

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 422 288**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **08.11.2006 E 06824255 (1)**

97 Fecha y número de publicación de la concesión europea: **24.04.2013 EP 1960541**

54 Título: **Método para el cribado ultrarrápido de poblaciones de transposón etiquetado e identificación masiva de la secuencia en paralelo de los sitios de inserción**

30 Prioridad:

**14.11.2005 US 735878 P**  
**10.10.2006 US 850274 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**10.09.2013**

73 Titular/es:

**KEYGENE N.V. (100.0%)**  
**AGRO BUSINESS PARK 90**  
**6708 PW WAGENINGEN, NL**

72 Inventor/es:

**VAN EIJK, MICHAEL, JOSEPHUS, THERESIA;**  
**GERATS, ANTONIUS, GERARDUS, MARIE;**  
**VAN TUNEN, ADRIANUS, JOHANNES y**  
**VANDEBUSSCHE, MICHIEL, MARCEL, ALBERT**

74 Agente/Representante:

**MANRESA VAL, Manuel**

**ES 2 422 288 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

Método para el cribado ultrarrápido de poblaciones de transposón etiquetado e identificación masiva de la secuencia en paralelo de los sitios de inserción.

5

Campo de la invención

La presente invención se refiere a los campos de la biología molecular y la genética. La presente invención se refiere a estrategias mejoradas para identificar mutantes de genes en poblaciones, basándose en la utilización de técnicas de secuenciación ultrarrápida.

10

Antecedentes de la invención

En la investigación genómica actual en vegetales se utilizan poblaciones de transposones marcados para identificar los genes que afectan a rasgos de importancia agronómica o general mediante métodos de genética inversa.

15

Suponen herramientas complementarias para el descubrimiento de genes, ya que las poblaciones de transposones se utilizan habitualmente para identificar el gen responsable de un fenotipo observado, el método denominado genética directa. Este se distingue en la técnica del método de genética inversa en que se identifican las mutaciones en secuencias (genes) de interés. La etapa limitante de dichos métodos es el trabajo de cribado relacionado con la identificación del individuo que presenta una mutación en el gen o secuencia de interés. A continuación, se describen más detalladamente los principios de las poblaciones de transposones y los métodos de cribado y se presentan métodos de cribado más eficientes que aumentan el valor de dichas herramientas en el descubrimiento de genes.

20

25

Los transposones son elementos genéticos móviles que se producen, de un modo natural o por genomanipulación, en una pluralidad de copias en el genoma. Son inestables ya que su posición en el genoma puede cambiar por escisión e inserción en sitios nuevos, por lo general en cualquier momento dado del ciclo de vida. Las poblaciones de transposones son valiosas en el descubrimiento de genes ya que pueden alterar la función genética si se insertan en las secuencias de genes o en sus regiones reguladoras. Se conocen las secuencias de muchos transposones utilizadas en la reproducción de vegetales, pero una vez que se observa una planta con un fenotipo interesante, se desconoce qué gen se ve afectado por la inserción de un transposón. En general, se desconoce asimismo si, y en caso afirmativo, que transposón es el responsable del fenotipo. Dependiendo del organismo y del transposón, el número de copias de transposones en las poblaciones de transposones varía de varias decenas a cientos de transposones por planta.

30

35

Los métodos actuales de cribado para analizar secuencias mutantes fenotípicas provocadas por transposones comprenden métodos relacionados con la PCR mediada por ligación para obtener las secuencias flanqueadoras de los sitios de integración del transposón específicos de la secuencia. Una limitación de la PCR mediada por ligación es que la determinación de las secuencias flanqueadoras requiere la escisión de bandas a partir de geles de secuenciación, lo que resulta lento, difícil de automatizar y con un rendimiento relativamente bajo (no se puede adaptar fácilmente a miles de bandas) (Salland et al. 2003, *Theor.Appl.Genet.* 106: 1396-1408). Se mejorará el cribado de poblaciones de transposones si se dispone de un método para recoger secuencias flanqueadoras de todos o por lo menos una parte de los transposones, integradas en el genoma. Pretendemos proporcionar un método eficaz para analizar y utilizar las inserciones en secuencias preferidas.

40

45

Definiciones

En la siguiente descripción y ejemplos se utilizan diversos términos. Para proporcionar una comprensión clara y coherente de la memoria descriptiva y de las reivindicaciones, comprendiendo el alcance que debe darse a dichos términos, se proporcionan las definiciones siguientes. Excepto si se indica lo contrario en la presente memoria, todos los términos técnicos y científicos utilizados tienen el mismo significado que entiende comúnmente un experto ordinario en la materia a la que pertenece la presente invención.

50

55

Transposón: Los transposones son secuencias de ADN que se pueden desplazar a distintas posiciones dentro del genoma de una célula simple, un proceso denominado transposición. En el proceso, se producen mutaciones y cambios en la cantidad de ADN del genoma. Los transposones se denominan asimismo "genes saltadores" o "elementos genéticos móviles". Existen diversos elementos genéticos móviles; se pueden agrupar en función de su mecanismo de transposición. Los elementos genéticos móviles de clase I, o retrotransposones, se desplazan en el genoma al transcribirse a ARN y vuelven de nuevo al ADN mediante la transcriptasa inversa, mientras que los elementos genéticos móviles de clase II se desplazan directamente de una posición a otra dentro del genoma utilizando una transposasa para "cortar y pegar" los mismos en el genoma. La transposición puede ser replicativa, en la que una copia del elemento transponible permanece en el sitio donador y otra se inserta en el sitio diana; o la transposición se puede realizar de un modo conservador, en el que el elemento transponible se escinde de un sitio y se inserta en otro. El término comprende, pero son limitarse a los mismos, elementos transponibles que se

60

65

- encuentran en procariotas, tales como secuencias de inserción (IS), transposones (Tn) o bacteriófagos tales como Mu y D108. Los elementos transponibles eucariotas comprenden, pero sin limitarse a los mismos: los elementos copia tales como los que se encuentran en *D. melanogaster*; los elementos TY tales como los que se encuentran en la levadura; los elementos transponibles Ta1 y Tnt 1 tales como los que se encuentran en *Arabidopsis*; los IAP que se encuentran en ratones; los elementos transponibles Tam o Cin tales como los que se encuentran en la boca de dragón, y los elementos transponibles AC, Spm, Bs, Cin, Dt y Mutator tales como los que se encuentran en el maíz. El término comprende asimismo elementos transponibles sintéticos que pueden insertarse ya sea de un modo replicativo o conservador en un genoma hospedador y cuya transposición o escisión del genoma se puede controlar mediante la intervención humana. Por ejemplo, se puede construir un elemento transponible sintético que carezca de una transposasa funcional (la enzima que interviene en la transposición), sino que se suministre en trans mediante el enlace funcional del gen de la transposasa con un promotor inducible. Población de transposones: población individual de un organismo (por lo general vegetales, pero son posibles asimismo otros organismos, tales como la *Drosophila* y el ratón), en la que cada uno de los mismos presenta una pluralidad de transposones en su genoma y cada uno de dichos transposones puede afectar a uno o más genes, con lo que se obtienen distintos fenotipos. Normalmente las poblaciones de transposones se pueden obtener seleccionándose a partir de individuos o variedades que expresen la inestabilidad en una característica fenotípica. Las poblaciones de transposones pueden presentar un tamaño muy variable y, para ciertos fines, se pueden utilizar poblaciones parciales que contengan el 90, 80, 70, 60, 50, 40, 30 o incluso únicamente el 20% de la población original.
- Etiqueta: secuencia corta que se puede añadir a un cebador o incluir en la secuencia del mismo o utilizar de algún otro modo como marcador para proporcionar un identificador único. Dicho identificador de la secuencia puede ser una secuencia de bases única de longitud variable pero definida utilizada únicamente para identificar una muestra específica de ácido nucleico. Por ejemplo las etiquetas de 4 pb permiten  $4(\text{exp}4) = 256$  etiquetas distintas. Los ejemplos típicos son las secuencias ZIP, conocidas en la técnica (Iannone *et al. Cytometry* 39:131-140, 2000). Utilizando dicha etiqueta se puede determinar el origen de una muestra de PCR mediante procesamiento adicional. En el caso de combinar productos elaborados procedentes de muestras distintas de ácido nucleico, se identifican generalmente las muestras diferentes de ácido nucleico utilizando etiquetas distintas. En el caso de la presente invención, la adición de una etiqueta de secuencia única permite identificar las coordenadas del vegetal en particular en la mezcla de secuencias de los productos de amplificación. Se pueden utilizar etiquetas múltiples.
- Etiquetado: se refiere al proceso de la adición de una etiqueta o marcador a un ácido nucleico para poder distinguir el mismo de un segundo o más ácidos nucleicos. El etiquetado se puede realizar, por ejemplo, añadiendo un identificador de la secuencia durante la amplificación utilizando cebadores etiquetados o mediante cualquier otro medio conocido en la técnica.
- Endonucleasa de restricción: una endonucleasa de restricción o enzima de restricción es un enzima que reconoce una secuencia de nucleótidos específica (sitio diana) en una molécula de ADN de doble cadena y que escindirá ambas cadenas de la molécula de ADN en cada sitio diana.
- Fragmentos de restricción: las moléculas de ADN producidas mediante la digestión con una endonucleasa de restricción se denominan fragmentos de restricción. Cualquier genoma determinado (o ácido nucleico, con independencia de su origen) se digerirá con una endonucleasa de restricción particular en un conjunto discreto de fragmentos de restricción. Los fragmentos de ADN que se obtienen a partir de la escisión con una endonucleasa de restricción se pueden utilizar asimismo en diversas técnicas y se pueden detectar, por ejemplo, mediante electroforesis en gel.
- Ligación: la reacción enzimática catalizada por un enzima ligasa en la que dos moléculas de ADN de doble cadena se enlazan covalentemente entre sí se denomina ligación. En general, ambas cadenas de ADN se unen covalentemente entre sí, pero es posible asimismo evitar la ligación de una de las dos cadenas mediante la modificación química o enzimática de uno de los extremos de las cadenas. En dicho caso, el enlace covalente se producirá en únicamente una de las dos cadenas de ADN.
- Oligonucleótido sintético: las moléculas de ADN monocatenario que presentan preferentemente entre aproximadamente 10 y aproximadamente 50 bases, que se pueden sintetizar químicamente se denominan oligonucleótidos sintéticos. En general, dichas moléculas de ADN sintéticas se diseñan para que presenten una secuencia de nucleótidos única o pretendida, aunque resulta posible sintetizar familias de moléculas con secuencias relacionadas y que presenten distintas composiciones de nucleótidos en posiciones específicas dentro de la secuencia de nucleótidos. Se utilizará el término oligonucleótido sintético para hacer referencia a moléculas de ADN que presentan una secuencia de nucleótidos diseñada o pretendida.
- Adaptadores: moléculas de ADN de doble cadena corta con un número limitado de pares de bases, por ejemplo, entre aproximadamente 10 y aproximadamente 30 pares de bases de longitud, que se diseñan de tal modo que se pueden ligar a los extremos de fragmentos de restricción. Los adaptadores comprenden generalmente dos oligonucleótidos sintéticos que presentan unas secuencias de nucleótidos que son parcialmente complementarias entre sí. Cuando se mezclan dos oligonucleótidos sintéticos en disolución en las condiciones apropiadas, se

hibridarán entre sí formando una estructura bicatenaria. Tras hibridarse, un extremo de la molécula adaptador se diseña de tal modo que es compatible con el extremo de un fragmento de restricción y se puede ligar al mismo; el otro extremo del adaptador se puede diseñar de tal modo que no se pueda ligar, pero no es necesario que este sea el caso (adaptadores dobles ligados).

5 Fragmentos de restricción ligados a un adaptador: fragmentos de restricción que se han rematado con adaptadores.

10 Ácido nucleico: un ácido nucleico según la presente invención puede comprender cualquier polímero u oligómero de bases de pirimidina y purina, preferentemente citosina, timina y uracilo, y adenina y guanina, respectivamente (véase Albert L. Lehninger, *Principles of Biochemistry* ("Principios de Bioquímica"), en 793-800 (Worth Pub. 1982)). La presente invención contempla cualquier desoxirribonucleótido, ribonucleótido o péptido componente de ácido nucleico, y cualquier variante química de los mismos, tales como las formas metiladas, hidroximetiladas o glucosiladas de dichas bases, y similares. Los polímeros u oligómeros pueden ser heterogéneos u homogéneos en su composición, y se pueden aislar a partir de fuentes naturales o se pueden producir artificial o sintéticamente.

15 Además, los ácidos nucleicos pueden ser ADN o ARN, o una mezcla de los mismos, y pueden existir de modo permanente o transitorio en forma monocatenaria o bicatenaria, comprendiendo los estados homodúplex, heterodúplex e híbrido.

20 Secuenciación: El término secuenciación se refiere a la determinación del orden de nucleótidos (secuencias de bases) en una muestra de ácido nucleico, por ejemplo, ADN o ARN.

25 Alinear y alineación: con los términos "alinear" y "alineación" se entiende la comparación entre dos o más secuencias de nucleótidos basándose en la presencia de fragmentos cortos o largos de nucleótidos idénticos o similares. Se conocen en la técnica diversos métodos para alinear secuencias de nucleótidos, tal como se describirá posteriormente. A veces los términos "grupo" o "agrupación" se utilizan como sinónimos. Cribado ultrarrápido: el cribado ultrarrápido, a menudo abreviado como HTS, es un método de experimentación científica especialmente importante en los campos de la biología y la química. Combinando la robótica moderna y otros equipos de laboratorio especializados, el investigador puede cribar eficazmente grandes cantidades de muestras simultáneamente.

30 Cebadores: en general, el término cebador se refiere a una cadena de ADN que puede cebar la síntesis de ADN. La ADN polimerasa no puede sintetizar ADN nuevo sin cebadores: únicamente puede extender una cadena de ADN existente en una reacción en la que la cadena complementaria se utiliza como molde para dirigir el orden de nucleótidos a ensamblar. Nos referiremos a las moléculas de oligonucleótidos sintéticos que se utilizan en una reacción en cadena de la polimerasa (PCR) como cebadores.

35

40 Cebadores con afinidad aumentada: los cebadores que contienen nucleótidos modificados tales como PNA o LNA, que aumentan la estabilidad térmica de los mismos, lo que permite una amplificación más específica basándose en diferencias simples de la secuencia de nucleótidos. Para conseguirlo, se incorporan con frecuencia uno o más nucleótidos modificados, preferentemente en el extremo 3' del cebador.

45 Amplificación del ADN: el término amplificación del ADN se utiliza habitualmente para referirse a la síntesis *in vitro* de moléculas de ADN bicatenario utilizando la PCR. Cabe señalar que existen otros métodos de amplificación y que se pueden utilizar en la presente invención sin apartarse de la esencia de la misma. Hibridación selectiva: se refiere a la hibridación, en unas condiciones restrictivas de hibridación, de una secuencia de ácido nucleico con una secuencia diana de ácido nucleico especificada en un grado claramente superior (por ejemplo, preferentemente por lo menos 2 veces con respecto al contexto) a su hibridación con secuencias de ácido nucleico no diana y con la exclusión sustancial de los ácidos nucleicos no diana. Los términos "condiciones restrictivas" o "condiciones de hibridación restrictivas" se refieren a condiciones en las que una sonda se hibridará con su secuencia diana, en un grado claramente superior a otras secuencias (por ejemplo, preferentemente por lo 2 veces con respecto al contexto). Las condiciones restrictivas dependen de la secuencia y serán distintas en circunstancias diferentes. Controlando la restricción de la hibridación y/o las condiciones de lavado, se pueden identificar las secuencias diana que son 100% complementarias a la sonda (sondeo homóloga). Alternativamente, se pueden ajustar las condiciones de restricción para permitir algún desajuste en las secuencias de tal modo que se detecten grados inferiores de similitud (sondeo heteróloga). Generalmente, una sonda presenta menos de aproximadamente 100 nucleótidos de longitud, preferentemente no más de 50 o 25 nucleótidos de longitud. Normalmente, las condiciones de restricción serán aquellas en las que la concentración salina es inferior a aproximadamente 1,5 M del ion Na, normalmente una concentración de ion Na (u otras sales) comprendida aproximadamente entre 0,01 a 1,0 M a un pH comprendido aproximadamente entre 7,0 y 8,3 y a una temperatura que normalmente es de por lo menos aproximadamente 30 °C para las sondas cortas (por ejemplo, entre 10 y 50 nucleótidos) y normalmente por lo menos aproximadamente 60 °C para las sondas largas (por ejemplo, superiores a 50 nucleótidos). Se pueden alcanzar asimismo unas condiciones de restricción añadiendo inestabilizantes tales como la formamida.

50

55

60

65 Los ejemplos de condiciones de poca restricción comprenden la hibridación con una disolución amortiguadora de formamida al 30 a 35%, NaCl 1 M, SDS (dodecilsulfato de sodio) al 1% a 37 °C y un lavado entre 1 \* y 2 \* SSC

(20\*SSC = NaCl 3,0 M / citrato trisódico 0,3 M) entre 50 y 55 °C. Los ejemplos de condiciones de restricción moderada comprenden la hibridación en formamida al 40 a 45%, NaCl 1 M, SDS al 1% a 37 °C y un lavado entre 0,5 \* y 1 \* SSC entre 55 y 60 °C. Los ejemplos de condiciones de restricción elevada comprenden la hibridación en formamida al 50%, NaCl 1 M, SDS al 1% a 37 °C y un lavado a 0,1 \* SSC entre 60 y 65 °C. La especificidad es normalmente función de los lavados posteriores a la hibridación, siendo los factores críticos la fuerza iónica y la temperatura de la disolución de lavado final. En el caso de los híbridos ADN-ADN, la T<sub>m</sub> se puede aproximar a partir de la ecuación de Meinkoth y Wahl, *Anal.Biochem.*, 138:267-284 (1984):  $T_m = 81,5 \text{ °C} + 16,6 (\log M) + 0,41 (\% \text{ GC}) - 0,61 (\% \text{ form}) - 500 / L$ ; en la que M es la molaridad de los cationes monovalentes, %GC es el porcentaje de los nucleótidos de guanosina y citosina en el ADN, %form es el porcentaje de formamida en la disolución de hibridación y L es la longitud del híbrido en pares de bases. La T<sub>m</sub> es la temperatura (bajo una fuerza iónica y un pH definidos) a la que el 50% de una secuencia diana complementaria se hibrida con una sonda perfectamente emparejada. La T<sub>m</sub> se reduce en aproximadamente 1 °C por cada 1% de emparejamiento incorrecto; de este modo se pueden ajustar la T<sub>m</sub> y las condiciones de hibridación y/o de lavado para hibridarse con secuencias que presenten la identidad pretendida. Por ejemplo, si se buscan secuencias con aproximadamente > 90% de identidad, la T<sub>m</sub> se puede disminuir 10 °C. Generalmente, las condiciones restrictivas se seleccionan para que sean aproximadamente 5 °C inferiores al punto de fusión térmico (T<sub>m</sub>) para la secuencia específica y su complemento a una intensidad iónica y un pH definidos. Sin embargo, las condiciones muy restrictivas pueden utilizar una hibridación y/o un lavado a 1, 2, 3 o 4 °C por debajo del punto de fusión térmico (T<sub>m</sub>); las condiciones moderadamente restrictivas pueden utilizar una hibridación y/o un lavado a 6, 7, 8, 9 o 10 °C por debajo del punto de fusión térmico (T<sub>m</sub>); las condiciones poco restrictivas pueden utilizar una hibridación y/o un lavado a 11, 12, 13, 14, 15 o 20 °C por debajo del punto de fusión térmico (T<sub>m</sub>). Utilizando la ecuación, las composiciones de hibridación y de lavado, y la T<sub>m</sub> pretendida, los expertos ordinarios en la materia comprenderán que se describen de manera inherente variaciones en la restricción de las disoluciones de hibridación y/o de lavado. Si el grado pretendido de emparejamiento incorrecto tiene como resultado una T<sub>m</sub> inferior a 45 °C (disolución acuosa) o 32 °C (disolución de formamida), se prefiere aumentar la concentración de SSC de modo que se pueda utilizar una temperatura superior. En Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes* ("Técnicas de laboratorio en bioquímica y biología molecular - hibridación con sondas de ácido nucleico"), Parte 1, Capítulo 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" ("Visión general de los principios de la hibridación y estrategia en los ensayos con sondas de ácido nucleico"), Elsevier, N.Y. (1993); y en *Current Protocols in Molecular Biology* ("Protocolos actuales en biología molecular"), Capítulo 2, Ausubel, et al., eds., Greene Publishing and Wiley-Interscience, New York (1995), se proporciona una guía extensa para la hibridación de ácidos nucleicos.

#### Descripción de la invención

Los presentes inventores han descubierto que utilizando métodos de secuenciación ultrarrápida, se pueden alcanzar los objetivos mencionados anteriormente y las poblaciones de transposones, o poblaciones que comprenden miembros que comprenden fenotipos de interés originados a partir de inserciones de transposones, se pueden seleccionar eficientemente con respecto a la presencia de inserciones en los genes de interés.

#### Descripción detallada de la invención

La presente invención se refiere a un método para la identificación de una inserción asociada con un gen o fenotipo de interés en un miembro de una población de transposones, que comprende las etapas de:

- (a) aislar, individualmente o en mezclas, el ADN genómico de la población de transposones;
- (b) opcionalmente, mezclar el ADN obtenido en la etapa (a);
- (c) realizar la restricción del ADN utilizando una o más, preferentemente dos o más, más preferentemente dos, endonucleasas de restricción, siendo preferentemente por lo menos una de las mismas una endonucleasa de restricción de corte frecuente que no corta en transposones y preferentemente por lo menos una es una endonucleasa de restricción de corte raro que corta en transposones, realizar la ligación de los adaptadores con los fragmentos de restricción para preparar de este modo los fragmentos de restricción ligados a adaptadores;
- (d) realizar la amplificación de los fragmentos de restricción ligados a adaptadores con un par de cebadores (opcionalmente marcados), con lo que uno de los cebadores comprende una sección que es complementaria a parte de una secuencia de transposón conocida y, opcionalmente, comprende además un sitio de enlace con el cebador de la secuencia, siendo el otro cebador por lo menos complementario al adaptador, comprendiendo uno o ambos cebadores una etiqueta para correlacionar el cebador y el producto de amplificación con el miembro original de la población;
- (e) opcionalmente, mezclar los productos de la amplificación de la etapa (d) para crear una genoteca de productos de amplificación;
- (f) opcionalmente, fragmentar los productos de amplificación de la genoteca;
- (g) determinar por lo menos parte de la secuencia de nucleótidos de por lo menos parte de los fragmentos de (d), (e) o (f) utilizando la secuenciación ultrarrápida;
- (h) opcionalmente, recortar informáticamente la secuencia de los fragmentos para eliminar de este modo cualquier adaptador e/o información de la secuencia relacionada con el transposón;

(i) identificar uno o más fragmentos de las etapas (g) o (h) que se puedan alinear con las secuencias de nucleótidos de una base de datos, correlacionando de este modo los fragmentos identificados con un gen o fenotipo de interés;

(j) identificar el / los miembro(s) de la población de transposones que presenten el / los fragmento(s) de la etapa (i);

(k) opcionalmente, diseñar una sonda o par de cebadores de PCR basándose en los fragmentos de la etapa (i) y utilizarla para confirmar la inserción del transposón en el gen de interés del genoma del miembro identificado en (j).

El aislamiento de ADN para proporcionar muestras de ADN de cada miembro de la población se alcanza generalmente utilizando métodos comunes en la técnica tales como la recogida de tejido de un miembro de la población, la extracción de ADN (por ejemplo usando el kit de ADN rápido de Q-Biogene), la cuantificación y la normalización para obtener cantidades equivalentes de ADN por muestra. A título de ejemplo, se ilustra la presente invención basándose en una población de transposones de 1000 plantas. Normalmente, se aísla el ADN de cada miembro de la población que expresa el fenotipo de interés.

Según el método de la presente invención, se pueden separar los organismos individuales cuyo ADN genómico comprende por lo menos un gen etiquetado con un elemento transponible mediante la presencia o ausencia de un fenotipo mutante de interés. Por lo tanto, se proporciona un método apto para identificar y aislar una secuencia genética de un organismo, en el que la interrupción del ADN genómico de dicho organismo por un elemento transponible que flanquea dicha secuencia genética se asocia, directa o indirectamente, a un fenotipo mutante. El fenotipo mutante del organismo es preferentemente uno conocido o del que se presume que procede de la interrupción de un único gen mediante la inserción de un elemento transponible o, por lo menos, del que no se puede descartar dicha inserción. En la práctica, ello puede significar que se separe un grupo de organismos basándose en la presencia (o ausencia) del fenotipo mutante. Los expertos en la materia comprenderán que la mezcla de organismos a separar debe reproducirse o cultivarse en unas condiciones similares para evitar la separación de fenotipos procedentes de contribuciones no genéticas (por ejemplo, efectos ambientales). El método de la presente invención se puede aplicar a cualquier fenotipo que se pueda distinguir y clasificar como de tipo silvestre o mutante. Dichos fenotipos se pueden detectar con medios visuales, bioquímicos, agronómicos o morfológicos. Los expertos reconocerán que los términos "silvestre" y "mutante", tal como se utilizan en la presente memoria, son términos arbitrarios utilizados para diferenciar organismos en función de la presencia o ausencia de un fenotipo particular. Los organismos a los que se puede aplicar la presente invención pueden ser procariotas o eucariotas. Los organismos eucariotas pueden ser haploides o diploides cuando se utilizan en los métodos de la presente invención. Los organismos diploides que presentan el fenotipo de tipo silvestre se pueden obtener a partir de la generación F1, pero los fenotipos mutantes asociados a los genes etiquetados con transposones se presentan más comúnmente como mutantes recesivos y, por lo tanto, aparecen más habitualmente en la generación F2. Por lo tanto, en una forma de realización preferida de la presente invención, los organismos procederán de la generación F2 a partir de un cruce entre un individuo donador de un elemento transponible y un individuo receptor endogámico que no presente elementos transponibles activos. Preferentemente, los métodos de la presente invención se aplicarán a vegetales. En ciertas formas de realización, el vegetal preferido es una planta monocotiledónea tal como las de la familia *Gramineae*, comprendiendo a título de ejemplo especies tales como *Zea mays*. En ciertas formas de realización de la presente invención, los organismos que presentan elementos transponibles serán las plantas de maíz a partir de la generación F2 de un cruce entre un individuo donador de Mu que presente el elemento regulador Mu-DR (Chomet *et al.* (1991) *Genetics* 122:447457) y un número elevado de copias de elementos Mu y un individuo endogámico receptor que no presente elementos Mu activos. El ADN genómico de los organismos presentará por lo menos un elemento transponible y, preferentemente, una pluralidad de elementos transponibles tal como por lo menos 5, 10, 25, 50 o 100. Los elementos transponibles en el genoma pueden ser del mismo tipo o de distintos tipos. Los organismos que comprenden elementos transponibles se pueden obtener experimentalmente según métodos disponibles en la técnica. Véase, por ejemplo, Chomet (1994) en *The Maize Handbook* ("Manual del maíz"), ed. Freeling and Walbot (Springer-Verlag, Nueva York), p. 243-248. En una forma de realización preferida, el elemento transponible es Mutator (Mu). Robertson (1978) *Mutation Res.* 51:21-28, Chandler and Hardeman (1992) *Advances in Genetics* 30:77-122). El ADN con repeticiones invertidas terminales (TIR), presente en muchos de los elementos transponibles, entre ellos el Mu, resulta muy adecuado para la presente invención. La inserción de un elemento transponible se puede producir dentro o en la proximidad de una secuencia de ADN del gen etiquetado con el elemento transponible. El gen etiquetado con el elemento transponible a identificar con el método de la presente invención puede presentar un elemento transponible insertado en la secuencia de codificación del gen, de tal modo que se interrumpe la transcripción del producto funcional normal del gen, lo que produce un fenotipo mutante. Alternativamente, el gen etiquetado puede presentar un elemento transponible insertado en un intrón, de tal modo que se ve afectado el corte y empalme de ARN, lo que a su vez puede alterar el producto del gen funcional, produciendo de este modo un fenotipo mutante. Además, el gen etiquetado puede presentar un elemento transponible insertado dentro de una región de control del gen tal como un promotor o un elemento potenciador de tal modo que la expresión del gen se vea aumentada o disminuida originando un fenotipo mutante. Para cada fenotipo al que se aplica el método de la presente invención, se separan por lo menos un organismo que presenta un fenotipo de tipo silvestre y por lo menos uno de tipo mutante. Opcionalmente, por lo menos 2, 4, 5, 10, 15 o 20

organismos se encuentran presentes en la población de tipo silvestre separada y por lo menos 2, 4, 5, 10, 15 o 20 se encuentran presentes en la población mutante separada.

5 Se puede realizar la mezcla del ADN aislado utilizando por ejemplo un protocolo de mezcla tridimensional (Vandenbussche *et al.*, 2003, *The Plant Cell*, 15, 2680-2693). La mezcla se logra preferentemente utilizando cantidades equivalentes de ADN. El esquema de mezcla 3D puede comprender 10x10x10, obteniéndose 30 mezclas (10+10+10), que contienen 10 x 10 = 100 muestras de ADN distintas por grupo. Se pueden utilizar otras estrategias diversas de mezcla con la presente invención, siendo ejemplos de las mismas la mezcla multidimensional (comprendiendo la mezcla 3-D) o la mezcla en columna, fila o placa. En ciertas formas de realización, la mezcla se  
10 puede realizar asimismo antes de la extracción del ADN en la etapa de muestreo, reduciendo el número de preparaciones de ADN a 30 muestras en lugar de 1000 (etapa (a) del método).

15 La etapa de mezcla sirve normalmente para identificar el vegetal que contiene una inserción del transposón observada tras una ronda de cribado por PCR. La mezcla del ADN sirve además para normalizar los ADN antes de la amplificación por PCR a fin de proporcionar una representación más equivalente de las genotecas para la secuenciación. El ADN de las mezclas se restringe utilizando por lo menos una endonucleasa de restricción. Dependiendo del caso, es decir, del tamaño del genoma o del número de transposones, se pueden utilizar más endonucleasas. En ciertas formas de realización, se pueden utilizar 2 o más endonucleasas. En la mayoría de genomas, 2 endonucleasas resultan suficientes y, por lo tanto, es el modo más preferido. En ciertas formas de  
20 realización, en particular en genomas grandes o complejos, se pueden utilizar más endonucleasas. Preferentemente, la endonucleasa proporciona fragmentos de restricción relativamente cortos de aproximadamente 50 a 500 pb, pero ello no resulta esencial. Normalmente, se prefiere por lo menos una endonucleasa de corte frecuente, es decir, endonucleasas que presenten una secuencia de reconocimiento de 4 o 5 pares de bases. Uno de dichos enzimas es el MseI, pero se encuentran disponibles comercialmente y se puede utilizar muchos otros.  
25 Asimismo se pueden utilizar enzimas que cortan fuera de su secuencia de reconocimiento (tipo IIs) o enzimas que proporcionan fragmentos de restricción con extremos romos. Una combinación preferida utiliza un enzima de corte raro (secuencia de reconocimiento de 6 y más pares de bases) y uno de corte frecuente.

30 Tras la restricción de los ADN mezclados, o simultáneamente, los adaptadores se ligan a los fragmentos de restricción para proporcionar los fragmentos de restricción ligados a adaptadores. Se pueden utilizar uno o más adaptadores distintos, por ejemplo dos adaptadores, uno directo y uno inverso. Alternativamente, se puede utilizar un adaptador para todos los fragmentos o conjuntos de adaptadores que en el extremo saliente del adaptador comprendan permutaciones de nucleótidos que proporcionen conectores de indización que puedan permitir una etapa de preselección (Unrau *et al.*, *Gene*, 1994, 145, 163-169). Alternativamente, se pueden utilizar adaptadores de  
35 extremos romos, en el caso de fragmentos de restricción con extremos romos. La ligación con adaptadores resulta muy conocida en la técnica y se describe, entre otros, en el documento EP 534858. Tras la ligación con el adaptador, las mezclas de fragmentos de restricción ligados a adaptadores se pueden (pre)amplificar con un conjunto de cebadores que sean complementarios a los adaptadores. Ello puede servir para normalizar (más) la cantidad de ADN de cada vegetal en las mezclas o para aumentar la cantidad total de ADN en las mezclas a fin de  
40 permitir un análisis múltiple de las mezclas (es decir, fragmentar las muestras) y mejorar la relación entre la señal y el ruido.

45 Los fragmentos de restricción ligados al adaptador se amplifican, tras la preamplificación opcional, en la etapa (d) del método de la presente invención con un par de cebadores. Uno de los cebadores es complementario a por lo menos una parte del adaptador y puede ser además complementario a una parte del resto de la secuencia de reconocimiento de la endonucleasa y puede comprender además nucleótidos selectivos en su extremo 3' (seleccionados aleatoriamente), similares a los descritos en EP534858. El otro cebador del conjunto de cebadores se diseña para que pueda hibridarse con (una parte de) un borde de una secuencia de transposón. Normalmente, los cebadores se superponen a la secuencia de consenso del transposón y preferentemente al borde del mismo.  
50 Preferentemente, los cebadores pueden hibridarse selectivamente en unas condiciones restrictivas de hibridación con el elemento transponible o con el adaptador, respectivamente. Alternativamente, el cebador se puede superponer (es complementaria) al transposón en por lo menos el 50, 60, 70, 80, 85, 90, 95%. Con una longitud media de un cebador de aproximadamente 20 pb, ello equivale a una superposición de aproximadamente 10 a 19 bases. Ello puede ser una secuencia consenso o una secuencia conocida realmente de un transposón o familia de  
55 transposones de un organismo. Se conocen secuencias de transposones normales en plantas, véase por ejemplo: De Keukeleire *et al.* *Chromosome Research*, 2004, 12(2): 117-123; Van den Broeck *et al.*, *The Plant Journal*, 1998, 13(1), 121-129; Gerats *et al.*, *Plant Cell*, 1990, 2, 1121-1128 que describen el sistema de transposición dTph1 de 284 pb en las petunias. Dichas referencias demuestran que se conoce la secuencia de consenso de familias de transposones, en particular, en los bordes de los transposones. Teniendo en cuenta dicha secuencia de consenso,  
60 se puede realizar fácilmente el diseño de cebadores aptos. Por ejemplo, la familia Hat (Hobo, Ac y Tam3 en vegetales y animales. Se conocen elementos de transposones así como la secuencia a partir de las publicaciones siguientes: Atkinson PW, Warren WD, O'Brochta DA (1993) *The hobo transposable element of Drosophila can be cross-mobilized in houseflies and excises like the Ac element of maize* ("Se puede realizar la movilización cruzada del elemento transponible hobo de *Drosophila* en moscas domésticas y escindirse como el elemento Ac del maíz").  
65 *Proc Natl Acad Sci USA* 90: 9693-9697; Capy P, Vitalis R, Langin T, Higuier D, Bazin C (1996) *Relationships between*

5 *transposable elements based upon the integrase-transposase domains: is there a common ancestor?* ("Las relaciones entre los elementos transponibles basados en los dominios integrasa-transposasa: ¿existe un ancestro común?") *J Mol Evol* 42: 359-368; Esposito T, Gianfrancesco F, Ciccociola A *et al.* (1999) *A novel pseudoautosomal human gene encodes a putative protein similar to Ac-like transposases* ("Un nuevo gen pseudoautosómico humano codifica una presunta proteína similar a transposasas parecidas al Ac"). *Hum Mol Genet* 8: 61-67; Grappin P, Audeon C, Chupeau MC, Grandbastien MA (1996) *Molecular and functional characterization of Slide, an Ac-like autonomous transposable element from tobacco* ("Caracterización molecular y funcional del Slide, un elemento transponible autónomo del tabaco parecido al Ac"). *Mol Gen Genet* 252: 386-397; Handler AM, Gomez SP (1996) *The hobo transposable element excises and has related elements in tephritid species* ("El elemento transponible hobo se escinde y presenta elementos relacionados en especies de tefriditas."). *Genetics* 143: 1339-1347; Hehl R, Nacken WK, Krause A, Saedler H, Sommer H (1991) *Structural analysis of Tam3, a transposable element from Antirrhinum majus, reveals homologies to the Ac element from maize* ("El análisis estructural del Tam3, un elemento transponible de *Antirrhinum majus*, pone de manifiesto homología con el elemento Ac del maíz"); *Plant Mol Biol* 16: 369-371; Huttley GA, McRae AF, Clegg MT (1995) *Molecular evolution of the Ac/Ds transposable element family in pearl millet and other grasses* ("Evolución molecular de la familia del elemento transponible Ac/Ds en el mijo perla y otras gramíneas"). *Genetics* 139: 1411-1419; Kempken F, Windhofer F (2001) *The hAT family: a versatile transposon group common to plants, fungi, animals, and man* ("La familia Hat: un grupo versátil de transposones común a plantas, hongos, animales y el ser humano."). *Chromosoma* 110: 1-9. Warren WD, Atkinson PW, O'Brochta DA (1995) *The Australian bushfly Musca vetustissima contains a sequence related to transposons of the hobo, Ac and Tam3 family* ("La mosca australiana *Musca vetustissima* presenta una secuencia relacionada con transposones de la familia hobo, Ac y Tam3"). *Gene* 154: 133-134.

25 Preferentemente, el cebador dirigido a los transposones se orienta y de tal modo que se encara hacia el exterior del transposón diana. En una forma de realización destinada a mejorar la especificidad, uno o ambos cebadores, preferentemente el cebador dirigido a los transposones, puede comprender nucleótidos con una afinidad de enlace mejorada.

30 Una parte o segmento del fragmento de restricción ligado a adaptadores se amplifica utilizando un par de cebadores etiquetados, pudiendo estar marcados uno o ambos. Preferentemente, para cada mezcla de cada dimensión se utiliza un cebador distinto. En la ilustración anterior ello significa que se prefieren 30 cebadores directos y un solo cebador inverso. Uno de los cebadores directo e inverso se puede dirigir al adaptador y el otro cebador inverso y directo se puede dirigir al transposón diana. Preferentemente, cada par de cebadores (el cebador dirigido a adaptadores y el cebador dirigido a transposones) puede comprender además, de un modo dependiente, uno o más de los siguientes elementos:

- 35 (i) un sitio de enlace con un cebador de la secuencia que se puede utilizar en la etapa siguiente de secuenciación,  
 (ii) una etiqueta que sirve para correlacionar el cebador (y el producto de amplificación resultante) con el miembro original de la población, y  
 40 (iii) una secuencia de unión con perlas que permite la unión con la perla y que se utiliza en la etapa de secuenciación ultrarrápida.

45 En una forma de realización normal, el cebador dirigido a transposones puede presentar la estructura siguiente, tanto en la dirección 3'-5' como en la dirección 5'-3'

Sitio de enlace del cebador de la secuencia --- etiqueta opcional --- secuencia del cebador por PCR específica del transposón o  
Sitio de enlace de la perla --- etiqueta opcional --- secuencia del cebador por PCR específica del transposón

50 En una forma de realización normal, el cebador dirigido a adaptadores puede presentar la estructura siguiente, tanto en la dirección 3'-5' como en la dirección 5'-3'

Sitio de enlace del cebador de la secuencia --- etiqueta opcional --- secuencia del cebador por PCR específica del adaptador o  
 55 Sitio de enlace de la perla --- etiqueta opcional --- secuencia del cebador por PCR específica del adaptador

60 En ciertas formas de realización, tanto el cebador dirigido a transposones como el cebador dirigido a adaptadores pueden presentar entre 1 y 10 nucleótidos seleccionados aleatoriamente en el extremo 3' que, cuando se utilizan en la amplificación, pueden proporcionar para subconjuntos. Véase la figura 1.

65 La longitud del sitio de enlace con el cebador de la secuencia y la secuencia del cebador de PCR específica del transposón son las normales en la práctica de la PCR común, es decir, independientemente, de aproximadamente 10 a aproximadamente 30 pb prefiriéndose entre 15 a 25 pb. Preferentemente, la parte o segmento de la secuencia ligada a adaptadores que se amplifica corresponde a una longitud que se puede secuenciar en una serie utilizando técnicas de secuenciación ultrarrápida que se describirán a continuación. En ciertas formas de realización, la parte o

segmento presenta una longitud comprendida entre aproximadamente 50 pb y aproximadamente 500 pb, preferentemente entre aproximadamente 75 pb y aproximadamente 300 pb y más preferentemente entre aproximadamente 90 pb y aproximadamente 250 pb. Tal como se indicó anteriormente, dicha longitud puede variar según la técnica de secuenciación empleada comprendiendo las que todavía no se han desarrollado.

5 La amplificación con este conjunto de cebadores proporcionará fragmentos de restricción ligados a adaptadores amplificados (amplicones) de las secuencias flanqueadoras del transposón diana en múltiple.

10 Al utilizar cebadores (directos y/o inversos) que comprenden una secuencia etiquetada que es única para cada uno de los cebadores que representan todas las dimensiones de la mezcla, se conoce el origen de la mezcla específica de cada secuencia etiquetada ya que la secuencia del cebador se hibrida en dirección 5' de la etiqueta y como consecuencia de ello, la secuencia etiquetada se encuentra presente en cada producto de amplificación.

15 En ciertas formas de realización, se etiquetan ambos cebadores directo e inverso. En otras formas de realización, se etiqueta únicamente uno de los cebadores directo o inverso. La elección entre una o dos etiquetas depende de las circunstancias y depende de la longitud de lectura de la reacción de secuenciación ultrarrápida y/o de la necesidad de validación independiente. En el caso de, por ejemplo, los productos de la PCR de 100 pb que se secuencian en una dirección, únicamente se necesita una etiqueta. En el caso de un producto de la PCR de 200 pb y una longitud de lectura de 100 pb, el doble etiquetado es útil junto con una secuenciación bidireccional ya que se mejora la eficiencia 2 veces. Se proporciona además la posibilidad de una validación independiente en la misma etapa. Cuando un producto de la PCR de 100 pb se secuencian en dos direcciones con dos cebadores etiquetados, todas las trazas, independientemente de la orientación, proporcionarán información sobre la mutación. Por lo tanto, ambos cebadores proporcionan "información de la dirección" sobre qué planta contiene cierta mutación.

25 La etiqueta puede presentar cualquier número de nucleótidos, pero preferentemente comprende 2, 3, 4 o 5 nucleótidos. Con 4 nucleótidos permutados, son posibles 256 etiquetas, mientras que 3 nucleótidos permutados proporcionan 64 etiquetas distintas. En la ilustración utilizada, las etiquetas difieren preferentemente en > 1 base, por lo que se prefieren etiquetas de 4 pares de bases de longitud. La amplificación utilizando dichos cebadores produce una genoteca de productos de amplificación etiquetados.

30 En ciertas formas de realización, se puede utilizar un sistema de etiquetas en el que el proceso de amplificación comprenda la utilización de

35 (1) un cebador largo que comprende (a) una sección constante 5' enlazada a (b) una sección redundante de la etiqueta (NNNN), enlazada a (c) una sección 3' específica de transposones o adaptadores

40 (2) un cebador corto en amplificaciones posteriores que consiste en (a) la sección constante 5' enlazada a (b) una sección 3' no redundante etiquetada (es decir, una selección entre NNNN). El cebador largo se utiliza preferentemente en una medida corta y el cebador corto se utiliza en un exceso. La sección etiquetada no redundante puede ser única para cada muestra mezclada, por ejemplo, ACTG para la muestra mezclada 1, AATC para la muestra mezclada 2, etc. El cebador corto se hibrida con un subconjunto del cebador largo. La sección constante del cebador puede utilizarse como cebador de la secuencia. La genoteca comprende preferentemente cantidades equivalentes de productos de la PCR amplificados a partir de todas las mezclas. En el ejemplo ilustrativo, la genoteca comprende 1000 vegetales x 100 pb = secuencia de 100 kb a determinar para cada sitio de inserción de transposones. En la etapa (e) del método, se pueden mezclar los productos de amplificación, preferentemente en cantidades equivalentes o normalizadas para crear de este modo una genoteca de productos de amplificación. A título de ejemplo, la complejidad de la genoteca será de 1000 vegetales X 250-500 pb = secuencia de 0,25-0,5 Mb para cada sitio de inserción de transposones.

50 Los productos de amplificación de la genoteca se pueden fragmentar aleatoriamente antes de la secuenciación de los fragmentos. Se puede realizar la fragmentación mediante técnicas físicas, es decir, cizalladura, sonicación u otros métodos de fragmentación aleatoria. En la etapa (g), se determina por lo menos una parte de, pero preferentemente toda, la secuencia, de nucleótidos de por lo menos una parte de, pero preferentemente todos, los fragmentos de las etapas (d) o (f). En ciertas formas de realización, la etapa de fragmentación de los productos amplificados es opcional. Por ejemplo, cuando la longitud de lectura de la técnica de secuenciación y las longitudes de los fragmentos de la PCR son aproximadamente las mismas no es necesaria la fragmentación. Asimismo en el caso de los productos de la PCR más grandes, puede no ser necesaria la fragmentación de los productos amplificados si se acepta que únicamente una parte de los mismos están secuenciados. Por ejemplo, en caso de un producto de la PCR de 500 pb y de una longitud de lectura de 100 (desde cada lado) permanecen 300 pb sin secuenciar en caso de que no se realice la fragmentación antes de la secuenciación. La necesidad de fragmentación disminuye al aumentar la longitud de lectura de la técnica de secuenciación.

65 La secuenciación se puede realizar, en principio, mediante cualquier medio conocido en la técnica, tal como el método didesoxi de terminación de la cadena (secuenciación de Sanger). Sin embargo, se prefiere y es más ventajoso que la secuenciación se realice utilizando métodos de secuenciación ultrarrápida, tales como los métodos descritos en los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO

2004/070007 y WO 2005/003375 (todos a nombre de 454 Life Sciences), por Seo *et al.* (2004) *Proc.Natl.Acad.Sci.* USA 101:5488-93, y las técnicas de Helios, Solexa, US Genomics. Se prefiere más que la secuenciación se realice utilizando el aparato y/o el método que se dan a conocer en los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007 y WO 2005/003375 (todos a nombre de 454 Life Sciences). La técnica descrita permite actualmente la secuenciación de hasta 40 millones de bases en una sola serie y es 100 veces más rápida y más económica que la técnica de la competencia. Ello aumentará cuando aumente la longitud de lectura por reacción y/o aumente el número de reacciones en paralelo. La técnica de secuenciación comprende aproximadamente 5 etapas: La técnica de secuenciación comprende aproximadamente 5 etapas: 1) fragmentación del ADN y ligación del adaptador específico para crear una genoteca de ADN monocatenario (ADNss); 2) hibridación del ADNss con perlas, emulsificación de las perlas en microrreactores de agua en aceite y realización de la PCR en emulsión para amplificar las moléculas individuales de ADNss en las perlas; 3) selección / enriquecimiento de las perlas que contienen moléculas de ADNss amplificadas en la superficie de las mismas 4) deposición del ADN que presenta las perlas en una PicoTiterPlate®; y 5) secuenciación simultánea en 100.000 pocillos mediante la generación de una señal luminosa de pirofosfato.

En una forma de realización preferida, la secuenciación comprende las etapas de:

- (1) hibridar fragmentos ligados a adaptadores de secuenciación con perlas, hibridándose cada perla con un único fragmento;
- (2) emulsionar las perlas en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una única perla;
- (3) realizar la PCR en emulsión para amplificar fragmentos ligados con adaptadores en la superficie de las perlas
- (4) seleccionar / enriquecer las perlas que comprenden fragmentos amplificados ligados a adaptadores
- (6) cargar las perlas en pocillos, comprendiendo cada pocillo una única perla; y
- (7) generar una señal de pirofosfato.

En la primera etapa (1), los adaptadores que se encuentran presentes en los fragmentos de restricción ligados a adaptadores se hibridan con las perlas. Tal como se ha indicado anteriormente en la presente memoria, el adaptador de secuenciación comprende por lo menos una región "clave" para hibridarse con una perla, una región del cebador de la secuenciación y una región del cebador de la PCR. En particular, los fragmentos de restricción amplificados ligados a adaptadores comprenden ahora en uno de los extremos la secuencia siguiente 5'-sitio de enlace con el cebador de la secuencia --- etiqueta --- secuencia del cebador por PCR específica del transposón-3', mientras que en el otro extremo se encuentra presente un segmento que puede ser del modo siguiente: 5'-secuencia de hibridación con la perla --- etiqueta --- secuencia específica del adaptador --- secuencia específica del sitio de restricción (opcional) --- secuencia selectiva (aleatoriamente) (opcional)-3'. Puede resultar evidente que se pueden intercambiar el sitio de enlace del cebador de la secuencia y la secuencia de hibridación con la perla. Dicha secuencia de hibridación con la perla se puede utilizar ahora en la hibridación de los fragmentos con la perla, presentando la perla una secuencia de nucleótidos con ese fin.

De este modo, los fragmentos adaptados se hibridan con perlas, hibridándose cada perla con un único fragmento adaptado. Para la mezcla de fragmentos adaptados, se añaden suficientes perlas en exceso para garantizar la hibridación de un fragmento adaptado único por perla para la mayoría de las perlas (distribución de Poisson).

En una forma de realización preferida, para aumentar más la eficiencia del cribado de transposones, es beneficioso para amplificar direccionalmente el producto de la PCR obtenido transpuesto en la perla para la secuenciación. Ello se puede realizar para efectuar la PCR del transposón con cebadores de la PCR prolongados con adaptadores de las que una cadena del adaptador en el sitio MseI (u otro enzima de restricción) es complementario al oligonucleótido acoplado a las perlas de la secuencia. Por lo tanto la reacción de secuenciación se cebará desde el sitio del transposón (ya que la secuenciación se produce hacia la perla), con lo que se obtienen secuencias que se originan desde el transposón hacia fuera.

En una etapa siguiente, las perlas se emulsionan en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una única perla. Los reactivos de la PCR se encuentran presentes en los microrreactores de agua en aceite que permiten realizar una reacción de la PCR en los microrreactores. Posteriormente, los microrreactores se rompen y se enriquecen las perlas que comprenden ADN (perlas positivas de ADN).

En la etapa siguiente, las perlas se cargan en pocillos, comprendiendo cada pocillo una única perla. Los pocillos forman parte preferentemente de una placa PicoTiter™ que permite la secuenciación simultánea de una gran cantidad de fragmentos.

Tras la adición de perlas que presentan enzimas, se determina la secuencia de los fragmentos utilizando la pirosecuenciación. En etapas sucesivas, la placa PicoTiter™ y las perlas así como las perlas con enzimas en las mismas se someten a distintos desoxirribonucleótidos en presencia de reactivos de secuenciación convencionales y,

tras la incorporación de un desoxirribonucleótido, se genera una señal luminosa que se registra. La incorporación del nucleótido correcto generará una señal de pirosecuenciación que se puede detectar. La propia pirosecuenciación resulta conocida en la técnica y se describe, entre otros, en [www.biotagebio.com](http://www.biotagebio.com); [www.pyrosequencing.com / section technology](http://www.pyrosequencing.com/section%20technology). La tecnología se aplica asimismo en, por ejemplo en los documentos WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007 y WO 2005/003375 (todos a nombre de 454 Life Sciences).

Tras la secuenciación, las secuencias de los fragmentos que se obtienen directamente de la etapa de secuenciación se pueden recortar, preferentemente informáticamente, para eliminar cualquier secuencia de hibridación con la perla, el cebador de la secuenciación, el adaptador o información relacionada con el transposón. Ello puede tener como resultado una mejor hibridación en la etapa siguiente con secuencias conocidas procedentes de la base de datos para identificar las posibles coincidencias. Al realizarlo informáticamente, se puede conservar la información proporcionada por la etiqueta en un campo independiente de la base de datos a fin de conectar más tarde el gen mutado descubierto con la dirección en las mezclas de ADN.

Normalmente, la hibridación o la agrupación se realiza en los datos de la secuencia que se ha recortado con respecto a cualquier adaptador / cebador y/o secuencias de identificador añadido es decir, utilizando únicamente los datos de la secuencia procedentes de los fragmentos originados a partir de la muestra de ácido nucleico.

Los métodos de hibridación de secuencias para fines de comparación resultan muy conocidos en la técnica. Diversos programas y algoritmos de hibridación se describen en: Smith and Waterman (1981) *Adv. Appl.Math.* 2:482; Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443; Pearson and Lipman (1988) *Proc.Natl.Acad.Sci. USA* 85:2444; Higgins and Sharp (1988) *Gene* 73:237-244; Higgins and Sharp (1989) *CABIOS* 5:151-153; Corpet et al. (1988) *Nucl.Acids Res.* 16:10881-90; Huang et al. (1992) *Computer Appl. in the Biosci.* 8:155-65; and Pearson et al. (1994) *Meth.Mol.Biol.* 24:307-31. Altschul et al. (1994) *Nature Genet.* 6:119-29 presentan una consideración detallada de los métodos de hibridación de secuencias y cálculos de homología. La *NCBI Basic Local Alignment Search Tool* (BLAST) (Altschul et al., 1990) se encuentra disponible a partir de varias fuentes, entre ellas el Centro Nacional de Información Biológica (NCBI, Bethesda, Md.) y en Internet, para su utilización en relación con el análisis de la secuencia de los programas blastp, blastn, blastx, tblastn y tblastx. Se puede acceder en <http://www.ncbi.nlm.nih.gov/BLAST/>. Una descripción sobre cómo determinar la identidad de la secuencia con dicho programa se encuentra disponible en [http://www.ncbi.nlm.nih.gov/BLAST/blast help.html](http://www.ncbi.nlm.nih.gov/BLAST/blast%20help.html). La base de datos comprende preferentemente secuencias EST, secuencias genómicas de las especies de interés y/o la base de datos de secuencias no redundante de GenBank o bases de datos de secuencias similares. Se pueden utilizar métodos de secuenciación ultrarrápida tal como se describe en Shendure et al. *Science*, Vol. 309, n.º 5741, 1728-1732. Constituyen ejemplos de los mismos la secuenciación microelectroforética, la secuenciación por hibridación (SBH), la secuenciación en matriz cíclica en moléculas amplificadas, la secuenciación en matriz cíclica en moléculas simples, no cíclica, de una sola molécula, métodos en tiempo real, tales como por ejemplo, la secuenciación con la polimerasa, la secuenciación con la exonucleasa, secuenciación de nanoporos.

Para un resultado óptimo, interesa que los fragmentos o los productos amplificados se secuencien con una redundancia suficiente. La redundancia es lo que permite realizar la distinción entre los errores de la secuenciación y las secuencias auténticas del genoma. En ciertas formas de realización, se prefiere una redundancia de la secuenciación de por lo menos 4, más preferentemente por lo menos 5 pero, tal como se puede observar en la ilustración, las redundancias superiores a 6, preferentemente superiores 8 o incluso superiores a 10 se consideran ventajosas, aunque no resulta esencial para el concepto inventivo.

En la etapa (i) del método, se identifican los fragmentos que coincidieron en la base de datos y que, por lo tanto, pueden estar ligados a un gen o un fenotipo de interés. Basándose en esta información, se pueden utilizar etiquetas para identificar la mezcla y/o el vegetal. Basándose en la coincidencia con la base de datos, se puede diseñar una sonda de tal modo que permita la identificación del gen de interés.

#### Descripción de las figuras

**Figura 1:** en el análisis de la distribución de las secuencias flanqueadoras del transposón dtph1, se describe la construcción general de una etiqueta de secuencia, comprendiendo (de derecha a izquierda) una secuencia genómica única, la secuencia de transposón (repetición invertida) y la etiqueta 3D. La población de 100 plantas se organiza según una cuadrícula 3D (10 x 10 x 10) en la que cada planta se identifica según unas coordenadas 3D únicas (x, y, z) que reflejan su posición en los ejes x, y y z. Las dimensiones X1 a X10 corresponden a los números de etiqueta de secuencia 1 a 10, y de un modo similar en Y y Z. Los códigos de etiqueta en las figuras se corresponden a tag # en el nombre de la secuencia, por ejemplo, AGAC corresponde a tag07. La imagen muestra la coincidencia 3D en la planta con las coordenadas de la mezcla (3, 17, 24)

**La figura 2** describe el resultado de una búsqueda blast con una secuencia génica específica, el gen NAM-like 3 del factor de transcripción *Petunia* (sin meristemas apicales como [gjl21105733|gb|AF509866.1](http://www.ncbi.nlm.nih.gov/blast/)), contra la base de datos de secuencias flanqueadoras de la inserción; una coincidencia de inserción se identifica con las

coordinadas 2, 12, 30. Dicho resultado demuestra que se pueden trazar las inserciones en las secuencias codificantes homólogas específicas.

**La figura 3** describe el resultado de una búsqueda BLAST con una secuencia génica específica pero heteróloga, un gen de la secuencia AGL62 MADS de *Arabidopsis*, contra la base de datos; se identifica una coincidencia de inserción en la planta en 9, 17, 29; dicha coincidencia específica un potencial gen de la secuencia MADS desconocido hasta ahora en *Petunia* y su correspondiente mutante. Dicho resultado demuestra que se trazan satisfactoriamente las inserciones en las secuencias codificantes heterólogas específicas.

**La figura 4** proporciona un análisis de la secuencia en la que un subconjunto de 230.000 de las 318.000 secuencias disponibles se ha ordenado completamente según tres niveles:

- 1) Identidad de la secuencia de las secuencias flanqueadoras (ordenar según el sitio de inserción). Todas las secuencias que identifican la misma inserción se denominan un grupo.
- 2) Dentro de los grupos, según sus secuencias de etiquetas 3D distintas.
- 3) Según el número de copias de secuencias que pertenecen a un grupo.

Se han extrapolado los siguientes gráficos basándose en el análisis del 20% de 230.000 secuencias ordenadas (de un total de 318.000 secuencias). Para facilitar la interpretación de dichos gráficos, en esta figura se muestran 3 grupos de secuencias, que representan 3 sitios de inserción de transposones independientes. En el primer ejemplo se identifican cuatro secuencias, con las etiquetas 3D correspondientes que comprenden las posiciones 5 a 8, seguido por la secuencia de repetición invertida del transposón, que finaliza en la posición 22, seguido por un fragmento de secuencia genómica. Las coordenadas 6-20-29 definen esta secuencia como perteneciente a la planta en las coordenadas particulares de la población. Tag01 a tag10: Dimensión X, Tag11 a tag20: Dimensión Y, Tag21 a tag30: Dimensión Z.

**La figura 5** proporciona un gráfico de la distribución dimensional relativa con respecto a la frecuencia del número de copias.

**Figura 6:** de las 3.500 etiquetas de secuencias que presentaban tres copias, 294 tenían 3 coordenadas únicas, lo que significa que estas podrían trazar dichas secuencias con respecto a su planta de origen. En el caso de las otras clases de copia, dichos números eran 532 para la clase de cuatro copias; 622 para la clase de cinco copias; 478 para la clase de seis copias; y 1500 para las clases restantes. Ello implica que en total se han identificado más de 3000 etiquetas de secuencias (de 230.000 y 318.000 disponibles) que podrían estar relacionadas de nuevo con su planta de origen.

**Figura 7:** clases de número de copias 4 y su contribución relativa al número total estimado de coincidencias 3D y el número total de secuencias en 3D en la genoteca de transposones 454.

**Figura 8:** número de sitios de inserción # (grupos) con respecto al número de copias (intervalo completo). Un análisis del número de copias por etiqueta de secuencia demostró que entre los 230.000 del subconjunto analizado, aproximadamente 16.000 eran fragmentos únicos; 7.500 fragmentos presentaban dos copias; 3500 presentaban tres copias; 2.500 presentaban cuatro copias; 1500 presentaban cinco copias; 1.000 presentaban seis copias, 1.350 presentaban 7 u 8 copias, 1.100 presentaban de 9 a 11 copias; 1.400 presentaban de 12 a 20 copias, 950 presentaban de 21 a 40 copias, mientras que el resto presentaba las copias restantes.

**Figura 9:** proporciona un gráfico de algunos resultados. Del subconjunto de 253.394 secuencias analizadas (total 318.000), únicamente el 1% de las secuencias no comprendía una etiqueta reconocible (representada como???, columna derecha). Un análisis del 20% del subconjunto de 230.000 etiquetas de secuencia indica una buena distribución de las etiquetas de secuencia en las distintas muestras mezcladas de la población variando desde más de 6.000 para las coordenadas 23 hasta aproximadamente 30.000 para las coordenadas 15; la media es de aproximadamente 8500. Menos del 1% de los fragmentos no se pudo asignar a una coordenada específica.

**Figura 10:** representación esquemática de un fragmento de restricción EcoRI-MseI dirigido a un transposón utilizando un cebador dirigido a adaptadores y un cebador dirigido a transposones que presentan una etiqueta y una secuencia de hibridación de perlas.

**Figura 11:** representación esquemática de un fragmento amplificado ligado a adaptadores hibridado con una perla mediante una secuencia de hibridación con una perla (B). El fragmento comprende las etiquetas (T1 y/o T2), el adaptador (AD), restos finales del sitio de restricción (RE), la secuencia del propio fragmento (SEC), la secuencia del cebador específico del transposón (TR) y el sitio de enlace con el cebador de la secuencia (SPBS) utilizado en la iniciación de la etapa de secuenciación.

### Ejemplos

La presente invención se ilustra mediante los ejemplos siguientes que proporcionan la ilustración del principio. Se avanza en el cribado de una población de transposones utilizando nuevos métodos de secuenciación ultrarrápida, tales como 454 Life Sciences. Con el estado actual de la técnica, la tecnología de 454 Life Sciences produce aproximadamente hasta 40 Mb de secuencia en una sola serie de secuenciación. Una limitación en la actualidad es que las longitudes de lectura son aproximadamente de 100 a 200 pb / lectura. Suponiendo que el cribado de una población que comprende 3072 plantas que presentan una media de 200 transposones para identificar el etiquetado de transposones de un gen particular, el método es el siguiente:

- 1) se aísla el ADN genómico de 3072 plantas de la población de transposones;
- 2) se prepara un esquema de mezcla tridimensional de cantidades equivalentes de ADN por planta (por ejemplo, 15 x 15 x 14), obteniéndose 44 mezclas (15 + 15 + 14 = 44) que contienen  $3072/14 = 219$  o  $3072/15 = 205$  muestras de ADN distintas (Vandenbussche *et al.*, 2003);  
Esta etapa de mezcla sirve para poder identificar la planta individual que contiene una inserción directamente a partir de los datos de la secuencia. La mezcla del ADN genómico sirve además para normalizar los ADN antes de la amplificación por PCR a fin de aumentar las posibilidades de que todos los ADN estén representados por igual en la genoteca de secuencias;
- 3) se preparan plantillas de fragmentos de restricción ligados a adaptadores (plantillas AFLP, véase el documento EP534858, Vos *et al.*, NAR 1995, 23, 4407) a partir de los 44 ADN mezclados utilizando un enzima de restricción único que corta el genoma cada 250 - 500 pb (por ejemplo, utilizando un cortador 4 o 5; por ejemplo MseI);
- 4) se realiza la amplificación por PCR unidireccional utilizando un cebador de PCR dispuesto en el borde de la secuencia del transposón y orientado hacia el exterior, y un cebador adaptador no selectivo destinado a amplificar las secuencias flanqueadoras de todos los transposones en múltiplex. Para una planta que contiene 200 transposones, ello produce 200 x aproximadamente 250 pb = 50 kb que flanquean la secuencia por lado del borde, de las que 20 kb se secuenciarán en el caso de longitudes de lectura de 100 pb. En el caso de 3072 plantas ello equivale a una secuencia flanqueadora de 153 Mb de las que 61 Mb se pueden secuenciar en el caso de longitudes de lectura de 100 pb;
- 5) se mezclan cantidades iguales de los productos de la PCR de los 44 pocillos para crear una genoteca de productos de la PCR mezclados;
- 6) se secuencian la genoteca de productos de la PCR mezclados utilizando la técnica de secuenciación por síntesis de 454 Life Sciences sin más fraccionamiento de los productos de la PCR. La salida es aproximadamente de 200.000 secuencias de 100 pb, que representan el 0,33 X (20/61 Mb) de cobertura media de todas las secuencias flanqueadoras de 3072 plantas. Por lo tanto se necesitan por lo menos 3 series de secuenciación para identificar la gran mayoría de todas las secuencias flanqueadoras de todas las 3072 plantas;
- 7) se utiliza BLAST con las secuencias resultantes para identificar coincidencias con EST o secuencias del genoma;
- 8) las plantas portadoras de inserciones de transposones en los genes de interés se identifican basándose en sus etiquetas y, opcionalmente, se generan sondas o cebadores de la PCR para confirmarlo.

#### Ejemplo 1:

Se obtuvieron muestras de una población de 1000 plantas de *Petunia W138* según la estrategia tridimensional tal como describen Vandenbussche *et al.* (2003) y otros, obteniéndose 30 muestras mezcladas (X1 - X10, Y1 - Y10 y Z1 - Z10), lo que cubre todos los individuos de la población entera con tres coordenadas. Ello permitió trazar el origen de cualquier producto específico de la PCR para la planta de origen en la población.

Las muestras de ADN se digirieron con un enzima que cortó en el interior del transposón y un enzima que cortó en una posición específica pero aleatoria del ADN genómico flanqueador.

A continuación se ligaron adaptadores para permitir la amplificación posterior por PCR de todos los fragmentos digeridos. Un adaptador biotinilado se ligó al sitio del transposón interno.

Las muestras de ADN se purificaron a continuación y los fragmentos biotinilados se recogieron añadiendo perlas de estreptavidina y utilizando un imán. Todas las secuencias flanqueadoras de todas las inserciones de transposones presentes en cada mezcla de ADN se amplificaron a continuación utilizando un protocolo adaptado de visualización de transposones (VandenBroeck *et al.*, 1998).

Para cada muestra mezclada, X1 - X10, Y1 - Y10 y Z1 - Z10, en el ejemplo, se utilizó un cebador de transposones distinto, incorporando la coordenada de mezcla correspondiente como un código de 4 nucleótidos en su extremo 5' (3D-tag).

A continuación se mezclaron todos los productos de la PCR en tres supermezclas, una para cada dimensión, a fin de permitir la normalización de las muestras, según los procedimientos descritos en la técnica; con esta etapa, se disminuye el número de fragmentos que se encuentran presentes en cada individuo y, por lo tanto, en todas las muestras. Esto evita la representación excesiva de fragmentos en las muestras a secuenciar.

Los grupos obtenidos a partir de moléculas monocatenarias se convirtieron en moléculas bicatenarias mediante una sola serie de amplificación por PCR con un cebador específico, que presentaba un sitio MunI.

Los productos obtenidos se digirieron con MnlI/MseI para habilitar la ligación posterior de secuencias adaptadoras que permitiesen una amplificación adicional o una secuenciación directa 454 (G-20).

5 A continuación se mezclaron las tres muestras en una supermezcla y se sometieron al procedimiento de secuenciación GS20/454 de Roche tal como describe el fabricante.

Se desarrolló un protocolo para la amplificación de secuencias flanqueadoras de transposones mediante la visualización de transposones y la posterior secuenciación ultrarrápida a partir de una población de 1000 plantas.

10 Visión general del procedimiento

A continuación se proporciona una visión general del procedimiento:

- 15 - Preparación del ADN (1000 muestras de plantas obtenidas del tipo 3D, obteniéndose 30 mezclas de ADN)
- Digestión con MnlI/MseI (aproximadamente 5µg de ADN mezclado)
- Ligación del adaptador Bio-Mun y Mse
- Purificación (columnas de purificación de PCR, para retirar los adaptadores de bio-Mun y los fragmentos muy pequeños)
- 20 - Extracción de las perlas (enriquecimiento de fragmentos Mun/Mse)
- Amplificaciones por PCR para la visualización de transposones:
- Preamplificación con cebadores MunACAC y Mse+0 (enriquecimiento de secuencias flanqueadoras de transposones)
- PCR selectiva con cebadores específicos IR\*\*outw y Mse+0 mezclados (amplificación de secuencias flanqueadoras de transposones)
- 25 - Segunda agrupación en mezclas "bloque", "fila" y "columna"
- Normalización
- Conversión a moléculas bicatenarias
- Digestión MnlI/MseI
- Ligación de adaptadores 454-Mun-B y 454-Mse
- 30 - Amplificación por PCR con cebadores bio-AmpB y AmpA
- Mezcla final en una muestra
- Secuenciación 454

35 **Preparación del ADN**

1000 muestras de plantas del tipo 3D, obteniéndose 30 muestras de ADN que representan 100 plantas cada una; procedimiento según Vandebussche *et al.*, *Plant Cell* 15 (11): 2680-2693 (2003)

**Digestión MnlI/MseI** (aproximadamente 5µg) 30 muestras

- 40 Aprox. 5 µg de ADN en 50 µl H<sub>2</sub>O
- Añadir 20µl de la mezcla: 2 µl MnlI (10 U/µl reserva)
- 2 µl MseI (10 U/µl reserva)
- 7 µl NEB 4 (10x reserva)
- 0,7 µl BSA (100x reserva)
- 45 H<sub>2</sub>O hasta 20 µl

Incubación: 1,5 h a 37 °C

**Ligación de los adaptadores**

- 50 Añadir 30µl de mezcla
- 8 µl de adaptador bio-MunI (5 pmol/µl reserva)
- 8 µl de adaptador MseI (50 pmol/µl reserva)
- 3 µl NEB 4 (10x reserva)
- 0,3 µl BSA (100x reserva)
- 3 µl ATP (10 mM reserva)
- 55 3 µl ADN ligasa T4 (5 UWeiss/pl reserva)
- H<sub>2</sub>O hasta 30 µl

Incubación: 4 h a 37 °C

**Secuencias de los adaptadores:**

- 60 Adaptador (bio) Mun I: bio-5'-CTCGTAGACTGCGTACG-3'
- 3'-CTGACGCATGCTTAA-5'
- Adaptador Mse I: 5'-GACGATGAGTCCTGAG-3'
- 3'-TACTCA GGACTCAT-5'

65 **Purificación** 30 muestras

## ES 2 422 288 T3

Se purifican los ADN utilizando el kit de purificación de PCR Qiagen. Se eluye con 55 µl de disolución amortiguadora EB (5 µl en gel de agarosa al 1,5%).

### **Extracción de las perlas 30 muestras**

- 5 Se lavan las perlas con 25 µl de estreptavidina (aproximadamente 0,1 mg de perlas MyOne, estreptavidina C1) una vez en 200 µl de STEX y se vuelven a suspender en 100 µl de disolución amortiguadora de enlace.

<u>STEX:</u>	<u>Disolución amortiguadora de enlace:</u>
10 mM Tris.Cl (pH 8,0)	10 mM Tris.Cl (pH 8,0)
NaCl 1 M	NaCl 2 M
Ácido edético 1 mM	Ácido edético 1 mM
Triton X-100 al 0,1%	Triton X-100 al 0,1%

- 10 Se añaden 100 µl de perlas de estreptavidina diluidas (y lavadas) a 500 µl de mezcla de restricción / ligación y se procede a la incubación durante 60 minutos en un rotor, a temperatura ambiente. Se recogen las perlas, utilizando el imán, y se elimina el sobrenadante. Se lavan las perlas con 200 µl de STEX y se transfieren a otro tubo. Se lavan las perlas tres veces con 200 µl de STEX y por último se vuelven a suspender las perlas en 50 µl de T<sub>01</sub>E, y se transfieren a otro tubo (eliminando el STEX).

- 15 T<sub>01</sub>E:  
10 mM Tris.Cl (pH 8,0)  
Ácido edético 0,1 mM

### **Amplificaciones por PCR para la visualización de transposones: Preamplificación 30 muestras**

- 20 Se toman 2 µl de la plantillade ADN (mezclando bien las perlas, los fragmentos de ADN continúan conectados) y se añade:

- 25 18 µl de mezcla: 0,6 µl cebador Mun + ACAC (10 µM)  
0,6 µl cebador Mse + 0 (10 µM)  
0,8 µl dNTP (5 mM)  
2 µl disolución amortiguadora PCR10x  
2 µl MgCl<sub>2</sub> (25 mM)  
30 0,6 U ADN polimerasa Red Hot Taq  
H<sub>2</sub>O hasta 18 µl

y se incuban según el siguiente perfil de PCR (PE 9600):

- 35 30" 94 °C  
15" 94 °C

Gradiente de disminución de la temperatura:

- 40 30" 65 °C >> 56 °C (Δt = - 0,7 °C/ciclo) | 13 ciclos  
60" 72 °C |  
15" 94 °C |  
30" 56 °C | 22 ciclos  
60" 72 °C |

- 45 Secuencias de cebadores:

Mun I + ACAC: 5'-AGACTGTGTACGAATTGACAC-3'  
Mse I + 0: 5'-GACGATGAGTCCTGAGTAA-3'

- 50 Se analizan 5µl en gel de agarosa al 1,5% y se diluyen las muestras 10 veces con H<sub>2</sub>O y se realiza una amplificación selectiva por PCR:

### **Amplificaciones por PCR para la visualización de transposones:**

- 55 **Amplificación selectiva 30 muestras**

- 60 Se toman 5 µl de la plantilla de ADN y se añaden: 45 µl de la mezcla: 1,5 µl cebador IR<sub>out</sub>W (10 µM)\*  
1,5 µl cebador Mse + 0 (10 µM)  
2 µl dNTP (5 mM)

5 µl disolución amortiguadora PCR10x  
 5 µl MgCl<sub>2</sub> (25 mM)  
 1 U ADN polimerasa Red Hot Taq  
 H<sub>2</sub>O hasta 45 µl

5 Y se incuban según el siguiente perfil de PCR (PE 9600):

30" 94 °C  
 15" 94 °C

10

Gradiente de disminución de la temperatura:

30" 65 °C >> 56 °C (Δt = - 0,7 °C/ciclo) | 13 ciclos

15

60" 72 °C |  
 15" 94 °C |  
 30" 56 °C | 22 ciclos  
 60" 72 °C |

20

Secuencias de cebadores:

IR<sub>outw</sub>\* 5'-CATATATTAANNNGTAGCTCCGCCCTG-3'

se amplifica cada muestra mezclada con un cebador IR<sub>outw</sub> único, especificado por las posiciones NNNN, lo que permite asignar las secuencias obtenidas a sus coordenadas de origen.

25

Mse I + 0: 5'-GACGATGAGTCCTGAGTAA-3'

**Segunda mezcla 30 muestras en 3 muestras**

30

Se mezclan los productos de la PCR de las diez muestras de cada dimensión para crear 3 muestras: columna / fila / bloque

**Normalización.**

35

Para mejorar la cantidad de fragmentos únicos contra un fondo de fragmentos compartidos por muchos o todos los individuos, se normalizan las segundas muestras mezcladas, basándose en los procedimientos conocidos convencionalmente. El procedimiento implica una hibridación y una etapa de purificación para obtener moléculas monocatenarias.

40

**Hibridación** (aproximadamente: 10 µg cada muestra) 3 muestras

Se precipitan los ADN de las muestras mezcladas y se disuelven en 15 – 35µl  
 Se añaden (volúmenes relativos) a 15 µl de formamida:

45

4,5 µl TE  
 3µl H<sub>2</sub>O

Se calienta hasta 80 °C en aceite mineral durante 3 minutos  
 Se añaden

50

3 µl de disolución amortiguadora A  
 4,5 µl H<sub>2</sub>O

Se incuban la sonda durante toda la noche a 30 °C

55

Disolución amortiguadora A:

0,1 M Tris.Cl (pH 8,0)  
 NaCl 1,2 M  
 Ácido edético 50 mM

60

**Purificación por cromatografía HAP 3 muestras**

Se seleccionan moléculas de cadena sencilla por cromatografía HAP estándar tal como describe Fatima Bonaldo *et al.*, *Genome Research*, 6: 791-806 (1996) y posteriormente se convierten en moléculas bicatenarias.

65

**Conversión a moléculas bicatenarias 3 muestras**

Un ciclo de PCR con cebador "Mse+0 con sitio Mun"  
añadir a 50 µl de muestra:

5  
25 µl mezcla: 5 µl MIBUS 796 (10 µM)  
4 µl dNTP (5 mM)  
7,5 µl disolución amortiguadora PCR10x  
10 2,5 µl MgCl<sub>2</sub> (50 mM)  
0,2 µl de ADN polimerasa PlatinumTaq  
H<sub>2</sub>O hasta 25 µl

**Secuencias de cebadores:**

15 MIBUS 796: 5'-CATATACAATTGGACGATGAGTCCTGAGTAA-3'

Y se incuban según el siguiente perfil (PE 9600):

20  
2' 94 °C  
1' 56 °C  
10' 72 °C

**Digestión Munl/Mse 3 muestras**

25 Plantilla de ADN en 65µl de H<sub>2</sub>O

Se añaden 25µl mezcla: 2 µl Munl (10 U/µl reserva)  
2 µl Msel (10 U/µl reserva)  
30 9 µl NEB 4 (10x reserva)  
0,9 µl BSA (100x reserva)  
H<sub>2</sub>O hasta 25 µl

35 Incubación: 1,5 h a 37 °C

**Ligación del adaptador 454**

Añadir 4 µl de adaptador B bio-Munl (50 pmol/µl reserva)  
4 µl de adaptador A - Msel (50 pmol/µl reserva)  
40 2 µl NEB 4 (10x reserva)  
0,2 µl BSA (100x reserva)  
3 µl ATP (10 mM reserva)  
3 µl ADN ligasa T4 (5 UWeiss/µl reserva)  
45 H<sub>2</sub>O hasta 20 µl

Incubación: 4 h a 37 °C

**Secuencias de los adaptadores:**

50 Adaptador B Mun I:

MIBUS 803  
5'- CCTATCCCCTGTGTGCCTTGCCTATCCCCTGTTGCGTGTCTCAG -3'  
MIBUS795  
55 3'- AGGGGACACACGGAACGGATAGGGGACAACGCACAGAGTCTTAA -5'

Adaptador A Mse I:

MIBUS 800  
60 5'- CCATCTCATCCCTGCGTGTCCCATCTGTTCCCTCCCTGTCTCAG -3'  
MIBUS 801  
3'-GAGTAGGGACGCACAGGGTAGACAAGGGAGGGACAGAGTCAT-5'

**Amplificación por PCR para la secuenciación 454 3 muestras**

65

Cebador adaptador amplificador A y B:

MIBUS 803 bio-5'- CCTATCCCCTGTGTGCCTTG -3'  
 MIBUS 802 5'- CCATCTCATCCCTGCGTGTC -3'

5

**Mezcla final 3 muestras en 1 muestra**

Se mezclan las muestras para crear 1 supermezcla, lista para la secuenciación ultrarrápida

10

**Secuenciación 454 1 muestra**

**Clonación pGEM-T para la prueba de distribución del tamaño de inserción 1 muestra**

15

Para probar la eficacia del procedimiento de normalización, se aislaron aleatoriamente 22 fragmentos a fin de determinar su distribución de tamaños. Se toma 1 µl de mezcla de la PCR (de la muestra de la supermezcla para la secuenciación 454)

20

Se añaden 4µl mezcla: 1 µl pGEM-T (diluido 4 veces)  
 2,5 µl disolución amortiguadora de ligación 2 x rápida  
 0,25 µl Ligasa  
 H<sub>2</sub>O hasta 4 µl

25

Incubación: 3 h a 37 °C  
 Transformación en *E. coli* (células DH5α)  
 Sembrar 100µl en placas LB- Amp  
 Incubar: durante toda la noche a 37 °C  
 Seleccionar 22 colonias  
 Realizar una PCR en preparaciones hervidas  
 con los cebadores AmpA/AmpB

30

y una serie en gel de agarosa al 2%

35

**Resultados:**

Se obtuvo una base de datos de 318.000 etiquetas de secuencia de una media de 102 pares de bases. Un subconjunto de 230.000 secuencias se ordenó completamente según tres niveles:

40

- 1) Identidad de secuencia de la secuencia, flanqueando la repetición invertida del transposón (que finaliza con CCGCCCCTG). Todas las secuencias que identifican la misma inserción se denominan un grupo.
- 2) Dentro de cada grupo, las secuencias se ordenan según las distintas etiquetas 3D en su secuencia 5'.
- 3) Según el número de copias de secuencias que pertenecen a un grupo.

45

Los datos se han extrapolado basándose en el análisis del 20% de 230.000 secuencias ordenadas (de un total de 318.000 secuencias). El análisis se describe en las figuras 1-9.

TABLA	secuencia	SEC ID #
Adaptador (bio) Mun I	bio-5'-CTCGTAGACTGCGTACG-3'	1
	3'-CTGACGCATGCTTAA-5'	2
Adaptador Mse I	5'-GACGATGAGTCCTGAG-3'	3
	3'-TACTCAGGACTCAT-5'	4
Secuencias de cebadores		
Mun I + ACAC:	5'-AGACTGTGTACGAATTGACAC-3'	5
Mse I + 0:	5'-GACGATGAGTCCTGAGTAA-3'	6
IR <sub>outv</sub> : *	5'-CATATATTAANNNGTAGCTCCGCCCTG-3'	7
Mse I + 0:	5'-GACGATGAGTCCTGAGTAA-3'	8
MIBUS 796:	5'-CATATACAATTGGACGATGAGTCCTGAGTAA-3'	9

ES 2 422 288 T3

Secuencias de los adaptadores		
Adaptador B Mun I: MIBUS 803	5'-CCTATCCCCTGTGTGCCTTGCCTATCCCCTGTTGCGTGTC TCAG-3'	10
MIBUS795	3'-AGGGGACACACGGAACGGATAGGGGACAACGCACAGAGTC TTAA -5	11 5
Adaptador A Mse I: MIBUS 800	5'-CCATCTCATCCCTGCGTGTCCCATCTGTTCCCTCCCTGTC TCAG-3'	12
MIBUS 801	3'-GAGTAGGGACGCACAGGGTAGACAAGGGAGGGACAGAGTC AT-5'	13 10
Cebador adaptador amplificador A y B:		
MIBUS 803	bio-5'- CCTATCCCCTGTGTGCCTTG -3	14 15
MIBUS 802	5'- CCATCTCATCCCTGCGTGTC -3'	15

LISTADO DE LA SECUENCIA

20 <110> Keygene NV

<120> Método para el cribado ultrarrápido de poblaciones de transposón etiquetado e identificación masiva de la secuencia en paralelo de los sitios de inserción

25 <130> P27927PC00

<150> USA60/735,878

<151>2005-11-14

30 <160> 15

<170> PatentIn version 3.3

35 <210> 1  
<211> 17  
<212> ADN  
<213> artificial

40 <220>  
<223> adaptador o cebador

<400> 1  
ctcgtagact gcgtacg 17

45 <210> 2  
<211> 15  
<212> ADN  
<213> artificial

50 <220>  
<223> adaptador o cebador

<400> 2  
ctgacgcatg cttaa 15

55 <210> 3  
<211> 16  
<212> ADN  
<213> artificial

60 <220>  
<223> adaptador o cebador

65 <400> 3

	gacgatgagt cctgag	16
	<210> 4	
	<211> 14	
5	<212> ADN	
	<213> artificial	
	<220>	
	<223> adaptador o cebador	
10	<400> 4 14	
	tactcaggac tcat	14
	<210> 5	
	<211> 21	
15	<212> ADN	
	<213> artificial	
	<220>	
20	<223> adaptador o cebador	
	<400> 5	
	agactgtgta cgaattgaca c	21
25	<210> 6	
	<211> 19	
	<212> ADN	
	<213> artificial	
30	<220>	
	<223> adaptador o cebador	
	<400> 6	
35	gacgatgagt cctgagtaa	19
	<210> 7	
	<211> 29	
	<212> ADN	
	<213> artificial	
40	<220>	
	<223> adaptador o cebador	
	<220>	
45	<221> misc_feature	
	<222> (11)..(14)	
	<223> N= A, C, T o G	
	<400> 7	
50	catatattaa nnnngtagct cgcgccctg	29
	<210> 8	
	<211> 19	
	<212> ADN	
55	<213> artificial	
	<220>	
	<223> adaptador o cebador	
60	<400> 8	
	gacgatgagt cctgagtaa	19
	<210> 9	
	<211> 31	
65	<212> ADN	

<213> artificial  
 <220>  
 <223> adaptador o cebador  
 5 <400> 9  
 catatacaat tggacgatga gtctcgagta a 31  
 <210> 10  
 10 <211> 44  
 <212> ADN  
 <213> artificial  
 <220>  
 15 <223> adaptador o cebador  
 <400> 10  
 cctatcccct gtgtgccttg cctatcccct gttgcgtgtc tcag 44  
 20 <210> 11  
 <211> 44  
 <212> ADN  
 <213> artificial  
 25 <220>  
 <223> adaptador o cebador  
 <400> 11  
 30 aggggacaca cggaacggat aggggacaac gcacagagtc ttaa 44  
 <210> 12  
 <211> 44  
 <212> ADN  
 <213> artificial  
 35 <220>  
 <223> adaptador o cebador  
 <400> 12  
 40 ccatctcatc cctgcgtgtc ccatctgttc cctccctgtc tcag 44  
 <210> 13  
 <211> 42  
 <212>. ADN  
 45 <213> artificial  
 <220>  
 <223> adaptador o cebador  
 50 <400> 13  
 gagtagggac gcacaggga gacaaggag ggacagagtc at 42  
 <210> 14  
 <211> 20  
 55 <212> ADN  
 <213> artificial  
 <220>  
 <223> adaptador o cebador  
 60 <400> 14  
 cctatcccct gtgtgccttg 20  
 <210> 15  
 65 <211> 20

<212> ADN  
 <213> artificial

5 <220>  
 <223> adaptador o cebador

<400> 15  
 ccatctcatc cctgcgtgtc 20

10 LISTADO DE LA SECUENCIA

< 110> Keygene NV  
 <120> Método para el cribado ultrarrápido de poblaciones de transposones e identificación masiva de la secuencia en paralelo de los sitios de inserción

15 <130> P27927PC00  
 <150> US60/735,878

20 <151> 2005-11-14  
 <160> 15  
 <170> PatentIn version 3.3

25 <210> 1  
 <211> 17  
 <212> ADN  
 <213> artificial

30 <220>  
 <223> adaptador o cebador

35 <400> 1  
 ctcgtagact gcgtacg 17

40 <210> 2  
 <211> 15  
 <212> ADN  
 <213> artificial

<220>  
 <223> adaptador o cebador

45 <400> 2  
 ctgacgcatg ctaa 15

50 <210> 3  
 <211> 16  
 <212> ADN  
 <213> artificial

<220>  
 <223> adaptador o cebador

55 <400> 3  
 gacgatgagt cctgag 16

60 <210> 4  
 <211> 14  
 <212> ADN  
 <213> artificial

65 <220>  
 <223> adaptador o cebador

<400> 4  
 tactcaggac tcat 14

5 <210> 5  
 <211> 21  
 <212> ADN  
 <213> artificial

10 <220>  
 <223> adaptador o cebador

<400> 5  
 agactgtgta cgaattgaca c 21

15 <210> 6  
 <211> 19  
 <212> ADN  
 <213> artificial

20 <220>  
 <223> adaptador o cebador

<400> 6  
 gacgatgagt cctgagtaa 19

25 <210> 7  
 <211> 29  
 <212> ADN  
 <213> artificial

30 <220>  
 <223> adaptador o cebador

35 <220>  
 <221> misc\_feature  
 <222> (11)..(14)  
 <223> N= A, C, T o G

<400> 7  
 catatattaa nnnngtagct cgcgccctg 29

40 <210> 8  
 <211> 19  
 <212> ADN  
 <213> artificial

45 <220>  
 <223> adaptador o cebador

50 <400> 8  
 gacgatgagt cctgagtaa 19

55 <210> 9  
 <211> 31  
 <212> ADN  
 <213> artificial

60 <220>  
 <223> adaptador o cebador

<400> 9  
 catatacaat tggacgatga gtctctgagta a 31

65 <210> 10

<211> 44  
 <212> ADN  
 <213> artificial  
 5 <220>  
 <223> adaptador o cebador  
 <400> 10  
 cctatcccct gtgtgccttg cctatcccct gttgcgtgtc tcag 44  
 10 <210> 11  
 <211> 44  
 <212> ADN  
 <213> artificial  
 15 <220>  
 <223> adaptador o cebador  
 <400> 11  
 20 aggggacaca cggaacggat aggggacaac gcacagagtc ttaa 44  
 <210> 12  
 <211> 44  
 <212> ADN  
 25 <213> artificial  
 <220>  
 <223> adaptador o cebador  
 30 <400> 12  
 ccatctcatc cctgcgtgtc ccatctgttc cctccctgtc tcag 44  
 <210> 13  
 <211> 42  
 35 <212> ADN  
 <213> artificial  
 <220>  
 <223> adaptador o cebador  
 40 <400> 13  
 gagtagggac gcacagggta gacaaggag ggacagagtc at 42  
 <210> 14  
 <211> 20  
 <212> ADN  
 <213> artificial  
 <220>  
 50 <223> adaptador o cebador  
 <400> 14  
 cctatcccct gtgtgccttg 20  
 55 <210> 15  
 <211> 20  
 <212> ADN  
 <213> artificial  
 60 <220>  
 <223> adaptador o cebador  
 <400> 15  
 65 ccatctcatc cctgcgtgtc 20

## REIVINDICACIONES

- 5 1. Método para la identificación de una inserción asociada a un gen o fenotipo de interés en un miembro de una población de transposones, que comprende las etapas de:
- 10 (a) aislar, individualmente o en mezclas, el ADN genómico de miembros de la población de transposones;
- (b) realizar la restricción del ADN utilizando una o más endonucleasas de restricción, realizar la ligación de los adaptadores con los fragmentos de restricción para preparar de este modo los fragmentos de restricción ligados a adaptadores;
- 15 (c) realizar la amplificación de los fragmentos de restricción ligados a adaptadores con un par de cebadores, con lo que uno de los cebadores comprende una sección que es complementaria a parte de una secuencia de transposón conocida, siendo el otro cebador por lo menos complementario al adaptador, comprendiendo uno o ambos cebadores una etiqueta para correlacionar el cebador y el producto de amplificación con el miembro original de la población;
- (d) determinar por lo menos parte de la secuencia de nucleótidos de por lo menos parte de los fragmentos de (c) utilizando la secuenciación ultrarrápida;
- 20 (e) identificar uno o más fragmentos de la etapa (d) que se puedan alinear con las secuencias de nucleótidos de una base de datos, correlacionando de este modo los fragmentos identificados con un gen o fenotipo de interés;
- (f) identificar el/los miembro(s) de la población de transposones que comprenden el/los fragmento(s) de la etapa (e);
- 25 2. Método según la reivindicación 1, en el que en la etapa (a) el ADN obtenido se mezcla posteriormente.
3. Método según cualquiera de las reivindicaciones 1 o 2, en el que en la etapa (c) los productos amplificados se mezclan posteriormente para crear una genoteca de productos de amplificación.
- 30 4. Método según la reivindicación 3, en el que la genoteca de productos de amplificación de la etapa (c) se fragmenta posteriormente.
5. Método según cualquiera de las reivindicaciones 1 a 4, en el que en la etapa (d) la secuencia de los fragmentos se recorta informáticamente para eliminar de este modo cualquier adaptador e/o información de la secuencia relacionada con el transposón.
- 35 6. Método según cualquiera de las reivindicaciones 1 a 5, en el que en la etapa (b) se utilizan dos o más, preferentemente dos, endonucleasas de restricción.
- 40 7. Método según la reivindicación 6, en el que en la etapa (b) preferentemente por los menos una de las endonucleasas de restricción es una endonucleasa de restricción de corte frecuente que no corta en transposones y por lo menos una de las endonucleasas de restricción es una endonucleasa de restricción de corte raro que corta en transposones.
- 45 8. Método según cualquiera de las reivindicaciones 1 a 7, en el que uno de los cebadores de la etapa (c) comprende además un sitio de enlace con un cebador de la secuencia.
9. Método según cualquiera de las reivindicaciones 1 a 8, en el que se etiquetan los cebadores de la etapa (c).
- 50 10. Método según cualquiera de las reivindicaciones 1 a 9, que comprende la etapa adicional (g) de diseñar una sonda o par de cebadores de PCR basándose en los fragmentos de la etapa (e) y utilizarla para confirmar la inserción del transposón en el gen de interés del genoma del miembro identificado en (f).
- 55 11. Método según cualquiera de las reivindicaciones 1 a 10, en el que la mezcla comprende una estrategia de mezcla 3D.
12. Método según cualquiera de las reivindicaciones 1 a 11, en el que la base de datos comprende secuencias EST, secuencias genómicas de las especies de interés y/o la base de datos de secuencias no redundantes de GenBank o bases de datos de secuencias similares.
- 60 13. Método según cualquiera de las reivindicaciones 1 a 12, en el que la secuenciación ultrarrápida es una secuenciación por síntesis, preferentemente una pirosecuenciación.
- 65 14. Método según cualquiera de las reivindicaciones 1 a 13, en el que la secuenciación se realiza sobre un soporte sólido tal como una perla.

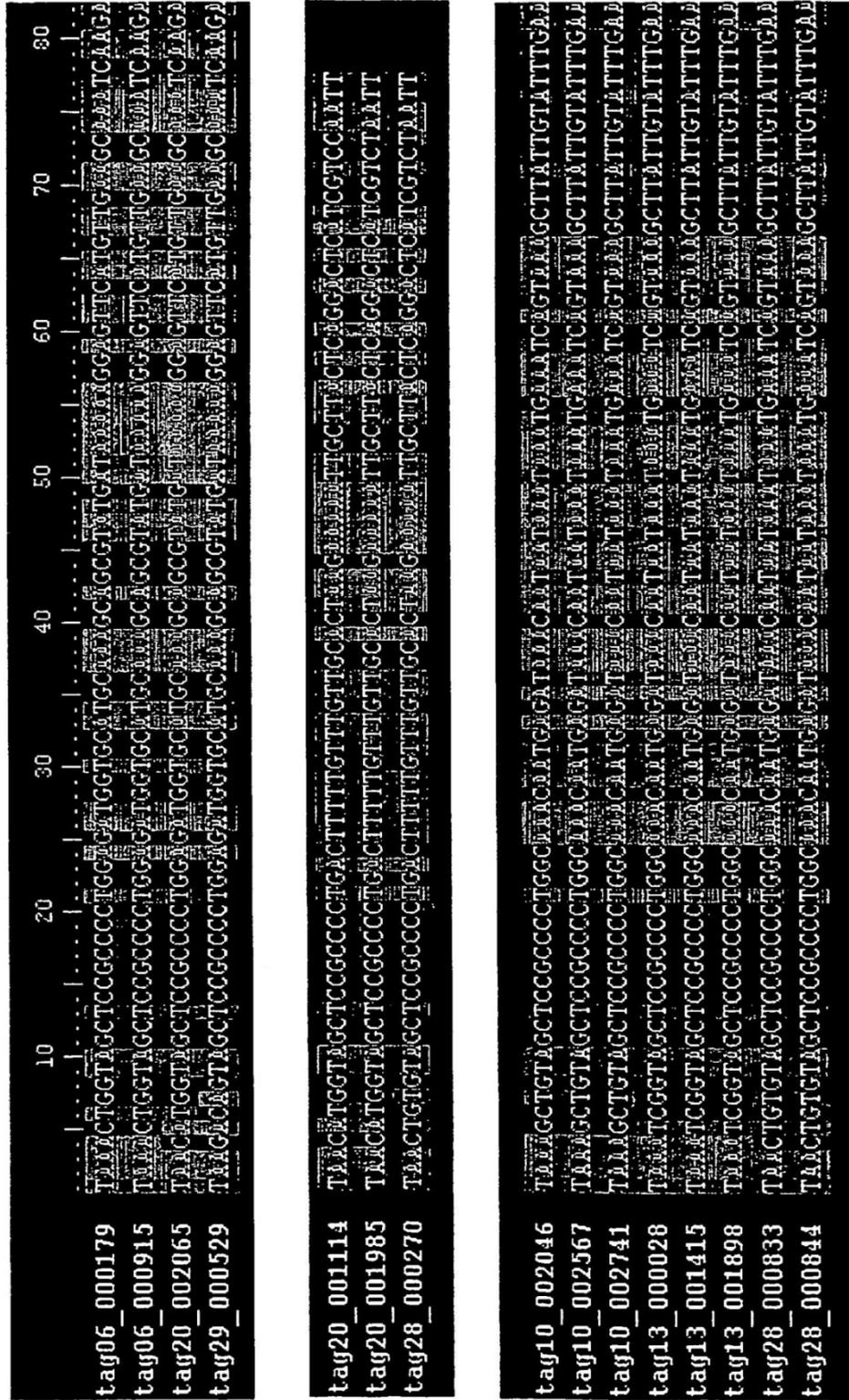
15. Método según la reivindicación 14, en el que la secuenciación comprende las etapas de:
- 5 - hibridar fragmentos ligados a adaptadores de secuenciación con perlas, hibridándose cada perla con un único fragmento;
  - emulsionar las perlas en microrreactores de agua en aceite, comprendiendo cada microrreactor de agua en aceite una única perla;
  - realizar la PCR en emulsión para amplificar fragmentos ligados con adaptadores en la superficie de las perlas
  - 10 - seleccionar / enriquecer las perlas que comprenden fragmentos amplificados ligados a adaptadores
  - cargar las perlas en pocillos, comprendiendo cada pocillo una única perla; y
  - generar una señal de pirofosfato.
- 15 16. Método según cualquiera de las reivindicaciones 1 a 15, en el que por lo menos uno de los cebadores comprende uno o más nucleótidos con una afinidad de enlace mejorada.





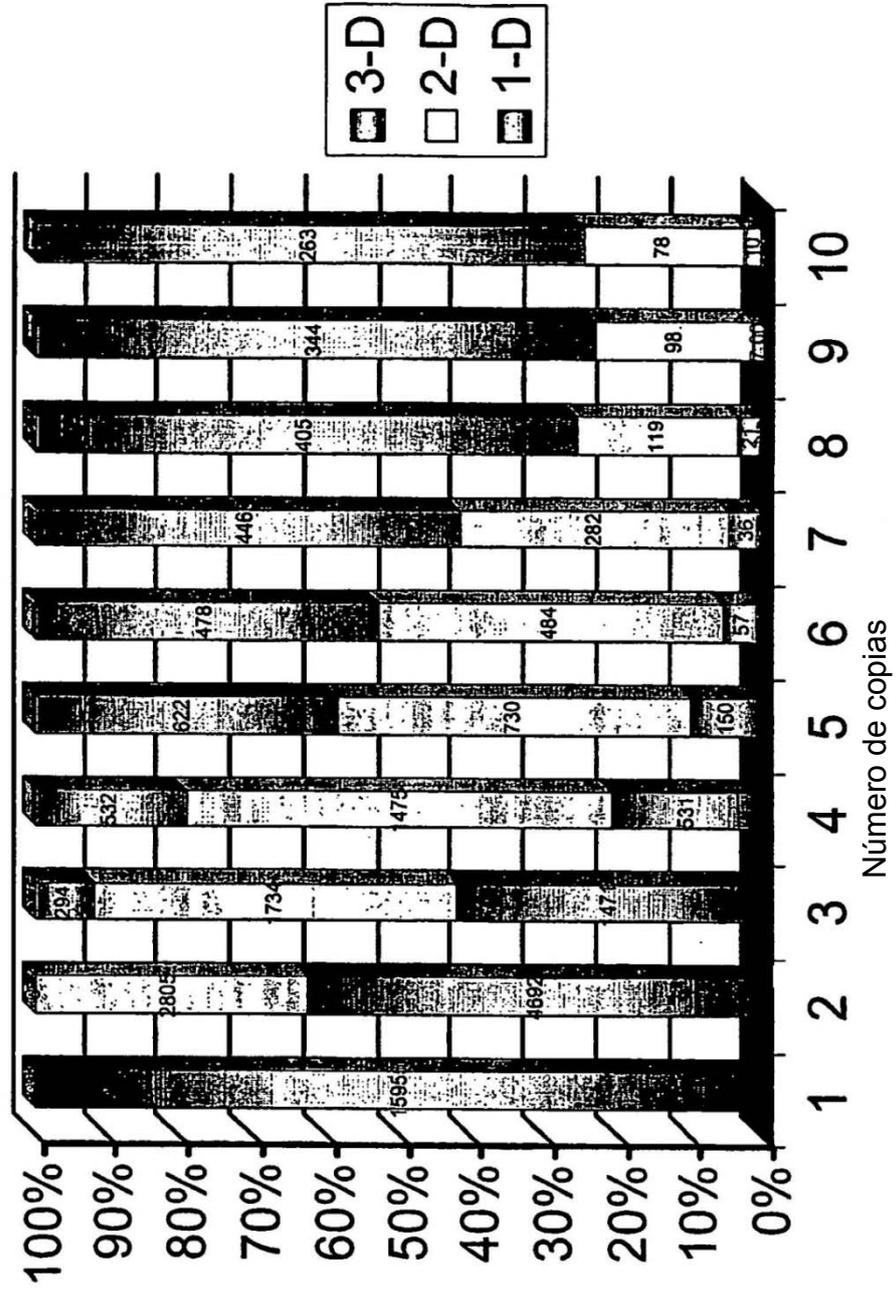


FIG. 4



**FIG. 5**

Distribución dimensional (relativa) con respecto a la frecuencia del número de copias (1 a 10)



**FIG. 6**

Distribución dimensional de coincidencias 3D (absoluta) con respecto al número de copias (3 a 10)

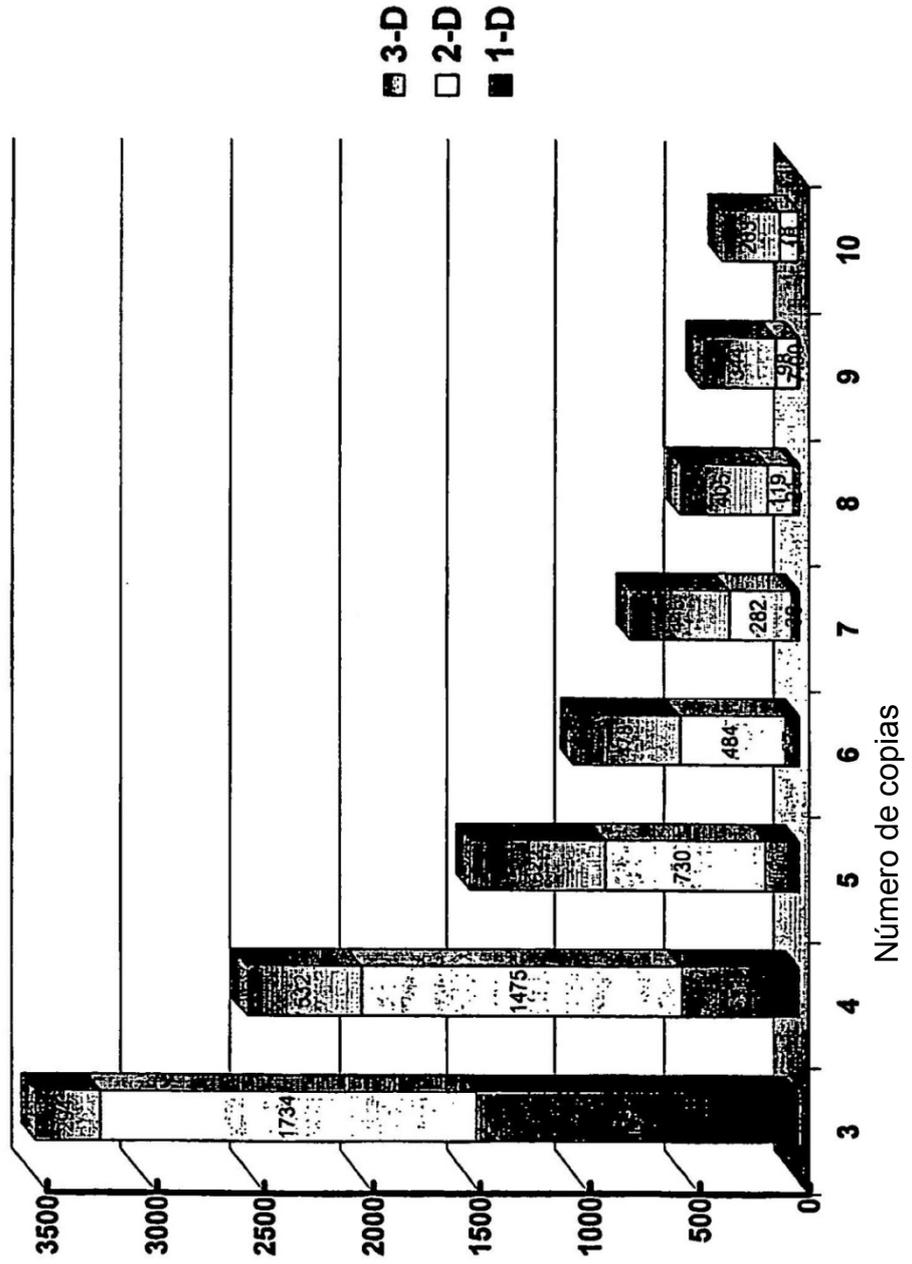
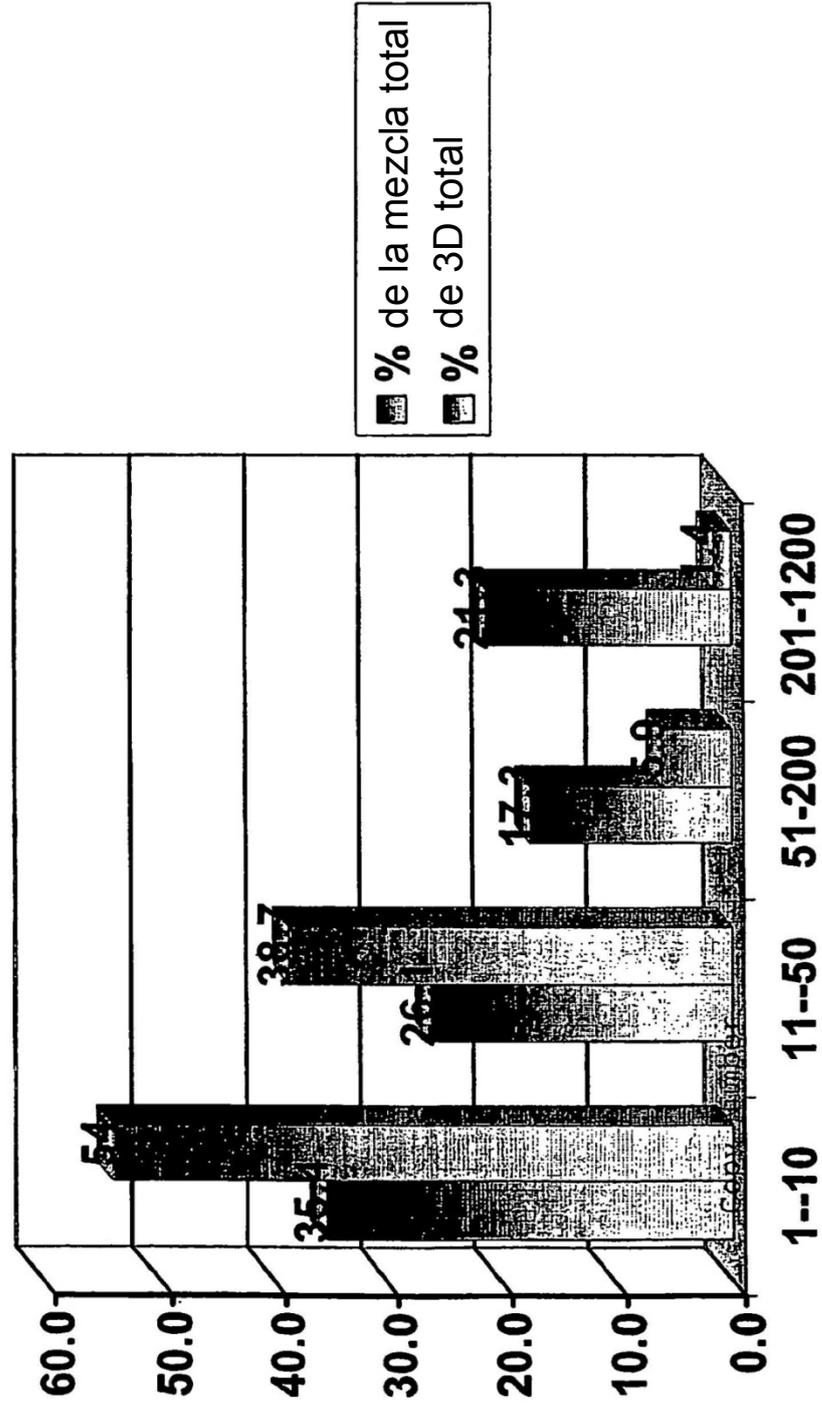
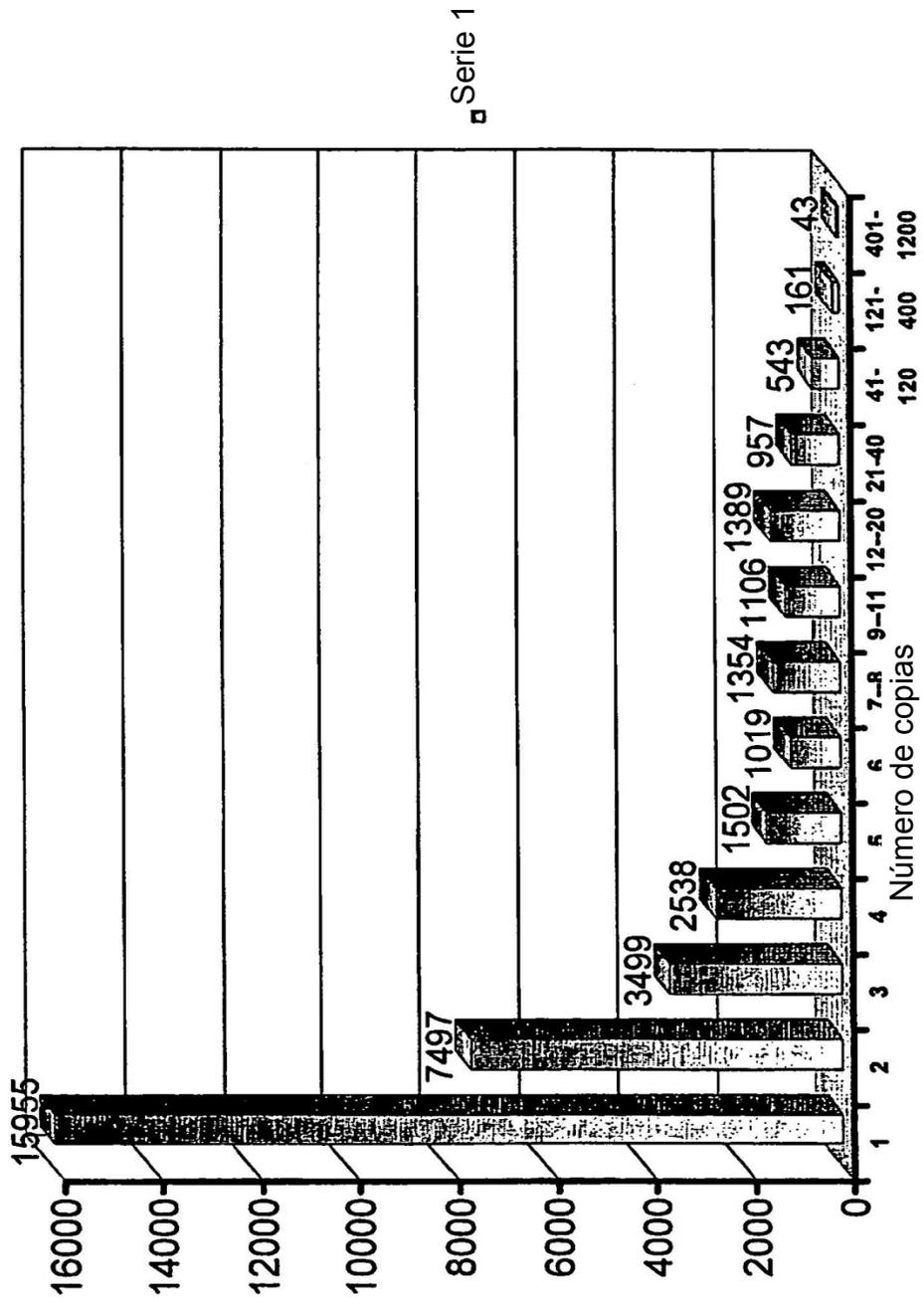


FIG. 7



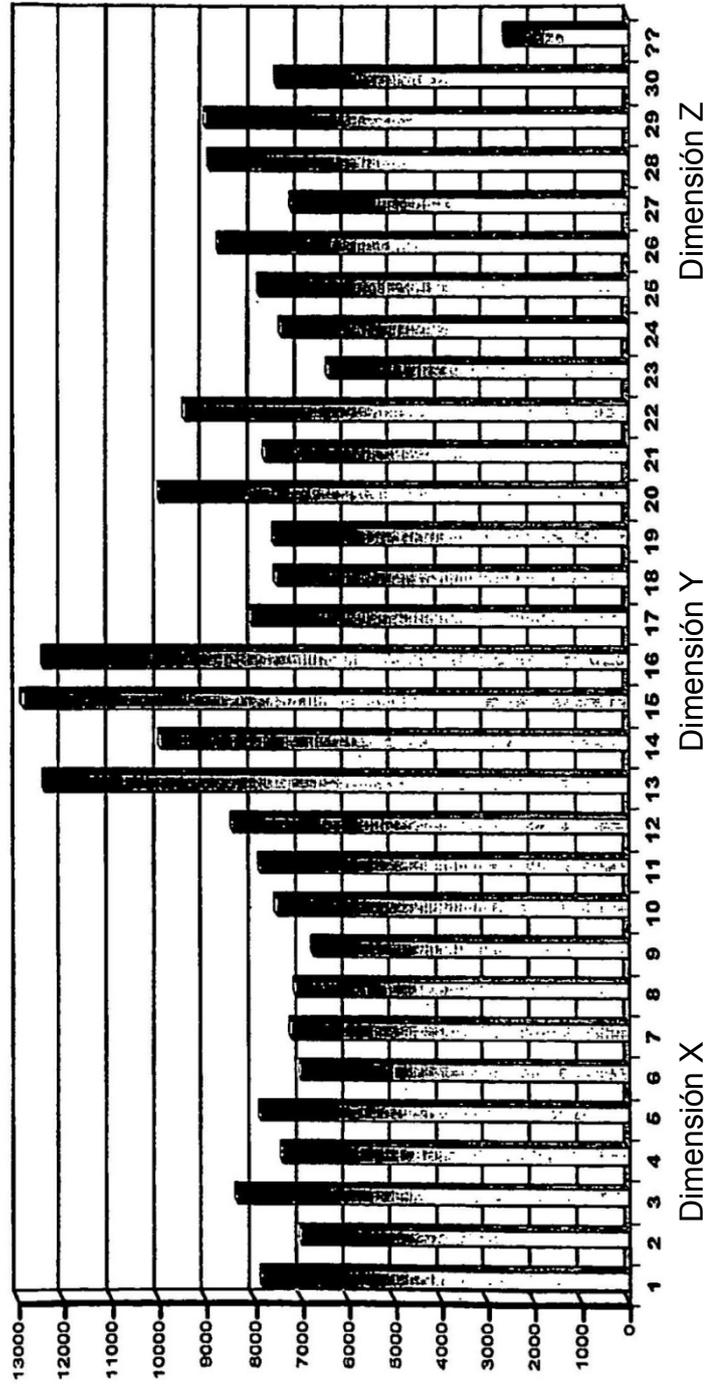
**FIG. 8**

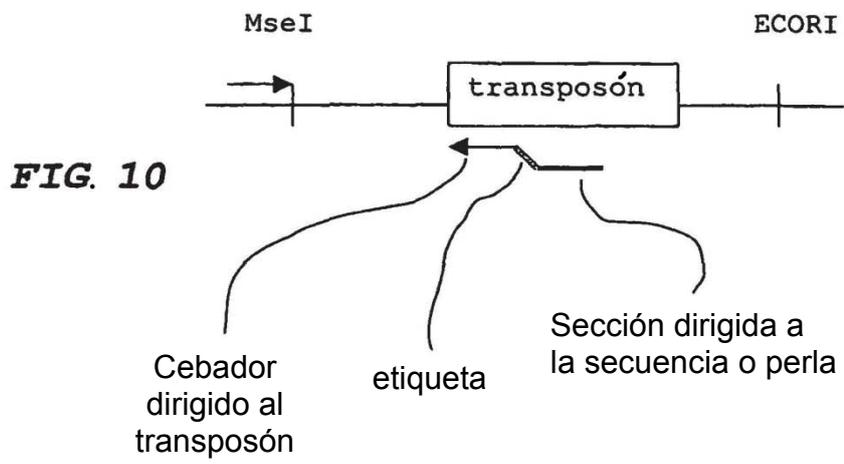
Número de sitios de inserción # (grupos) con respecto al número de copias (todo el intervalo)



**FIG. 9**

Número de secuencias según las coordenadas 3D-TAG 1-30





**FIG 11**

