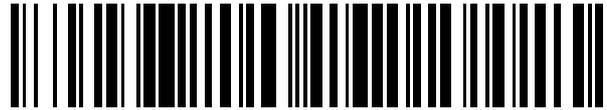


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 562 159**

51 Int. Cl.:

**C12N 15/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.08.2010 E 10768061 (3)**

97 Fecha y número de publicación de la concesión europea: **06.01.2016 EP 2467479**

54 Título: **Composiciones y métodos para el reordenamiento de ácido nucleico molecular**

30 Prioridad:

**20.08.2009 US 235595 P**  
**21.12.2009 US 288792 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**02.03.2016**

73 Titular/es:

**POPULATION GENETICS TECHNOLOGIES LTD.**  
**(100.0%)**  
**Babraham Research Campus**  
**Cambridge CB3 22AT, GB**

72 Inventor/es:

**BRENNER, SYDNEY;**  
**MIKAWA, GI;**  
**OSBORNE, ROBERT y**  
**SLATTER, ANDREW**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 562 159 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

Composiciones y métodos para el reordenamiento de ácido nucleico molecular

### Antecedentes

5 Los autores de la presente invención han descrito previamente métodos que permiten el etiquetado de cada una de las poblaciones de genomas fragmentados y la posterior combinación de los mismos para crear una "biblioteca de población" que puede ser procesada y, eventualmente secuenciada en forma de una mezcla. Las etiquetas de población permiten que el soporte lógico de análisis analice las lecturas de secuencias en archivos que se pueden atribuir a un genoma concreto en la población. Una de las limitaciones del procedimiento global deriva de las limitaciones de las tecnologías de secuenciación de ADN existentes. En particular, si los fragmentos en las regiones de interés del genoma son más largos que las longitudes que pueden ser secuenciadas por una tecnología concreta, tales fragmentos no serán analizados por completo (puesto que la secuenciación avanza desde un extremo de un fragmento hacia adentro). Además, una desventaja de cualquier tecnología de secuenciación dependiente de la fragmentación es que los cambios de secuencia en una parte de una región genómica concreta pueden no ser susceptibles de ser vinculados a los cambios de secuencia en otras partes del mismo genoma (p. ej., el mismo cromosoma) porque los cambios de secuencia residen en diferentes fragmentos. (Véase la Figura 5 y su descripción a continuación).

El documento WO 2004/061083 describe ARN de interferencia.

El documento EP 1.841.879 describe la amplificación isotérmica de ADN.

El documento US 2007/0238113 describe un método para detectar mutaciones y/o polimorfismos.

20 El documento EP 1.672.064 describe un método de preparación de una estructura repetida invertida de ADN.

El documento WO 2007/018601 describe composiciones y métodos para el procesamiento y la amplificación de ADN, incluyendo el uso de enzimas en una única reacción.

Yan et al., (2009) describen la construcción en una etapa, rápida de horquillas de ADN (Biochem. Biophys. Res. Comm. 383: 464-8).

25 La presente invención elimina las limitaciones impuestas por las tecnologías de secuenciación actuales, además de ser útil en otros numerosos análisis de ácido nucleico.

### Compendio de la invención

De acuerdo con un aspecto de la presente invención, se proporciona un método de análisis de muestras como se especifica en la reivindicación 1.

30 Se describen los procedimientos para mover una región de interés en un polinucleótido desde una primera posición a una segunda posición con respecto a un dominio dentro del polinucleótido, también conocido como "método réflex" (o procedimiento réflex, procedimiento de secuencia réflex, reacción réflex, y similares). En ciertas realizaciones, el método réflex da como resultado el movimiento de una región de interés en proximidad funcional a los elementos de dominios específicos presentes en el polinucleótido (p. ej., sitios de cebadores y/o MID). También se describen composiciones, kits y sistemas que encuentran uso en la realización de los procedimientos réflex descritos en la presente memoria.

### Breve descripción de los dibujos

40 La invención se comprende mejor a partir de la siguiente descripción detallada cuando se lee conjuntamente con los dibujos adjuntos. Se destaca que, de acuerdo con la práctica común, las diversas características de los dibujos no están a escala. De hecho, las dimensiones de las diversas características están arbitrariamente ampliadas o reducidas para una mayor claridad. Las siguientes figuras están incluidas en los dibujos:

45 Figura 1: el Panel A es un diagrama esquemático que ilustra el movimiento de un primer dominio de un sitio a otro en una molécula de ácido nucleico utilizando una secuencia réflex. El Panel B es un diagrama esquemático que representa la posición relativa de los pares de cebadores (cebadores  $A_n$ - $B_n$ ) que encuentran uso en los aspectos del procedimiento réflex descrito en la presente memoria.

La Figura 2 muestra una realización ilustrativa de la utilización de pares de compañeros de unión (biotina/estreptavidina) para aislar polinucleótidos de cadena sencilla de interés.

La Figura 3 es un diagrama esquemático que ilustra una realización ilustrativa para mover un sitio de cebador y un MID a una localización específica en un ácido nucleico de interés.

50 La Figura 4 muestra un diagrama esquemático que ilustra un uso ilustrativo del procedimiento réflex para la

generación de una muestra enriquecida en fragmentos que tienen una región de interés (p. ej., de una población de polinucleótidos etiquetados asimétricamente y fragmentados al azar).

5 La Figura 5 muestra una comparación de métodos para la identificación de polimorfismos de ácidos nucleicos en ácidos nucleicos homólogos en una muestra (p. ej., la misma región derivada de un par de cromosomas de una célula diploide o genomas/transcritos virales). El esquema superior muestra dos moléculas de ácido nucleico en una muestra (1 y 2) que tienen un surtido diferente de polimorfismos en sitios polimórficos A, B y C (A1, B1, C1 y C2). Los métodos de secuenciación convencionales que utilizan la fragmentación (lado izquierdo) pueden identificar los polimorfismos en estos ácidos nucleicos, pero no retienen la información de la conexión. Empleando el procedimiento réflex descrito en la presente memoria para identificar polimorfismos (lado derecho) se mantiene la información de la conexión.

Figura 6: el Panel A es un esquema que muestra estructuras y tamaños esperados de especies de ácidos nucleicos en el procedimiento réflex; el Panel B es un gel de poliacrilamida que muestra las especies de ácidos nucleicos producidas en el procedimiento réflex descrito en el Ejemplo 1.

15 Figura 7: el Panel A es un esquema que muestra la estructura del ácido nucleico y el competidor utilizados en el procedimiento réflex; el Panel B es un gel de poliacrilamida que muestra las especies de ácidos nucleicos producidas en el procedimiento réflex descrito en el Ejemplo 1.

La Figura 8 muestra un diagrama de flujo de un procedimiento réflex (izquierda) en el que la etapa de la exonucleasa T7 es opcional. El gel de la derecha muestra el producto resultante del procedimiento réflex, ya sea sin la etapa de la exonucleasa T7 (calle 1) o con la etapa de la exonucleasa T7 (calle 2).

20 La Figura 9 muestra un flujo de trabajo del procedimiento réflex ilustrativo con indicaciones a la derecha en cuanto a dónde se emplea la purificación de productos de reacción (p. ej., utilizando cuentas Agencourt para eliminar oligonucleótidos cebadores).

25 La Figura 10 muestra el material de partida (panel izquierdo) y el producto resultante generado (panel derecho) utilizando un procedimiento réflex sin emplear una etapa de exonucleasa T7 (como se describe en el Ejemplo II). El sitio réflex en el material de partida es una secuencia normalmente presente en el polinucleótido que se está procesando (también llamado un sitio réflex "no-artificial"). Esta figura muestra que el ácido nucleico de partida de 755 pares de bases era procesado al producto esperado de 461 pares de bases, lo que confirma que un sitio réflex "no-artificial" es eficaz en la transferencia de un dominio adaptador desde un lugar a otro en un polinucleótido de interés de una manera específica de la secuencia.

30 La Figura 11 muestra un esquema y los resultados de un experimento en el que el procedimiento réflex se lleva a cabo sobre un único molde inicial grande (un fragmento "parental") para generar cinco productos diferentes (productos "hijos") que tienen cada uno una región diferente de interés (es decir, se producen productos hijos que tienen cualquiera de las regiones 1, 2, 3, 4 ó 5).

35 La Figura 12 muestra un esquema y los resultados de experimentos llevados a cabo para determinar la prevalencia del reordenamiento intramolecular durante el procedimiento réflex (si se desea) vs. reordenamiento intermolecular (conmutación MID).

La Figura 13 muestra un diagrama de flujos de trabajo ilustrativos para la preparación de material y para la realización del procedimiento réflex.

### Definiciones

40 A menos que se definan de otro modo, todos los términos técnicos y científicos utilizados en la presente memoria tienen el mismo significado comúnmente comprendido por un experto normal en la técnica a la que pertenece esta invención. Sin embargo, ciertos elementos se definen en aras de claridad y facilidad de referencia.

45 Los términos y símbolos de la química de ácidos nucleicos, la bioquímica, la genética y la biología molecular utilizados en la presente memoria siguen los de tratados y textos convencionales en el campo, p. ej., Kornberg y Baker, DNA Replication, Segunda Edición (W.H. Freeman, Nueva York, 1992); Lehninger, Biochemistry, Segunda Edición (Worth Publishers, Nueva York, 1975); Strachan y Read, Human Molecular Genetics, Segunda Edición (Wiley-Liss, Nueva York, 1999); Eckstein, editor, Oligonucleotides and Analogs: A Practical Approach (Oxford University Press, Nueva York, 1991); Gait, editor, Oligonucleotide Synthesis: A Practical Approach (IRL Press, Oxford, 1984); y similares.

50 "Amplicon" representa el producto de una reacción de amplificación de polinucleótidos. Es decir, es una población de polinucleótidos, generalmente de doble hebra, que replican a partir de una o más secuencias de partida. Las una o más secuencias de partida pueden ser una o más copias de la misma secuencia, o pueden ser una mezcla de diferentes secuencias. Los amplicones se pueden producir por medio de una variedad de reacciones de amplificación cuyos productos son múltiples repeticiones de uno o más ácidos nucleicos diana. Generalmente, las reacciones de amplificación que producen amplicones son "dirigidas por moldes" ya que el emparejamiento de bases

de los reaccionantes, ya sean nucleótidos u oligonucleótidos, tienen complementos en un polinucleótido molde que son necesarios para la creación de productos de reacción. En un aspecto, las reacciones dirigidas por moldes son extensiones de cebadores con una ácido nucleico polimerasa o ligaciones de oligonucleótidos con una ácido nucleico ligasa. Tales reacciones incluyen, pero no se limitan a, reacciones en cadena de la polimerasa (PCR), reacciones de polimerasas lineales, amplificaciones basadas en secuencias de ácido nucleico (NASBA), amplificaciones de círculo rodante, y similares, descritas en las siguientes referencias: Mullis et al., Patentes de los Estados Unidos Núm. 4.683.195; 4.965.188; 4.683.202; 4.800.159 (PCR); Gelfand et al., Patente de los Estados Unidos Núm. 5.210.015 (PCR en tiempo real con sondas "TAQMAN™"); Wittwer et al., Patente de los Estados Unidos Núm. 6.174.670; Kacian et al., Patente de los Estados Unidos Núm. 5.399.491 ("NASBA"); Lizardi, Patente de los Estados Unidos Núm. 5.854.033; Aono et al., Publ. de Patente Japonesa. JP 4-262799 (amplificación por círculo rodante); y similares. En un aspecto, los amplicones de la invención son producidos por PCR. Una reacción de amplificación puede ser una amplificación en "tiempo real" si se encuentra disponible una química de detección que permita medir un producto de reacción a medida que progresa la reacción de amplificación, p. ej., "PCR en tiempo real" descrita a continuación, o "NASBA en tiempo real", como describen Leone et al., en *Nucleic Acids Research*, 26: 2150-2155 (1998), y referencias similares. Según se utiliza en la presente memoria, el término "amplificación" se refiere a la realización de una reacción de amplificación. Una "mezcla de reacción" significa una solución que contiene todos los reactivos necesarios para realizar una reacción, que puede incluir, pero no se limita a, agentes tamponadores para mantener el pH a un nivel seleccionado durante una reacción, sales, cofactores, captadores, y similares.

El término "valoración" incluye cualquier forma de medición, e incluye la determinación de si un elemento está presente o no. Los términos "determinación", "medición", "evaluación", "valoración" y "análisis" se utilizan indistintamente, e incluyen determinaciones cuantitativas y cualitativas. La valoración puede ser relativa o absoluta. La "valoración de la presencia de" incluye la determinación de la cantidad de algo presente, y/o la determinación de si está presente o ausente. Según se utilizan en la presente memoria, los términos "determinación", "medición" y "valoración" y "análisis" se utilizan indistintamente e incluyen determinaciones tanto cuantitativas como cualitativas.

Los polinucleótidos que están "etiquetados asimétricamente" tienen dominios adaptadores a la derecha y a la izquierda que no son idénticos. Este procedimiento es referido genéricamente como anclaje de adaptadores asimétricamente o etiquetado asimétrico de un polinucleótido, p. ej., un fragmento de polinucleótido. La producción de polinucleótidos que tienen extremos adaptadores asimétricos puede conseguirse de cualquier manera conveniente. Los adaptadores asimétricos ilustrativos se describen en: las Patentes de los Estados Unidos Núms. 5.712.126 y 6.372.434; las Publicaciones de Patente de los Estados Unidos 2007/0128624 y 2007/0172839; y la Publicación PCT WO/2009/032167. En ciertas realizaciones, los adaptadores asimétricos empleados son los descritos en la Patente de los Estados Unidos 2009/0275087.

Como ejemplo, un usuario de la presente invención puede utilizar un adaptador asimétrico para etiquetar polinucleótidos. Un "adaptador asimétrico" es uno que, cuando se liga a ambos extremos de un fragmento de ácido nucleico de doble hebra, da lugar a la producción de productos de extensión o amplificación de cebadores que tienen secuencias no idénticas que flanquean la inserción genómica de interés. La ligación está seguida generalmente de etapas de procesamiento posteriores con el fin de generar las secuencias adaptadoras terminales no idénticas. Por ejemplo, la replicación de uno o varios fragmentos anclados al adaptador asimétrico da como resultado productos de polinucleótidos en los que hay al menos una diferencia en la secuencia de ácido nucleico, o una modificación de nucleótido/nucleósido, entre las secuencias adaptadoras terminales. El anclaje de adaptadores asimétricamente a polinucleótidos (p. ej., fragmentos de polinucleótidos) da como resultado polinucleótidos que tienen una o más secuencias adaptadoras en un extremo (p. ej., una o más regiones o dominios, p. ej., un sitio de cebador) que, o bien no están presentes o bien tienen una secuencia de ácido nucleico diferente en comparación con la secuencia adaptadora en el otro extremo. Se observa que un adaptador que se denomina un "adaptador asimétrico" no es necesariamente en sí estructuralmente asimétrico, ni el mero acto de anclar un adaptador asimétrico a un fragmento de polinucleótido lo hace inmediatamente asimétrico. Más bien, un polinucleótido anclado a un adaptador asimétrico, que tiene un adaptador asimétrico idéntico en cada extremo, produce productos de replicación (o polinucleótidos de hebra sencilla aislados) que son asimétricos con respecto a las secuencias adaptadoras en los extremos opuestos (p. ej., después de al menos una ronda de la amplificación/extensión de cebadores).

Cualquier adaptador asimétrico conveniente, o procedimiento para la fijación de adaptadores asimétricamente, se pueden emplear en la práctica de la presente invención. Los adaptadores asimétricos ilustrativos se describen en: las Patentes de los Estados Unidos Núm. 5.712.126 y 6.372.434; las Publicaciones de Patente de los Estados Unidos 2007/0128624 y 2007/0172839; y la Publicación PCT WO/2009/032167. En ciertas realizaciones, los adaptadores asimétricos empleados son los descritos en el documento US 2009/0275087.

"Complementario" o "sustancialmente complementario" se refiere a la hibridación o al emparejamiento de bases o a la formación de un dúplex entre nucleótidos o ácidos nucleicos, tales como, por ejemplo, entre las dos hebras de una molécula de ADN de doble hebra o entre un cebador de oligonucleótido y un sitio del cebador en un ácido nucleico de hebra sencilla. Los nucleótidos complementarios son, generalmente, A y T (o A y U), o C y G. Se dice que dos moléculas de ARN o ADN de hebra sencilla son sustancialmente complementarias cuando los nucleótidos de una

hebra, óptimamente alineados y comparados y con inserciones o deleciones de nucleótidos apropiadas, se emparejan con al menos aproximadamente 80% de los nucleótidos de la otra hebra, habitualmente al menos aproximadamente 90% a 95%, y más preferiblemente de aproximadamente 98 a 100%. Alternativamente, existe complementariedad sustancial cuando una hebra de ARN o ADN hibrida en condiciones de hibridación selectivas con su complemento. Típicamente, la hibridación selectiva se producirá cuando haya al menos aproximadamente 65% complementarios sobre un tramo de al menos 14 a 25 nucleótidos, preferiblemente al menos aproximadamente 75%, más preferiblemente al menos aproximadamente 90% complementarios. Véase, M. Kanehisa *Nucleic Acids Res.* 12: 203 (1984).

"Dúplex" representa al menos dos oligonucleótidos y/o polinucleótidos que son totalmente o parcialmente complementarios que se someten a emparejamiento de bases de tipo Watson-Crick entre todos o la mayoría de sus nucleótidos de manera que se forma un complejo estable. Los términos "emparejamiento" e "hibridación" se utilizan indistintamente para referirse a la formación de un dúplex estable. "Perfectamente emparejado" en referencia a un dúplex significa que las hebras de poli- u oligo-nucleótidos que constituyen el dúplex forman una estructura de doble hebra entre sí de tal manera que cada nucleótido de cada hebra se somete a emparejamiento de bases de Watson-Crick con un nucleótido de la otra hebra. Un dúplex estable puede incluir emparejamiento de bases de Watson-Crick y/o emparejamiento de bases distinto del de Watson-Crick entre las hebras del dúplex (donde el emparejamiento de bases representa la formación de enlaces de hidrógeno). En ciertas realizaciones, un par de bases de Watson-Crick incluye un análogo de nucleósido, tal como desoxiinosina, 2,6-diaminopurina, PNA, LNA y similares. En ciertas realizaciones, un par de bases distinto de uno de Watson-Crick incluye una "base de balanceo", tal como desoxiinosina, 8-oxo-dA, 8-oxo-dG y similares, en donde por "base de balanceo" se entiende una base de ácido nucleico que puede emparejar bases con una primera base de nucleótido en una hebra de ácido nucleico complementaria, pero que, cuando se emplea como una hebra molde para la síntesis de ácido nucleico, conduce a la incorporación de una segunda base nucleotídica diferente en la hebra de síntesis (las bases de balanceo se describen con más detalle a continuación). Un "emparejamiento erróneo" en un dúplex entre dos oligonucleótidos o polinucleótidos significa que un par de nucleótidos del dúplex no logra experimentar unión de Watson-Crick.

"Locus genético", "locus", o "locus de interés", en referencia a un genoma o polinucleótido diana, significa una sub-región o segmento contiguos del genoma o el polinucleótido diana. Según se utiliza en la presente memoria, locus genético, locus o locus de interés pueden hacer referencia a la posición de un nucleótido, un gen o una porción de un gen en un genoma, incluyendo ADN mitocondrial u otro ADN no cromosómico (p. ej., plásmido bacteriano), o pueden hacer referencia a cualquier porción contigua de la secuencia genómica esté o no esté dentro, o asociada con, un gen. Un locus genético, locus o locus de interés puede ser de un solo nucleótido a un segmento de unos pocos cientos o unos pocos miles de nucleótidos de longitud o más. En general, un lugar de interés tendrá una secuencia réflex asociada a ella (véase la descripción de "secuencia réflex" de más abajo).

"Kit" se refiere a cualquier sistema de liberación para liberar materiales o reactivos para llevar a cabo un método de la invención. En el contexto de los análisis de reacción, tales sistemas de liberación incluyen sistemas que permiten el almacenamiento, transporte o liberación de reactivos de reacción (p. ej., sondas, enzimas, etc. en los envases apropiados) y/o materiales de soporte (p. ej., tampones, instrucciones escritas para realizar el análisis, etc.) de una ubicación a otra. Por ejemplo, los kits incluyen uno o más recintos (p. ej., cajas) que contienen los reactivos de reacción pertinentes y/o materiales de soporte. Tales contenidos se pueden liberar al receptor deseado juntos o por separado. Por ejemplo, un primer envase puede contener una enzima para su uso en un análisis, mientras que un segundo envase contiene las sondas.

"Ligación" significa para formar un enlace o conexión covalente entre los extremos de dos o más ácidos nucleicos, p. ej., oligonucleótidos y/o polinucleótidos, en una reacción dirigida por un molde. La naturaleza del enlace o conexión puede variar ampliamente y la ligación se puede llevar a cabo enzimáticamente o químicamente. Según se utiliza en la presente memoria, las ligaciones se llevan a cabo normalmente enzimáticamente para formar una conexión fosfodiéster entre un carbono 5' de un nucleótido terminal de un oligonucleótido con un carbono 3' de otro oligonucleótido. Una variedad de reacciones de ligación dirigidas por moldes se describe en las siguientes referencias: Whiteley et al., Patente de los Estados Unidos Núm. 4.883.750; Letsinger et al., Patente de los Estados Unidos Núm. 5.476.930; Fung et al., Patente de los Estados Unidos Núm. 5.593.826; Kool, Patente de los Estados Unidos Núm. 5.426.180; Landegren et al., Patente de los Estados Unidos Núm. 5.871.921; Xu y Kool, *Nucleic Acids Research*, 27: 875-881 (1999); Higgins et al., *Methods in Enzymology*, 68: 50-71 (1979); Engler et al., *The Enzymes*, 15: 3 a 29 (1982); y Namsaraev, Publicación de la Patente de los Estados Unidos 2004/0110213.

"Identificador múltiple" (MID) según se utiliza en la presente memoria se refiere a una etiqueta o una combinación de etiquetas asociadas con un polinucleótido cuya identidad (p. ej., la secuencia de ADN etiquetada) se puede utilizar para diferenciar polinucleótidos en una muestra. En ciertas realizaciones, el MID en un polinucleótido se utiliza para identificar la fuente de la que deriva el polinucleótido. Por ejemplo, una muestra de ácido nucleico puede ser una reserva de polinucleótidos derivados de diferentes fuentes (p. ej., polinucleótidos derivados de diferentes individuos, diferentes tejidos o células, o polinucleótidos aislados en diferentes momentos puntuales), donde los polinucleótidos de cada fuente diferente están etiquetados con un MID único. Como tal, un MID proporciona una correlación entre un polinucleótido y su fuente. En ciertas realizaciones, se emplean MID para etiquetar de forma exclusiva cada polinucleótido individual en una muestra. La identificación del número de MID únicos en una muestra puede proporcionar una lectura de la cantidad de polinucleótidos individuales que están presentes en la muestra (o

de a partir de cuántos polinucleótidos originales deriva una muestra de polinucleótidos manipulados; véase, p. ej., la Patente de los Estados Unidos Núm. 7.537.897, presentada el 26 de Mayo de 2009. Los MID puede variar en longitud de 2 a 100 bases de nucleótidos o más, y pueden incluir múltiples subunidades, donde cada MID diferente tiene una identidad y/o un orden de las subunidades distintos. Las etiquetas de ácidos nucleicos ilustrativas que encuentran uso como MID se describen en la Patente de los Estados Unidos Núm. 7.544.473, presentada el 6 de Junio de 2009, y titulada "Nucleic Acid Analysis Using Sequence Tokems", así como la Patente de los Estados Unidos Núm. 7.393.665, presentada el 1 de Julio de 2008, y titulada "Methods and Compositions for Tagging and Identifying Polynucleotides". En ciertas realizaciones, un conjunto de MID empleados para etiquetar una pluralidad de muestras no tienen por qué tener ninguna propiedad común concreta (p. ej., Tm, longitud, composición de bases, etc.), ya que los métodos descritos en la presente memoria pueden adaptarse a una amplia variedad de conjuntos de MID únicos. Se destaca aquí que los MID sólo necesitan ser únicos dentro de un determinado experimento. Así, el mismo MID se puede utilizar para etiquetar una muestra diferente que se está siendo procesada en un experimento diferente. Además, en ciertos experimentos, un usuario puede utilizar el mismo MID para etiquetar un subconjunto de muestras diferentes dentro del mismo experimento. Por ejemplo, todas las muestras derivadas de los individuos que tienen un fenotipo específico pueden ser etiquetadas con el mismo MID, p. ej., todas las muestras derivadas de sujetos de control (o de tipo salvaje) se pueden etiquetar con un primer MID mientras que los sujetos que tienen un estado de enfermedad pueden ser etiquetados con un segundo MID (diferente del primer MID). Como otro ejemplo, puede ser deseable etiquetar diferentes muestras derivadas de la misma fuente con diferentes MID (p. ej., muestras derivadas en el tiempo o derivadas de diferentes sitios dentro de un tejido). Además, los MID pueden ser generados de varias maneras diferentes, p. ej., mediante un enfoque combinatorio de etiquetado en el que un MID está unido mediante ligación y un segundo MID está unido por extensión del cebador. Por lo tanto, los MID se pueden diseñar y aplicar MID de varias maneras diferentes para realizar un seguimiento de los fragmentos de polinucleótidos durante el procesamiento y el análisis, y por lo tanto no se pretende ninguna limitación en este sentido.

"Nucleósidos" según se utiliza en la presente memoria incluye los nucleósidos naturales, incluyendo las formas 2'-desoxi y 2'-hidroxilo, p. ej., como describen Kornberg y Baker, en DNA Replication, 2ª Ed. (Freeman, San Francisco, 1992). "Análogos" en referencia a los nucleósidos incluye nucleósidos sintéticos que tienen radicales base modificados y/o radicales azúcar modificados, p. ej., descritos por Scheit, Nucleotide Analogs (John Wiley, Nueva York, 1980); Uhlman y Peyman, Chemical Reviews, 90: 543-584 (1990), o similares, con la condición de que sean susceptibles de hibridación específica. Tales análogos incluyen nucleósidos sintéticos diseñados para mejorar propiedades de unión, reducir la complejidad, aumentar la especificidad, y similares. Los polinucleótidos que comprenden análogos con propiedades de hibridación o de resistencia a nucleasa mejoradas son descritos por Uhlman y Peyman (citado anteriormente); Crooke et al., Exp. Opin. Ther. Patents, 6: 855-870 (1996); Mesmaeker et al., Current Opinion in Structural Biology, 5: 343-355 (1995); y similares. Los tipos ilustrativos de polinucleótidos que son capaces de mejorar la estabilidad del dúplex incluyen N3'→ P5' fosforamidatos de oligonucleótidos (denominados en la presente memoria "amidatos"), ácidos peptidonucleicos (denominados en la presente memoria "PNA"), oligo-2'-O-alkilribonucleotidos, polinucleótidos que contienen propinilpirimidinas C-5, ácidos nucleicos bloqueados ("LNA"), y compuestos similares. Tales oligonucleótidos se encuentran disponibles comercialmente o se pueden sintetizar utilizando métodos descritos en la bibliografía.

"Reacción en Cadena de la Polimerasa", o "PCR", significa una reacción para la amplificación in vitro de secuencias de ADN específicas por medio de la extensión simultánea con cebador de hebras complementarias de ADN. En otras palabras, la PCR es una reacción para elaborar múltiples copias o réplicas de un ácido nucleico diana flanqueado por sitios de los cebadores, comprendiendo semejante reacción una o más repeticiones de las siguientes etapas: (i) desnaturalización del ácido nucleico diana, (ii) emparejamiento de los cebadores a los sitios de los cebadores, y (iii) extensión de los cebadores mediante una polimerasa de ácido nucleico en presencia de nucleósidos trifosfato. Por lo general, la reacción se cicla a través de diferentes temperaturas optimizadas para cada etapa en un aparato termociclador. Las temperaturas concretas, las duraciones en cada etapa, y las tasas de cambio entre las etapas dependen de muchos factores bien conocidos para los expertos en la técnica, p. ej., ilustrados por las referencias: McPherson et al., editores, PCR: A Practical Approach y PCR2: A Practical Approach (IRL Press, Oxford, 1991 y 1995, respectivamente). Por ejemplo, en una PCR convencional utilizando ADN polimerasa Taq, un ácido nucleico diana de doble hebra se puede desnaturalizar a una temperatura >90°C, emparejar los cebadores a una temperatura en el intervalo de 50-75°C, y extender los cebadores a una temperatura en el intervalo de 72-78°C. El término "PCR" abarca formas derivadas de la reacción, incluyendo, pero no limitadas a, RT-PCR, PCR en tiempo real, PCR anidada, PCR cuantitativa, PCR multiplexada, y similares. Los volúmenes de reacción van desde unos pocos cientos de nanolitros, p. ej., 200 nL, a unos pocos cientos de µL, p. ej., 200 µL. La "PCR con transcripción inversa", o "RT-PCR", significa una PCR que está precedida por una reacción de transcripción inversa que convierte un ARN diana en un ADN complementario de hebra sencilla, que se amplifica a continuación, p. ej., Tecott et al., Patente de los Estados Unidos Núm. 5.168.038. La "PCR en tiempo real" significa una PCR para la que la cantidad de producto de reacción, es decir amplicón, se controla a medida que avanza la reacción. Hay muchas formas de PCR en tiempo real que difieren principalmente en las químicas de detección utilizadas para controlar el producto de reacción, p. ej., Gelfand et al., Patente de los Estados Unidos Núm. 5.210.015 ("TAQMAN™"); Wittwer et al., Patentes de los Estados Unidos Núm. 6.174.670 y 6.569.627 (colorantes intercalantes); Tyagi et al., Patente de los Estados Unidos Núm. 5.925.517 (balizas moleculares). Las químicas de detección para la PCR en tiempo real son revisadas por Mackay et al., en Nucleic Acids Research, 30: 1292-1305 (2002). La "PCR anidada" significa una PCR

de dos etapas en la que el amplicón de una primera PCR se convierte en la muestra para la segunda PCR utilizando un nuevo conjunto de cebadores, al menos uno de los cuales se une a un lugar interior del primer amplicón. Según se utiliza en la presente memoria, "cebadores iniciales" en referencia a una reacción de amplificación anidada significa los cebadores utilizados para generar un primer amplicón, y "cebadores secundarios" significa los uno o más cebadores utilizados para generar un segundo amplicón o amplicón anidado. "PCR multiplexada" significa una PCR en donde múltiples secuencias diana (o una sola secuencia diana y una o más secuencias de referencia) se llevan a cabo simultáneamente en la misma mezcla de reacción, p. ej., Bernard et al., *Anal. Biochem.*, 273: 221-228 (1999) (PCR de dos colores en tiempo real). Por lo general, se emplean distintos conjuntos de cebadores para cada secuencia que se está amplificando.

"PCR cuantitativa" significa una PCR diseñada para medir la abundancia de una o más secuencias diana específicas en una muestra o espécimen. La PCR cuantitativa incluye tanto la cuantificación absoluta como la cuantificación relativa de tales secuencias diana. Las mediciones cuantitativas se realizan utilizando una o más secuencias de referencia que se pueden analizar por separado o junto con una secuencia diana. La secuencia de referencia puede ser endógena o exógena con respecto a una muestra o espécimen, y en este último caso, puede comprender uno o más moldes competidores. Las secuencias de referencia endógenas típicas incluyen segmentos de transcritos de los siguientes genes:  $\beta$ -actina, GAPDH,  $\beta_2$ -microglobulina, ARN ribosomal, y similares. Los mecanismos para la PCR cuantitativa son bien conocidos por los expertos normales en la técnica, como se ilustra en las siguientes referencias: Freeman et al., *Biotechniques*, 26: 112-126 (1999); Becker-Andre et al., *Nucleic Acids Research*, 17: 9437-9447 (1989); Zimmerman et al., *Biotechniques*, 21: 268-279 (1996); Diviacco et al., *Gene*, 122: 3013-3020 (1992); Becker-Andre et al., *Nucleic Acids Research*, 17: 9437-9446 (1989); y similares.

"Polinucleótido" o "oligonucleótido" se usan de forma indistinta y cada uno significa un polímero lineal de monómeros de nucleótidos. Los monómeros que constituyen los polinucleótidos y oligonucleótidos son capaces de unirse específicamente a un polinucleótido natural, por medio de un patrón regular de interacciones monómero a monómero, tales como el emparejamiento de bases de tipo Watson-Crick, el apilamiento de bases, los emparejamientos de bases de Hoogsteen o de Hoogsteen inverso, el emparejamiento de bases por balanceo, o similares. Como se describe con detalle a continuación, por "base de balanceo" se entiende una base de ácido nucleico que puede emparejar bases con una primera base de nucleótido en una hebra de ácido nucleico complementaria, pero que, cuando se emplea como una hebra molde para la síntesis de ácido nucleico, conduce a la incorporación de una segunda base de nucleótido diferente, en la hebra de síntesis. Tales monómeros y sus conexiones internucleosídicas pueden ser de origen natural o pueden ser análogos de los mismos, p. ej., análogos de origen natural o de origen no natural. Los análogos de origen no natural pueden incluir ácidos peptidonucleicos (PNAS, p. ej., como se describe en la Patente de los Estados Unidos Núm. 5.539.082), ácidos nucleicos bloqueados (LNA, p. ej., como se describe en la Patente de los Estados Unidos Núm. 6.670.461), conexiones internucleosídicas fosforotioato, bases que contienen grupos de conexión que permiten el anclaje de marcas, tales como fluoróforos, o haptenos y similares. Siempre que el uso de un oligonucleótido o polinucleótido requiera un procesamiento enzimático, tal como la extensión por una polimerasa, la ligación por una ligasa, o similar, un experto en la técnica entenderá que los oligonucleótidos o polinucleótidos en esos casos no contendrán ciertos análogos de conexiones internucleosídicas, radicales azúcar, o bases en cualquiera o en alguna posición. Los polinucleótidos normalmente varían en tamaño desde unas pocas unidades monoméricas, p. ej., 5 - 40, cuando son referidos generalmente como "oligonucleótidos", a varios miles de unidades monoméricas. Cada vez que un polinucleótido u oligonucleótido está representado por una secuencia de letras (mayúsculas o minúsculas), tales como "ATGCCTG", se entenderá que los nucleótidos están en un orden 5'  $\rightarrow$  3', de izquierda a derecha y que "A" denota desoxiadenosina, "C" indica desoxicidina, "G" indica desoxiguanosina, y "T" indica timidina, "I" denota desoxiinosina, "U" indica uridina, salvo que se indique lo contrario o resulte evidente a partir del contexto. A menos que se indique lo contrario la terminología y las convenciones de numeración de átomos seguirán las descritas por Strachan y Read, en *Human Molecular Genetics 2* (Wiley-Liss, Nueva York, 1999). Por lo general, los polinucleótidos comprenden los cuatro nucleósidos naturales (p. ej., desoxiadenosina, desoxicidina, desoxiguanosina, desoxitimidina para el ADN o sus contrapartes de ribosa para el ARN) unidos por enlaces fosfodiéster; sin embargo, también pueden comprender análogos de nucleótidos no naturales, p. ej., incluyendo bases modificadas, azúcares, o enlaces internucleosídicos. Es evidente para los expertos en la técnica que, cuando una enzima tiene requisitos específicos de sustratos de oligonucleótidos o polinucleótidos para su actividad, p. ej., ADN de hebra sencilla, dúplex de ARN/ADN, o similar, en ese caso la selección de la composición apropiada para los sustratos de oligonucleótidos o polinucleótidos se encuentra dentro del conocimiento de un experto normal, especialmente con la orientación de los tratados, tales como Sambrook et al., *Molecular Cloning, Segunda Edición* (Cold Spring Harbor Laboratory, Nueva York, 1989), y referencias similares.

"Cebador" significa un oligonucleótido, ya sea natural o sintético, que es capaz, después de formar un dúplex con un molde de polinucleótido, de actuar como un punto de iniciación de la síntesis de ácidos nucleicos y de extenderse desde su extremo 3' a lo largo del molde de modo que se forme un dúplex extendido. La secuencia de nucleótidos añadida durante el procedimiento de extensión está determinada por la secuencia del polinucleótido molde. Por lo general, los cebadores son extendidos por una polimerasa de ADN. Los cebadores tienen generalmente una longitud compatible con su utilización en la síntesis de productos de extensión del cebador, y se encuentra en el intervalo entre 8 y 100 nucleótidos de longitud, tal como de 10 a 75, de 15 a 60, de 15 a 40, de 18 a 30, de 20 a 40, de 21 a 50, de 22 a 45, de 25 a 40, y así sucesivamente, más típicamente en el intervalo entre 18-40, 20-35, 21-30

nucleótidos de longitud, y cualquier longitud entre los intervalos indicados. Los cebadores típicos pueden estar en el intervalo entre 10-50 nucleótidos de longitud, tal como 15-45, 18-40, 20-30, 21-25 y así sucesivamente, y cualquier longitud entre los intervalos indicados. En algunas realizaciones, los cebadores tienen por lo general no más de aproximadamente 10, 12, 15, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, o 70 nucleótidos de longitudes

Los cebadores son generalmente de hebra sencilla para una máxima eficacia en la amplificación, pero pueden ser alternativamente de doble hebra. Si son de doble hebra, el cebador generalmente se trata primero para separar sus hebras antes de ser utilizado para preparar productos de extensión. Esta etapa de desnaturalización resulta afectada típicamente por el calor, pero alternativamente se puede llevar a cabo utilizando álcali, seguido de neutralización. Por lo tanto, un "cebador" es complementario a un molde, y forma complejos por enlace de hidrógeno o hibridación con el molde para proporcionar un complejo de cebador/molde para la iniciación de la síntesis por una polimerasa, que se extiende por la adición de bases unidas covalentemente conectadas a su extremo 3' complementario al molde en el procedimiento de síntesis de ADN.

Un "par de cebadores" según se utiliza en la presente memoria se refiere a un primer y un segundo cebadores que tienen una secuencia de ácido nucleico adecuada para la amplificación basada en el ácido nucleico de un ácido nucleico diana. Tales pares de cebadores incluyen generalmente un primer cebador que tiene una secuencia que es la misma o similar a la de una primera porción de un ácido nucleico diana, y un segundo cebador que tiene una secuencia que es complementaria a una segunda porción de un ácido nucleico diana para proporcionar la amplificación del ácido nucleico diana o un fragmento del mismo. La referencia al "primer" y "segundo" cebadores en la presente memoria es arbitraria, a menos que se indique específicamente lo contrario. Por ejemplo, el primer cebador puede ser diseñado como un "cebador directo" (que inicia la síntesis de ácido nucleico a partir de un extremo 5' del ácido nucleico diana) o como un "cebador inverso" (que inicia la síntesis de ácido nucleico a partir de un extremo 5' del producto de extensión producido a partir de la síntesis iniciada desde el cebador directo). Del mismo modo, el segundo cebador puede ser diseñado como un cebador directo o un cebador inverso.

"Sitio del cebador" (p. ej., un sitio de cebador de secuenciación, y un sitio de amplificación de cebador, etc.) tal como se utiliza en la presente memoria, se refiere a un dominio de un polinucleótido que incluye la secuencia de un cebador (p. ej., un cebador de secuenciación) y/o la secuencia complementaria de un cebador. Cuando está presente en forma de hebra sencilla (p. ej., en un polinucleótido de hebra sencilla), un sitio del cebador puede tener o bien la secuencia idéntica de un cebador o bien la secuencia complementaria de un cebador. Cuando está presente en forma de doble hebra, un sitio del cebador contiene la secuencia de un cebador hibridado a la secuencia complementaria del cebador. Por lo tanto, un sitio del cebador es una región de un polinucleótido que o bien es idéntica o bien es complementaria a la secuencia de un cebador (cuando está en forma de hebra sencilla) o una región de doble hebra formada entre una secuencia de cebador y su complemento. Los sitios de los cebadores pueden estar presentes en un adaptador anclado a un polinucleótido. La orientación específica de un sitio del cebador puede ser deducida por un experto normal en la técnica a partir de las características estructurales del polinucleótido relevante y/o el contexto en el que se utiliza.

"Lectura" significa un parámetro, o parámetros, que se miden y/o detectan y que se pueden convertir en un número o valor. En algunos contextos, la lectura se puede referir a una representación numérica real de tales datos recogidos o registrados. Por ejemplo, una lectura de las señales de intensidad de fluorescencia de una micromatriz es la dirección y la intensidad de fluorescencia de una señal que se genera en cada sitio de hibridación de la micromatriz; por lo tanto, una lectura de este tipo puede ser registrada o almacenada de varias formas, por ejemplo, como una imagen de la micromatriz, como una tabla de números, o similar.

"Sitio réflex", "secuencia réflex" y equivalentes se utilizan para indicar secuencias en un polinucleótido que se emplean para mover un dominio intramolecularmente desde su ubicación inicial a una ubicación diferente en el polinucleótido. La secuencia de un sitio réflex se puede añadir a un polinucleótido de interés (por ejemplo, presente en un adaptador ligado al polinucleótido), basarse en una secuencia presente naturalmente dentro del polinucleótido de interés (p. ej., una secuencia genómica en el polinucleótido), o una combinación de ambos. La secuencia réflex se elige de manera que sea distinta de otras secuencias del polinucleótido (es decir, con poca homología de secuencia con otras secuencias que puedan estar presentes en el polinucleótido, p. ej., secuencias genómicas o sub-genómicas que vayan a ser procesadas). Como tal, una secuencia réflex se debe seleccionar de manera que no hibride con cualquier secuencia a excepción de su complemento en las condiciones empleadas en los procedimientos réflex descritos en la presente memoria. Como se describe más adelante en esta solicitud, el complemento de la secuencia réflex se inserta en la misma hebra del polinucleótido (p. ej., la misma hebra de un polinucleótido de doble hebra o en el mismo polinucleótido de hebra sencilla) en un lugar concreto con el fin de facilitar un evento de unión intramolecular en dicha hebra en particular. Las secuencias réflex empleadas en el procedimiento réflex descrito en la presente memoria pueden tener por lo tanto una amplia gama de longitudes y secuencias. Las secuencias réflex pueden variar de 5 a 200 bases de nucleótidos de longitud.

"Soporte sólido", "soporte" y "soporte en fase sólida" se utilizan indistintamente y se refieren a un material o grupo de materiales que tienen una superficie o superficies rígidas o semi-rígidas. En muchas realizaciones, al menos una superficie del soporte sólido será sustancialmente plana, aunque en algunas realizaciones puede ser deseable separar físicamente las regiones de síntesis para diferentes compuestos, por ejemplo, con pocillos, regiones

elevadas, pivotes, hoyos grabados al agua fuerte, o similares. De acuerdo con otras realizaciones, el soporte o los soportes sólidos adoptarán la forma de cuentas, resinas, geles, microesferas, u otras configuraciones geométricas. Las micromatrices comprenden normalmente al menos un soporte en fase sólida plana, tal como un portaobjetos de microscopio de vidrio.

5 "Específico" o "especificidad", en referencia a la unión de una molécula a otra molécula, tal como una secuencia diana marcada para una sonda, significa el reconocimiento, el contacto y la formación de un complejo estable entre las dos moléculas, junto con sustancialmente menos reconocimiento, contacto o formación de complejos de esa molécula con otras moléculas. En un aspecto, "específico" en referencia a la unión de una primera molécula a una  
10 segunda molécula significa que en la medida en la que la primera molécula reconoce y forma un complejo con otra molécula en una reacción o muestra, ésta forma el mayor número de los complejos con la segunda molécula. Preferiblemente, este número más grande es al menos cincuenta por ciento. Generalmente, las moléculas implicadas en un evento de unión específica tienen zonas sobre sus superficies o en cavidades que dan lugar al reconocimiento específico entre las moléculas de unión entre sí. Los ejemplos de unión específica incluyen las interacciones anticuerpo-antígeno, las interacciones enzima-sustrato, la formación de dúplex o tríplex entre  
15 polinucleótidos y/u oligonucleótidos, las interacciones biotina-avidina o biotina-estreptavidina, las interacciones receptor-ligando, y similares. Según se utiliza en la presente memoria, "contacto" en referencia a la especificidad o unión específica significa que dos moléculas están lo suficientemente cerca como para que las interacciones químicas débiles no covalentes, tales como las fuerzas de Van der Waals, los enlaces de hidrógeno, las interacciones de apilamiento de bases, las interacciones iónicas e hidrófobas, y similares, dominen la interacción de  
20 las moléculas.

Según se utiliza en la presente memoria, el término " $T_m$ " se utiliza en referencia a la "temperatura de fusión". La temperatura de fusión es la temperatura (p. ej., medida en °C) a la que una población de moléculas de ácido nucleico de doble hebra medió disociadas se convierte en hebras sencillas. Se conocen en la técnica varias ecuaciones para calcular la  $T_m$  de ácidos nucleicos (véase, p. ej., Anderson y Young, Quantitative Filter  
25 Hybridization, in Nucleic Acid Hybridization (1985). Otras referencias (p. ej., Allawi, H.T. y SantaLucia, J., Jr., Biochemistry 36, 10581-94 (1997)) incluyen métodos alternativos de cálculo que tienen en consideración características estructurales y ambientales, así como de secuencia para el cálculo de la  $T_m$ .

"Muestra" significa una cantidad de material de una fuente biológica, ambiental, médica, o de un paciente en el que se busca la detección, medición, o el marcaje de los ácidos nucleicos diana. Por un lado, se pretende incluir un espécimen o cultivo (p. ej., cultivos microbiológicos). Por otro lado, se pretende incluir las muestras tanto biológicas como ambientales. Una muestra puede incluir un espécimen de origen sintético. Las muestras biológicas pueden ser animales, incluidos los seres humanos, fluidos, sólidos (p. ej., heces) o tejidos, así como alimentos y piensos líquidos y sólidos e ingredientes tales como productos lácteos, verduras, carne y subproductos cárnicos, y residuos. Las muestras biológicas pueden incluir materiales tomados de un paciente, incluyendo, pero no limitados a cultivos,  
30 sangre, saliva, líquido cefalorraquídeo, líquido pleural, la leche, linfa, esputo, semen, aspirados con aguja, y similares. Las muestras biológicas se pueden obtener de todas las diversas familias de animales domésticos, así como animales asilvestrados o salvajes, incluyendo, pero no limitados a, animales tales como ungulados, osos, peces, roedores, etc. Las muestras ambientales incluyen material ambiental tal como materia de superficie, suelo, agua y muestras industriales, así como muestras obtenidas de los instrumentales de procesamiento, aparatos, equipamientos, utensilios de alimentos y productos lácteos, artículos desechables y no desechables. Estos ejemplos no deben interpretarse como limitantes de los tipos de muestras aplicables a la presente invención.  
40

Los términos "aguas arriba" y "aguas abajo" en la descripción de la orientación y/o polimerización de las moléculas de ácido nucleico se utilizan en la presente memoria como entiende un experto en la técnica. Como tal, "aguas abajo" generalmente significa que prosigue en la dirección 5' a 3', es decir, la dirección en la que una nucleótido polimerasa extiende normalmente una secuencia, y "aguas arriba" significa generalmente lo contrario. Por ejemplo, un primer cebador que hibrida "aguas arriba" de un segundo cebador en la misma molécula de ácido nucleico diana está situado en el lado 5' del segundo cebador (y de ese modo la polimerización del ácido nucleico a partir del primer cebador prosigue hacia el segundo cebador).  
45

Se observa adicionalmente que las reivindicaciones se pueden redactar para excluir cualquier elemento opcional. Como tal, esta declaración tiene la intención de servir como base antecedente para el uso de una terminología exclusiva tal como "único", "solo" y similares en relación con la citación de los elementos de las reivindicaciones, o el uso de una limitación "negativa".  
50

### Descripción detallada de la invención

La invención se refiere a composiciones y métodos para el reordenamiento intramolecular de ácido nucleico que encuentran uso en diversas aplicaciones de análisis genético, incluyendo la secuenciación, así como las manipulaciones biológicas moleculares generales de las estructuras de los polinucleótidos.  
55

Antes de describir la presente invención, de debe entender que esta invención no está limitada a las realizaciones concretas descritas, ya que estas pueden, por supuesto, variar. También se debe entender que la terminología utilizada en la presente memoria tiene el propósito de describir realizaciones concretas solamente, y no se pretende

que sea limitante, ya que el alcance de la presente invención estará limitado únicamente por las reivindicaciones adjuntas.

5 Cuando se proporciona un intervalo de valores, se entiende que cada valor intermedio, hasta la décima de la unidad del límite inferior, a menos que el contexto indique claramente lo contrario, entre los límites superior e inferior de ese intervalo también se describe específicamente. Cada intervalo más pequeño entre cualquier valor indicado o valor intermedio en un intervalo indicado y cualquier otro valor indicado o intermedio en ese intervalo establecido está englobado dentro de la invención. Los límites superior e inferior de estos intervalos más pequeños pueden estar incluidos o excluidos del intervalo independientemente, y cada intervalo en el que cualquiera, ninguno o ambos límites se incluyen en los intervalos más pequeños también se engloban dentro de la invención, sujeta a cualquier límite excluido específicamente en el intervalo establecido. Cuando el intervalo indicado incluye uno o ambos límites, los intervalos que excluyen cualquiera o ambos de esos límites incluidos también se incluyen en la invención.

15 A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en la presente memoria tienen el mismo significado comúnmente entendido por un experto normal en la técnica a la que esta invención pertenece. Aunque se puede utilizar en la práctica o el ensayo de la presente invención cualquier método y material similar o equivalente a los descritos en la presente memoria, algunos métodos y materiales posibles y preferidos se describen a continuación. Todas las publicaciones mencionadas en la presente memoria revelan y describen los métodos y/o materiales en relación con los cuales se citan las publicaciones. Se entiende que la presente descripción reemplaza cualquier divulgación de una publicación citada en la medida en la que haya una contradicción.

20 Se debe tener en cuenta que, según se utiliza en la presente memoria y en las reivindicaciones adjuntas, las formas en singular "un", "uno", "una", y "el" y "la" incluyen referentes en plural a menos que el contexto indique claramente lo contrario. Así, por ejemplo, la referencia a "un ácido nucleico" incluye una pluralidad de tales ácidos nucleicos y la referencia a "el compuesto" incluye la referencia a uno o más compuestos y equivalentes de los mismos conocidos por los expertos en la técnica, y así sucesivamente.

25 La práctica de la presente invención puede emplear, a menos que se indique lo contrario, mecanismos convencionales y descripciones de la química orgánica, la tecnología de polímeros, la biología molecular (incluyendo técnicas recombinantes), la biología celular, la bioquímica, y la inmunología, que están dentro de la experiencia de la técnica. Tales mecanismos convencionales incluyen la síntesis de matrices de polímeros, la hibridación, la ligación, y la detección de la hibridación utilizando una marca. Las ilustraciones específicas de los mecanismos adecuados se pueden tomar como referencia para el siguiente ejemplo de la presente memoria. Sin embargo, también se pueden utilizar, por supuesto, otros procedimientos convencionales equivalentes. Dichos mecanismos y descripciones convencionales se pueden encontrar en los manuales de laboratorio convencionales, tales como *Genome Analysis: A Laboratory Manual Series* (Vol. I-IV.), *Using Antibodies: A Laboratory Manual*, *Cells: A Laboratory Manual*, *PCR Primer: A Laboratory Manual and Molecular Cloning: A Laboratory Manual* (todos de Cold Spring Harbor Laboratory Press), Stryer, L. (1995) *Biochemistry* (4ª Ed.) Freeman, Nueva York, Gait, "Oligonucleotide Synthesis: A Practical Approach" 1984, IRL Press, Londres, Nelson y Cox (2000), Lehninger, A., *Principles of Biochemistry* 3ª Ed., W.H. Freeman Pub., Nueva York, N.Y. y Berg et al. (2002) *Biochemistry*, 5ª Ed., W. H. Freeman Pub., Nueva York, N.Y.

30 Las publicaciones comentadas en la presente memoria se proporcionan únicamente para su descripción antes de la fecha de presentación de la presente solicitud. Nada en la presente memoria debe interpretarse como una admisión de que la presente invención no tiene derecho a anteceder a dicha publicación en virtud de invención anterior. Adicionalmente, las fechas de publicación proporcionadas pueden ser diferentes de las fechas de publicación reales que pueden necesitar ser confirmadas de forma independiente.

35 Como se resume anteriormente, los aspectos de la presente invención se interesan por el uso de una secuencia "réflex" presente en un polinucleótido (p. ej., en una estructura de adaptador del polinucleótido, en una región genómica del polinucleótido, o una combinación de ambos) para mover un dominio del polinucleótido intramolecularmente desde una primera ubicación a una segunda ubicación. El procedimiento réflex descrito en la presente memoria encuentra uso en cualquiera de las numerosas aplicaciones, p. ej., colocando los elementos funcionales de un polinucleótido (p. ej., sitios de secuenciación del cebador y/o etiquetas MID) en la proximidad con una sub-región de interés deseada.

#### 50 Ácidos nucleicos

El procedimiento réflex (como se describe con detalle más adelante) se puede emplear para la manipulación y el análisis de secuencias de ácidos nucleicos de interés a partir de prácticamente cualquier fuente de ácido nucleico, incluyendo pero sin limitarse a ADN genómico, ADN complementario (ADNc), ARN (p. ej., ARN mensajero, ARN ribosomal, ARN de interferencia corto, microARN, etc.), ADN plasmídico, ADN mitocondrial, ADN sintético, etc. Además, cualquier organismo, material orgánico o sustancia que contiene ácido nucleico se puede utilizar como una fuente de ácidos nucleicos para ser procesada de acuerdo con la presente invención incluyendo, pero no limitados a, plantas, animales (p. ej., reptiles, mamíferos, insectos, gusanos, peces, etc.), muestras de tejidos, bacterias, hongos (p. ej., levadura), fagos, virus, tejido de cadáver, muestras arqueológicas/antiguas, etc. En ciertas realizaciones, los ácidos nucleicos de la muestra de ácido nucleico derivan de un mamífero, donde en determinadas realizaciones el

mamífero es un ser humano.

En ciertas realizaciones, las secuencias de ácidos nucleicos se enriquecen antes del procedimiento de la secuencia réflex. Por enriquecido se quiere significar que el ácido nucleico está sometido a un procedimiento que reduce la complejidad de los ácidos nucleicos, generalmente mediante el aumento de la concentración relativa de determinadas especies de ácidos nucleicos en la muestra (p. ej., que tienen un locus específico de interés, que incluye una secuencia de ácido nucleico específica, que carece de un locus o secuencia, que está dentro de un intervalo de tamaño específico, etc.). Existe una amplia variedad de formas de enriquecer ácidos nucleicos que tienen una o varias características o secuencias específicas, y como tal se puede emplear cualquier método conveniente para lograr esto. El enriquecimiento (o reducción de la complejidad) puede tener lugar en cualquiera de una serie de etapas del procedimiento, y será determinado por los deseos del usuario. Por ejemplo, el enriquecimiento puede tener lugar en las muestras parentales individuales (p. ej., ácidos nucleicos no etiquetados antes de la ligación del adaptador) o en muestras multiplexadas (p. ej., ácidos nucleicos etiquetados con sitios de cebadores, MID y/o secuencias réflex y agrupadas; MID se describe con más detalle a continuación).

En ciertas realizaciones, los ácidos nucleicos de la muestra de ácido nucleico se amplifican antes del análisis. En algunas de estas realizaciones, la reacción de amplificación también sirve para enriquecer una muestra de ácido nucleico de partida en una secuencia o locus de interés. Por ejemplo, una muestra de ácido nucleico de partida puede ser sometida a una reacción en cadena de la polimerasa (PCR) que amplifica una o más regiones de interés. En ciertas realizaciones, la reacción de amplificación es una reacción de amplificación exponencial, mientras que en otras ciertas realizaciones, la reacción de amplificación es una reacción de amplificación lineal. Cualquier método conveniente para llevar a cabo reacciones de amplificación en una muestra de ácido nucleico de partida se puede utilizar en la práctica de la presente invención. En ciertas realizaciones, la polimerasa de ácido nucleico empleada en la reacción de amplificación es una polimerasa que tiene capacidad de corrección (p. ej., ADN polimerasa phi29, ADN polimerasa de *Thermococcus litoralis*, ADN polimerasa de *Pyrococcus furiosus*, etc.).

En ciertas realizaciones, la muestra de ácido nucleico que se está analizando deriva de una única fuente (p. ej., un solo organismo, virus, tejido, célula, sujeto, etc.), mientras que en otras realizaciones, la muestra de ácido nucleico es una reserva de ácidos nucleicos extraída de una pluralidad de fuentes (p. ej., una reserva de ácidos nucleicos de una pluralidad de organismos, tejidos, células, sujetos, etc.), donde por "pluralidad" se entiende dos o más. Así, en ciertas realizaciones, una muestra de ácido nucleico puede contener ácidos nucleicos de 2 o más fuentes, 3 o más fuentes, 5 o más fuentes, 10 o más fuentes, 50 o más fuentes, 100 o más fuentes, 500 o más fuentes, 1000 o más fuentes, 5000 o más fuentes, hasta e incluyendo aproximadamente 10.000 o más fuentes.

En ciertas realizaciones, los fragmentos de ácidos nucleicos que se van a reunir con fragmentos de ácidos nucleicos derivados de una pluralidad de fuentes (p. ej., una pluralidad de organismos, tejidos, células, sujetos, etc.), donde por "pluralidad" se entiende dos o más. En tales realizaciones, los ácidos nucleicos derivados de cada fuente incluyen un identificador múltiple (MID) de manera que la fuente de la que derivó cada fragmento de ácido nucleico etiquetado pueda ser determinada. En tales realizaciones, cada fuente de muestra de ácido nucleico se correlaciona con un MID único, donde por MID único se entiende que cada MID diferente empleado puede ser diferenciado de cualquier otro MID empleado en virtud de al menos una característica, p. ej., la secuencia de ácido nucleico del MID. Se puede utilizar cualquier tipo de MID, incluyendo, pero no limitado, a los descritos en el documento US 2007/0259357 titulado "Nucleic Acid Analysis Using Sequence Tokens", así como la Patente de los Estados Unidos 7.393.665, presentada el 1 de Julio de 2008, y titulada "Methods and Compositions for Tagging and Identifying Polynucleotides", ambos los cuales describen etiquetas de ácido nucleico y su uso en la identificación de polinucleótidos. En ciertas realizaciones, un conjunto de MID empleado para etiquetar una pluralidad de muestras no tienen por qué tener ninguna propiedad común concreta (p. ej.,  $T_m$ , longitud, composición de bases, etc.), ya que los métodos de etiquetado asimétricos (y muchos métodos de lectura de etiquetas, incluyendo pero no limitados a la secuenciación de la etiqueta o la medición de la longitud de la etiqueta) pueden acomodar una amplia variedad de conjuntos de MID únicos.

En ciertas realizaciones, cada polinucleótido individual (p. ej., de doble hebra o de hebra sencilla, según corresponda a los detalles metodológicos empleados) de una muestra a analizar está etiquetado con un MID único de manera que el destino de cada polinucleótido puede ser rastreado en posteriores procedimientos (donde, como se señaló anteriormente, se pretende que el MID único indique que cada MID diferente empleado puede ser diferenciado de cualquier otro MID empleado en virtud de al menos una característica, p. ej., la secuencia de ácido nucleico del MID). Por ejemplo (y como se describe a continuación), tener cada ácido nucleico etiquetado con un MID único permite el análisis de la secuencia de cada ácido nucleico individual utilizando los métodos de secuencias réflex descritos en la presente memoria. Esto permite la conexión de la información de la secuencia para grandes fragmentos de ácido nucleico que no pueden ser secuenciados en una única ronda de secuenciación.

Procedimiento de secuencia réflex

Como se ha resumido anteriormente, los aspectos de la presente invención incluyen métodos y composiciones para mover un dominio de un polinucleótido desde una primera ubicación a una segunda ubicación en el polinucleótido. Una realización ilustrativa se muestra en la Figura 1A.

La Figura 1A muestra un polinucleótido de hebra sencilla 100 que comprende, en una orientación 5' a 3', un primer dominio (102; el dominio que se va a mover); una secuencia reflex 104; una secuencia de ácido nucleico 106 que tiene un sitio distal con respecto al primer dominio (Sitio A), y un complemento de la secuencia reflex 108 (situado en el extremo 3' del polinucleótido). Las etapas del método reflex descrito a continuación moverán el primer dominio más cerca del Sitio A. Se observa aquí que la denominación cebador en la Figura 1A denota una secuencia complementaria de un dominio. Por ejemplo, Primer Dominio' es el complemento del Primer Dominio.

En la etapa 1, la secuencia reflex y su complemento en el polinucleótido se recuecen intramolecularmente para formar la estructura del polinucleótido 112, plegándose el polinucleótido sobre sí mismo e hibridando para formar una región de complementariedad (es decir, reflex de doble hebra/región reflex'). En esta configuración, el extremo 3' del complemento de la secuencia reflex puede servir como un sitio de cebado para la síntesis de ácido nucleico. La síntesis de ácido nucleico a partir de este sitio se realiza a continuación en la Etapa de extensión 2 que produce un complemento del primer dominio en el extremo 3' de la extensión del ácido nucleico (mostrado en el polinucleótido 114; la extensión se indica con la flecha discontinua marcada con "extender").

La desnaturalización del polinucleótido 114 (p. ej., por el calor) genera el polinucleótido de hebra sencilla lineal 116. Como se muestra en la Figura 1, el polinucleótido resultante 116 contiene un complemento del primer dominio en una posición próxima al Sitio A (es decir, separado sólo por el complemento de la secuencia reflex). Este polinucleótido resultante puede ser utilizado para cualquier etapa de análisis o procesamiento subsiguiente según desee el usuario (p. ej., la secuenciación, como un molde para la amplificación (lineal, PCR, etc.), la extracción específica de la secuencia, etc.).

En realizaciones alternativas, el primer dominio y la secuencia reflex se retiran del extremo 5' de la región de doble hebra del polinucleótido 114 (mostrado en el polinucleótido 118, la eliminación se muestra en la flecha discontinua marcada con "eliminar"). La eliminación de esta región se puede lograr por cualquier método conveniente, incluyendo, pero no limitado a, tratamiento (en condiciones de incubación apropiadas) de la estructura del polinucleótido 114 con exonucleasa de T7 o mediante tratamiento con exonucleasa Lambda; la exonucleasa Lambda se puede emplear siempre que el extremo 5' del polinucleótido esté fosforilado. Si la región se elimina enzimáticamente, el polinucleótido resultante 118 se utiliza en lugar del polinucleótido 116 en las etapas posteriores (p. ej., haciendo copias para invertir la polaridad).

En ciertas realizaciones, se utiliza el polinucleótido 116 o 118 como molde para producir un polinucleótido de doble hebra, por ejemplo mediante la realización de una reacción de síntesis de ácido nucleico con un cebador que ceba en el complemento del primer dominio. Esta etapa se refiere a veces como copiado para invertir la polaridad de un polinucleótido de hebra sencilla, y en algunos casos, el producto de doble hebra intermedio de este copiado no se muestra (véase, p. ej., la Figura 3). Por ejemplo, el copiado para invertir la polaridad del polinucleótido 116 da como resultado un polinucleótido de hebra sencilla 120 que tiene, en una orientación 5' a 3', el primer dominio (122); la secuencia reflex (124); el complemento del polinucleótido 106 (orientado con el complemento del Sitio A (Sitio A'; 126) próximo a la secuencia reflex); el complemento de la secuencia reflex (128); y el complemento del primer dominio (130).

En ciertas realizaciones, el primer dominio del polinucleótido comprende uno o más elementos que encuentran uso en una o más etapas de procesamiento o análisis posteriores. Tales secuencias incluyen, pero no se limitan a, sitios de enzimas de restricción, sitios de los cebadores de PCR, sitios de los cebadores de amplificación lineales, sitios de los cebadores de transcripción inversa, sitios del promotor de la ARN polimerasa (por ejemplo para ARN polimerasa T7, T3 o SP6), etiquetas MID, sitios de cebadores de secuenciación, etc. Cualquier elemento conveniente puede ser incluido en el primer dominio y, en ciertas realizaciones, está determinado por los deseos del usuario de los métodos descritos en la presente memoria.

Como una realización ilustrativa, supongamos que los autores de la presente invención quieren secuenciar una región de polinucleótido específica de múltiples genomas en una muestra reunida en donde la región de polinucleótido es demasiado larga para secuenciarla en una sola reacción. Por ejemplo, la secuenciación de una región de polinucleótidos que tiene 2 kilobases o más de longitud utilizando la tecnología Roche 454 (Branford, Conn.), donde la longitud de una sola ronda de secuenciación es de aproximadamente 400 bases. En este escenario, los autores de la presente invención pueden diseñar un conjunto de cebadores a mano izquierda ( $A_n$ ) y cebadores a mano derecha ( $B_n$ ) específicos para la región de polinucleótido que se colocan de tal manera que los autores de la presente invención pueden obtener secuencias directas de todas las partes del inserto, como se muestra en La Figura 1B. Obsérvese que el polinucleótido mostrado en la Figura 1B (140) tiene un dominio (142) que contiene un sitio cebador y un MID que indica a partir de cual muestra original deriva el polinucleótido. El sitio 142 por lo tanto representa un ejemplo de un sitio para el Primer Dominio identificado como 122 en la Figura 1A. El polinucleótido también incluye un sitio reflex (144), que puede ser parte de la propia región del polinucleótido (p. ej., una secuencia genómica), añadido en un dominio adaptador ligado junto con el sitio cebador y el MID (una secuencia artificial), o una combinación de ambos (una secuencia que abarca la unión adaptador/polinucleótido).

Se hace notar aquí que el polinucleótido 140 puede ser categorizado como un precursor del polinucleótido 100 en la Figura 1A, ya que no incluye una secuencia reflex 3' complementaria al sitio reflex (dominio 108 en la Figura 1A). Como se detalla a continuación, el polinucleótido 140 se puede convertir en un polinucleótido que tiene la

configuración estructural del polinucleótido 100, un polinucleótido adecuado como sustrato para el procedimiento réflex descrito en la presente memoria (p. ej., por la extensión del cebador utilizando un cebador  $B_n$  y la inversión de la polaridad).

5 En una realización ilustrativa, cada par de cebadores  $A_n$ - $B_n$  define una región de ácido nucleico que tiene aproximadamente 400 bases de longitud o menos. Este intervalo de tamaño está dentro de la longitud de la lectura de una única ronda de secuenciación de la plataforma de secuenciación Roche 454 actual; se puede utilizar un intervalo de tamaño diferente para la región de ácido nucleico definida para una plataforma de secuenciación diferente. De este modo, cada producto de cada procedimiento réflex puede ser secuenciado en una sola ronda. Se hace notar aquí que los pares de cebadores que se muestran en la Figura 1B se pueden utilizar para definir las regiones 1 a 5 mostradas en la Figura 3 (descrita con más detalle a continuación).

10 En ciertas realizaciones, para obtener la primera parte de la secuencia de la región de polinucleótido (es decir, en la estructura original, aquella parte del polinucleótido más cercana al primer dominio), los autores de la presente invención solamente necesitan un cebador a la derecha (p. ej.,  $B_0$ ) y no necesitan transferir el MID, ya que está al alcance de este cebador de secuenciación (es decir, el MID está a 400 bases del cebador de secuenciación  $B_0$ ). Todos los demás cebadores  $B_n$  tienen la secuencia réflex añadida a sus extremos 5' (el elemento de "R" aparece en los cebadores B) de manera que leen 5'-réflex- $B_n$ . Sin embargo, en ciertas realizaciones, el cebador  $B_0$  sí incluye la secuencia réflex y se utiliza en el procedimiento réflex (junto con un cebador  $A_0$  correspondiente) tal como se detalla a continuación.

20 Como se describió anteriormente, los autores de la presente invención obtienen un polinucleótido de hebra sencilla que tiene, en la orientación 5' a 3', un sitio de cebador (p. ej., para la secuenciación por Roche 454), un MID, una secuencia réflex y el polinucleótido que se va a secuenciar. Se han descrito numerosos métodos para obtener polinucleótidos de una sola hebra de interés y se conocen en la técnica, incluyendo en la Patente de los Estados Unidos 7.217.522, presentada el 15 de mayo de 2007; el documento US 2006/0211030; y el documento US 2009/0275087. Por ejemplo, se puede producir producto de hebra sencilla utilizando la amplificación lineal con un cebador específico para el sitio del cebador del molde. En ciertas realizaciones, el cebador incluye un radical de unión para facilitar el aislamiento del ácido nucleico de hebra sencilla de interés, p. ej., para inmovilizar la hebra superior sobre un compañero de unión del radical de unión inmovilizado sobre un soporte sólido. La eliminación de una hebra no biotinilada, hibridada por desnaturalización usando calor o pH elevado (o cualquier otro método conveniente) sirve para aislar la hebra biotinilada. Los radicales de unión y sus correspondientes compañeros de unión se denominan a veces en la presente memoria pares de compañeros de unión. Se puede utilizar cualquier par de compañeros de unión conveniente, incluyendo pero no limitados a pares de biotina/avidina (o estreptavidina), antígeno/anticuerpo, etc.

30 Se observa aquí que mientras que las Figuras y la descripción del procedimiento réflex proporcionadas en la presente memoria representan manipulaciones con respecto a un polinucleótido de hebra sencilla, no se requiere necesariamente que el polinucleótido de hebra sencilla descrito o representado en las figuras esté presente en la muestra en una forma aislada (es decir, aislado de su cadena complementaria). En otras palabras, los polinucleótidos de doble hebra se pueden usar cuando se describe/representa solamente una hebra, que generalmente será determinada por el usuario.

40 La implementación de una etapa de aislamiento de una sola hebra utilizando los métodos descritos anteriormente o variaciones de los mismos (o cualquier otra etapa de aislamiento de una sola hebra conveniente) generalmente se basará en los deseos del usuario. Un ejemplo de aislamiento de polinucleótidos de hebra sencilla se muestra en la Figura 2. En esta Figura, se desnaturaliza un molde de doble hebra de partida (con una orientación 5' a 3' mostrada como una flecha) y se ceba con un cebador de síntesis biotinilado específico para el sitio del cebador. Después de la extensión del cebador (es decir, la síntesis del ácido nucleico), la muestra se pone en contacto con un soporte sólido que tiene estreptavidina unida al mismo. El radical de biotina (es decir, el compañero de unión de la estreptavidina) en las hebras extendidas se unirá a la estreptavidina en fase sólida. A continuación se realizan la desnaturalización y el lavado para eliminar todas las hebras de polinucleótido no biotiniladas. Si se desea, el polinucleótido unido, que se puede utilizar en etapas del procedimiento réflex posteriores (p. ej., como un molde para las reacciones de extensión del cebador  $B_n$ ), se puede hacer eluir del soporte con estreptavidina. Alternativamente, el polinucleótido unido se puede emplear en etapas posteriores del procedimiento deseado mientras todavía está unido al soporte sólido (p. ej., en reacciones de extensión de fase sólida utilizando cebadores  $B_n$ ). Este procedimiento, con pequeñas variaciones dependiendo del molde que se esté utilizando y de la identidad del polinucleótido de hebra sencilla deseado, se puede emplear en cualquiera de una serie de etapas en las que se va a aislar un producto de hebra sencilla. Se hace notar que en ciertas realizaciones, el polinucleótido biotinilado unido al sustrato se puede utilizar para producir y aislar los productos de hebra sencilla no biotinilados aislados (es decir, haciendo eluir los productos no biotinilados mientras se dejan los moldes biotinilados unidos a la estreptavidina sobre el soporte sólido). Por lo tanto, los detalles de cómo se utilizan los compañeros unión para aislar polinucleótidos de hebra sencilla de interés variarán dependiendo de los parámetros del diseño experimental.

60 Los métodos adicionales de producción/aislamiento de una sola hebra incluyen la PCR aismétrica, la degradación enzimática específica de la hebra, y el uso de la transcripción in vitro, seguido de la transcriptasa inversa (IVT-RT), con la consiguiente destrucción de la hebra de ARN. Como se señaló anteriormente, se puede emplear cualquier

método de producción/aislamiento de una hebra sencilla conveniente.

Para el polinucleótido de hebra sencilla mostrado en la Figura 1B los autores de la presente invención recuecen uno de los cebadores  $B_n$  que tienen la secuencia réflex anexa, indicada con una "R" mayúscula (p. ej.,  $B_1$ ) y extienden el cebador en condiciones de síntesis de ácidos nucleicos para producir una copia del polinucleótido que tiene una secuencia réflex en su extremo 5'. A continuación se produce una copia de hebra sencilla de este polinucleótido para invertir la polaridad utilizando un cebador específico para el sitio del cebador en el primer dominio' (complemento del primer dominio 102). El ácido nucleico resultante tiene la estructura 100 que se muestra en la Figura 1A, donde el primer dominio 102 incluye el sitio del cebador y el MID. El sitio A (110) en la Figura 1 se determina por la especificidad del réflex 5'-cebador  $B_n$  utilizado.

El procedimiento réflex (p. ej., como se muestra en la Figura 1) se realiza a continuación para producir un producto en el que el sitio del cebador y el MID están ahora en estrecha proximidad al sitio deseado (o región de interés (ROI)) dentro del polinucleótido original (es decir, el sitio definido por el cebador utilizado, p. ej.,  $B_1$ ). El polinucleótido resultante se puede utilizar en los análisis posteriores según lo deseado por el usuario (p. ej., la tecnología de secuenciación Roche 454).

Se observa aquí que, aunque no se muestra en las Figuras 1A y 1B, se puede utilizar cualquier método conveniente para añadir adaptadores a un polinucleótido que se va a procesar como se describe en la presente memoria en la práctica del procedimiento réflex (adaptadores que contienen, p. ej., sitios de cebadores, sitios de polimerasa, MID, sitios de enzimas de restricción y secuencias réflex). Por ejemplo, se pueden añadir adaptadores en una posición concreta por medio de ligación. Para los polinucleótidos de doble hebra, se puede configurar un adaptador para que sea ligado a un sitio de corte de una enzima de restricción concreta. Cuando se emplea un polinucleótido de hebra sencilla, se puede utilizar un constructo adaptador de doble hebra que posee un saliente configurado para unirse al extremo del polinucleótido de hebra sencilla. Por ejemplo, en este último caso, el extremo de un polinucleótido de hebra sencilla puede ser modificado para incluir bases de nucleótidos específicas que son complementarias al saliente en el adaptador de doble hebra utilizando una transferasa terminal y nucleótidos específicos. En otras realizaciones, se emplean la PCR o métodos de amplificación lineales utilizando cebadores conjugados con adaptadores para añadir un adaptador a un sitio de interés. Una vez más, se puede emplear cualquier método conveniente para la producción de un polinucleótido de partida en la práctica de los métodos de la presente invención.

En ciertas realizaciones, el ácido nucleico se puede secuenciar directamente utilizando un cebador de secuenciación específico para el sitio del cebador. Esta reacción de secuenciación leerá a través del MID y el sitio deseado en el inserto.

En ciertas realizaciones, el polinucleótido se puede aislar (o fraccionar) utilizando un cebador  $A_n$  apropiado (p. ej., cuando se utiliza  $B_1$  como el primer cebador, se puede utilizar el cebador  $A_1$ ). En ciertas realizaciones, el polinucleótido cebado con  $A_n$  se somete a condiciones de síntesis de ácidos nucleicos para producir una copia del fragmento producido en el procedimiento réflex. En algunas de estas realizaciones, el cebador  $A_n$  tiene anexado en su extremo 5' un sitio de cebador que se puede utilizar en las etapas posteriores, incluyendo reacciones de secuenciación. Proporcionar un sitio del cebador en el cebador  $A_n$  permite la amplificación y/o secuenciación desde ambos extremos del fragmento resultante: desde el sitio del cebador en el primer dominio 102 y el sitio del cebador en el cebador  $A_n$  (no mostrado en la Figura 1B). Debido a la posición de los sitios de los cebadores y a su distancia de separación (es decir, menos de una ronda de secuenciación de diferencia), la secuenciación desde ambos extremos capturará por lo general la secuencia del sitio deseado (o ROI) y la secuencia del MID, que se puede utilizar para los análisis bioinformáticos posteriores, p. ej., para identificar positivamente la muestra de origen. Se observa aquí que, si bien la secuenciación en ambas direcciones es posible, no es necesaria, ya que la secuenciación desde cualquiera de los sitios del cebador solo capturará la secuencia del ROI, así como su secuencia de MID correspondiente.

Obsérvese que en ciertas realizaciones, el primer fragmento obtenido mediante amplificación/extensión a partir del cebador  $B_0$  directamente, la polaridad del ROI en el fragmento resultante está invertida en comparación con el ROI en los fragmentos obtenidos por los cebadores  $B_1$ - $B_n$ . Esto se debe a que el fragmento generado por  $B_0$ , a diferencia de los fragmentos generados por  $B_1$ - $B_n$ , no ha sido sometido a un procedimiento réflex que invierte la orientación de la secuencia ROI con respecto al primer dominio/secuencia réflex (como se ha descrito anteriormente). Por lo tanto, el cebador  $B_0$  puede tener anexado a él un sitio del cebador (p. ej., en su extremo 5') que se puede utilizar para posteriores reacciones de amplificación y/o secuenciación (p. ej., en el sistema de secuenciación Roche 454) en lugar de una secuencia réflex como con los cebadores  $B_1$ - $B_n$ . Sin embargo, en ciertas realizaciones, como se ha señalado anteriormente, el procedimiento réflex puede ser utilizado con un par de cebadores  $B_0$ - $A_0$  correspondiente como se ha descrito anteriormente, es decir, utilizando un cebador  $B_0$  que tiene una secuencia réflex 5' y un cebador  $A_0$  correspondiente con su correspondiente dominio adaptador 5' (p. ej., un sitio de cebador).

Se observa aquí que, debido a que las secciones concretas de secuencias que se van a analizar se definen por los pares de cebadores  $A_n$ - $B_n$  (como se muestra y se describe más arriba), se consigue una especificidad de secuencia mucho mayor en comparación con el uso de los métodos de extracción anteriores que emplean solamente un evento

de unión de oligo único (p. ej., utilizando sondas sobre una micromatriz).

La Figura 3 proporciona un diagrama de flujo detallado de una realización ilustrativa que emplea secuencias réflex para su uso en la secuenciación de múltiples regiones específicas en un polinucleótido (es decir, las regiones 1, 2, 3, 4 y 5 en una región de 11 kb del ADN de lambda).

5 Se genera un único fragmento de ADN parental 202 que incluye dominios de adaptadores (es decir, un sitio de cebador de secuenciación Roche 454, un único MID, y una secuencia réflex) y la secuencia de interés. En el ejemplo mostrado, la secuencia de interés es de ADN de lambda y la secuencia réflex está presente en la hebra superior (con su complemento mostrado en la hebra inferior). Se puede utilizar cualquier método conveniente para la producción de este fragmento de ADN parental, incluyendo la amplificación con un cebador que incluye los dominios de adaptadores (p. ej., utilizando la PCR), la clonación del fragmento en un vector que incluye los dominios de adaptadores (p. ej., un vector con los dominios de adaptadores adyacentes a un sitio de clonación), o anclando los adaptadores a los fragmentos de polinucleótidos (p. ej., fragmentos elaborados por fragmentación aleatoria, por digestión con enzimas de restricción específica de la secuencia, o combinaciones de los mismos). Mientras que sólo se muestra un único fragmento con un solo MID, las etapas de la Figura 3 son aplicables a muestras que tienen múltiples fragmentos diferentes cada uno con un MID diferente, p. ej., una muestra que tiene una población de fragmentos homólogos de cualquiera de numerosas fuentes diferentes (p. ej., individuos diferentes). La Figura 3 describe las etapas enzimáticas posteriores implicadas en la creación de los cinco fragmentos hijos en los que las regiones 1, 2, 3, 4 y 5 (mostradas en el polinucleótido 204) se reordenan para ser colocadas dentro de una distancia funcional de los dominios de adaptadores (es decir, suficientemente cerca de los dominios de adaptadores que se van a secuenciar en una única reacción de secuenciación Roche 454). Obsérvese que ciertas etapas se muestran para la región 4 solamente (206).

En la etapa 1, las cinco regiones de interés se definen dentro del fragmento parental (marcadas de 1 a 5 en el polinucleótido 204) y se diseñan los correspondientes pares de cebadores para cada una. La distancia de cada región de interés de la secuencia réflex se muestra debajo del polinucleótido 204. Los pares de cebadores se diseñan como se ha descrito y mostrado en la Figura 1B (es decir, el par de cebadores  $A_n$ - $B_n$ ). Para mayor claridad, solamente se muestran los sitios de cebadores para la región 4 en la Figura 3 ("sitios de cebadores" que rodean la región 4). En la etapa 2, se realizan las extensiones de cebadores específicos de secuencia (solamente se muestra la región 4) con los correspondientes cebadores  $B_n$  para producir polinucleótidos de hebra sencilla que tienen la estructura 208 (es decir, que tienen la secuencia réflex en el extremo 5'). Como se muestra, el cebador  $B_n$  para la región 4 incluirá un sitio de cebador específico de la secuencia que ceba en el sitio del cebador más 3' observado para la región 4 (donde "más 3'" se refiere a la hebra molde, que en la Figura 3 es la hebra superior). Este polinucleótido se vuelve a copiar de nuevo para producir el polinucleótido 210 que tiene una polaridad invertida (p. ej., copiado utilizando un cebador que hibrida con el dominio 454A'). El polinucleótido 210 tiene una estructura similar a la del polinucleótido 100 mostrado en la parte superior de la Figura 1. La etapa 4 representa el resultado del cebado intramolecular entre la secuencia réflex y su complemento seguido por la extensión para producir las estructuras del MID' y 454 A' en el extremo 3' (polinucleótido 212). En las realizaciones mostradas en la Figura 3, el polinucleótido 212 se trata con exonucleasa T7 para eliminar el ADN de doble hebra del extremo 5' (como se indicó anteriormente, esta etapa es opcional). El polinucleótido formado para la región 4 se muestra como 216 mostrando también los polinucleótidos para las otras regiones (214).

40 Se observa aquí que la formación de cada uno de los polinucleótidos 214 puede llevarse a cabo ya sea en reacciones separadas (es decir, la estructura con la región 1 en la proximidad de los dominios de adaptadores está en una primera muestra, la estructura con la región 2 en la proximidad de la región de adaptador está en una segunda muestra, etc.) o en una o más muestras combinadas.

45 En la etapa 6 se copian los polinucleótidos 214 para invertir la polaridad para formar polinucleótidos 218. En la etapa 7, cada uno de estos productos se ceba a continuación con el segundo cebador del par de cebadores específico (véanse los cebadores  $A_n$  como se muestra en la Figura 1B) teniendo cada uno un segundo sitio del cebador Roche 454 (454 B) anclado en el extremo 5', y extendido para formar los productos 220. Las etapas 6 y 7 se pueden combinar (p. ej. en una única PCR u otras reacciones de amplificación).

50 En resumen, la Figura 3 muestra cómo se puede emplear el procedimiento réflex para producir cinco fragmentos hijos 220 de longitud similar (p. ej., ~ 500 pb) cada uno de los cuales contiene secuencias de ADN que difieren en su distancia de la secuencia réflex en la estructura de partida 202 mientras se mantiene el MID original.

La Figura 4 muestra otro uso ilustrativo del procedimiento réflex descrito en la presente memoria. En la realización mostrada en la Figura 4, una secuencia diana (es decir, que contiene la región de interés "E") se enriquece a partir de una reserva de fragmentos anclados a adaptadores. En ciertas realizaciones, los fragmentos cortan al azar, se seleccionan para un cierto intervalo de tamaño (p. ej., ADN que tiene una longitud de 100 a 5000 pares de bases), y se etiquetan con adaptadores (p. ej., adaptadores asimétricos, p. ej., como se describe en el documento 2009/0275087). El adaptador asimétrico empleado en la Figura 4 contiene un sitio de cebador de secuenciación (454A, como se utiliza en la plataforma de secuenciación Roche 454), un MID, una secuencia de X, y una región tallo interna (ISR), que indica la región de complementariedad para el adaptador asimétrico que es adyacente al sitio de anclaje del adaptador (véase, p. ej., la descripción en el documento US 2009/0275087). La

secuencia X puede ser cualquier secuencia que pueda servir como un sitio de unión para un polinucleótido que contenga el complemento de la secuencia X (similar a un sitio del cebador). Como se describe a continuación, la secuencia X permite el emparejamiento de un oligonucleótido que tiene un saliente 5' que puede servir como un molde para la extensión del extremo 3' del oligonucleótido adaptador. La dirección de secuenciación del sitio de cebador de secuenciación (sitio del cebador 454A en la estructura 401 de la Figura 4) está orientada de tal manera que la amplificación del fragmento ligado al adaptador utilizando el sitio del cebador de secuenciación continúa lejos del inserto genómico ligado. Esto tiene el efecto de hacer que la biblioteca ligada al adaptador asimétrico inicial sea "inerte" a la amplificación utilizando este cebador, p. ej., en una reacción de PCR.

Para extraer una región de interés (la región "E"), la biblioteca se mezcla con un oligonucleótido (403) que contiene una secuencia x' 3' y una secuencia de cebado específica de la diana (la secuencia 1') en condiciones de hibridación/emparejamiento. La secuencia específica de la diana 1' está diseñada para flanquear un lado de la región de interés (secuencia 1' adyacente a E en el inserto genómico; obsérvese que sólo se muestra en la Figura 4 el fragmento de polinucleótido que contiene E), al igual que un cebador de PCR. Después de emparejar el cebador 403, el complejo hibridado se extiende, por medio de lo cual todos los fragmentos adaptadores etiquetados obtendrán el complemento de la secuencia específica de la diana (es decir, la secuencia 1) en el extremo 3' (véase la estructura 405; las flechas indican la dirección de extensión).

Los productos extendidos 405 se desnaturalizan a continuación y se deja que las regiones 1/1' hibriden intramolecularmente en un evento de cebado del procedimiento réflex, después de lo cual se realiza la extensión de ácido nucleico para formar la estructura 407 (la extensión es desde el sitio de cebado 1; mostrado con una flecha). Esta reacción réflex crea un producto (407) que, a diferencia de su estructura parental (405), tiene un sitio de cebador de secuenciación (454 A) que está orientado de tal manera que la extensión utilizando esta secuencia de cebador prosigue hacia la región de interés. Por lo tanto, en ausencia de reacción réflex de cebado y extensión, la extensión con un cebador de secuenciación no generará un producto que contenga la región de interés (la región E). En otras palabras, sólo los polinucleótidos diana que contienen la región E tendrán una secuencia 454 A que pueda amplificar el material genómico (estructura 407).

Después de completar el procedimiento réflex (utilizando 1/1' como secuencias réflex), se realiza una reacción de amplificación por PCR para amplificar la región de interés (con dominios de adaptadores asociados). Sin embargo, antes de realizar la reacción de PCR, se "inactiva" la extensión adicional de la muestra de fragmento utilizando transferasa terminal y ddNTP. Esta inactivación impide que las moléculas etiquetadas no diana realicen la extensión del cebador desde el sitio 1 del cebador 3'. Una vez inactivado, se realiza una reacción de PCR utilizando un cebador de secuenciación (es decir, cebador 454 409) y un segundo cebador que ceba y se extiende desde el lado opuesto de la región de interés (es decir, el cebador 411, que incluye un sitio del cebador de secuenciación 454 B 5' y una región "2" 3' que ceba el lado opuesto de E desde la región 1). Solamente los fragmentos que han sido sometidos al procedimiento réflex y contienen la región E serán moldes adecuados para la reacción de PCR y producirán el producto deseado (413).

Así, el procedimiento ilustrado en la Figura 4 permite el movimiento de un dominio de adaptador (p. ej., que contiene elementos funcionales y/o MID) en la proximidad a una región de interés deseada.

El procedimiento réflex descrito en la presente memoria puede ser utilizado para realizar un potente análisis de ligamiento combinándolo con métodos de recuento de ácidos nucleicos. Se puede emplear cualquier método conveniente para el etiquetado y/o el recuento de moléculas de ácido nucleico individuales con etiquetas únicas (véase, p. ej., la Patente de los Estados Unidos 7.537.897, presentada el 26 de mayo de 2009; la Patente de los Estados Unidos 7.217.522, presentada el 15 de mayo de 2007; el documento US 2006/0211030; y el documento US 2009/0275087). Todo esto se puede llevar a cabo en paralelo con el consiguiente ahorro en el coste de mano de obra, tiempo y materiales.

En una realización ilustrativa, se etiqueta una gran colección de secuencias con MID de tal manera que cada molécula de polinucleótido de la muestra tenga un MID único. En otras palabras, cada polinucleótido de la muestra (p. ej., cada polinucleótido de doble hebra o de hebra sencilla individual) se etiqueta con un MID que es diferente de todos los demás MID de cualquier otro polinucleótido de la muestra. En general, para llevar a cabo semejante etiquetado molecular el número de etiquetas MID distintas que se van a utilizar debe ser muchas veces mayor que el número real de moléculas a analizar. Esto dará como resultado que la mayoría de las moléculas de ácido nucleico individuales estarán marcadas con una etiqueta ID única (véase, p. ej., Brenner et al, Proc. Natl. Acad. Sci. USA. 2000 97(4): 1665-70). A continuación, cualquier secuencia que resulte del procedimiento réflex en esa molécula concreta (p. ej., como se ha descrito anteriormente) estará marcada con la misma etiqueta MID única y por lo tanto intrínsecamente conectada. Obsérvese que una vez que todas las moléculas en una muestra están etiquetadas individualmente, pueden ser manipuladas y amplificadas tanto como sea necesario para el procesamiento, siempre que la etiqueta MID se mantenga en los productos generados.

Por ejemplo, los autores de la presente invención podrían querer secuenciar mil genomas virales (o una región genómica específica) o mil copias de un gen presente en las células somáticas. Después de etiquetar cada polinucleótido de la muestra con un sitio de cebador de secuenciación, MID y secuencia réflex (como se muestra en las figuras y se ha descrito anteriormente), los autores de la presente invención utilizan el procedimiento réflex para

romper cada polinucleótido en longitudes apropiadas para el procedimiento de secuenciación que se está utilizando, transfiriendo el sitio del cebador de secuenciación y el MID a cada fragmento (como se ha descrito anteriormente). Se puede utilizar la obtención de información de la secuencia de todas las muestras sometidas al procedimiento réflex para determinar la secuencia de cada polinucleótido individual de la muestra de partida, usando la secuencia de MID para definir las relaciones de ligamiento entre las secuencias de diferentes regiones en el polinucleótido que se está secuenciando. Utilizando una plataforma de secuenciación con longitudes de lectura más largas se puede minimizar el número de cebadores a utilizar (y fragmentos réflex generados).

Las ventajas mencionadas anteriormente se muestran en la Figura 5. Esta Figura muestra una comparación de métodos para la identificación de polimorfismos de ácidos nucleicos en los ácidos nucleicos homólogos en una muestra (p. ej., la misma región derivada de un par de cromosomas de una célula diploide o genomas/transcritos virales). El esquema superior muestra dos moléculas de ácido nucleico de una muestra (1 y 2) que tienen una variedad diferente de polimorfismos en sitios polimórficos A, B y C (A1, B1, C1 y C2). Los métodos de secuenciación convencionales utilizando la fragmentación (lado izquierdo) pueden identificar los polimorfismos en estos ácidos nucleicos, pero no retener la información de ligamiento. Empleando el procedimiento réflex descrito en la presente memoria para identificar polimorfismos (lado derecho) se mantiene la información de ligamiento. Cabe señalar que, por simplificar, no se muestran en el procedimiento réflex todas las estructuras de dominio y las etapas.

#### Kits y sistemas

También se describen kits y sistemas para la práctica de los procedimientos objeto, como se ha descrito anteriormente, tales como vectores configurados para añadir secuencias réflex a insertos de ácidos nucleicos de interés y reactivos para llevar a cabo todas las fases de la clonación o el procedimiento réflex descritos en la presente memoria (p. ej., enzimas de restricción, nucleótidos, polimerasas, cebadores, exonucleasas, etc.). Los diversos componentes de los kits pueden estar presentes en recipientes separados o ciertos componentes compatibles pueden ser combinados previamente en un solo recipiente, según se desee.

Los sistemas y kits también pueden incluir uno o más de otros reactivos para la preparación o el procesamiento de una muestra de ácido nucleico de acuerdo con los presentes métodos. Los reactivos pueden incluir una o más matrices, disolventes, reactivos de preparación de muestras, tampones, reactivos de desalación, reactivos enzimáticos, reactivos desnaturalizantes, donde también se pueden proporcionar los patrones de calibración, tales como los controles positivos y negativos. Como tales, los kits pueden incluir uno o más recipientes tales como viales o botellas, conteniendo cada envase un componente separado para llevar a cabo una etapa de procesamiento o preparación de la muestra y/o para llevar a cabo una o más etapas de un análisis de aislamiento de variantes de ácidos nucleicos de acuerdo con la presente invención.

Además de los componentes mencionados anteriormente, los kits típicamente incluyen adicionalmente instrucciones para utilizar los componentes del kit para poner en practicar los presentes métodos, p. ej., para preparar muestras de ácido nucleico para realizar el procedimiento réflex de acuerdo con los aspectos de los presentes métodos. Las instrucciones para la práctica de los presentes métodos se registran generalmente en un medio de registro adecuado. Por ejemplo, las instrucciones se pueden imprimir sobre un sustrato, tal como papel o plástico, etc. Como tales, las instrucciones pueden estar presentes en los kits como un inserto en el paquete, en el etiquetado del recipiente del kit o sus componentes (es decir, asociado con el envase o sub-embalaje) etc. En otros aspectos, las instrucciones están presentes como un archivo electrónico de almacenamiento de datos presente en un medio de almacenamiento legible por ordenador adecuado, por ejemplo, CD-ROM, disquete, etc. En otros aspectos más, las instrucciones reales no están presentes en el kit, pero se proporcionan los medios para obtener las instrucciones a partir de una fuente remota, por ejemplo a través de Internet. Un ejemplo es un kit que incluye una dirección web en la que se puede ver las instrucciones y/o desde la que se pueden descargar las instrucciones. Al igual que con las instrucciones, este medio para la obtención de las instrucciones se registra en un sustrato adecuado.

Además de la base de datos, los programas y las instrucciones, los kits también pueden incluir una o más muestras de control y reactivos, p. ej., dos o más muestras de control para su uso en las pruebas del kit.

#### Utilidad

El procedimiento reflex descrito en la presente memoria proporciona ventajas significativas en numerosas aplicaciones, algunas de las cuales se indican a continuación (como también se ha descrito anteriormente).

Por ejemplo, como se ha descrito anteriormente, ciertos aspectos del procedimiento réflex definen secciones concretas de la secuencia que va a ser analizada por un par de cebadores, como en la PCR (p. ej., los dos oligos mostrados como A<sub>n</sub>-B<sub>n</sub> en la Figura 1B). Esto da como resultado una mayor especificidad de secuencia en comparación con otros métodos de extracción (p. ej., utilizando sondas sobre una micromatriz) que solamente utilizan una única secuencia de oligo. La separación de las sondas define una longitud que puede ser relativamente uniforme (haciendo de ese modo la posterior manipulación, incluyendo la amplificación, más uniforme) y también puede ser adaptada a la plataforma de secuenciación concreta que se esté empleando.

Además, como se ha descrito anteriormente, los aspectos de la presente invención se pueden utilizar para analizar ubicaciones genómicas homólogas en una muestra multiplexada (es decir, una muestra que tiene polinucleótidos de

diferentes muestras genómicas) en la que los polinucleótidos se etiquetan con el MID. Esto es posible porque el procedimiento *réflex*, que opera intramolecularmente, mantiene el MID conectando de ese modo cualquier fragmento concreto, a la muestra de la que se origina.

5 Finalmente, como los procedimientos *réflex* descritos en la presente memoria funcionan intramolecularmente, se puede determinar el ligamiento genético entre las diferentes regiones en el mismo fragmento grande que están demasiado separadas para ser secuenciadas en una lectura de secuencia. Semejante determinación del ligamiento puede ser de gran valor en la genética de plantas o animales (p. ej., para decidir si un determinado conjunto de variaciones están ligadas entre sí en el mismo tramo del cromosoma) o en estudios virales (p. ej., para determinar si las variaciones concretas están ligadas juntas en el mismo tramo de un genoma/transcrito viral, p. ej., VIH, virus de la hepatitis, etc.).

### Ejemplos

#### Ejemplo 1

Las Figuras 6 y 7 proporcionan datos experimentales y la validación del procedimiento *réflex* descrito en la presente memoria utilizando sustratos de polinucleótidos sintéticos.

#### 15 Métodos

##### Sustrato:

El sustrato oligonucleotídico de 100 bases (mostrado esquemáticamente en la Figura 6A) se sintetizó con fluoresceína-dT interna situada entre las secuencias *RÉFLEX* y *RÉFLEX'*. Esta marca proporciona un método de detección conveniente y sensible de especies de oligonucleótidos utilizando la electroforesis en gel de poli(acrilamida).

#### 20 Reacciones de extensión:

Se prepararon reacciones que contenían 1  $\mu$ M del sustrato oligonucleotídico de 100 bases, dNTP 200  $\mu$ M, presencia o ausencia de oligonucleótido competidor 1  $\mu$ M, 0,5  $\mu$ l de cada ADN polimerasa ("DNAP"): Vent (NEB, 2 unidades/ $\mu$ l), Taq (Qiagen HotStarTaq 5 unidades/ $\mu$ l) y Herculase (Stratagene), y se completaron hasta 50  $\mu$ l con los tampones comerciales adecuados para cada polimerasa y dH<sub>2</sub>O. Para las titulaciones de Taq se utilizó enzima 0,5  $\mu$ l, 1  $\mu$ l, 2  $\mu$ l, y 3  $\mu$ l en el mismo volumen de 50  $\mu$ l. Las reacciones se calentaron en un termociclador Biometra a 95°C durante 15 minutos (Taq) o 5 minutos (Herculase, Vent), seguido de 55°C o 50°C durante 30 segundos, y una incubación final a 72°C durante 10 minutos.

##### Digestiones con exonucleasa T7:

30 Se prepararon reacciones con las reacciones de extensión de 10  $\mu$ l anteriores, 0,5  $\mu$ l de exonucleasa T7 (NEB, 10 unidades/ $\mu$ l), y se completaron hasta 50  $\mu$ l utilizando Tampón 4 NEB y dH<sub>2</sub>O. Las reacciones se incubaron a 25°C durante 30 minutos.

##### Análisis de electroforesis en gel:

35 Se utilizó un gel de poli(acrilamida) desnaturalizante al 8% para analizar las especies de reacción. Se cargaron 0,4  $\mu$ l de reacciones de extensión, y 2  $\mu$ l de reacciones de digestión y se ejecutaron a 800V durante ~ 1,5 horas. Se analizaron los geles para determinar la fluoresceína utilizando un generador de imágenes Amersham/General Electric Typhoon.

#### Resultados

40 La Figura 6A muestra la estructura de cada fase de procesamiento de la secuencia *reflex* con el tamaño esperado de ácido nucleico mostrada a la izquierda. El ácido nucleico de hebra sencilla inicial que tiene un sitio de cebador de secuenciación (el sitio del cebador de secuenciación Roche 454; enumerado como 454A); un MID; una secuencia *reflex*; el inserto; y un complemento de la secuencia *reflex* tiene 100 nucleótidos de longitud. Después de el auto-emparejamiento y la extensión, se espera que el producto tenga 130 nucleótidos de longitud. Después de la eliminación de la región de doble hebra a partir del extremo 5', se espera que el ácido nucleico tenga 82 bases de longitud.

45 La Figura 6B muestra los resultados de tres experimentos utilizando tres diferentes polimerasas de ácido nucleico (Vent, Herculase y Taq, indicadas en la parte superior de las calles). La temperatura a la que se llevó a cabo el emparejamiento se muestra en la parte superior de cada calle (a 50°C o 55°C). Los tamaños de los tres ácidos nucleicos como se señaló anteriormente se indican en la parte izquierda y derecha del gel.

50 Como se muestra en la Figura 6B, la extensión parece ser muy eficiente en las condiciones utilizadas con Herculase (Herculase es una mezcla de dos enzimas: DNAP Pfu modificada y Archaemax (dUTPasa)). La mayor parte (o la totalidad) del ácido nucleico de 100 pares de bases inicial se convierte en el producto de 130 pares de bases (véanse las calles 6 y 7). Sin embargo, después de la digestión con la exonucleasa T7, la actividad exonucleasa 3'-5'

de la Herculase da como resultado una digestión parcial del producto de 82 bases deseado (obsérvense las bandas en y por debajo de los 82 pares de bases en las calles 8 y 9).

Taq, que carece de actividad exonucleasa 3'-5', muestra una banda más fuerte en el tamaño esperado del producto final después de la digestión con exonucleasa T7 (véase la calle 13).

- 5 La Figura 7 muestra el efecto sobre el procedimiento réflex de las cantidades crecientes de polimerasa Taq, así como el uso de un competidor de la secuencia réflex (mostrado esquemáticamente en la Figura 7A).

10 Como se muestra en las calles 2 a 5, el aumento de la concentración de Taq mejora la extensión a una conversión de ~ 90% del ácido nucleico de partida (véase la calle 5). Las calles 7 a 8 muestran que la digestión con la exonucleasa T7 no deja un producto de 82 bases perfecto. Esto puede ser debido al colapso del ADNdh cuando la exonucleasa T7 casi ha completado su digestión a partir del extremo 5' en la región de doble hebra de la estructura plegada. Se observa que en muchas realizaciones, la eliminación de unas pocas bases adicionales desde el extremo 5' del polinucleótido no interferirá en los análisis posteriores, ya que las bases nucleótídicas en el extremo 5' se eliminan a menudo durante las etapas subsiguientes.

15 Como se muestra en las calles 11-14, la adición de un competidor (que puede interferir en el emparejamiento de las secuencias réflex para formar la estructura plegada) da como resultado sólo una pequeña disminución (~5-10%) de producto completamente extendido. Por lo tanto, como se esperaba, la reacción intramolecular está fuertemente favorecida. Aunque no se muestra, los autores de la presente invención han observado que el oligonucleótido competidor también se extendió en la misma cantidad (~5-10%).

20 La concentración de competidor, la concentración del sustrato réflex, y la complejidad genética en general, afectarán todas probablemente a los resultados específicos. Los experimentos mostrados en las Figuras 6 y 7 demuestran que las partes centrales de los procedimientos réflex descritos en la presente memoria son funcionales y se pueden implementar.

#### Ejemplo II

25 La Figura 8 muestra el flujo de trabajo réflex (diagrama a la izquierda) y los resultados ilustrativos del flujo de trabajo (gel de la derecha) para una región específica de interés (ROI). El material de partida es una molécula de ácido nucleico de doble hebra (700) que contiene un sitio de cebador 454A, un MID, un sitio réflex, y un polinucleótido de interés que tiene tres ROI (2, 3 y 4) en diferentes localizaciones en la misma. Este material de partida se somete a procedimientos réflex (como se describe más arriba) específicos para la ROI 2 como se muestra en el diagrama de la izquierda de la figura, con y sin el uso de una etapa de exonucleasa T7 (la etapa de exonucleasa T7 mostrada en la diagrama se indica como "Opcional").

30 La finalización de todas las etapas que se muestran en el procedimiento réflex debería dar como resultado un polinucleótido de doble hebra de 488 pares de bases (702) con o sin la etapa de la exonucleasa T7.

Como se muestra en el gel a la derecha de la Figura 8, el producto de 488 pares de bases fue producido en los procedimientos réflex con y sin la etapa de exonucleasa T7.

35 La Figura 9 muestra un protocolo ilustrativo para un procedimiento réflex basado en los resultados discutidos anteriormente. El diagrama muestra las etapas del procedimiento réflex específicas con indicaciones a la derecha en cuanto a dónde se emplea la purificación de los productos de reacción (p. ej., usando cuentas Agencourt SPRI para eliminar oligonucleótidos cebadores). Una de las razones para llevar a cabo tales etapas de purificación es reducir el potencial para la generación de productos secundarios en una reacción (p. ej., amplicones indeseables). Mientras la Figura 9 indica tres etapas de purificación, se pueden emplear menos o más etapas de purificación dependiendo de los deseos del usuario. Cabe señalar que las etapas de inversión de polaridad, cebado réflex y extensión, y "estiramiento" (o desnaturalización)/segunda etapa de reversión de la polaridad se pueden realizar sin etapas de purificación intermedias.

El protocolo mostrado en la Figura 9 incluye las siguientes etapas:

45 emparejamiento de un primer cebador que contiene una secuencia réflex 5' (o cola réflex, como se señala en la figura) específica para el sitio de cebador 3' para la región R' con el polinucleótido de partida y que se extiende (el cebador se recucece con la hebra superior en el sitio de cebador a la derecha de R en el polinucleótido 902, indicado con un \*; esta etapa representa el primer procedimiento de desnaturalización, emparejamiento y extensión indicados a la derecha);

50 después de la purificación, adición de un cebador 454A y realización de tres ciclos de desnaturalización, emparejamiento y extensión: el primer ciclo da como resultados la copia del cebador 454 para invertir la polaridad de la hebra recién sintetizada; el segundo ciclo rompe la estructura de doble hebra producida, permite que se forme la estructura réflex y se extienda; el tercer ciclo da como resultado otra copia utilizando el mismo cebador 454 originalmente añadido;

después de la purificación, adición de un segundo cebador específico para el segundo sitio de cebador para la región R' que tiene una cola 454B 5' (este cebador se recuece con el sitio 3' del cebador de la región R' en el polinucleótido 904, indicado con un \*) y desnaturalización, emparejamiento y extensión dando como resultado un producto polinucleotídico que tiene sitios 454A y 454B que rodean el MID, la secuencia réflex, y R'. Obsérvese que el primer cebador específico para la región R' y el segundo cebador específico para la región R' definen sus límites, como se ha descrito y representado anteriormente en la Figura 1B);

después de otra purificación, adición de cebadores 454A y 454B y realización de una reacción de amplificación por PCR.

#### Ejemplo III

Como se describió anteriormente, una secuencia réflex puede ser una secuencia "artificial" añadida a un polinucleótido como parte de un adaptador o puede estar basada en una secuencia presente en el polinucleótido de interés que esté siendo analizado, p. ej., una secuencia genómica (o "no artificial").

Los datos mostrados en los Ejemplos anteriores utilizaban sitios réflex "artificiales". En este Ejemplo, el sitio réflex es una secuencia genómica presente en el polinucleótido que está siendo analizado.

El material de partida es un ADN de doble hebra que contiene un sitio 454A, un MID y un polinucleótido que se va a analizar. El 454A y el MID fueron añadidos por medio de la ligación de un adaptador a fragmentos del polinucleótido parental, seguido por el enriquecimiento del polinucleótido que se va a analizar mediante una reacción de extracción basada en la hibridación y la posterior amplificación por PCR secundaria (véase la Ruta 1 en la Figura 13). Por lo tanto, el sitio réflex empleado en este ejemplo es una secuencia normalmente presente en el extremo 5' del polinucleótido objeto (una secuencia genómica). El polinucleótido que se va a analizar incluye una región de interés distal con respecto a las secuencias 454A y MID que tiene 354 pares de bases de longitud.

Este ácido nucleico de doble hebra de partida tiene 755 pares de bases de longitud. Basándose en la longitud de cada uno de los dominios relevantes en este ácido nucleico de partida, el procedimiento réflex debe dar como resultado un producto de 461 pares de bases.

La Figura 10 muestra el material de partida para el procedimiento réflex (panel izquierdo) y el producto resultante generado utilizando el procedimiento réflex (panel derecho; el procedimiento réflex se realizó como se describe en el Ejemplo II, sin necesidad de utilizar una etapa de exonucleasa T7). Se incluye una escala de tamaños en la calle de la izquierda de cada gel para permitir la estimación del tamaño del material de ensayo. Esta figura muestra que el ácido nucleico de 755 pares de bases de partida era procesado al producto esperado de 461 pares de bases, lo que confirma que un sitio reflex "no-artificial" es eficaz para mover un dominio de adaptador de un lugar a otro en un polinucleótido de interés en una manera específica de la secuencia.

#### Ejemplo IV

La Figura 11 muestra un esquema de un experimento en el que se realiza el procedimiento réflex sobre un único molde inicial grande (un fragmento "parental") para generar 5 productos diferentes (productos "hijo") teniendo cada uno una región diferente de interés (es decir, se producen productos hijo que tienen cualquier región 1, 2, 3, 4 o 5). El esquema de la Figura 11 muestra el fragmento de partida (11.060 pares de bases) y los productos resultantes (488 pares de bases cada uno) generados a partir de cada una de las diferentes regiones de las reacciones réflex específicas de interés (las reacciones réflex se realizan como se describe más arriba). El panel (gel) en la base de la Figura 11 muestra el fragmento de partida más grande (Calle 1) y los productos hijos resultantes para cada reacción réflex específica de la región (calles 2 a 6, con la región de interés señalada en cada una en el recuadro), donde los fragmentos de partida e hijos tienen las longitudes esperadas. La secuenciación de los productos confirmó la identidad de la región de interés en cada uno de los productos réflex mostrados en el gel. Estos resultados demuestran que se pueden generar múltiples productos réflex diferentes a partir de un único fragmento parental etiquetado asimétricamente a la vez que se mantienen los dominios del adaptador (p. ej., los sitios de los cebadores y MID).

#### Ejemplo V

La Figura 12 detalla los experimentos realizados para determinar la prevalencia del reordenamiento intramolecular (según se desee en el procedimiento réflex) vs. el reordenamiento intermolecular. El reordenamiento intermolecular no es deseable, ya que puede conducir a la transferencia de un MID de un fragmento a otro (también denominado conmutación de MID). La conmutación de MID se puede producir si una secuencia réflex en un primer fragmento hibrida con su complemento en un segundo fragmento durante el procedimiento réflex, lo que conduce a la anexión del MID del segundo fragmento al primer fragmento. Por lo tanto, el reordenamiento intermolecular o conmutación de MID, debe reducirse al mínimo para evitar la transferencia de un MID de un fragmento de la muestra a otro, lo que podría conducir a una representación errónea de la fuente de un fragmento.

Para medir la prevalencia de la conmutación de MID bajo diferentes condiciones réflex, se generaron fragmentos que tenían diferentes tamaños que incluían dos MID diferentes, como se muestra en el panel superior de la Figura

12. La secuencia común en estos fragmentos sirve como el sitio de cebado para la primera reacción de extensión para agregar la segunda secuencia réflex (véase, p. ej., la etapa 2 de la Figura 3). En la Figura 12 se muestran tres fragmentos ilustrativos para cada tamaño de fragmento diferente (es decir, 800 pares de bases con una combinación MIDB y MIDA, 1900 pares de bases con una combinación MIDC y MIDA, y 3000 pares de bases con una combinación MIDD y MIDA). Para cada familia de MID (A, B, C y D), existen 10 miembros diferentes (es decir, MIDA tiene 10 miembros diferentes, MIDB tiene 10 miembros diferentes, etc.). Se generó un conjunto de 10 fragmentos de MID duales para cada fragmento de diferente tamaño (es decir, 800, 1900 y 3000 pares de bases), en donde los pares de MID (es decir, MIDA/MIDB, MIDA/MIDC y MIDA/MIDD) fueron designados como 1/1, 2/2, 3/3, 4/4, 5/5, 6/6, 7/7, 8/8, 9/9 y 10/10. Los 10 fragmentos del mismo tamaño se mezclaron después y se realizó un protocolo réflex.

Debido a la estructura de los dominios de los fragmentos, un procedimiento réflex satisfactorio da como resultado que los dos MID para cada fragmento se muevan a una proximidad suficientemente cercana como para ser secuenciados en una sola lectura utilizando la plataforma de secuenciación Roche 454 (véanse los productos réflex mostrados en el esquema de la Figura 12). Las reacciones réflex para cada tamaño de fragmento se realizaron a cuatro concentraciones diferentes de fragmentos para determinar el efecto de este parámetro, así como la longitud del fragmento, en la prevalencia de la conmutación de MID. Los productos réflex de cada reacción realizada se sometieron a secuenciación 454 para determinar la identidad de ambos MID en cada fragmento, y por lo tanto la proporción de conmutación de MID que se había producido.

El panel en la parte inferior izquierda de la Figura 12 muestra la tasa de conmutación de MID (eje Y, que se muestra en % de pares de MID incorrecto (o conmutado)) para cada fragmento de longitud diferente en cada concentración diferente (eje X; 300, 30, 3 y 0,3 nM). Como se muestra en este panel, la tasa de conmutación de MID disminuye con concentraciones más bajas, como sería de esperar, porque los eventos de unión intermolecular, en contraposición a los intramoleculares, son dependientes de la concentración (es decir, las concentraciones más bajas conducen a la reducción de la hibridación/unión intermolecular). Además, la tasa de conmutación de MID disminuye ligeramente con la longitud. Esto es algo inesperado ya que los extremos de los fragmentos de ADN más largos están efectivamente a una concentración más baja con respecto a la otra. Las razones por las cuales los autores de la presente invención no observan esto es probablemente debido a que la producción de intermedios de cebado réflex continúa durante la PCR final, lo que significa que las reacciones de cebado réflex están sucediendo continuamente, lo que contribuye a la conmutación de MID. Este es probablemente el caso en el que los productos réflex más cortos son capaces de experimentar una mayor tasa de réflex de 'fondo', y por lo tanto aumenta un poco la tasa general de conmutación de MID.

Estos resultados demuestran que la conmutación de MID puede ser minimizada (p. ej., por debajo de 2%, por debajo de 1% o incluso a niveles casi indetectables) mediante la alteración de ciertos parámetros de la reacción, p. ej., mediante la reducción de la concentración de fragmentos y/o la longitud de los fragmentos.

El panel en la parte inferior derecha de la Figura 12 muestra la frecuencia de conmutación de MID en el procedimiento réflex para los fragmentos de 800 pares de base (es decir, fragmentos que contienen MIDA/MIDB). En esta figura, el área de cada círculo es proporcional al número de lecturas que contienen las correspondientes especies MIDA y MIDB (p. ej., MIDA1/MIDB1; MIDA1/MIDB2; etc.). Por lo tanto, un círculo que representa 200 lecturas será 40 veces más grande en términos de área que un círculo que representa 5 lecturas.

Como se señaló anteriormente, las combinaciones MIDA/MIDB que tienen el mismo número (que se muestra en los ejes X e Y, respectivamente) representan las combinaciones MIDA / MIDB presentes en la muestra antes del procedimiento réflex que se va a realizar (es decir, las combinaciones MIDA/MIDB 1/1, 2/2, 3/3, 4/4, 5/5, 6/6, 7/7 8/8, 9/9, y 10/10 estaban presentes en la muestra de partida). Todas las demás combinaciones MIDA/MIDB identificadas por medio de secuenciación con Roche 454 fueron el resultado de una conmutación de MID.

Esta figura muestra que la conmutación de MID que se produce durante el procedimiento réflex es aleatoria, es decir, que la conmutación de MID no está sesgada en base a la identidad de los MID de la reacción).

#### Protocolos réflex ilustrativos

La Figura 13 muestra un diagrama de protocolos ilustrativos para realizar el procedimiento réflex en reservas de ácidos nucleicos, por ejemplo, reservas de ácidos nucleicos de diferentes individuos, cada uno de los cuales está marcado con un MID único. En la Ruta 3, una biblioteca extendida reunida y etiquetada se somete directamente a un procedimiento réflex. En la Ruta 2, la biblioteca reunida se enriquece mediante la hibridación específica de la diana seguido de la realización del procedimiento réflex. En la Ruta 1 se emplea el enriquecimiento mediante amplificación por PCR. Como se muestra en la Figura 13, el enriquecimiento por PCR se puede realizar directamente sobre la biblioteca extendida etiquetada agrupada o en una reacción de PCR secundaria después de haber realizado una etapa de enriquecimiento basada en la hibridación (como en la Ruta 2) para generar un sustrato de amplicón que es adecuado para el procedimiento réflex. Otras rutas para preparar una muestra de polinucleótido para llevar a cabo un procedimiento reflex pueden ser implementadas (p. ej., con etapas de amplificación, purificación y/o enriquecimiento adicionales), que serán generalmente dependientes de los deseos del usuario.

Aunque la invención anterior se ha descrito con cierto detalle a modo de ilustración y ejemplo para propósitos de

claridad de comprensión, resulta fácilmente evidente para los expertos normales en la técnica a la luz de las enseñanzas de esta invención que se pueden realizar ciertos cambios y modificaciones en la misma sin apartarse del alcance de las reivindicaciones adjuntas.

5 Por consiguiente, lo anterior simplemente ilustra los principios de la invención. Se apreciará que los expertos en la técnica serán capaces de idear diversas disposiciones que, aunque no se describen explícitamente ni se muestran en la presente invención, plasman los principios de la invención y se incluyen dentro de su alcance. Por otra parte, todos los ejemplos y el lenguaje condicional citados en la presente memoria tienen como principal objetivo ayudar al lector en la comprensión de los principios de la invención y los conceptos aportados por los autores de la presente invención para la promoción de la técnica, y se debe interpretar que no tienen limitación para tales ejemplos y condiciones citados específicamente. No se pretende que el alcance de la presente invención, por lo tanto, se limite a las realizaciones ilustrativas mostradas y descritas en la presente memoria. Más bien, el alcance de la actual invención está plasmado en las reivindicaciones adjuntas.

10

**REIVINDICACIONES**

1. Un método de análisis de una muestra, que comprende:
  - a) emparejar una secuencia réflex en una hebra de polinucleótido de una muestra con un complemento de la secuencia réflex presente en la hebra de polinucleótido, en donde la hebra de polinucleótido comprende lo siguiente en una orientación 5' a 3':
    - un primer dominio que comprende:
      - i) un sitio de cebador; y
      - ii) un identificador múltiplex (MID) cuya identidad se puede utilizar para diferenciar los polinucleótidos de la muestra;
    - la secuencia réflex;
    - un polinucleótido de interés que tiene un sitio deseado que es distal con respecto al primer dominio; y
    - el complemento de la secuencia réflex, en donde el complemento de la secuencia réflex está presente en el extremo 3' de la hebra de polinucleótido;
  - b) extender desde el extremo 3' del complemento de la secuencia réflex a través del primer dominio para producir una región de doble hebra que comprende un complemento del primer dominio en el extremo 3' de la hebra de polinucleótido extendida, moviendo de ese modo el primer dominio a la proximidad del sitio deseado;
  - c) replicar la hebra de polinucleótido de la Etapa b) con un primer cebador de síntesis de ácido nucleico que se recuece en el extremo 3' del complemento del primer dominio para producir un primer producto de extensión;
  - d) emparejar un segundo cebador de síntesis de ácido nucleico al primer producto de extensión, en donde el segundo cebador de síntesis de ácido nucleico comprende un dominio de cebado 3' que es complementario a un sitio dentro del polinucleótido de interés en el primer producto de extensión, estando el sitio 3' con respecto al sitio deseado;
  - e) extender el segundo cebador de síntesis para producir un segundo producto de extensión; y
  - f) secuenciar el MID y el sitio deseado utilizando un cebador complementario al sitio del cebador en el complemento del primer dominio.
2. Un método como se reivindica en la reivindicación 1, en donde la secuencia réflex tiene de 10 a 50 nucleótidos de longitud.
3. Un método como se reivindica en la reivindicación 1, en donde el primer dominio comprende un sitio de ARN polimerasa.
4. Un método como se reivindica en la reivindicación 1, que comprende eliminar el primer dominio y la secuencia réflex del extremo 5' de la región de doble hebra.
5. Un método como se reivindica en la reivindicación 4, en donde la etapa de eliminación comprende el tratamiento del ácido nucleico de la Etapa b) con exonucleasa T7
6. Un método como se reivindica en la reivindicación 1, en donde el primer dominio está presente en una región de adaptador anclada al segmento del genoma.
7. Un método como se reivindica en la reivindicación 6, en donde la secuencia réflex está presente en la región de adaptador.
8. Un método como se reivindica en la reivindicación 6, en donde la secuencia réflex es una secuencia presente dentro de la región 5' del segmento del genoma.
9. Un método como se reivindica en la reivindicación 1, en donde la muestra es una muestra multiplexada, en donde el primer dominio de cada una de las hebras de polinucleótidos en la muestra multiplexada comprende un MID correspondiente a su muestra parental.
10. Un método como se reivindica en la reivindicación 1, en donde la muestra comprende múltiples hebras de polinucleótidos, y en donde cada hebra de polinucleótido de la muestra comprende un único MID en el primer dominio.
11. Un método como se reivindica en la reivindicación 1, en donde la hebra de polinucleótido de la Etapa a) es producida mediante:

emparejamiento de un cebador de síntesis de ácido nucleico a una hebra de polinucleótido parental, en donde la hebra de polinucleótido parental comprende lo siguiente en una orientación 5' a 3':

el primer dominio;

la secuencia réflex; y

5 un segundo dominio que comprende el segmento del genoma;

en donde el cebador de síntesis de ácido nucleico comprende:

un dominio de cebado 3' que es complementario a un sitio adyacente a y 3' del polinucleótido de interés en el segundo dominio; y

una secuencia réflex 5';

10 extensión del cebador de síntesis de ácido nucleico para producir un producto de extensión; y

replicación del producto de extensión resultante con un segundo cebador de síntesis de ácido nucleico que se recuece en el extremo 3' del complemento del primer dominio, produciendo de ese modo la hebra de polinucleótido.

15 12. Un método como se reivindica en la reivindicación 1, en donde el segundo cebador de síntesis comprende un dominio 5' que no es complementario al polinucleótido de interés en el primer producto de extensión.

13. Un método como se reivindica en la reivindicación 12, en donde el primer dominio comprende un primer sitio de cebador y el dominio 5' del segundo cebador de síntesis comprende un segundo sitio de cebador.

14. Un método como se reivindica en cualquiera de las reivindicaciones anteriores, en donde el polinucleótido de interés deriva de un virus.

20

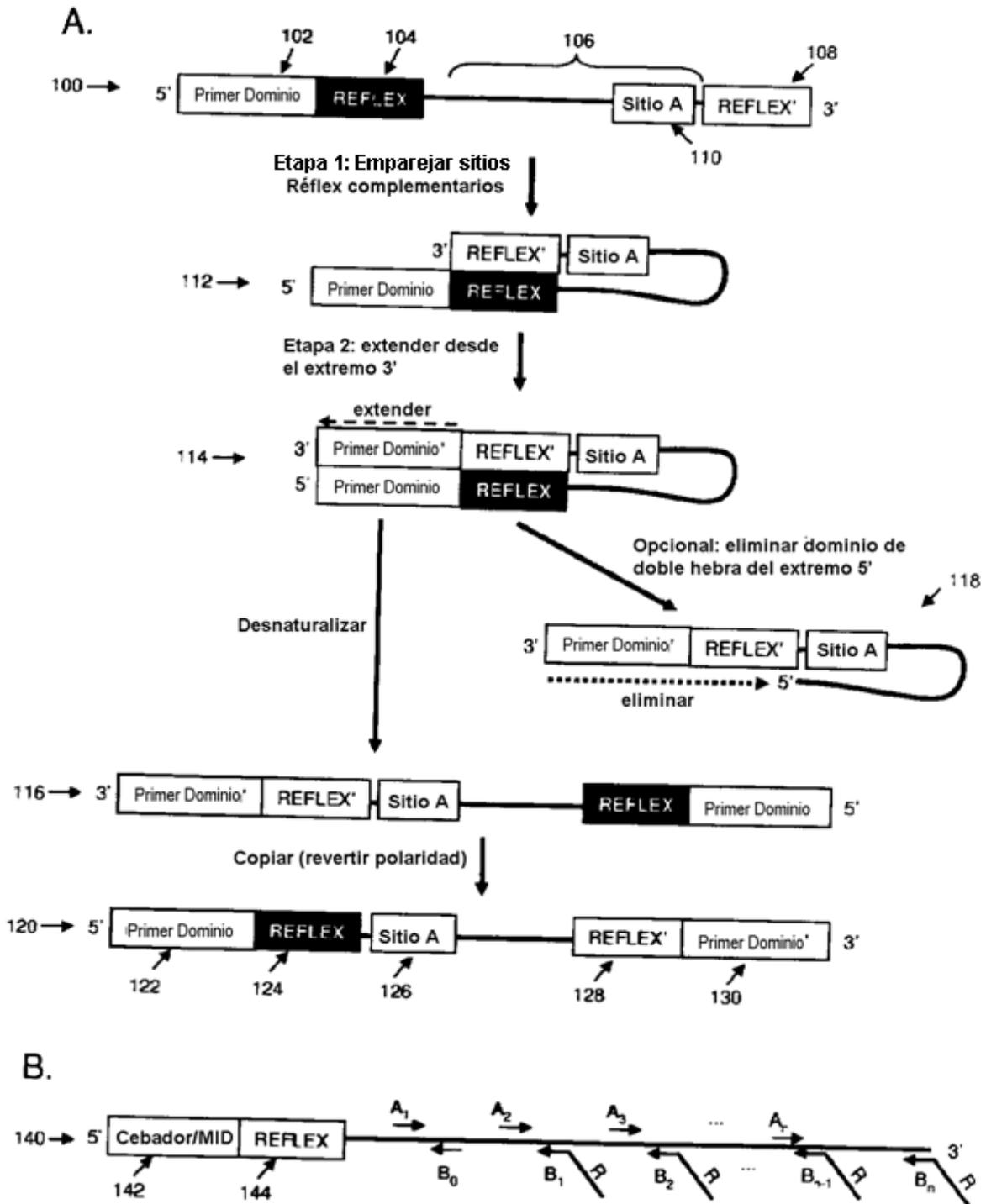


Fig. 1

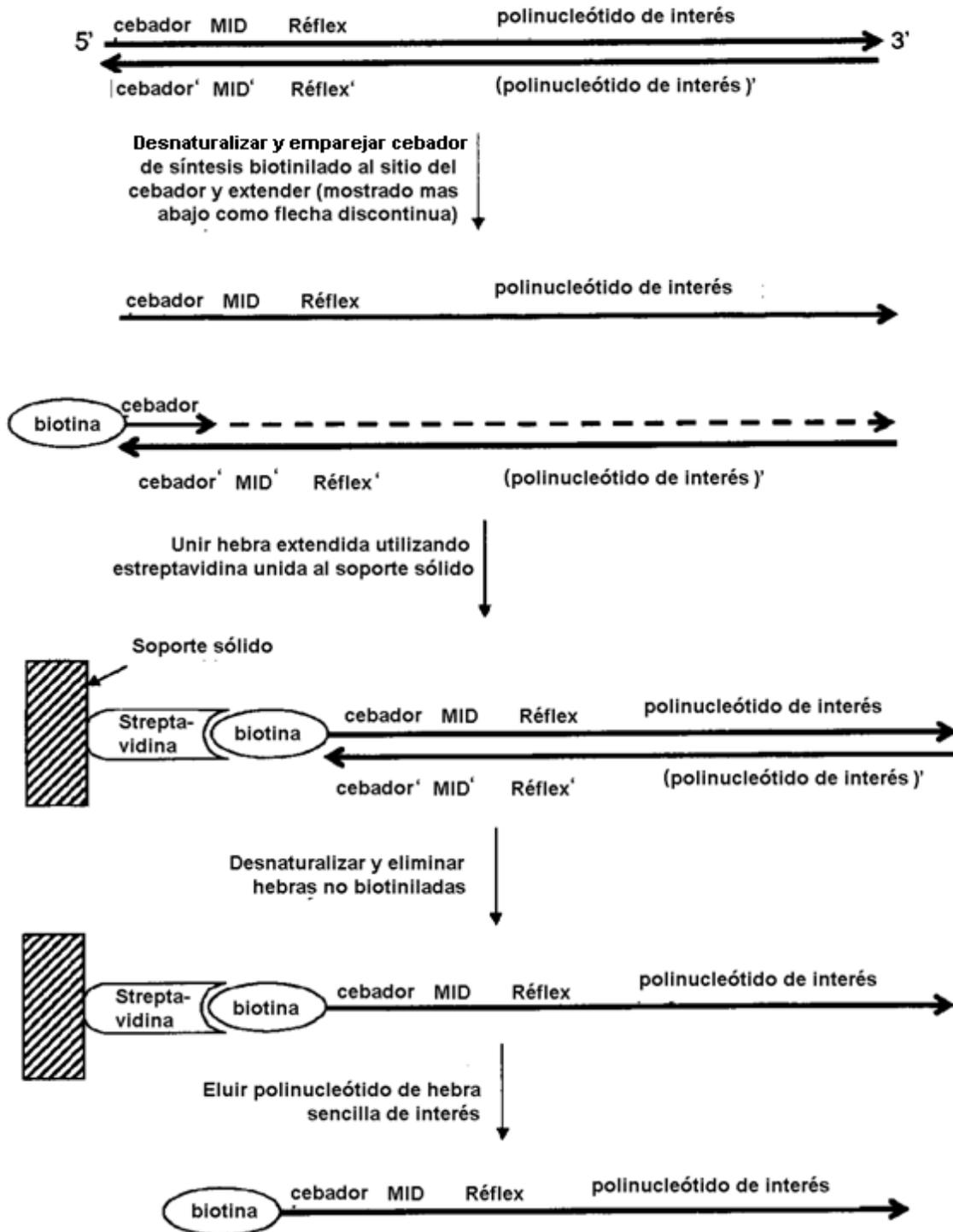
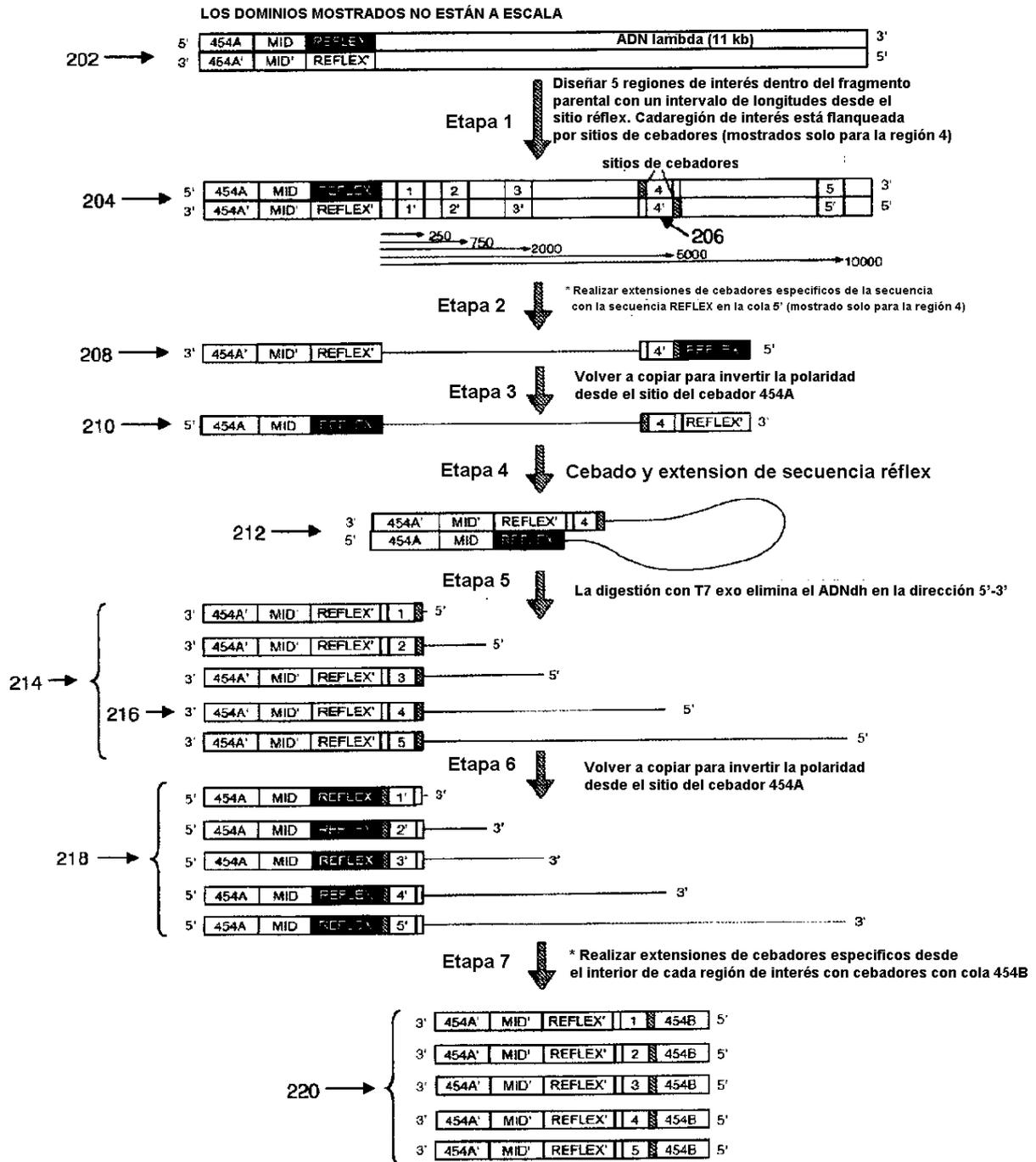


Fig. 2



Las reacciones de extensión de cebadores con \* se pueden realizar de manera que el aislamiento de la especie de hebra sencilla se facilite (p. ej., utilizando cebadores con radicales de unión y/o múltiples ciclos de extensión)

Fig. 3

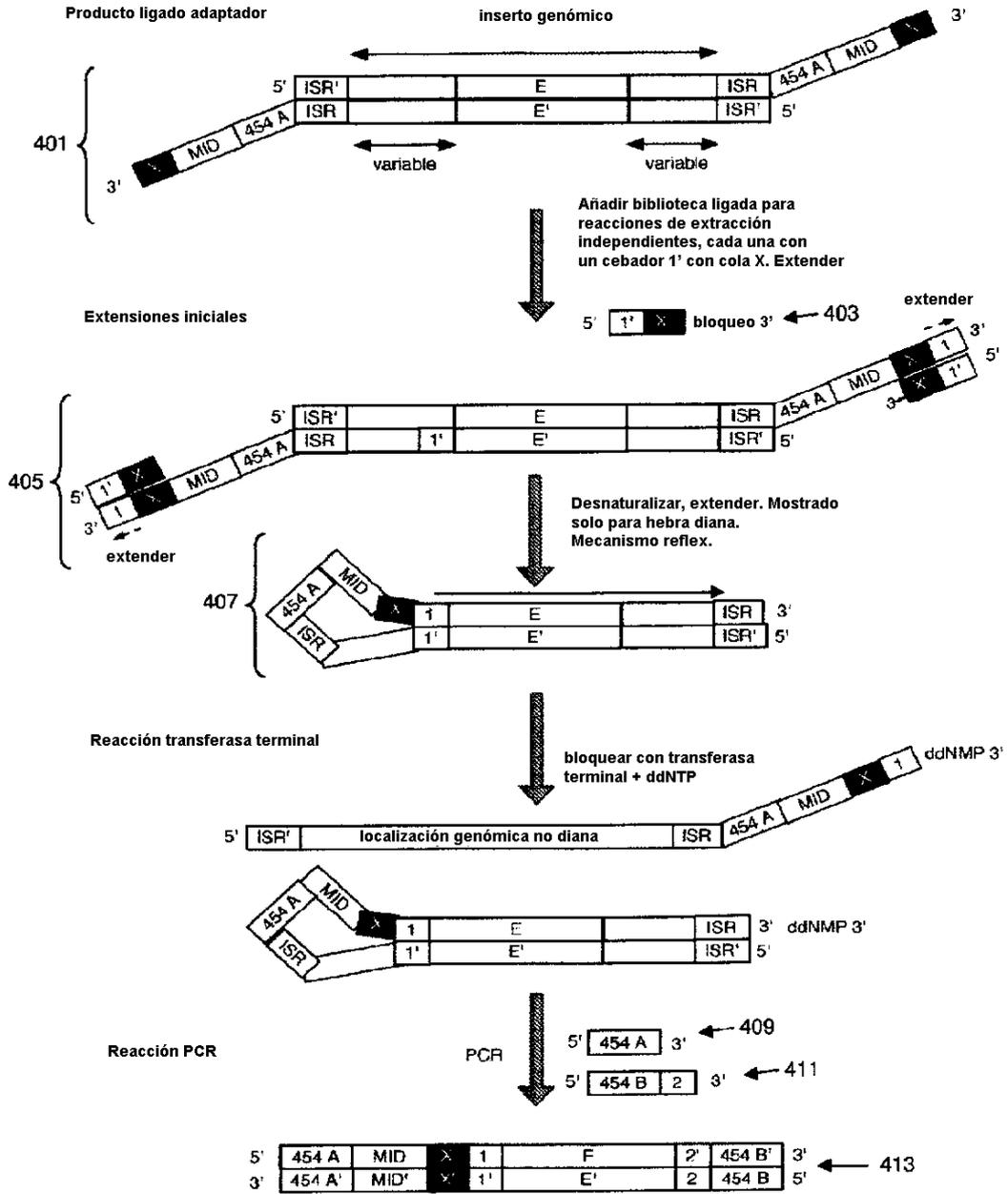


Fig. 4

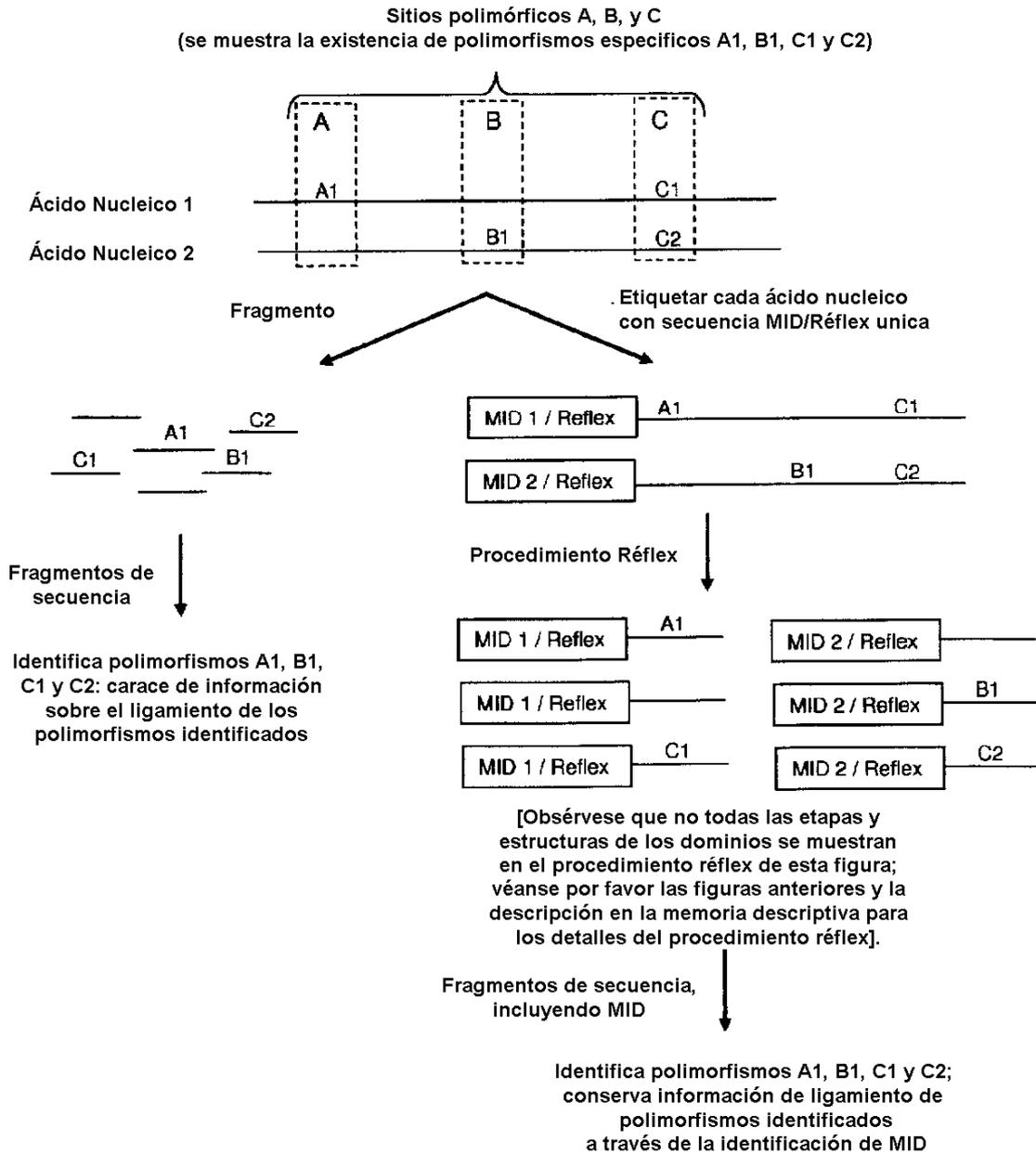
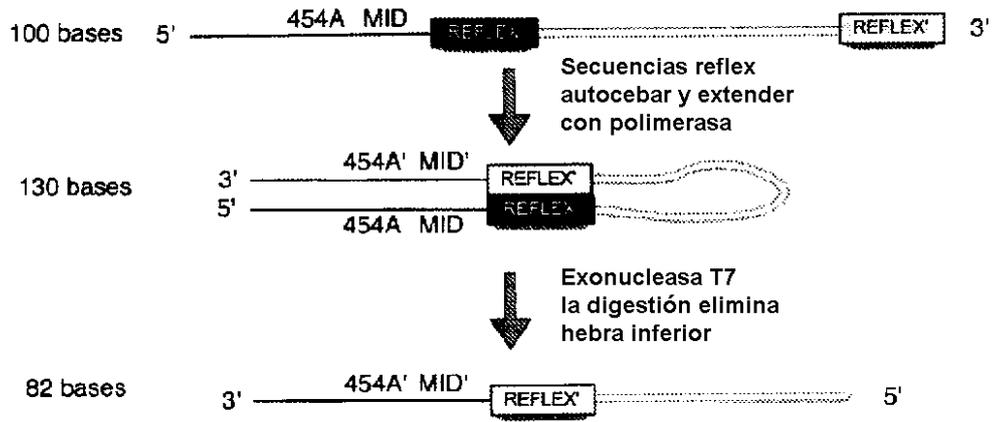


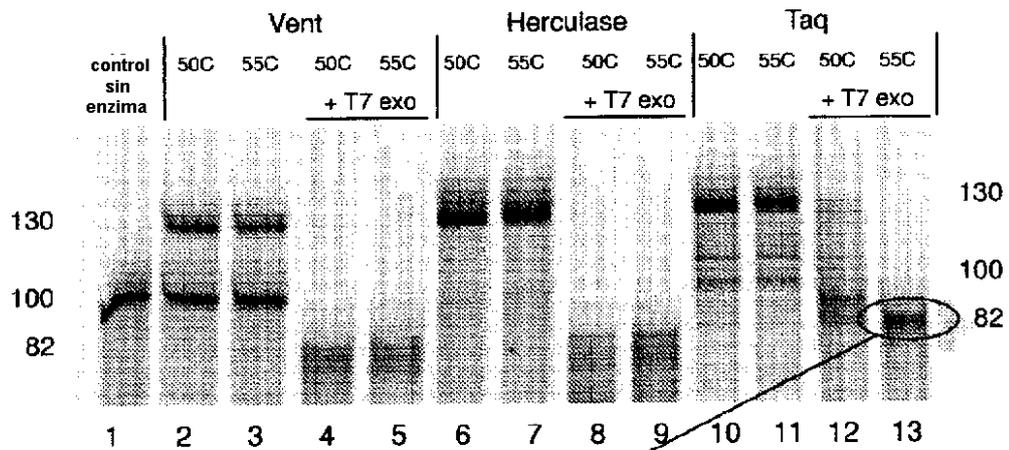
Fig. 5

A.



B.

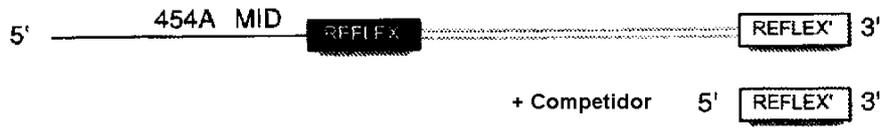
Ensayo con polimerasa a dos temperaturas de emparejamiento



La extensión es mejor con Herculase, pero la actividad exonucleasa 3'-5' da como resultado la digestión parcial del producto de 82 bases deseado. Taq, que carece de actividad exonucleasa 3'-5', muestra una banda más fuerte en el tamaño esperado del producto final.

Fig. 6

A.



B.

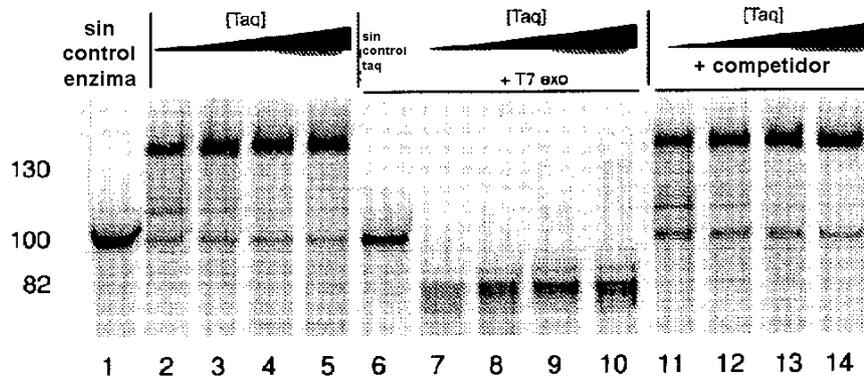
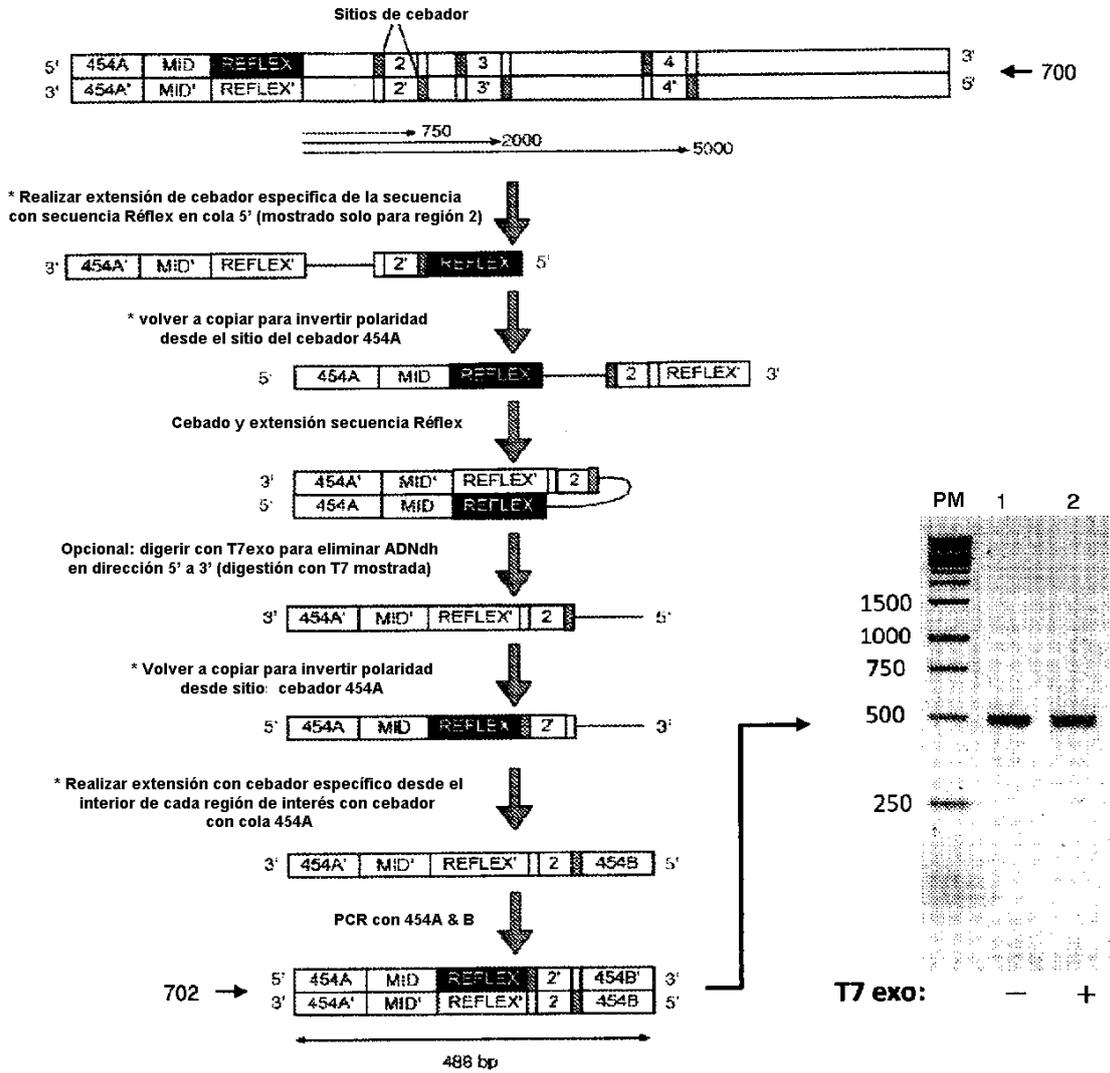


Fig. 7



Las reacciones de extensión de cebador con \* se pueden realizar de manera que se facilita el aislamiento de especies de hebra sencilla (p. ej., utilizando cebadores con radicales de unión y/o múltiples ciclos de extensión)

Fig. 8

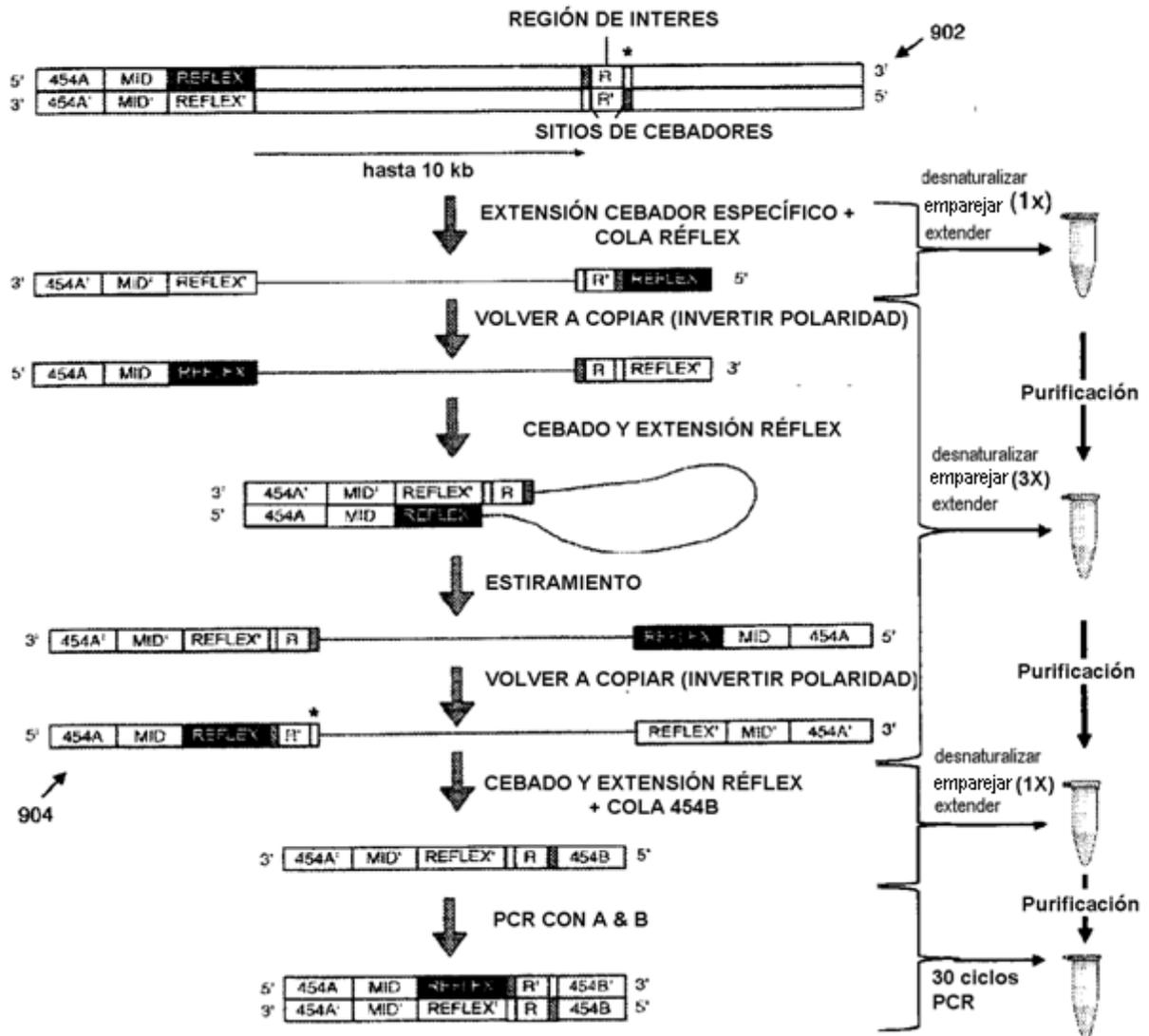
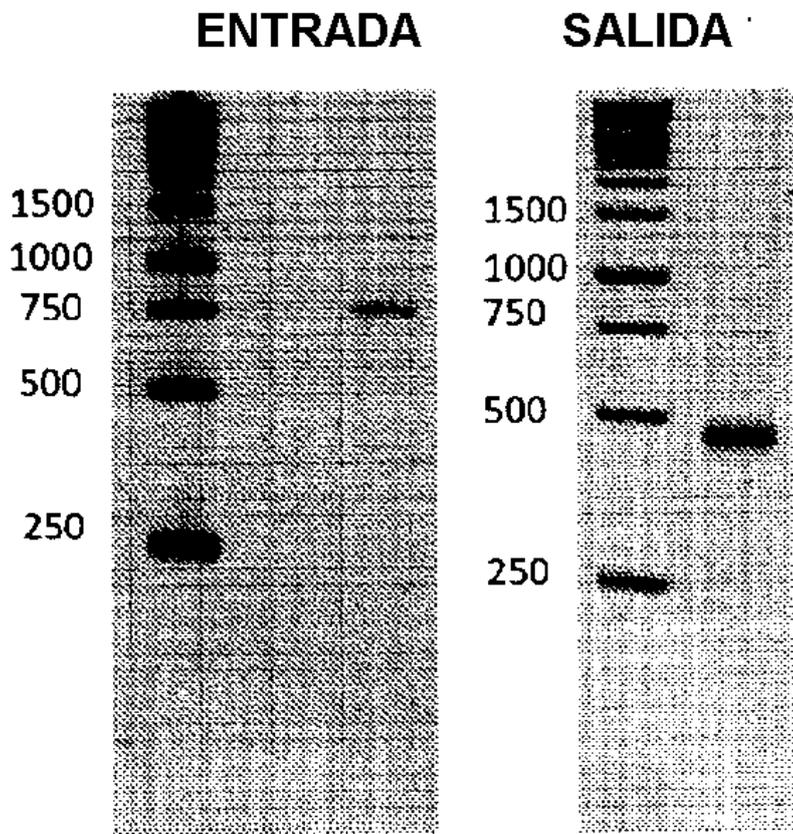
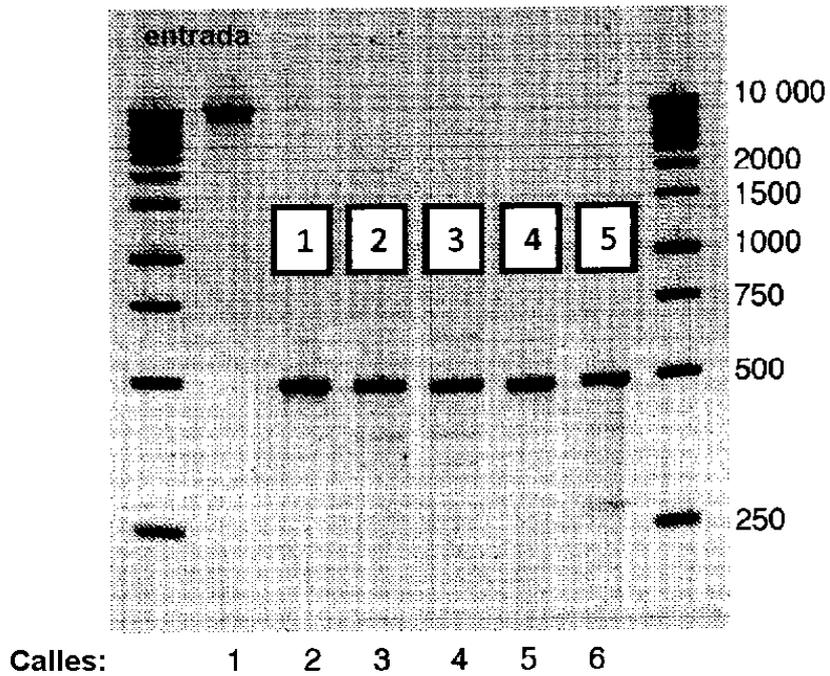
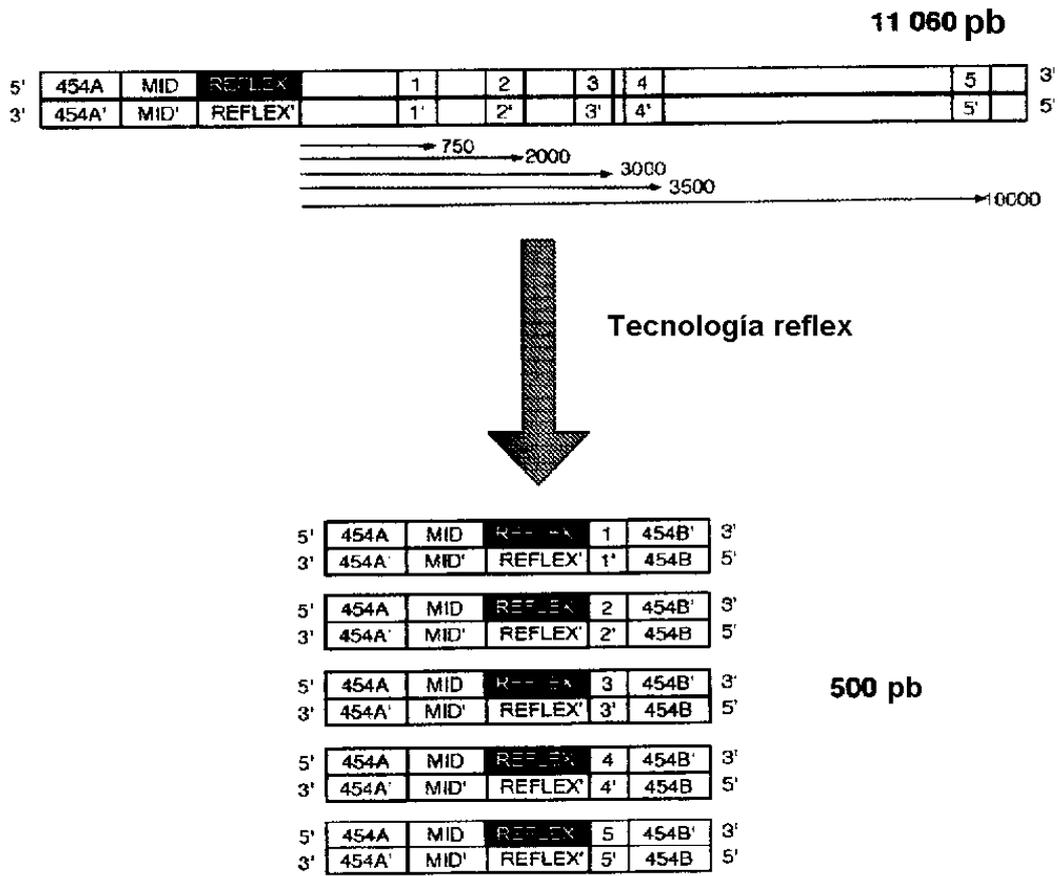


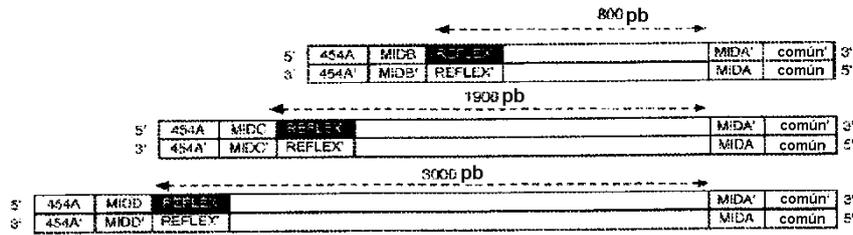
Fig. 9



**Fig. 10**

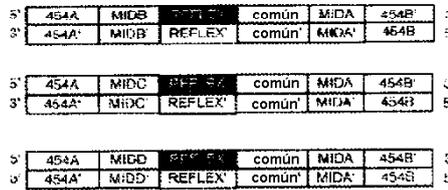


**Fig. 11**



Para cada longitud hay 10 pares de MID

Tecnología réflex a cuatro concentraciones diferentes



Secuenciación Roche 454

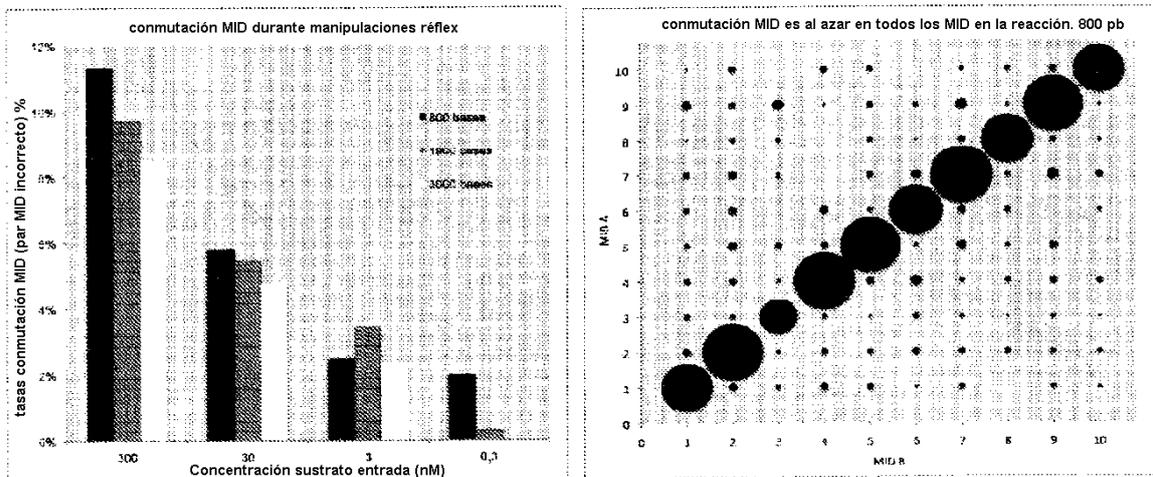


Fig. 12

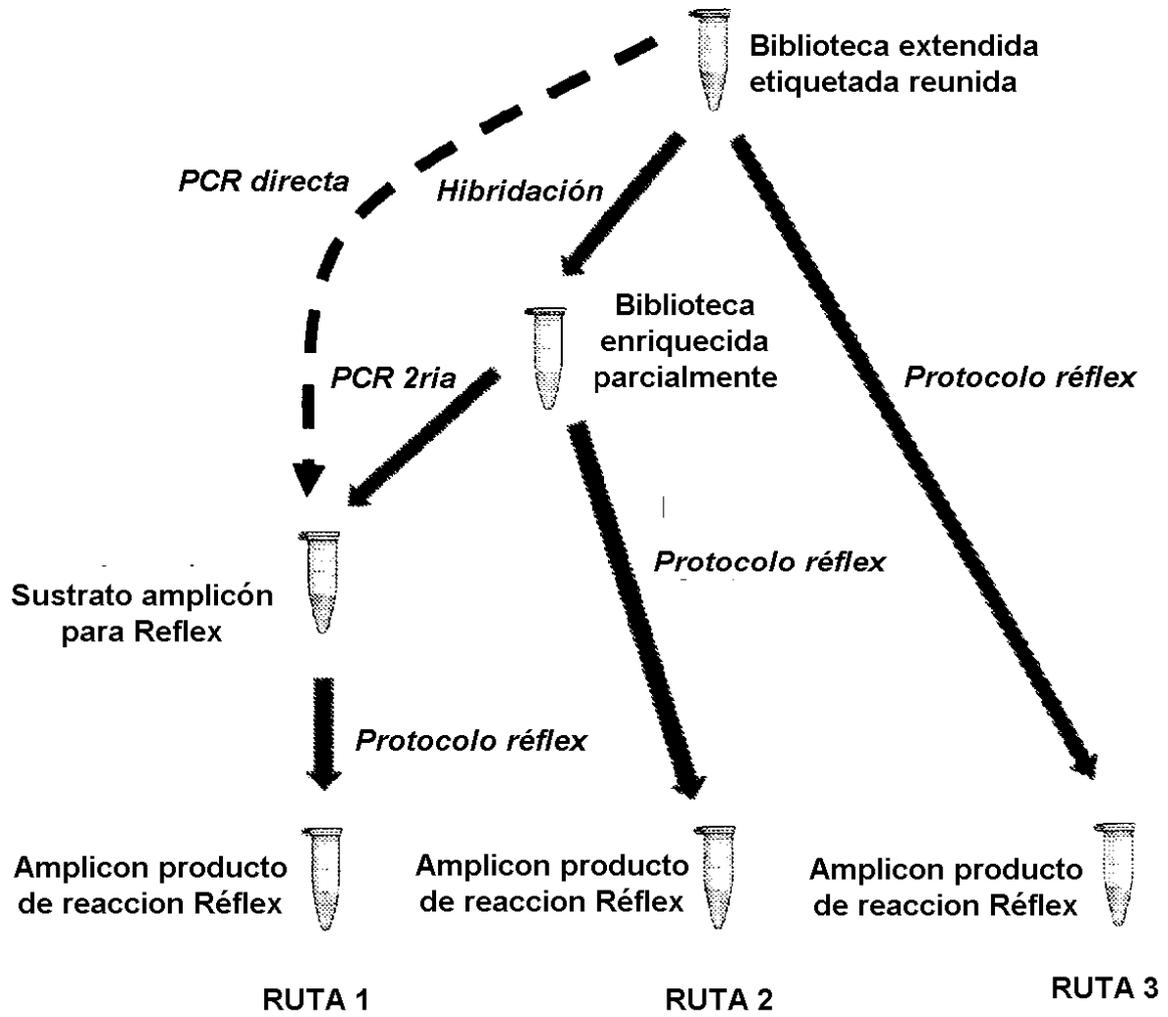


Fig. 13