heterogénea también puede incluir ácidos nucleicos diana individuales que tienen en los extremos 3', y opcionalmente en los extremos 5', una secuencia de unión de captura universal que tiene el 100 % de complementariedad con la secuencia de captura. Las relaciones molares de las diferentes secuencias de unión de captura universal en una población heterogénea pueden ser iguales o estar alteradas. En las realizaciones donde la relación molar no es igual, se prefieren relaciones molares más altas de las secuencias de unión de captura universales que tienen una eficiencia de amplificación más alta. En consecuencia, en las realizaciones donde la población heterogénea incluye una secuencia de unión de captura universal que tiene el 100 % de complementariedad con la secuencia de captura, la secuencia de unión de captura universal que tiene el 100 % de complementariedad puede estar presente en una proporción mayor que cualquier otro miembro de la población heterogénea.

20 La región de cadenas de ácido nucleico monocatenario no complementario típicamente también incluye al menos un sitio de unión de cebador universal. Un sitio de unión de cebador universal es una secuencia universal que puede usarse para la amplificación y/o la secuenciación de un ácido nucleico diana ligado al adaptador universal.

30 La región de cadenas de ácido nucleico monocatenario no complementario también puede incluir al menos un índice. Se puede utilizar un índice como marcador característico de la fuente de un ácido nucleico diana particular en una matriz. Generalmente, el índice es una secuencia sintética de nucleótidos que forma parte del adaptador universal que se añade a los ácidos nucleicos diana como parte del paso de preparación de la biblioteca. En consecuencia, un índice es una secuencia de ácido nucleico que está unida a cada una de las moléculas diana de una muestra particular, cuya presencia es indicativa de, o se utiliza para identificar, la muestra o fuente de la que se aislaron las moléculas diana.

35 Preferiblemente, el índice puede ser de hasta 20 nucleótidos de longitud, más preferiblemente de 1 - 10 nucleótidos y muy preferiblemente de 4 - 8 nucleótidos de longitud. Por ejemplo, un índice de cuatro nucleótidos ofrece la posibilidad de multiplexar 256 (4⁴) muestras en la misma matriz, mientras que un índice de seis bases permite procesar 4.096 (4⁶) muestras en la misma matriz.

40 En una realización, la secuencia de unión de captura universal es parte del adaptador universal cuando se liga a los fragmentos diana bicatenarios, y en otra realización el sitio de unión de extensión de cebador universal se añade al adaptador universal después de que el adaptador universal se liga a los fragmentos diana bicatenarios. La adición se puede lograr usando métodos rutinarios, incluidos métodos basados en PCR.

45 La secuencia nucleotídica precisa de los adaptadores universales generalmente no es material para la invención y puede seleccionarla el usuario de forma que los elementos de secuencia deseados se incluyan finalmente en las secuencias comunes de la pluralidad de diferentes ácidos nucleicos diana modificados, por ejemplo, para proporcionar las secuencias de unión de captura universales y los sitios de unión para conjuntos particulares de cebadores de amplificación universales y/o cebadores de secuenciación. Pueden incluirse elementos de secuencia adicionales, por ejemplo, para proporcionar sitios de unión para cebadores de secuenciación que se utilizarán en última instancia en la secuenciación de ácidos nucleicos diana en la biblioteca, o productos procedentes de la amplificación de los ácidos nucleicos diana en la biblioteca, por ejemplo, en un soporte sólido.

50 Aunque la secuencia de nucleótidos precisa del adaptador universal es generalmente no limitante para la descripción, las secuencias de las cadenas individuales en la región no emparejada deben ser de tal manera que ninguna cadena individual presente ninguna autocomplementariedad interna que pudiera conducir a una autohibridación, formación de estructuras de horquilla, etc., en condiciones de hibridación estándar. La autohibridación de una cadena en la región no emparejada debe impedirse ya que esto puede evitar o reducir la unión específica de un cebador de amplificación a esta cadena.

60 Los adaptadores con emparejamiento erróneo se forman preferiblemente a partir de dos cadenas de ADN, pero pueden incluir mezclas de nucleótidos naturales y no naturales (p. ej., uno o más ribonucleótidos) unidos mediante una mezcla de enlaces de cadena principal de enlaces fosfodiéster y no fosfodiéster.

65

Ligamiento y amplificación

Los métodos de ligamiento son conocidos en la técnica y utilizan métodos estándar. Dichos métodos utilizan enzimas ligasa, tales como la ADN ligasa, para efectuar o catalizar la unión de los extremos de las dos cadenas de polinucleótido de, en este caso, el adaptador universal y los ácidos nucleicos diana bicatenarios, de forma que se formen enlaces covalentes. El adaptador universal puede contener un resto 5'-fosfato para facilitar el ligamiento al 3'-OH presente en el fragmento diana. El ácido nucleico diana bicatenario contiene un resto 5'-fosfato, bien residual del proceso de corte, bien añadido utilizando un paso de tratamiento enzimático, y ha sido reparado en sus extremos, y opcionalmente extendido en una base o en bases protuberantes, para dar un 3'-OH adecuado para el ligamiento. En este contexto, unión significa enlace covalente de cadenas de polinucleótido que no estaban previamente unidas covalentemente. En un aspecto particular de la descripción, dicha unión tiene lugar mediante la formación de un enlace fosfodiéster entre las dos cadenas de polinucleótido, pero pueden usarse otros medios de enlace covalente (por ejemplo, enlaces de cadena principal de enlaces no fosfodiéster).

Como se analiza en la presente memoria, en una realización los adaptadores universales usados en el ligamiento están completos e incluyen una secuencia de unión de captura universal y otras secuencias universales, por ejemplo, un sitio de unión de cebador universal y una secuencia índice. La pluralidad resultante de ácidos nucleicos diana se puede usar para preparar muestras inmovilizadas para secuenciación.

Además, como se analiza en la presente memoria, en una realización los adaptadores universales usados en el ligamiento incluyen un sitio de unión de cebador universal y una secuencia índice, y no incluyen una secuencia de unión de captura universal. La pluralidad resultante de ácidos nucleicos diana modificados se puede modificar además para incluir secuencias específicas, tales como una secuencia de unión de captura universal. Los métodos para la adición de secuencias específicas, tales como una secuencia de unión de captura universal, a cebadores universales que se ligan a fragmentos diana bicatenarios incluyen métodos basados en PCR, y son conocidos en la técnica y se describen, por ejemplo, en Bignell y col. (documento US 8,053,192) y Gunderson y col. (documento WO2016/130704).

En las realizaciones donde se modifica un adaptador universal, se prepara una reacción de amplificación. Los expertos en la técnica conocen el contenido de una reacción de amplificación e incluyen sustratos apropiados (tales como los dNTP), enzimas (p. ej., una ADN polimerasa) y componentes de tampón necesarios para una reacción de amplificación. Generalmente, las reacciones de amplificación precisan al menos dos cebadores de amplificación, frecuentemente denominados cebadores "directos" e "inversos" (oligonucleótidos cebadores) que son capaces de hibridarse específicamente con una parte de la secuencia de polinucleótido que se va a amplificar, p. ej., un ácido nucleico diana, bajo condiciones encontradas en el paso de hibridación del cebador de cada ciclo de una reacción de amplificación. Se apreciará que si los cebadores contienen cualquier secuencia de nucleótidos que no hibrida con los ácidos nucleicos diana modificados en el primer ciclo de amplificación, entonces esta secuencia puede copiarse a los productos de amplificación. Por ejemplo, en el uso de cebadores que tienen secuencias de unión de captura universales, es decir, secuencias que no hibridan con los ácidos nucleicos diana modificados, las secuencias de unión de captura universales se incorporarán al amplicón resultante.

Los cebadores de amplificación son generalmente estructuras de polinucleótido monocatenarios. También pueden contener una mezcla de bases naturales y no naturales y también enlaces de cadena principal naturales y no naturales, siempre que cualquier modificación no natural no impida la función como cebador, que se define como la capacidad de hibridar con una cadena de polinucleótido molde durante las condiciones de la reacción de amplificación y para actuar como punto de inicio para la síntesis de una nueva cadena de polinucleótido complementaria a la cadena molde. Los cebadores pueden incluir de forma adicional modificaciones químicas no nucleotídicas, por ejemplo fosforotioatos para aumentar la resistencia a exonucleasas, siempre que las modificaciones no impidan la función del cebador.

Preparación de muestras inmovilizadas para secuenciación

Un método de la presente memoria puede incluir hacer reaccionar un reactivo de amplificación (una matriz de sitios de amplificación y una pluralidad de ácidos nucleicos diana modificados diferentes) para producir una pluralidad de sitios de amplificación que incluyen cada uno una población clonal de amplicones de un ácido nucleico diana individual que ha sembrado el sitio. En las reacciones estándar, la amplificación de exclusión se produce debido a la velocidad relativamente lenta de siembra de ácido nucleico diana (p. ej., difusión o transporte relativamente lentos) frente a la velocidad relativamente rápida a la que se produce la amplificación para llenar el sitio con copias de la semilla de ácido nucleico. En los métodos descritos en la presente memoria, la amplificación de exclusión puede producirse debido a un retraso cinético en la formación de una primera copia de un ácido nucleico diana que ha sembrado un sitio frente a la velocidad relativamente rápida a la que se fabrican las copias posteriores para llenar el sitio. Por ejemplo, un sitio individual puede haber sido sembrado con varios ácidos nucleicos diana diferentes, teniendo cada uno una secuencia de unión de captura universal diferente (p. ej., una pluralidad de ácidos nucleicos diana modificados diferentes incluye una población heterogénea de secuencias de unión de captura universales). Sin embargo, se espera que la formación de la primera copia de cualquier ácido nucleico diana dado dependa de la eficiencia de amplificación de su secuencia de unión de captura universal, de forma que la velocidad de formación promedio de la primera copia sea relativamente lenta en comparación con la velocidad a la que se generan las copias posteriores. En este caso, aunque un sitio individual puede haber sido sembrado con varios ácidos nucleicos diana diferentes, sólo uno comenzará la amplificación en primer lugar, y la amplificación de exclusión permitirá típicamente que sólo ese ácido nucleico diana llene el sitio de amplificación. Más específicamente, una vez que un primer ácido nucleico

diana comienza la amplificación, el sitio se llenará rápidamente con sus copias, impidiendo de este modo que se fabriquen copias de un segundo ácido nucleico diana en el sitio.

En algunas realizaciones, la clonalidad aparente puede lograrse incluso si un sitio de amplificación no se llena hasta su capacidad antes de que un segundo ácido nucleico diana comience la amplificación en el sitio. En algunas condiciones, la amplificación de un primer ácido nucleico diana puede continuar hasta un punto en que se produce una cantidad suficiente de copias para superar o saturar eficazmente la producción de copias de un segundo ácido nucleico diana que se transporta al sitio. Por ejemplo, en una realización en que se utiliza un proceso de amplificación en puente en una casilla circular con un diámetro menor de 500 nm, se ha determinado que después de 14 ciclos de amplificación exponencial para un primer ácido nucleico diana, la contaminación de un segundo ácido nucleico diana en el mismo sitio producirá una cantidad insuficiente de amplicones contaminantes que afecte negativamente al análisis de secuenciación por síntesis en una plataforma de secuenciación Illumina.

No es necesario que los sitios de amplificación en una matriz sean completamente clonales en todas las realizaciones. En cambio, para algunas aplicaciones, un sitio de amplificación individual puede estar predominantemente poblado con amplicones de un primer ácido nucleico diana y también puede tener un nivel bajo de amplicones contaminantes de un segundo ácido nucleico diana. Una matriz puede tener uno o más sitios de amplificación que tengan un nivel bajo de amplicones contaminantes siempre que el nivel de contaminación no tenga un impacto inaceptable en un uso posterior de la matriz. Por ejemplo, cuando la matriz va a utilizarse en una aplicación de detección, un nivel aceptable de contaminación sería un nivel que no altere la relación señal a ruido o la resolución de la técnica de detección de una forma inaceptable. Por consiguiente, en general, la clonalidad aparente será relevante para un uso o aplicación particular de una matriz creada mediante los métodos expuestos en la presente memoria. Niveles de contaminación ilustrativos que pueden ser aceptables en un sitio de amplificación individual para aplicaciones particulares incluyen, aunque no de forma limitativa, como máximo 0,1 %, 0,5 %, 1 %, 5 %, 10 % o 25 % de amplicones contaminantes. Una matriz puede incluir uno o más sitios de amplificación que tengan estos niveles ilustrativos de amplicones contaminantes. Por ejemplo, hasta un 5 %, 10 %, 25 %, 50 %, 75 % o incluso 100 % de los sitios de amplificación en una matriz pueden tener algunos amplicones contaminantes.

Aunque el uso de cebadores diferencialmente activos para provocar diferentes velocidades de formación del primer amplicón y de amplicones posteriores se ha ilustrado anteriormente para una realización donde los ácidos nucleicos diana están presentes en los sitios de amplificación antes de la amplificación, el método también se puede llevar a cabo en condiciones en donde los ácidos nucleicos diana se transportan (p. ej., mediante difusión) a los sitios de amplificación a medida que se produce la amplificación. Por lo tanto, la amplificación de exclusión puede aprovechar tanto una velocidad de transporte relativamente lenta como una producción relativamente lenta del primer amplicón con respecto a la formación de amplicones posteriores. Por lo tanto, se puede llevar a cabo una reacción de amplificación expuesta en la presente memoria de manera que los ácidos nucleicos diana se transporten de la solución a los sitios de amplificación simultáneamente con (i) la producción de un primer amplicón, y (ii) la producción de los amplicones posteriores en otros sitios de amplificación de la matriz. En realizaciones particulares, la velocidad promedio a la que se generan los amplicones posteriores en los sitios de amplificación puede superar la velocidad promedio a la que los ácidos nucleicos diana se transportan de la solución a los sitios de amplificación. En algunos casos, se puede generar un número suficiente de amplicones a partir de un único ácido nucleico diana en un sitio de amplificación individual para llenar la capacidad del sitio de amplificación respectivo. La velocidad a la que se generan amplicones para llenar la capacidad de los respectivos sitios de amplificación puede, por ejemplo, superar la velocidad a la que los ácidos nucleicos diana individuales se transportan de la solución a los sitios de amplificación.

Un reactivo de amplificación que se usa en un método expuesto en la presente memoria es preferiblemente capaz de fabricar rápidamente copias de ácidos nucleicos diana en sitios de amplificación. Típicamente, un reactivo de amplificación usado en un método de la presente memoria incluirá una polimerasa y nucleótidos trifosfato (los NTP). Se puede usar cualquiera de una diversidad de polimerasas conocidas en la técnica, pero en algunas realizaciones, puede ser preferible usar una polimerasa que sea exonucleasa negativa. Los NTP pueden ser desoxirribonucleótidos trifosfato (los dNTP) para realizaciones donde se realizan copias de ADN. Típicamente, las cuatro especies naturales, dATP, dTTP, dGTP y dCTP, estarán presentes en un reactivo de amplificación de ADN; sin embargo, se pueden utilizar análogos si se desea. Los NTP pueden ser ribonucleótidos trifosfato (rNTP) para realizaciones donde se realizan copias de ARN. Típicamente, las cuatro especies naturales, rATP, rUTP, rGTP y rCTP, estarán presentes en un reactivo de amplificación de ARN; sin embargo, se pueden utilizar análogos si se desea.

Un reactivo de amplificación puede incluir componentes adicionales que faciliten la formación de amplicones y, en algunos casos, aumenten la velocidad de formación de amplicones. Un ejemplo es una proteína de carga de recombinasa. La recombinasa puede facilitar la formación de amplicones al permitir la invasión/extensión repetida. Más específicamente, la recombinasa puede facilitar la invasión de un ácido nucleico diana por la polimerasa y la extensión de un cebador por la polimerasa utilizando el ácido nucleico diana como molde para la formación de amplicones. Este proceso puede repetirse como una reacción en cadena donde los amplicones producidos a partir de cada ronda de invasión/extensión sirven como moldes en una ronda posterior. El proceso puede producirse más rápidamente que la PCR estándar, ya que no se requiere un ciclo de desnaturalización (por ejemplo, mediante calentamiento o desnaturalización química). Como tal, la amplificación facilitada por recombinasa puede llevarse a cabo isotérmicamente. De forma general, es deseable incluir ATP u otros nucleótidos (o, en algunos casos, análogos no hidrolizables de los mismos) en un reactivo de amplificación facilitada por recombinasa para facilitar la amplificación. Es particularmente útil una mezcla de recombinasa, proteína de unión

monocatenaria (SSB) y proteína accesoria. Formulaciones ilustrativas para la amplificación facilitada por recombinasa incluyen las vendidas comercialmente como kits TwistAmp por TwistDx (Cambridge, Reino Unido). Los componentes útiles del reactivo de amplificación facilitada por recombinasa y las condiciones de reacción se exponen en la patente estadounidense N.º 5.223.414 y la patente estadounidense N.º 7.399.590.

5 Otro ejemplo de un componente que puede incluirse en un reactivo de amplificación para facilitar la formación de amplicones y en algunos casos para aumentar la velocidad de formación de amplicones es una helicasa. La helicasa puede facilitar la formación de amplicones al permitir una reacción en cadena de la formación de amplicones. El proceso puede producirse
10 más rápidamente que la PCR estándar, ya que no se requiere un ciclo de desnaturalización (por ejemplo, mediante calentamiento o desnaturalización química). Como tal, la amplificación facilitada por helicasa puede llevarse a cabo isotérmicamente. Una mezcla de helicasa y proteína de unión monocatenaria (SSB) es especialmente útil ya que la SSB puede facilitar aún más la amplificación. Las formulaciones ilustrativas para la amplificación facilitada por helicasa incluyen las comercializadas como kits de IsoAmp de Biohelix (Beverly, Mass). Además, se describen ejemplos de formulaciones
15 útiles que incluyen una proteína helicasa en la patente estadounidense N.º 7.399.590 y la patente estadounidense N.º 7.829.284.

Otro ejemplo de componente que puede incluirse en un reactivo de amplificación para facilitar la formación de amplicones y, en algunos casos, aumentar la velocidad de formación de amplicones, es una proteína de unión al origen.

20 La presencia de reactivos de aglutinamiento molecular en la solución se puede utilizar para ayudar a la amplificación de exclusión. Los ejemplos de reactivos de aglutinamiento molecular útiles incluyen, pero no se limita a, polietilenglicol (PEG), Ficoll®, dextrano o alcohol polivinílico. Las formulaciones y reactivos de aglutinamiento molecular ilustrativos se exponen en la patente estadounidense N.º 7.399.590.

25 La velocidad a la que se produce una reacción de amplificación se puede aumentar aumentando la concentración o cantidad de uno o más de los componentes activos de una reacción de amplificación. Por ejemplo, se puede aumentar la cantidad o concentración de la polimerasa, los nucleótidos trifosfato, cebadores, recombinasa, helicasa o SSB para aumentar la velocidad de amplificación. En algunos casos, uno o más componentes activos de una reacción de amplificación que
30 aumentan en cantidad o concentración (o se manipulan de otro modo en un método expuesto en la presente memoria) son componentes no de ácido nucleico de la reacción de amplificación.

La velocidad de amplificación también se puede aumentar en un método expuesto en la presente memoria ajustando la temperatura. Por ejemplo, la velocidad de amplificación en uno o más sitios de amplificación se puede aumentar aumentando
35 la temperatura en el sitio o sitios hasta una temperatura máxima donde la velocidad de reacción disminuye debido a la desnaturalización u otros acontecimientos adversos. Las temperaturas óptimas o deseadas pueden determinarse a partir de propiedades conocidas de los componentes de amplificación en uso o empíricamente para una mezcla de reacción de amplificación determinada. Dichos ajustes se pueden realizar basándose en predicciones anteriores de la temperatura de fusión (T_m) del cebador o empíricamente.

40 La velocidad a la que se produce una reacción de amplificación se puede aumentar aumentando la actividad de uno o más reactivos de amplificación. Por ejemplo, se puede añadir un cofactor que aumente la velocidad de extensión de una polimerasa a una reacción donde se utiliza la polimerasa. En algunas realizaciones, se pueden añadir cofactores metálicos tales como magnesio, cinc o manganeso a una reacción de polimerasa o se puede añadir betaina.

45 En algunas realizaciones de los métodos expuestos en la presente memoria, se prefiere utilizar una población de ácidos nucleicos diana que sea bicatenaria. Se ha observado que la formación de amplicones en una matriz de sitios en condiciones de amplificación de exclusión es eficaz para ácidos nucleicos diana bicatenarios. Por ejemplo, se puede producir más eficientemente una pluralidad de sitios de amplificación que tienen poblaciones clonales de amplicones a partir de ácidos
50 nucleicos diana bicatenarios (en comparación con ácidos nucleicos diana monocatenarios a la misma concentración) en presencia de recombinasa y proteína de unión monocatenaria. Sin embargo, se entenderá que se pueden usar ácidos nucleicos diana monocatenarios en algunas realizaciones de los métodos expuestos en la presente memoria.

Un método expuesto en la presente memoria puede utilizar cualquiera de una diversidad de técnicas de amplificación. Las técnicas ilustrativas que se pueden usar incluyen, aunque no de forma limitativa, la reacción en cadena de la polimerasa
55 (PCR), la amplificación de círculo rodante (RCA), la amplificación de desplazamiento múltiple (MDA) o la amplificación primaria aleatoria (RPA). En algunas realizaciones, la amplificación se puede llevar a cabo en solución, por ejemplo, cuando los sitios de amplificación son capaces de contener amplicones en un volumen que tiene una capacidad deseada. Preferiblemente, una técnica de amplificación utilizada en condiciones de amplificación de exclusión en un método de la presente memoria se llevará a cabo en fase sólida. Por ejemplo, uno o más cebadores usados para la amplificación pueden
60 unirse a una fase sólida en el sitio de amplificación. En realizaciones de PCR, uno o ambos cebadores utilizados para la amplificación se pueden unir a una fase sólida. Los formatos que utilizan dos especies de cebadores unidos a la superficie se denominan frecuentemente amplificación puente porque los amplicones bicatenarios forman una estructura similar a un puente entre los dos cebadores unidos a la superficie que flanquean la secuencia molde que se ha copiado. Los reactivos y condiciones ilustrativos que se pueden usar para la amplificación puente se describen, por ejemplo, en las patentes
65 estadounidense N.º 5,641,658; publicación de patente estadounidense N.º 2002/0055100; patente estadounidense N.º 7.115.400; publicación de patente estadounidense N.º 2004/0096853; publicación de patente estadounidense N.º

2004/0002090; publicación de patente estadounidense N.º 2007/0128624; y publicación de patente estadounidense N.º 2008/0009420. La amplificación por PCR en fase sólida también se puede llevar a cabo con uno de los cebadores de amplificación unido a un soporte sólido y el segundo cebador en solución. Un formato ilustrativo que usa una combinación de un cebador unido a una superficie y un cebador soluble es la PCR en emulsión como se describe, por ejemplo, en Dressman y col., Proc. Natl. Acad. Sci. USA 100:8817-8822 (2003), WO 05/010145, o publicación de patente estadounidense N.º 2005/0130173 o 2005/0064460. La PCR en emulsión es ilustrativa del formato y se entenderá que para los fines de los métodos expuestos en la presente memoria, el uso de una emulsión es opcional y, de hecho, para varias realizaciones no se utiliza una emulsión. Las técnicas de PCR descritas se pueden modificar para amplificación no cíclica (p. ej., amplificación isotérmica) usando componentes ilustrados en otros lugares de la presente memoria para facilitar o aumentar la velocidad de amplificación. Por consiguiente, las técnicas de PCR descritas se pueden utilizar en condiciones de amplificación de exclusión.

Las técnicas de RCA pueden modificarse para su uso en un método de la presente divulgación. Los componentes ilustrativos que se pueden usar en una reacción de RCA y los principios por los cuales la RCA produce amplicones se describen, por ejemplo, en Lizardi y col., Nat. Genet. 19:225-232 (1998) y el documento US 2007/0099208 A1. Los cebadores utilizados para la RCA pueden estar en solución o unidos a una superficie de soporte de apoyo sólida en un lugar de amplificación. Las técnicas de RCA ilustradas en las referencias anteriores se pueden modificar según las enseñanzas de la presente memoria, por ejemplo, para aumentar la velocidad de amplificación para adaptarse a aplicaciones particulares. Por lo tanto, las técnicas de RCA se pueden utilizar en condiciones de amplificación de exclusión.

Las técnicas de MDA pueden modificarse para su uso en un método de la presente divulgación. Algunos principios básicos y condiciones útiles para la MDA se describen, por ejemplo, en Dean y col., Proc Natl. Acad. Sci. USA 99:5261-66 (2002); Lage y col., Genome Research 13:294-307 (2003); Walker y col., Molecular Methods for Virus Detection, Academic Press, Inc., 1995; Walker y col., Nucl. Acids Res. 20:1691-96 (1992); patente estadounidense N.º 5.455.166; patente estadounidense N.º 5.130.238; y la publicación de patente estadounidense N.º 6.214.587. Los cebadores utilizados para la MDA pueden estar en solución o unidos a una superficie de soporte de apoyo sólida en un lugar de amplificación. Las técnicas de MDA ilustradas en las referencias anteriores se pueden modificar según las enseñanzas de la presente memoria, por ejemplo, para aumentar la velocidad de amplificación para adaptarse a aplicaciones particulares. En consecuencia, las técnicas de MDA se pueden utilizar en condiciones de amplificación de exclusión.

En realizaciones particulares, se puede usar una combinación de las técnicas de amplificación descritas para fabricar una matriz en condiciones de amplificación de exclusión. Por ejemplo, se pueden usar RCA y MDA en una combinación en donde se utiliza RCA para generar un amplicón concatémico en solución (p. ej., usando cebadores en fase de solución). A continuación, el amplicón puede utilizarse como molde para MDA utilizando cebadores que se unen a una superficie de apoyo sólida en un sitio de amplificación. En este ejemplo, los amplicones producidos después de los pasos combinados de RCA y MDA se unirán a la superficie del sitio de amplificación.

Como se ilustra con respecto a varias de las realizaciones anteriores, un método de la presente memoria no necesita utilizar una técnica de amplificación cíclica. Por ejemplo, la amplificación de ácidos nucleicos diana se puede llevar a cabo en sitios de amplificación sin un ciclo de desnaturalización. Los ciclos de desnaturalización ilustrativos incluyen la introducción de desnaturalizantes químicos en una reacción de amplificación y/o el aumento de la temperatura de una reacción de amplificación. Por lo tanto, la amplificación de los ácidos nucleicos diana no necesita incluir un paso de sustitución de la solución de amplificación por un reactivo químico que desnaturalice los ácidos nucleicos diana y los amplicones. Similarmente, la amplificación de los ácidos nucleicos diana no necesita incluir el calentamiento de la solución a una temperatura que desnaturalice los ácidos nucleicos diana y los amplicones. En consecuencia, la amplificación de ácidos nucleicos diana en los sitios de amplificación se puede llevar a cabo de forma isotérmica durante la duración de un método expuesto en la presente memoria. De hecho, un método de amplificación expuesto en la presente memoria puede producirse sin una o más manipulaciones cíclicas que se llevan a cabo para algunas técnicas de amplificación en condiciones estándar. Adicionalmente, en algunas técnicas de amplificación en fase sólida estándar se lleva a cabo un lavado después de cargar los ácidos nucleicos diana en un sustrato y antes de que se inicie la amplificación. Sin embargo, en realizaciones de los presentes métodos, no es necesario llevar a cabo un paso de lavado entre el transporte de los ácidos nucleicos diana a los sitios de reacción y la amplificación de los ácidos nucleicos diana en los sitios de amplificación. En lugar de ello, se permite que el transporte (p. ej., mediante difusión) y la amplificación se produzcan simultáneamente para proporcionar una amplificación de exclusión.

En algunas realizaciones, puede ser deseable repetir un ciclo de amplificación que se produce en condiciones de amplificación de exclusión. Por lo tanto, aunque se pueden fabricar copias de un ácido nucleico diana en un sitio de amplificación individual sin manipulaciones cíclicas, se puede tratar cíclicamente una matriz de sitios de amplificación para aumentar el número de sitios que contienen amplicones después de cada ciclo. En realizaciones particulares, las condiciones de amplificación se pueden modificar de un ciclo al siguiente. Por ejemplo, una o más de las condiciones expuestas anteriormente para alterar la velocidad de transporte o alterar la velocidad de amplificación se pueden ajustar entre ciclos. Como tal, la tasa de transporte se puede aumentar de un ciclo a otro, la tasa de transporte se puede disminuir de un ciclo a otro, la tasa de amplificación se puede aumentar de un ciclo a otro o la tasa de amplificación se puede disminuir de un ciclo a otro.

Composiciones

Durante o después de un método de agrupamiento de amplificación descrito en la presente memoria, pueden producirse diferentes composiciones. En una realización, una composición incluye una matriz de sitios de amplificación. Cada sitio incluye ácidos nucleicos de captura primero y segundo que incluyen secuencias de captura primera y segunda, respectivamente, donde los ácidos nucleicos de captura primero y segundo están unidos a la superficie de los sitios. Los diferentes sitios de la matriz incluyen ácidos nucleicos diana hibridados a la primera secuencia de captura del primer ácido nucleico de captura. Cada uno de los ácidos nucleicos diana en los diferentes sitios incluye en el extremo 3' una secuencia de unión de captura universal que hibrida con la secuencia de captura. Están presentes secuencias de unión de captura universales que tienen menos afinidad por la secuencia de captura que una secuencia de unión de captura universal que tiene el 100 % de complementariedad con la primera secuencia de captura. En una realización, en cada sitio están presentes diferentes secuencias de unión de captura universales, p. ej., está presente una primera población heterogénea de secuencias de unión de captura universales. La primera población heterogénea puede incluir al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7 u al menos 8 secuencias de unión de captura universales diferentes.

En una realización, la primera secuencia de unión de captura universal incluye de 1 a 5 nucleótidos que no son complementarios con la primera secuencia de captura. La composición puede incluir algunos ácidos nucleicos diana que tienen una secuencia de unión de captura universal con el 100 % de complementariedad con la primera secuencia de captura. En una realización, los miembros de la primera población heterogénea que tienen el 100 % de complementariedad con la primera secuencia de captura están presentes en un número mayor que los otros miembros de la primera población heterogénea. La primera población heterogénea también puede incluir primeras secuencias de unión de captura universales individuales que tienen una longitud que es menor que la longitud de la primera secuencia de captura, tal como una longitud que es de 1 a 12 nucleótidos menor que la longitud de la primera secuencia de captura. Los miembros individuales de la primera población heterogénea pueden tener tanto una longitud que es menor que la longitud de la primera secuencia de captura e incluir de 1 a 5 nucleótidos que no son complementarios a la secuencia de la primera secuencia de captura, o el 100 % de complementariedad con la secuencia de la primera secuencia de captura.

El extremo 5' puede opcionalmente incluir una segunda secuencia de unión de captura universal que tiene un complemento que tiene menos afinidad por la segunda secuencia de captura que una segunda secuencia de unión de captura universal que tiene un complemento con el 100 % de complementariedad con la segunda secuencia de captura. En una realización, el complemento de la segunda secuencia de unión de captura universal incluye de 1 a 5 nucleótidos que no son complementarios con la segunda secuencia de captura. La composición puede incluir algunos ácidos nucleicos diana que tienen una segunda secuencia de unión de captura universal con un complemento que tiene el 100 % de complementariedad con la segunda secuencia de captura. En una realización, en cada sitio están presentes segundas secuencias de unión de captura universales diferentes, p. ej., está presente una segunda población heterogénea de segundas secuencias de unión de captura universales. La segunda población heterogénea puede incluir al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7 u al menos 8 segundas secuencias de unión de captura universales diferentes. En una realización, los miembros de una segunda población heterogénea con un complemento que tiene el 100 % de complementariedad con la segunda secuencia de captura están presentes en un número mayor que los otros miembros de la segunda población heterogénea. La segunda población heterogénea de secuencias de unión de captura universales en el extremo 5' también puede incluir segundas secuencias de unión de captura universales individuales que tienen una longitud que es menor que la longitud de la segunda secuencia de captura, tal como una longitud que es de 1 a 12 nucleótidos menos que la longitud de la segunda secuencia de captura. Los miembros individuales de la segunda población heterogénea pueden tener una longitud que es menor que la longitud de la segunda secuencia de captura e incluir un complemento que tiene de 1 a 5 nucleótidos que no son complementarios a la secuencia de la segunda secuencia de captura, o el 100 % de complementariedad con la secuencia de la segunda secuencia de captura.

Otra composición que puede producirse incluye una solución que incluye diferentes ácidos nucleicos diana bicatenarios de una única muestra o fuente, p. ej., una biblioteca, donde cada ácido nucleico diana incluye un adaptador universal unido en cada extremo. Los adaptadores universales incluyen una secuencia de unión de captura universal, y la secuencia de unión de captura universal es una población heterogénea. La población heterogénea puede incluir al menos 2, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7 u al menos 8 secuencias de unión de captura universales diferentes.

Uso en secuenciación/métodos de secuenciación

Una matriz de la presente memoria, que por ejemplo se ha producido mediante un método expuesto en la presente memoria y que incluye ácidos nucleicos diana amplificados en sitios de amplificación, se puede utilizar para cualquiera de una diversidad de aplicaciones. Una aplicación particularmente útil es la secuenciación de ácidos nucleicos. Un ejemplo es la secuenciación por síntesis (SBS). En la SBS, para determinar la secuencia de nucleótidos en el molde, se monitoriza la extensión de un cebador de ácido nucleico a lo largo de un molde de ácido nucleico (p. ej., un ácido nucleico diana o amplicón del mismo). El procesamiento químico subyacente puede ser polimerización (p. ej., catalizada por una enzima polimerasa). En una realización particular de SBS basada en polimerasa, se añaden nucleótidos etiquetados con fluorescencia a un cebador (extendiéndose de ese modo el cebador) de una manera dependiente del molde, de modo que, para determinar la secuencia del molde, pueda utilizarse la detección del orden y tipo de nucleótidos añadidos al cebador. Una pluralidad de moldes diferentes en sitios diferentes de una matriz expuesta en la presente memoria puede someterse a una técnica de SBS en condiciones en que los acontecimientos que se producen para distintos moldes pueden distinguirse debido a su ubicación en la matriz.

Las celdas de flujo proporcionan un formato conveniente para alojar una matriz que se produce por los métodos de la presente memoria y que se somete a una SBS u otra técnica de detección que implica el suministro repetido de reactivos en ciclos. Por ejemplo, para iniciar un primer ciclo de SBS, uno o más nucleótidos marcados, ADN polimerasa, etc., pueden fluir hacia/a través de una celda de flujo que aloja una matriz de moldes de ácidos nucleicos. Pueden detectarse aquellos sitios de una serie en los que la extensión del cebador provoca que se incorpore un nucleótido marcado. Opcionalmente, los nucleótidos pueden incluir además una propiedad de terminación reversible que finalice la extensión adicional del cebador una vez que se haya añadido un nucleótido a un cebador. Por ejemplo, a un cebador se le puede añadir un análogo de nucleótido que tenga una fracción de terminación reversible, de modo que no se pueda producir la extensión posterior hasta que se suministre un agente desbloqueante para eliminar la fracción. Por lo tanto, para realizaciones que usan terminación reversible, puede suministrarse un reactivo de desbloqueo a la célula de flujo (antes o después de que se produzca la detección). Los lavados pueden llevarse a cabo entre los diversos pasos de suministro. El ciclo puede repetirse a continuación n veces para extender el cebador en n nucleótidos, detectando de esta manera una secuencia de longitud n. Los procedimientos de SBS, sistemas fluidicos y plataformas de detección ilustrativos que pueden adaptarse fácilmente para su uso con una matriz producida por los métodos de la presente memoria se describen, por ejemplo, en Bentley y col., *Nature* 456:53-59 (2008), documento WO 04/018497; patente estadounidense N.º 7.057.026; WO 91/06678; WO 07/123.744; patente estadounidense N.º 7.329.492; patente estadounidense N.º 7.211.414; patente estadounidense N.º 7.315.019; patente estadounidense N.º 7.405.281, y la patente estadounidense N.º 8.343.746.

Pueden utilizarse otros procedimientos de secuenciación que utilizan reacciones cíclicas, como la pirosecuenciación. La pirosecuenciación detecta la liberación de pirofosfato inorgánico (PPI) a medida que se incorporan nucleótidos particulares en una cadena de ácido nucleico incipiente (Ronaghi, y col., *Analytical Biochemistry* 242(1), 84-9 (1996); Ronaghi, *Genome Res.* 11(1), 3-11 (2001); Ronaghi y col. *Science* 281(5375), 363 (1998); patente estadounidense N.º 6.210.891; patente estadounidense N.º 6.258.568 y la patente estadounidense US-6.274.320). En la pirosecuenciación, el PPI liberado se puede detectar convirtiéndolo inmediatamente en trifosfato de adenosina (ATP) por la ATP sulfurilasa, y el nivel de ATP generado se puede detectar mediante fotones producidos por luciferasa. Por lo tanto, la reacción de secuenciación se puede controlar a través de un sistema de detección de luminiscencia. Las fuentes de radiación de excitación utilizadas para los sistemas de detección a base de fluorescencia no son necesarias para los procedimientos de pirosecuenciación. Los sistemas fluidicos, detectores y procedimientos útiles que pueden utilizarse para la aplicación de la pirosecuenciación a las matrices de la presente memoria se describen, por ejemplo, en la patente publicada por la solicitud de patente publicada por la OMPI 2012/058096, documento US 2005/0191698 A1, patente estadounidense N.º 7.595.883, y la patente estadounidense N.º 7.244.559.

Las reacciones de secuenciación por ligación también son útiles, incluidas, por ejemplo, las descritas en Shendure y col. *Science* 309:1728-1732 (2005); patente estadounidense N.º 5.599.675; y la publicación de patente estadounidense N.º 5.750.341. Algunas realizaciones pueden incluir procedimientos de secuenciación por hibridación como se describe, por ejemplo, en Bains y col., *Journal of Theoretical Biology* 135(3), 303-7 (1988); Drmanac y col., *Nature Biotechnology* 16, 54-58 (1998); Fodor y col., *Science* 251(4995), 767-773 (1995); y el documento WO 1989/10977. Tanto en los procedimientos de secuenciación por ligamiento como en los de secuenciación por hibridación, los ácidos nucleicos molde (por ejemplo, un ácido nucleico diana o amplicones del mismo) que están presentes en los sitios de una matriz se someten a ciclos repetidos de administración y detección de oligonucleótidos. Los sistemas fluidicos para los métodos de SBS expuestos en la presente memoria o en las referencias citadas en la presente memoria se pueden adaptar fácilmente para el suministro de reactivos para procedimientos de secuenciación por ligadura o secuenciación por hibridación. Típicamente, los oligonucleótidos están marcados con fluorescencia y pueden detectarse usando detectores de fluorescencia similares a los descritos con respecto a los procedimientos de SBS en la presente memoria o en las referencias citadas en la presente memoria.

Algunas realizaciones pueden utilizar métodos que implican la monitorización en tiempo real de la actividad de la ADN polimerasa. Por ejemplo, las incorporaciones de nucleótidos pueden detectarse a través de interacciones de transferencia de energía por resonancia de fluorescencia (FRET) entre una polimerasa portadora de fluoróforo y nucleótidos marcados con γ -fosfato o con guías de ondas de modo cero (ZMW, por sus siglas en inglés). Se describen técnicas y reactivos de secuenciación basada en FRET, por ejemplo, en Levene y col. *Science* 299, 682-686 (2003); Lundquist y col. *Opt. Lett.* 33, 1026-1028 (2008); Korlach y col. *Proc. Natl. Acad. Sci. EE. UU.* 105, 1176-1181 (2008).

Algunas realizaciones de SBS (secuenciación por síntesis) incluyen la detección de un protón liberado tras la incorporación de un nucleótido en un producto de extensión. Por ejemplo, la secuenciación basada en la detección de protones liberados puede usar un detector eléctrico y técnicas asociadas que son comercializadas por Ion Torrent (Guilford, Conn., una subsidiaria de Life Technologies) o métodos y sistemas de secuenciación descritos en los documentos US 2009/0026082 A1; US 2009/0127589 A1; US 2010/0137143 A1; o US 2010/0282617 A1. Los métodos expuestos en la presente memoria para amplificar ácidos nucleicos diana utilizando amplificación de exclusión pueden aplicarse fácilmente a sustratos utilizados para detectar protones. Más específicamente, los métodos expuestos en la presente memoria pueden utilizarse para producir poblaciones clonales de amplicones en los sitios de las matrices que se utilizan para detectar protones.

Una aplicación útil para una matriz de la presente memoria, que por ejemplo se ha producido mediante un método expuesto en la presente memoria, es el análisis de la expresión génica. La expresión génica se puede detectar o cuantificar usando técnicas de secuenciación de ARN, tales como las denominadas secuenciación de ARN digital. Las técnicas de secuenciación de ARN se pueden realizar usando metodologías de secuenciación conocidas en la técnica, tales como las

expuestas anteriormente. La expresión génica también se puede detectar o cuantificar usando técnicas de hibridación llevadas a cabo mediante hibridación directa con una matriz o usando un ensayo múltiple, cuyos productos se detectan en una matriz. Una matriz de la presente memoria, por ejemplo, que se haya producido mediante un método expuesto en la presente memoria, también se puede usar para determinar genotipos para una muestra de ADN genómico de uno o más individuos. Los métodos ilustrativos para el análisis de expresión y genotipado basados en matriz que pueden llevarse a cabo en una matriz de la presente memoria se describen en las patentes estadounidenses N.º 7.582.420; 6.890.741; 6.913.884 o 6.355.431 o publicación de patente estadounidense N.º 2005/0053980 A1; 2009/0186349 A1 o US 2005/0181440 A1.

Otra aplicación útil para una matriz que se ha producido mediante un método expuesto en la presente memoria es la secuenciación de células individuales. Cuando se combina con métodos de indexación, la secuenciación de células individuales se puede utilizar en ensayos de accesibilidad a la cromatina para producir perfiles de elementos reguladores activos en miles de células individuales, y se pueden producir bibliotecas de genomas completos de células individuales. Los ejemplos de secuenciación de células individuales que se pueden llevar a cabo en una matriz de la presente memoria se describen en la solicitud de patente publicada estadounidense 2018/0023119 A1, números de serie de solicitudes provisionales estadounidense 62/673.023 y 62/680.259.

Una ventaja de los métodos expuestos en la presente memoria es que permiten la creación rápida y eficiente de matrices a partir de cualquiera de una diversidad de bibliotecas de ácidos nucleicos. En consecuencia, la presente memoria proporciona sistemas integrados capaces de fabricar una matriz usando uno o más de los métodos expuestos en la presente memoria y además capaces de detectar ácidos nucleicos en las matrices usando técnicas conocidas en la técnica tales como las ilustradas anteriormente. Por lo tanto, un sistema integrado de la presente memoria puede incluir componentes fluidicos capaces de suministrar reactivos de amplificación a una matriz de sitios de amplificación, tales como bombas, válvulas, depósitos, líneas fluidicas y similares. Un componente fluidoico particularmente útil es una celda de flujo. Se puede configurar y/o usar una celda de flujo en un sistema integrado para crear una matriz de la presente memoria y para detectar la matriz. Se describen celdas de flujo ilustrativas, por ejemplo, en el documento US 2010/0111768 A1 y en la patente estadounidense N.º 8.951.781. Como se ilustra en las cubetas de lectura, puede utilizarse uno o más de los componentes fluidicos de un sistema integrado para un método de amplificación y para un método de detección. Tomando como ejemplo una realización de secuenciación de ácidos nucleicos, puede utilizarse uno o más de los componentes fluidicos de un sistema integrado para un método de amplificación expuesto en la presente memoria y para el suministro de reactivos de secuenciación en un método de secuenciación como en los ilustrados anteriormente. Como alternativa, un sistema integrado puede incluir sistemas fluidicos distintos para llevar a cabo métodos de amplificación y para llevar a cabo métodos de detección. Los ejemplos de sistemas de secuenciación integrados que son capaces de crear matrices de ácidos nucleicos y también determinar la secuencia de los ácidos nucleicos incluyen, sin limitación, las plataformas MiSeq™, HiSeq™, NextSeq™, MiniSeq™, NovaSeq™ e iSeq™ (Illumina, Inc., San Diego, California) y dispositivos descritos en la patente estadounidense N.º 8.951.781. Dichos dispositivos se pueden modificar para fabricar matrices que utilicen amplificación de exclusión según las directrices expuestas en la presente memoria.

Un sistema capaz de llevar a cabo un método expuesto en la presente memoria no necesita estar integrado con un dispositivo de detección. Más bien, también es posible un sistema autónomo o integrado con otros dispositivos. En tales realizaciones se pueden usar componentes fluidicos similares a los ilustrados anteriormente en el contexto de un sistema integrado.

Un sistema capaz de llevar a cabo un método expuesto en la presente memoria, ya sea integrado con capacidades de detección o no, puede incluir un controlador de sistema que sea capaz de realizar un conjunto de instrucciones para realizar una o más pasos de un método, técnica o proceso expuesto en la presente memoria. Por ejemplo, las instrucciones pueden dirigir la realización de pasos para crear una matriz en condiciones de amplificación de exclusión. Opcionalmente, las instrucciones pueden dirigir adicionalmente la realización de los pasos para detectar ácidos nucleicos usando métodos expuestos anteriormente en la presente memoria. Un controlador del sistema útil puede incluir cualquier sistema basado en procesador o microprocesador, incluidos sistemas que utilicen microcontroladores, ordenadores de conjunto de instrucciones reducidas (RISC), circuitos integrados de aplicación específica (ASIC), matrices de puertas programables en campo (FPGA), circuitos lógicos y cualquier otro circuito o procesador capaz de realizar las funciones descritas en la presente memoria. Un conjunto de instrucciones para un controlador de sistema puede tener la forma de un programa de informático. Como se usa en la presente memoria, los términos “*software*” y “*firmware*” son intercambiables, e incluyen cualquier programa informático almacenado en la memoria para su ejecución por un ordenador, incluyendo memoria RAM, memoria ROM, memoria EPROM, memoria EEPROM y memoria RAM no volátil (NVRAM). El *software* puede adoptar varias formas, tal como un *software* de sistema o un *software* de aplicación. Además, el *software* puede adoptar la forma de una colección de programas independientes, o un módulo de programa dentro de un programa más amplio o una parte de un módulo de programa. El *software* también puede incluir programación modular en forma de programación orientada a objetos.

Varias aplicaciones para matrices de la presente memoria se han ilustrado anteriormente en el contexto de la detección por conjuntos, en donde se detectan juntos múltiples amplicones presentes en cada sitio de amplificación. En realizaciones alternativas, se puede detectar un único ácido nucleico, ya sea un ácido nucleico diana o un amplicón del mismo, en cada sitio de amplificación. Por ejemplo, un sitio de amplificación puede configurarse para contener una única molécula de ácido nucleico que tiene una secuencia de nucleótidos diana que se va a detectar y una pluralidad de ácidos nucleicos de relleno. En este ejemplo, los ácidos nucleicos de relleno actúan llenando la capacidad del sitio de amplificación y no necesariamente

están destinados a ser detectados. La molécula individual que se va a detectar se puede detectar mediante un método que sea capaz de distinguir la molécula individual en el fondo de los ácidos nucleicos de relleno. Puede usarse cualquiera de una diversidad de técnicas de detección de moléculas individuales, incluidas, por ejemplo, modificaciones de las técnicas de detección por conjuntos expuestas anteriormente para detectar los sitios con mayor ganancia o usando marcadores más sensibles. Otros ejemplos de métodos de detección de moléculas individuales que se pueden usar se exponen en los documentos US 2011/0312529 A1; patentes estadounidenses N.º 9.279.154; y documento U.S. 2013/0085073 A1.

Se puede crear una matriz útil para la detección de ácidos nucleicos de una sola molécula usando uno o más de los métodos expuestos en la presente memoria con las siguientes modificaciones. Se puede configurar una pluralidad de ácidos nucleicos diana diferentes para incluir tanto una secuencia de nucleótidos diana que se va a detectar como una o más secuencias de nucleótidos de relleno que se van a amplificar para crear amplicones de relleno. La pluralidad de diferentes ácidos nucleicos diana se puede incluir en un reactivo de amplificación, tal como los expuestos en otra parte de la presente memoria, y se pueden hacer reaccionar con una matriz de sitios de amplificación en condiciones de amplificación de exclusión tales que la secuencia o secuencias de nucleótidos de relleno llenen los sitios de amplificación. Las configuraciones ilustrativas que pueden usarse para permitir que las secuencias de relleno se amplifiquen al tiempo que prohíben la amplificación de la secuencia diana incluyen, por ejemplo, una molécula diana individual que tiene una primera región con secuencias de relleno flanqueadas por sitios de unión para cebadores de amplificación presentes en el sitio de amplificación y una segunda región que tiene una secuencia diana fuera de la región flanqueada. En otra configuración, un ácido nucleico diana puede incluir moléculas o cadenas separadas que portan la secuencia diana y la secuencia o secuencias de relleno, respectivamente. Las moléculas o cadenas separadas pueden estar unidas a una partícula o formadas como brazos de un dendrímero de ácido nucleico u otra estructura ramificada.

En una realización particular, se puede detectar una matriz que tiene sitios de amplificación que contienen cada uno secuencias de relleno y una secuencia diana usando un ensayo de extensión de cebadores o una técnica de secuenciación por síntesis. En tales casos, se puede lograr una extensión específica en la secuencia de nucleótidos diana en lugar de en la gran cantidad de secuencia de relleno mediante el uso de sitios de unión de cebadores ubicados apropiadamente. Por ejemplo, los sitios de unión para los cebadores de secuenciación pueden colocarse secuencia arriba de la secuencia diana y pueden estar ausentes en cualquiera de las secuencias de relleno. Alternativamente, o de forma adicional, la secuencia diana puede incluir uno o más análogos de nucleótidos no nativos que no son capaces de formar enlaces de hidrógeno con nucleótidos estándar. El o los nucleótidos no naturales pueden colocarse secuencia abajo del sitio de unión del cebador (p. ej., en la secuencia diana o en una región intermedia entre la secuencia diana y el sitio de unión del cebador) y como tales evitarán la extensión o la secuenciación por síntesis hasta que se añada un compañero nucleotídico apropiado (es decir, uno capaz de formar enlaces de hidrógeno con el análogo o los análogos no naturales en la secuencia diana). Los análogos de nucleótidos isocitosina (isoC) e isoguanina (isoG) son particularmente útiles ya que se emparejan específicamente entre sí pero no con otros nucleótidos estándar utilizados en la mayoría de las técnicas de extensión y secuenciación por síntesis. Un beneficio adicional de usar isoC y/o isoG en una secuencia diana o secuencia arriba de la secuencia diana es evitar la amplificación no deseada de la secuencia diana durante los pasos de amplificación al omitir el compañero respectivo en la mezcla de nucleótidos usada para la amplificación.

Se entenderá que una matriz de la presente memoria, por ejemplo, que se haya producido mediante un método expuesto en el presente documento, no necesita utilizarse para un método de detección. Más bien, la matriz se puede utilizar para almacenar una biblioteca de ácidos nucleicos. En consecuencia, la matriz se puede almacenar en un estado que conserve los ácidos nucleicos que contiene. Por ejemplo, una matriz se puede almacenar en estado desecado, congelado (p. ej., en nitrógeno líquido) o en una solución que proteja los ácidos nucleicos. Alternativamente, o de forma adicional, la matriz se puede utilizar para replicar una biblioteca de ácidos nucleicos. Por ejemplo, se puede usar una matriz para crear amplicones replicados de uno o más de los sitios en la matriz.

En la presente memoria se han ilustrado varias realizaciones de la descripción con respecto al transporte de ácidos nucleicos diana a sitios de amplificación de una matriz y a la fabricación de copias de los ácidos nucleicos diana capturados en los sitios de amplificación. Se pueden utilizar métodos similares para moléculas diana que no sean ácidos nucleicos. Por lo tanto, los métodos expuestos en la presente memoria se pueden usar con otras moléculas diana en lugar de los ácidos nucleicos diana ilustrados. Por ejemplo, se puede llevar a cabo un método de la presente memoria para transportar moléculas diana individuales de una población de moléculas diana diferentes. Cada molécula diana puede transportarse a (y en algunos casos capturarse en) un sitio individual de una matriz para iniciar una reacción en el sitio de captura. La reacción en cada sitio puede, por ejemplo, producir copias de la molécula capturada o la reacción puede alterar el sitio para aislar o secuestrar la molécula capturada. En cualquier caso, el resultado final pueden ser sitios de la matriz que sean puros con respecto al tipo de molécula diana que está presente en una población que contiene diferentes tipos de moléculas diana.

En realizaciones particulares que usan moléculas diana distintas de ácidos nucleicos, se puede crear una biblioteca de diferentes moléculas diana usando un método que aproveche la amplificación de exclusión. Por ejemplo, se puede preparar una matriz de moléculas diana en condiciones en que los sitios de la matriz se siembran al azar con moléculas diana de una solución y se generan copias de la molécula diana para llenar cada uno de los sitios sembrados hasta su capacidad. Según los métodos de amplificación de exclusión de la presente memoria, los procesos de siembra y copia pueden proceder simultáneamente en condiciones en que la velocidad a la que se fabrican las copias supera la velocidad de siembra. Como tal, la velocidad relativamente rápida a la que se fabrican las copias en un sitio que se ha sembrado por una primera molécula diana excluirá de forma efectiva que una segunda molécula diana siembre el sitio. En algunos casos, la siembra de una

molécula diana iniciará una reacción que llena un sitio al máximo de su capacidad mediante un proceso distinto de la copia de la molécula diana. Por ejemplo, la captura de una molécula diana en un sitio puede iniciar una reacción en cadena que eventualmente hace que el sitio sea incapaz de capturar una segunda molécula diana. La reacción en cadena puede producirse a una velocidad que supere la velocidad a la que se capturan las moléculas diana, teniendo lugar de este modo en condiciones de amplificación de exclusión.

Como se ilustra para los ácidos nucleicos diana, la amplificación de exclusión cuando se aplica a otras moléculas diana puede aprovechar una velocidad relativamente lenta para iniciar una reacción repetitiva (p. ej., una reacción en cadena) en un sitio de una matriz, frente a una velocidad relativamente rápida para continuar la reacción repetitiva una vez iniciada. En el ejemplo del párrafo anterior, la amplificación de exclusión se produce debido a la velocidad relativamente lenta de siembra de la molécula diana (p. ej., difusión relativamente lenta) frente a la velocidad relativamente rápida a la que se produce una reacción, por ejemplo, para llenar el sitio con copias de la semilla de la molécula diana. En otra realización ilustrativa, la amplificación de exclusión puede producirse debido a un retraso en la formación de una primera copia de una molécula diana que ha sembrado un sitio (por ejemplo, activación retardada o lenta) frente a la velocidad relativamente rápida a la que se fabrican las copias posteriores para llenar el sitio. En este ejemplo, es posible que se haya sembrado un sitio individual con varias moléculas diana diferentes. Sin embargo, la primera formación de copias para cualquier molécula diana dada puede activarse de al azar, de modo que la velocidad promedio de la primera formación de copias sea relativamente lenta en comparación con la velocidad a la que se generan copias posteriores. En este caso, aunque un sitio individual puede haberse sembrado con varias moléculas diana diferentes, la amplificación de exclusión sólo permitirá copiar una de esas moléculas diana.

En consecuencia, la presente memoria proporciona un método para fabricar una matriz de moléculas que puede incluir los pasos de (a) proporcionar un reactivo que incluya (i) una matriz de sitios, y (ii) una solución que tiene una pluralidad de moléculas diana diferentes, en donde el número de moléculas diana en la solución supera el número de sitios en la matriz, en donde las diferentes moléculas diana tienen acceso fluido a la pluralidad de sitios, y en donde cada uno de los sitios comprende una capacidad para varias moléculas diana en la pluralidad de diferentes moléculas diana; y (b) hacer reaccionar el reactivo para producir una pluralidad de sitios que tienen cada uno una única molécula diana de la pluralidad o para producir una pluralidad de sitios que tienen cada uno una población pura de copias de una molécula diana individual de la solución, en donde la reacción incluye simultáneamente (i) transportar las diferentes moléculas a los sitios a una velocidad de transporte promedio, e (ii) iniciar una reacción que llena el sitio hasta su capacidad a una velocidad de reacción promedio, en donde la velocidad de reacción promedio supera la velocidad de transporte promedio. En algunas realizaciones, el paso (b) se puede llevar a cabo haciendo reaccionar el reactivo para producir una pluralidad de sitios que tienen cada uno una única molécula diana de la pluralidad o para producir una pluralidad de sitios que tienen cada uno una población pura de copias de una molécula diana individual de la solución, en donde la reacción incluye (i) iniciar una reacción repetitiva (p. ej., una reacción en cadena) para formar un producto a partir de la molécula diana en cada uno de los sitios, y (ii) continuar la reacción en cada uno de los sitios para formar productos posteriores, en donde la velocidad promedio a la que se produce la reacción en los sitios supera la velocidad promedio a la que se inicia la reacción en los sitios.

En las realizaciones anteriores no de ácidos nucleicos, la molécula diana puede ser un iniciador de una reacción repetitiva que se produce en cada sitio de la matriz. Por ejemplo, la reacción repetitiva puede formar un polímero que impida que otras moléculas diana ocupen el sitio. Alternativamente, la reacción repetitiva puede formar uno o más polímeros que constituyen copias moleculares de una molécula diana que fue transportada al sitio.

Ejemplos

La presente invención se define mediante los siguientes ejemplos. Debe entenderse que los ejemplos, materiales, cantidades y procedimientos particulares deben interpretarse ampliamente según el alcance de la invención tal como se expone en la presente memoria.

Ejemplo 1

Métodos y condiciones generales de ensayo

Salvo que se indique lo contrario, se describen las condiciones generales de ensayo utilizadas en los Ejemplos descritos en la presente memoria.

Las bibliotecas de ácidos nucleicos se generaron comenzando con la preparación de bibliotecas Nextera™ estándar para introducir la parte universal del adaptador mediante tagmentación de ADN humano. Esta tagmentación universal se dividió luego en reacciones individuales, una para cada uno de los diferentes pares de adaptadores (PCR1 - PCR21). En cada una de estas reacciones, se utilizaron los reactivos de la preparación de la biblioteca Nextera™ XT (Illumina, Inc., San Diego, California) para introducir los adaptadores modificados a través de 12 ciclos de PCR, reemplazando los adaptadores P5/P7 estándar por los modificados. Se diseñaron adaptadores modificados con cambios de las secuencias de P5/P7 estándar (como se describe a continuación) y se sintetizaron por Integrated DNA Technologies (IDT Inc., Skokie, Illinois).

Ejemplo 2

Evaluación de mutantes adaptadores para retraso cinético

Se utilizó una gama de adaptadores modificados para generar bibliotecas que difieren de la biblioteca Nextera™ estándar (Tabla 1). Estos adaptadores eran ligeramente más cortos que los estándares (-4 pb o -9 pb) o tenían 1, 2 o 3 emparejamientos erróneos (bases “oscilantes”: 1W, 2W o 3W, respectivamente) introducidos a lo largo de las regiones P5/P7. La gama de mutaciones era de secuencias de P5 y P7 perfectas (PCR1) a 3 emparejamientos erróneos en ambos extremos (PCR 21).

A continuación, se cuantificó la concentración de cada biblioteca y las bibliotecas se normalizaron con respecto a concentraciones idénticas. A continuación, las bibliotecas con los diferentes adaptadores modificados se amplificaron por separado en un instrumento de qPCR (BioRad CFX384 Real-Time System) utilizando un kit KAPA Library Quantification kit para plataformas Illumina (Kapa Biosystems) con cebadores personalizados para simular las condiciones de la celda de flujo para generar los resultados de la Tabla 1. La eficiencia se calculó a partir del número de ciclos necesarios para alcanzar el Ct (ciclo umbral) como se define de forma estándar para la qPCR.

Tabla 1

		FCP5 FCP7
		Eficiencia
Número de adaptador	Identidad del adaptador	qPCR 3
1	PCR 1 - Control de Nex.	1,5079
2	PCR 2 $\alpha\beta$ -4	1,4875
3	PCR 3 $\alpha\beta$ -9	1,5229
4	PCR 4 - Nex P5 + 1W	1,4305
5	PCR 5 - Nex P5 + 2W	1,4160
6	PCR 6 - Nex P5 + 3W	1,3716
7	PCR 7 $\alpha\beta$ -4 + 1W	1,3577
8	PCR 8 $\alpha\beta$ -4 + 2W	1,3673
9	PCR 9 $\alpha\beta$ -4 + 3W	1,3273
10	PCR 10 $\alpha\beta$ -9 + 1W	1,4236
11	PCR 11 $\alpha\beta$ -9 + 2W	1,4179
12	PCR 12 $\alpha\beta$ -9 + 3W	1,3691
13	PCR 13 - 1 MM + 1MM	1,3692
14	PCR 14 - 1MM + 2W	1,3789
15	PCR 15 - 1MM + 3W	1,3488
16	PCR 16 - 2W + 1MM	1,3504
17	PCR 17 - 2W + 2W	1,2971
18	PCR 18 - 2W + 3W	1,2980
19	PCR 19 - 3W + 1MM	1,2861
20	PCR 20 - 3W + 2W	1,2498
21	PCR 21 - 3W + 3W	1,2261

$\alpha\beta$ -4 se refiere a cuatro pares de bases eliminados del adaptador; $\alpha\beta$ -9 se refiere a nueve pares de bases eliminados del adaptador; 1W, 2W, 3W se refieren a 1, 2 o 3 emparejamiento erróneos oscilantes, respectivamente, presentes solos en la longitud de una región que se une a un ácido nucleico de captura presente en la superficie de una matriz; 1MM, emparejamiento erróneo verdadero; FCP5, FCP7, cebadores P5 y P7 de longitud completa, respectivamente.

Ejemplo 3

Evaluación de mutantes adaptadores para el retraso cinético mediante secuenciación

Se eligieron diferentes adaptadores mutantes y se procesaron solos o en grupos en una celda de flujo HiSeq™X (Tabla 2). Toda la secuenciación se realizó en un instrumento HiSeq™X de Illumina utilizando kits de reactivos estándar.

Tabla 2

Carril de la celda de flujo	Los adaptadores utilizados en el carril
Carril 1	PCR 1
Carril 2	PCR 2
Carril 3	PCR 18
Carril 4	PCR 6
Carril 5	PCR 10
Carril 6	PCR 1-2, 4-6, 10
Carril 7	PCR 1-2, 7-8, 10, 11
Carril 8	PCR 2, 6, 8, 12

Resultados

Como se muestra en la **FIG. 3A**, los carriles 1, 2 y 4 fueron reacciones que utilizaron adaptadores con menos emparejamientos erróneos y mayor eficiencia y, como se esperaba, dieron como resultado una alta intensidad y un alto porcentaje de agrupamientos que pasaron el filtro. Los carriles 3 y 5 fueron reacciones que utilizaron adaptadores con más emparejamientos erróneos y eficiencias más bajas y, como se esperaba, dieron como resultado una intensidad baja y un porcentaje bajo de agrupamientos que pasaron el filtro. Los carriles 6, 7 y 8 eran mezclas de adaptadores con eficiencias altas y bajas. Contrariamente a la intuición, las mezclas no se desempeñaron como un promedio del desempeño de los componentes individuales (p. ej., a medio camino entre las eficiencias alta y baja), sino que superaron a todas las bibliotecas de tipo único tanto en intensidad como en agrupamientos que pasaron el filtro. Por lo tanto, el resultado sorprendente es que al reducir la homología promedio, se mejoró la velocidad de la llamada monoclonalidad de los nanopocillos, aunque se reduce la velocidad de amplificación promedio. La novedad es que se introduce un cierto grado de variabilidad en las secuencias adaptadoras, de modo que ahora existe una gama de eficiencias entre la población de moldes. De esta manera, cuando se siembran varios moldes en un parche, normalmente hay uno que tiene una ventaja sobre todos los demás, de modo que domina claramente el parche. Adicionalmente, la homología reducida se corrige en las copias hijas, de modo que el retraso se introduce sólo en la primera copia, sin afectar la eficiencia de la amplificación posterior.

Se eligieron diferentes adaptadores mutantes y se procesaron solos, o en grupos en una celda de flujo HiSeq™X (Tabla 3). Al igual que en el caso anterior, toda la secuenciación se realizó en un instrumento HiSeq™X de Illumina utilizando kits de reactivos estándar.

Tabla 3

Carril de la celda de flujo	Los adaptadores utilizados en el carril
Carril 1	PCR 1
Carril 2	PCR 3
Carril 3	PCR 10
Carril 4	PCR 16
Carril 5	PCR 1, 3, 6, 8, 9, 10, 16-18
Carril 6	PCR 21
Carril 7	PCR 1, 3, 6, 8, 9, 10, 16-18
Carril 8	PCR 1, 3, 10

Cuando un grupo de adaptadores mutantes se procesó en una mezcla (p. ej., el carril 7), se combinaron en concentraciones iguales. En un ciclo de secuenciación convencional, es decir, una que no utiliza los métodos propuestos, concentraciones iguales de bibliotecas mixtas de los mismos adaptadores darían como resultado relaciones iguales de lecturas en la celda de flujo. Como se muestra en la **FIG. 3B**, la mezcla de diferentes adaptadores mutantes dio como resultado una representación de los recuentos de lectura finales que era proporcional a su eficiencia y no a su concentración sembrada, demostrando por lo tanto la eficacia del método propuesto, es decir, los adaptadores con menor afinidad tenían retrasos cinéticos más prolongados, lo que dio como resultado en una menor proporción de las lecturas finales.

La descripción detallada y los ejemplos anteriores se han facilitado únicamente para facilitar la comprensión. De ello no se desprende ninguna limitación innecesaria. La descripción no se limita a los detalles exactos mostrados y descritos, ya que las variaciones obvias para un experto en la técnica se incluirán dentro de la descripción definida por las reivindicaciones.

5 Salvo que se indique lo contrario, se debe entender que todos los números que expresan cantidades de componentes, pesos moleculares, etc. utilizados en la memoria descriptiva y las reivindicaciones están modificados en todos los casos por el término “aproximadamente”. En consecuencia, salvo que se indique lo contrario, los parámetros numéricos expuestos en la memoria descriptiva y las reivindicaciones son aproximaciones que pueden variar dependiendo de las propiedades
10 de deseadas que se buscan obtener mediante la presente memoria. Como mínimo, y no como un intento de limitar la doctrina de equivalentes al alcance de las reivindicaciones, cada parámetro numérico debe considerarse al menos a la luz del número de dígitos significativos informados y aplicando técnicas de redondeo comunes.

A pesar de que los intervalos y parámetros numéricos que exponen el amplio alcance de la descripción son aproximaciones, los valores numéricos expuestos en los ejemplos específicos se indican de forma tan precisa como sea posible. Sin embargo,
15 todos los valores numéricos contienen inherentemente un intervalo que es el resultado necesariamente de la desviación estándar encontrada en sus respectivas mediciones de prueba.

Todos los encabezados son por comodidad de uso del lector y no deben usarse para limitar el significado del texto que sigue al encabezado, salvo que así se especifique.
20

REIVINDICACIONES

1. Un método para amplificar ácidos nucleicos, que comprende
- 5 (a)proporcionar un reactivo de amplificación que comprende
- (i)una matriz de sitios de amplificación,
- 10 en donde los sitios de amplificación comprenden dos poblaciones de ácidos nucleicos de captura, comprendiendo cada población una secuencia de captura, en donde una primera población comprende una primera secuencia de captura y una segunda población comprende una segunda secuencia de captura, y
- 15 (ii)una solución que comprende una pluralidad de diferentes ácidos nucleicos diana bicatenarios modificados, en donde los diferentes ácidos nucleicos diana modificados comprenden en el extremo 3' una primera secuencia de unión de captura universal que tiene menos afinidad por la primera secuencia de captura que una primera secuencia de unión de captura universal que tiene el 100 % de complementariedad con la primera secuencia de captura; y
- 20 (b)hacer reaccionar el reactivo de amplificación para producir una pluralidad de sitios de amplificación que comprenden cada uno una población clonal de amplicones de un ácido nucleico diana individual de la solución, opcionalmente en donde la reacción comprende
- 25 (i)producir un primer amplicón de un ácido nucleico diana individual que se transporta a cada uno de los sitios de amplificación, y
- (ii)producir amplicones posteriores del ácido nucleico diana individual que se transporta a cada uno de los sitios de amplificación o del primer amplicón,
- 30 en donde la velocidad promedio a la que se generan los amplicones posteriores en los sitios de amplificación es menor que la velocidad promedio a la que se genera el primer amplicón en los sitios de amplificación.
2. Un método para determinar secuencias de ácido nucleico, que comprende realizar un procedimiento de secuenciación que detecta una población aparentemente clonal de amplicones en cada uno de una pluralidad de sitios de amplicones en una matriz, en donde la matriz se fabrica mediante un proceso que comprende:
- 35 (a)proporcionar un reactivo de amplificación que comprende
- 40 (i)una pluralidad de sitios de amplificación,
- en donde los sitios de amplificación comprenden dos poblaciones de ácidos nucleicos de captura, comprendiendo cada población una secuencia de captura, en donde una primera población comprende una primera secuencia de captura y una
- 45 segunda población comprende una segunda secuencia de captura, y
- (ii)una solución que comprende una pluralidad de diferentes ácidos nucleicos diana modificados, en donde los diferentes ácidos nucleicos diana modificados comprenden en el extremo 3' una primera secuencia de unión de captura universal que tiene menos afinidad por la primera secuencia de captura que una primera secuencia de unión de captura universal que tiene el 100 % de complementariedad con la primera secuencia de captura; y
- 50 (b)hacer reaccionar el reactivo de amplificación.
- 55 3. El método de la reivindicación 1 o 2, en donde el número de diferentes ácidos nucleicos diana modificados en la solución supera el número de sitios de amplificación en la matriz,
- 60 en donde los diferentes ácidos nucleicos diana modificados tienen acceso fluido a la pluralidad de sitios de amplificación, y en donde cada uno de los sitios de amplificación comprende una capacidad para varios ácidos nucleicos en la pluralidad de ácidos nucleicos diferentes
- 65 4. El método de la reivindicación 1, en donde la reacción comprende simultáneamente

- (i)transportar los diferentes ácidos nucleicos diana modificados a los sitios de amplificación a una velocidad de transporte promedio, y
(ii)amplificar los ácidos nucleicos diana que están en los sitios de amplificación a una velocidad de amplificación promedio, en donde la velocidad de amplificación promedio es menor que la velocidad de transporte promedio.
- 5
5. El método de la reivindicación 2, en donde la pluralidad de diferentes ácidos nucleicos diana modificados en la solución está en una concentración que da como resultado, simultáneamente:
- 10 (i)el transporte de los diferentes ácidos nucleicos diana modificados desde la solución hasta los sitios de amplificación, y
(ii)la amplificación de los ácidos nucleicos diana que están en los sitios de amplificación a una velocidad de amplificación para producir una matriz de sitios de amplicones que comprenden cada uno la población aparentemente clonal de amplicones.
- 15
6. El método de una cualquiera de las reivindicaciones 1-2, en donde la primera secuencia de unión de captura universal tiene menos que el 100 % de complementariedad con la primera secuencia de captura; opcionalmente
- 20 a)en donde la primera secuencia de unión de captura universal comprende 1, 2 o 3 nucleótidos que no son complementarios con la primera secuencia de captura; o
b)en donde los diferentes ácidos nucleicos diana modificados comprenden una población heterogénea de primeras secuencias de unión de captura universales, en donde la población heterogénea comprende primeras secuencias de unión de captura universales individuales que tienen (i) 1, 2 o 3 nucleótidos que no son complementarios con la primera secuencia de captura, o (ii) el 100 % de complementariedad con la primera secuencia de captura, opcionalmente
- 25 en donde los miembros de la población heterogénea que tienen el 100 % de complementariedad con la primera secuencia de captura están presentes en un número mayor que los otros miembros de la población heterogénea, además opcionalmente
- 30 en donde la población heterogénea comprende además primeras secuencias de unión de captura universales individuales que tienen (iii) una longitud que es menor que la longitud de la primera secuencia de captura.
- 35
7. El método de una cualquiera de las reivindicaciones 1-2, en donde la primera secuencia de unión de captura universal tiene una longitud que es menor que la longitud de la primera secuencia de captura, opcionalmente en donde la primera secuencia de unión de captura universal tiene una longitud que es de 1 a 12 nucleótidos menor que la longitud de la primera secuencia de captura.
- 40
8. El método de la reivindicación 7, en donde los diferentes ácidos nucleicos diana modificados comprenden una población heterogénea de primeras secuencias de unión de captura universales, en donde la población heterogénea comprende primeras secuencias de unión de captura universales individuales que tienen de 1 a 12 nucleótidos menos que la longitud de la primera secuencia de captura.
- 45
9. El método de la reivindicación 6, en donde los miembros individuales de la población heterogénea que tienen una longitud que es menor que la longitud de la primera secuencia de captura comprende 1, 2 o 3 nucleótidos que no son complementarios con la secuencia de la primera secuencia de captura, o el 100 % de complementariedad con la secuencia de la primera secuencia de captura.
- 50
10. El método de una cualquiera de las reivindicaciones 1-2, en donde los diferentes ácidos nucleicos diana modificados comprenden en el extremo 5' una segunda secuencia de unión de captura universal que tiene un complemento que tiene menos afinidad por la segunda secuencia de captura que una segunda secuencia de unión de captura universal que tiene un complemento con el 100 % de complementariedad con la segunda secuencia de captura opcionalmente,
- 55 en donde el complemento de la segunda secuencia de unión de captura universal tiene menos del 100 % de complementariedad con la segunda secuencia de captura, además opcionalmente
- 60 a)en donde el complemento de la segunda secuencia de unión de captura universal comprende 1, 2 o 3 nucleótidos que no son complementarios con la segunda secuencia de captura; o
b)en donde los diferentes ácidos nucleicos diana modificados comprenden una población heterogénea de segundas secuencias de unión de captura universales, en donde la población heterogénea comprende segundas secuencias de unión de captura universales individuales que comprenden un complemento que tiene (i) 1, 2 o 3 nucleótidos que no son complementarios con la segunda secuencia de captura, o (ii) el 100 % de complementariedad con la segunda secuencia de captura, y opcionalmente
- 65

en donde los miembros de la población heterogénea que comprende un complemento que tiene el 100 % de complementariedad con la segunda secuencia de captura están presentes en un número mayor que los otros miembros de la población heterogénea, además opcionalmente,
 en donde la población heterogénea comprende además segundas secuencias de unión de captura universales individuales que tienen (iii) una longitud que es menor que la longitud de la segunda secuencia de captura.

11. El método de una cualquiera de las reivindicaciones 1-2, en donde la segunda secuencia de unión de captura universal tiene una longitud que es menor que la longitud de la segunda secuencia de captura opcionalmente,

en donde la segunda secuencia de unión de captura universal tiene una longitud que es de 1 a 12 nucleótidos menor que la longitud de la segunda secuencia de captura, además opcionalmente, en donde los diferentes ácidos nucleicos diana modificados comprenden una población heterogénea de segundas secuencias de unión de captura universales, en donde la población heterogénea comprende segundas secuencias de unión de captura universales individuales que tienen de 1 a 12 nucleótidos menos que la longitud de la segunda secuencia de captura.

12. El método de una cualquiera de las reivindicaciones 1-2, en donde la matriz de sitios de amplificación comprende una matriz de casillas en una superficie, opcionalmente

en donde el área para cada una de las casillas es mayor que el diámetro del volumen excluido de los ácidos nucleicos diana que se transportan a los sitios de amplificación, opcionalmente en donde las casillas no son contiguas y están separadas por regiones intersticiales de la superficie que carecen de agentes de captura.

13. Una composición que comprende una matriz de sitios de amplificación y al menos un ácido nucleico diana unido a un sitio de amplificación,

en donde los sitios de amplificación comprenden dos poblaciones de ácidos nucleicos de captura, comprendiendo cada población una secuencia de captura, en donde una primera población comprende una primera secuencia de captura y una segunda población comprende una segunda secuencia de captura, en donde el ácido nucleico diana comprende en el extremo 3' una primera secuencia de unión de captura universal que tiene menos afinidad por la primera secuencia de captura que una primera secuencia de unión de captura universal que tiene el 100 % de complementariedad con la primera secuencia de captura, en donde la secuencia de unión de captura universal del ácido nucleico diana hibrida con la primera secuencia de captura.

14. La composición de la reivindicación 13, en donde

a) la primera secuencia de unión de captura universal tiene menos del 100 % de complementariedad con la primera secuencia de captura, o
 b) la primera secuencia de unión de captura universal comprende 1, 2 o 3 nucleótidos que no son complementarios con la primera secuencia de captura, o
 c) al menos el 30 % de los sitios de amplificación de la matriz están ocupados por al menos un ácido nucleico diana opcionalmente, en donde la primera secuencia de unión de captura universal comprende una población heterogénea, en donde la población heterogénea comprende primeras secuencias de unión de captura universales individuales que tienen (i) 1, 2 o 3 nucleótidos que no son complementarios con la primera secuencia de captura, o (ii) el 100 % de complementariedad con la primera secuencia de captura, y en donde los miembros de la población heterogénea están unidos a diferentes sitios de amplificación, además opcionalmente en donde los miembros de la población heterogénea que tienen el 100 % de complementariedad con la primera secuencia de captura están presentes en un número mayor que los otros miembros de la población heterogénea.

15. La composición de la reivindicación 13, en donde la segunda secuencia de unión de captura universal tiene una longitud que es menor que la longitud de la segunda secuencia de captura, opcionalmente, en donde la segunda secuencia de unión de captura universal tiene una longitud que es de 1 a 12 nucleótidos menor que la longitud de la segunda secuencia de captura, además opcionalmente, en donde la composición comprende una pluralidad de diferentes ácidos nucleicos diana, comprendiendo los diferentes ácidos nucleicos diana una población heterogénea de primeras secuencias de unión de captura universales, en donde la población heterogénea comprende primeras secuencias de unión de captura universales individuales que tienen de 1 a 12 nucleótidos menos que la longitud de la primera secuencia de captura.

16. La composición de la reivindicación 13, en donde la población heterogénea comprende además primeras secuencias de unión de captura universales individuales que tienen (iii) una longitud que es menor que la longitud de la primera secuencia de captura, opcionalmente en donde
- 5 a) las primeras secuencias de unión de captura universales individuales que comprenden una longitud reducida tienen una longitud que es de 1 a 12 nucleótidos menor que la longitud de la primera secuencia de captura, o
- b) los miembros individuales de la población heterogénea que tienen una longitud reducida comprenden 1, 2 o 3 nucleótidos que no son complementarios con la secuencia de la primera secuencia de captura, o el 100 % de complementariedad con la secuencia de la primera secuencia de captura.
- 10
17. La composición de la reivindicación 13, en donde la composición comprende una pluralidad de diferentes ácidos nucleicos diana, comprendiendo los diferentes ácidos nucleicos diana en el extremo 5' una segunda secuencia de unión de captura universal que tiene un complemento que tiene menos afinidad por la segunda secuencia de captura que una segunda secuencia de unión de captura universal que tiene un complemento con el 100 % de complementariedad con la segunda secuencia de captura, opcionalmente en donde
- 15 el complemento de la segunda secuencia de unión de captura universal tiene menos del 100 % de complementariedad con la segunda secuencia de captura, además opcionalmente,
- 20 a) en donde el complemento de la segunda secuencia de unión de captura universal comprende 1, 2 o 3 nucleótidos que no son complementarios con la segunda secuencia de captura, o
- b) en donde la composición comprende una pluralidad de diferentes ácidos nucleicos diana, comprendiendo los diferentes ácidos nucleicos diana una población heterogénea de segundas secuencias de unión de captura universales, en donde la población heterogénea comprende segundas secuencias de unión de captura universales individuales que comprenden un complemento que tiene
- 25 (i) 1, 2 o 3 nucleótidos que no son complementarios con la segunda secuencia de captura, o
- (ii) el 100 % de complementariedad con la segunda secuencia de captura, o
- 30 c) en donde los miembros de la población heterogénea que comprende un complemento que tiene el 100 % de complementariedad con la segunda secuencia de captura están presentes en un número mayor que los otros miembros de la población heterogénea.
- 35
18. La composición de la reivindicación 13, en donde el ácido nucleico diana comprende en el extremo 5' una segunda secuencia de unión de captura universal que tiene una longitud que es menor que la longitud de la segunda secuencia de captura, opcionalmente
- 40 a) en donde la segunda secuencia de unión de captura universal tiene una longitud que es de 1 a 12 nucleótidos menor que la longitud de la segunda secuencia de captura, o
- b) en donde la composición comprende una pluralidad de diferentes ácidos nucleicos diana, comprendiendo los diferentes ácidos nucleicos diana una población heterogénea de segundas secuencias de unión de captura universales, en donde la población heterogénea comprende segundas secuencias de unión de captura universales individuales que tienen de 1 a 12 nucleótidos menos que la longitud de la segunda secuencia de captura.
- 45
19. La composición de la reivindicación 17, en donde la población heterogénea comprende además segundas secuencias de unión de captura universales individuales que tienen (iii) una longitud que es menor que la longitud de la segunda secuencia de captura.
- 50
20. La composición de la reivindicación 19, en donde las segundas secuencias de unión de captura universales individuales tienen una longitud que es de 1 a 12 nucleótidos menor que la longitud de la segunda secuencia de captura, opcionalmente
- 55 en donde los miembros individuales de la población heterogénea que tienen una longitud que es menor que la longitud de la segunda secuencia de captura comprende un complemento que comprende 1, 2 o 3 nucleótidos que no son complementarios con la secuencia de la segunda secuencia de captura, o el 100 % de complementariedad con la secuencia de la segunda secuencia de captura.
- 60

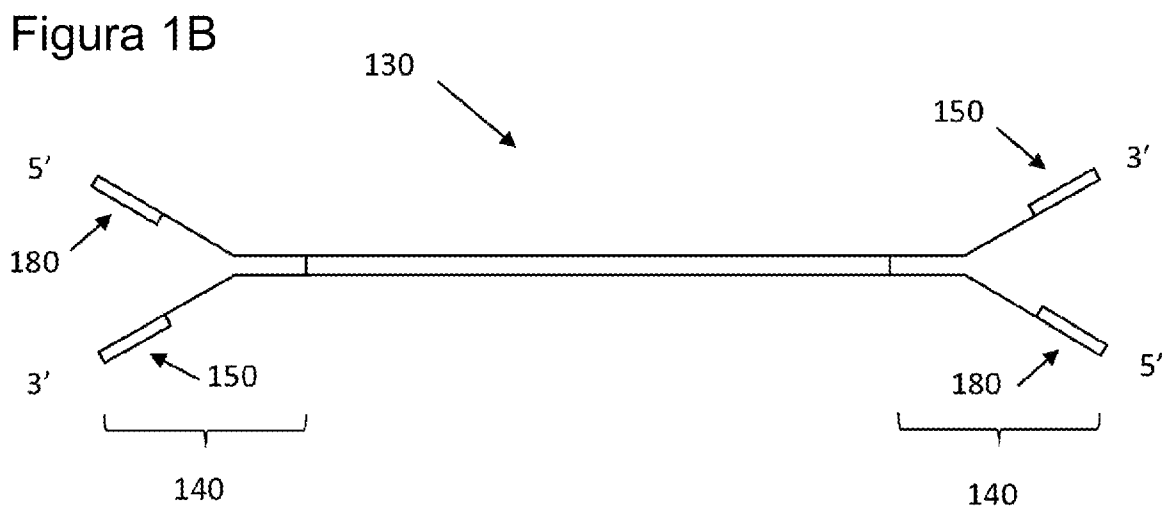
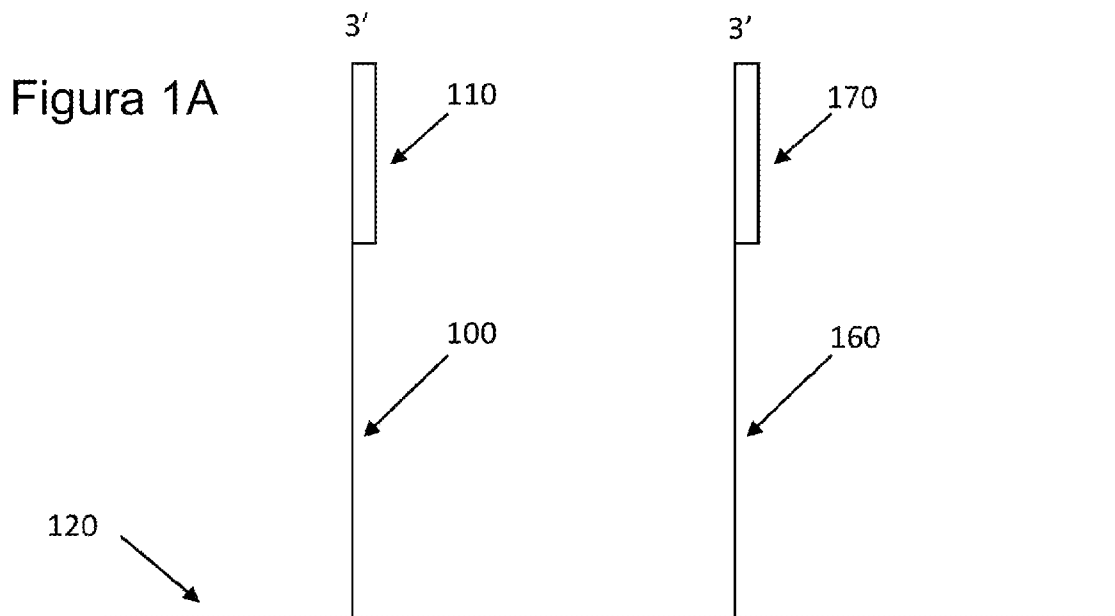


Figura 2

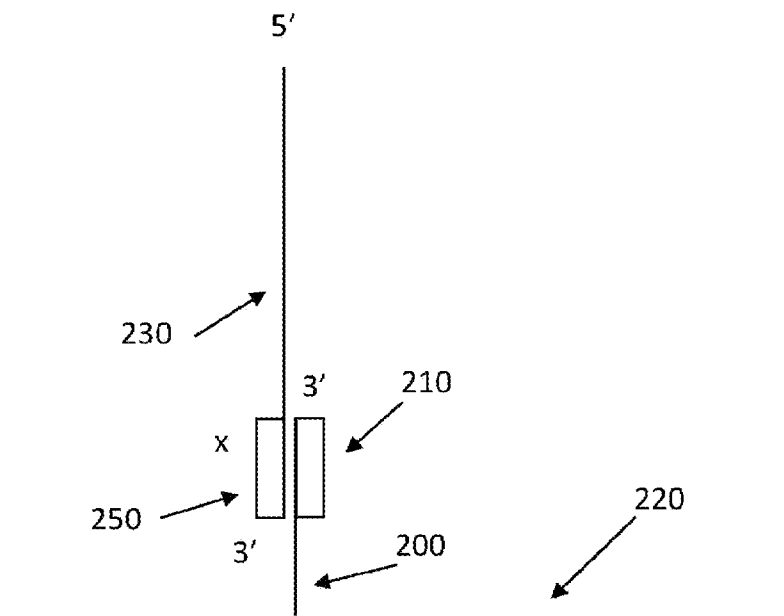


Figura 3A

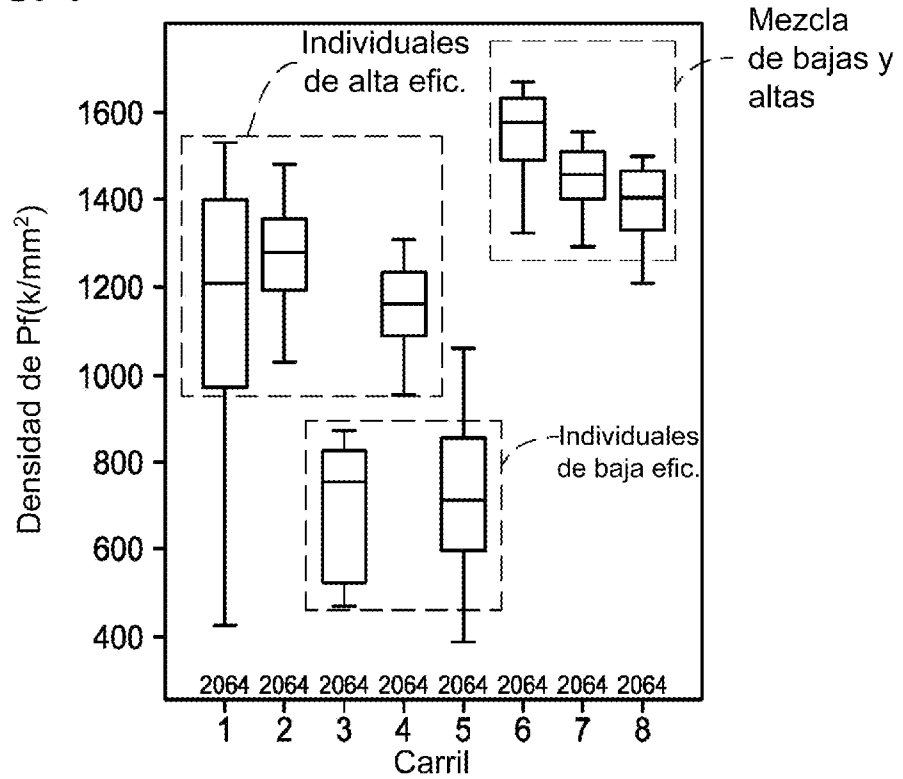
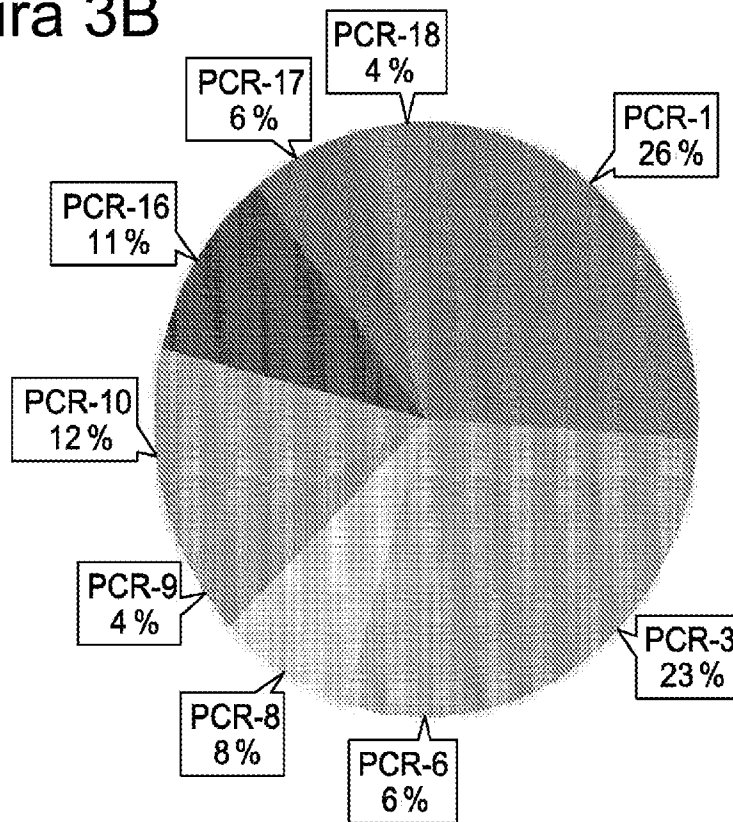


Figura 3B



Relación de lecturas finales provenientes de cada adaptador mutante

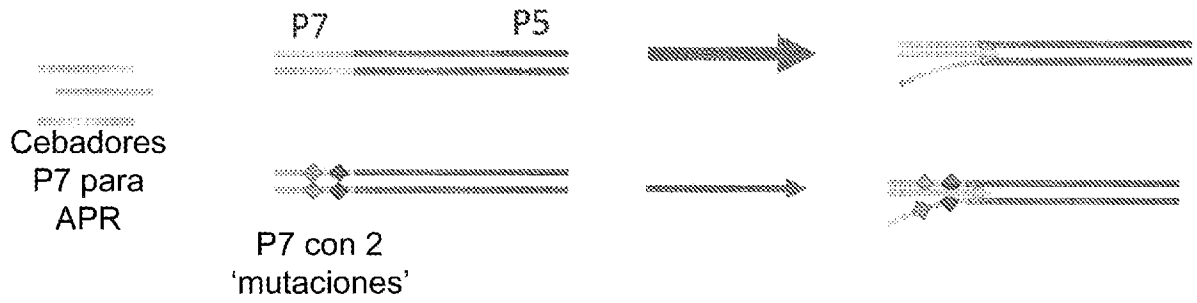


Figura 4A

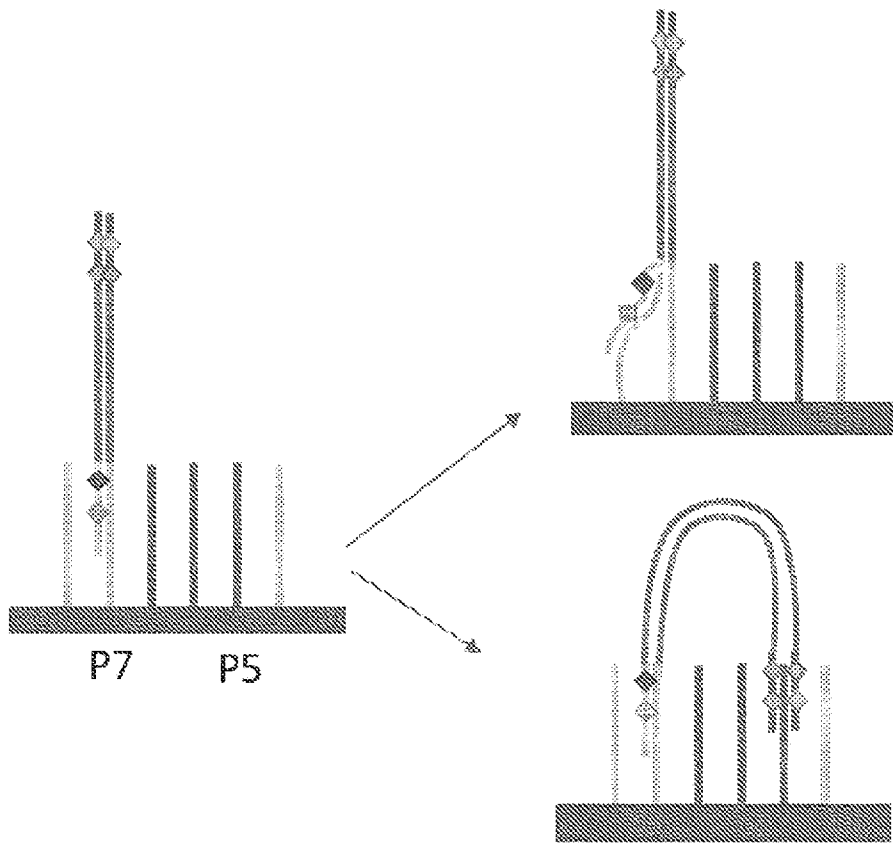


Figura 4B

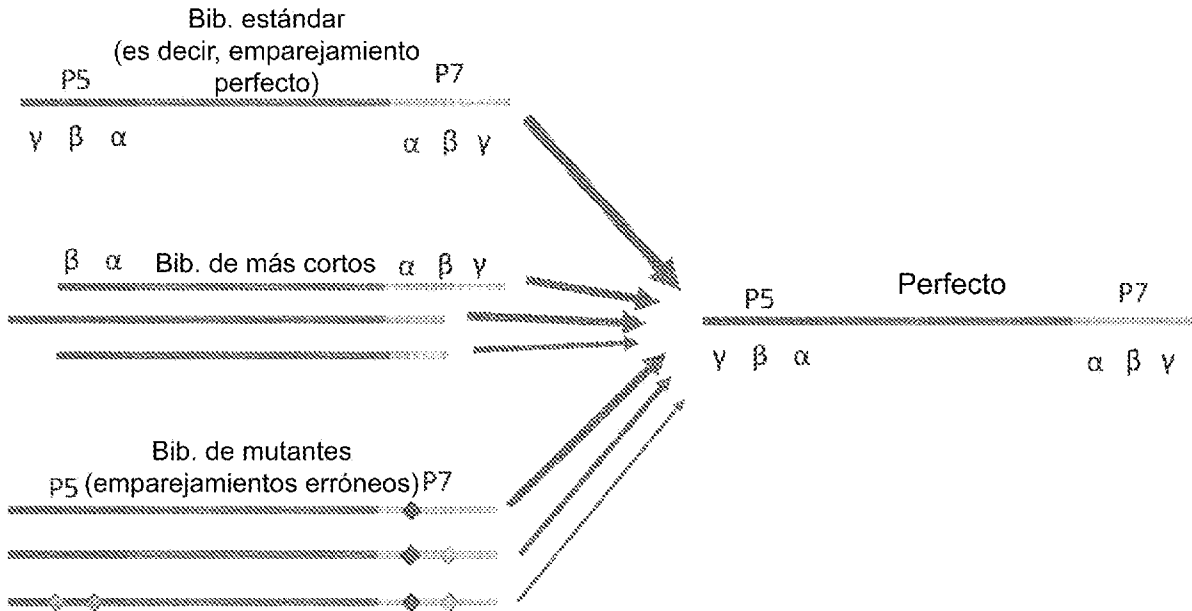


Figura 5A

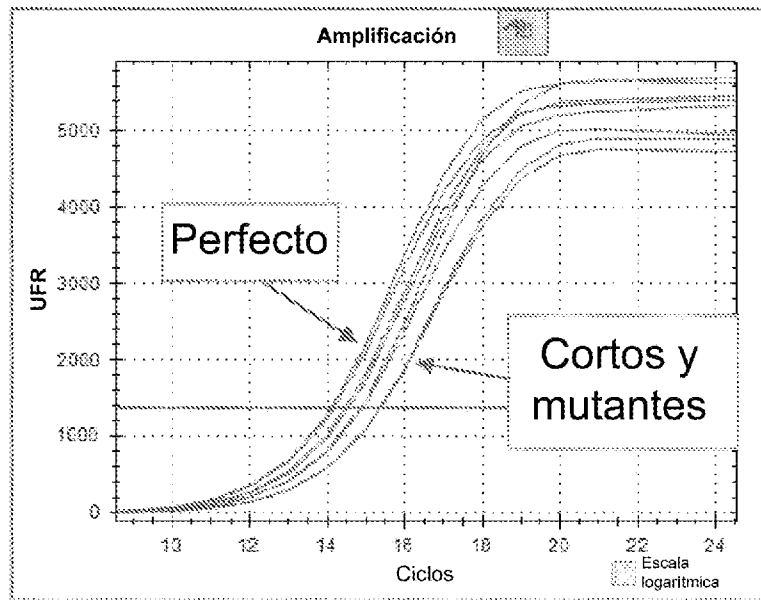


Figura 5B

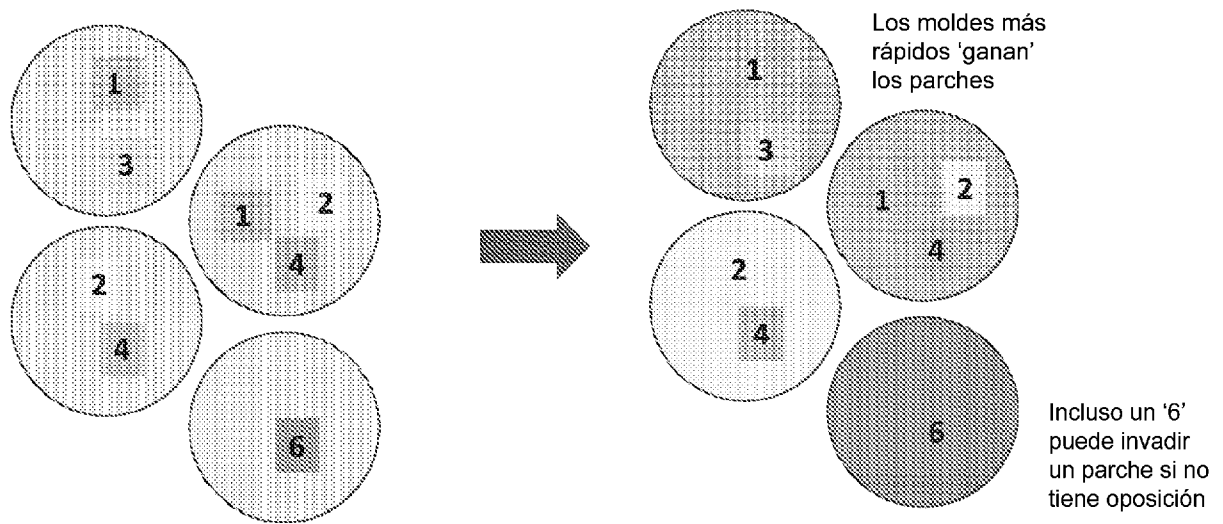


Figura 6