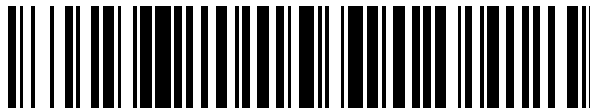


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 528 971**

51 Int. Cl.:

C12Q 1/68 (2006.01)

C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **17.03.2009 E 09721510 (7)**

97 Fecha y número de publicación de la concesión europea: **07.01.2015 EP 2268834**

54 Título: **Descubrimientos de genes vinculados a la expresión**

30 Prioridad:

17.03.2008 EP 08152859

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

13.02.2015

73 Titular/es:

**STICHTING GENETWISTER IP (100.0%)
Nieuwe Kanaal 7b
6709 PA Wageningen , NL**

72 Inventor/es:

**DE BOER, ANNE DOUWE;
EBSKAMP, MICHAËL JOHANNES MARCUS;
LANGEVELD, SIMON ALBERTUS;
LAROS, IVO y
VAN DE RHEE, MIRANDA DEBORA**

74 Agente/Representante:

LAZCANO GAINZA, Jesús

ES 2 528 971 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

Descubrimientos de genes vinculados a la expresión.

5 La invención se refiere al campo de la biología molecular y la biotecnología, más específicamente al campo de la
 secuenciación, la detección y la identificación de secuencias de ácidos nucleicos en el ADN genómico. Más en particular, la
 invención se refiere a la aplicación de un método en la identificación y/o detección de secuencias de nucleótido que
 representan la mayoría de las regiones transcritas y su entorno en un genoma y que están relacionados con una amplia
 10 variedad de rasgos genéticos, genes y combinaciones de los mismos. La invención se puede usar en el campo de la
 detección de alto rendimiento y la identificación de marcadores moleculares a partir de cualquier origen, ya sea de plantas,
 animales, humanos, artificial o de otra manera.

15 Las tecnologías de cría han evolucionado de la simple selección de rasgos visibles en métodos para la detección de rasgos
 multigénicos usando marcadores moleculares. En principio, cada diferencia genética entre diferentes líneas de una
 población de cruce puede representar un rasgo alterado. Sin embargo, debido a la complejidad de la mayoría de los
 genomas no es posible identificar cada diferencia que existe entre los genomas y vincularlo a un rasgo particular. En teoría,
 la secuenciación completa de los genomas revelaría todas las diferencias entre los genomas. Sin embargo, esto no se
 20 puede realizar de forma práctica, con tiempo y costos efectivos con las tecnologías de secuenciación actuales. Por lo tanto,
 los métodos para la detección de las diferencias genéticas principalmente se han basado en el principio de reducción de la
 complejidad que implica la secuenciación de una parte limitada pero completamente definida del ADN genómico de
 diferentes individuos. Con los avances en las tecnologías de secuenciación la reducción de la complejidad se ha hecho
 menos importante para algunas aplicaciones como análisis de transcriptomas que representan todas las secuencias de
 genes expresados. El tamaño de los genomas eucariotas que van desde unas pocas decenas a varios cientos de
 25 megabases, está sin embargo, más allá de la capacidad de las tecnologías actuales de secuenciación de alto rendimiento.
 Además, la gran mayoría del ADN genómico en los organismos eucariotas, especialmente aquellos con tamaños más
 grandes del genoma, no proporciona información valiosa para fines de reproducción, ya que nunca se expresa y por lo tanto
 no parece contribuir a la expresión de rasgos

30 Por lo tanto, para identificar marcadores moleculares, los métodos que se centran en las partes de un genoma que son más
 propensos a revelar los marcadores moleculares estrechamente vinculados a los rasgos tienen una ventaja sobre los
 métodos que analizan selecciones simplemente al azar de genomas que incluyen áreas no expresadas. Este problema se
 hace más agudo cuando el tamaño del genoma aumenta. El método descrito hace posible determinar secuencias en una
 parte seleccionada del ADN genómico que representa las regiones codificantes de la mayoría de los genes expresados y su
 entorno. La comparación de dichas partes seleccionadas entre diferentes individuos permite la identificación de sitios
 35 polimórficos que están dentro o en las cercanías de los genes expresados. Ya que la frecuencia de polimorfismos es
 superior en regiones no codificantes, más polimorfismos se pueden relacionar a los genes expresados que con las
 tecnologías actuales. Además grandes regiones no codificantes que rodean genes más conservados se pueden analizar por
 la presencia de polimorfismos. Esto puede en última instancia resultar en el descubrimiento de al menos un marcador por
 rasgo. El método de la invención hace posible enfocarse en la detección SNP en las áreas de codificación de genes y las
 40 áreas reguladoras de genes por la iluminación de partes bien definidas del genoma entre diferentes individuos y organismos,
 aun en organismos con genomas complejos y grandes

45 Los polimorfismos de las secuencias de nucleótidos, como SNPs se aplican ampliamente para construir mapas del genoma.
 Después de que los polimorfismos se vinculan a fenotipos en un proceso llamado mapeo genético, tales polimorfismos se
 pueden usar como marcadores en tecnologías de cría asistida por marcadores para detectar un fenotipo particular en
 cualquier etapa del desarrollo. Los polimorfismos de la secuencia de nucleótidos se identifican generalmente en el ADN
 genómico. Como el tamaño del genoma de todos los organismos eucarióticos supera con creces el número de nucleótidos
 que se pueden analizar con las tecnologías de secuenciación de alto rendimiento actuales, se necesitan procedimientos
 50 reproducibles para reducir la complejidad para analizar las partes seleccionadas de los genomas completos para encontrar
 diferencias genéticas entre individuos que se puedan usar para el mapeo genómico. Sin embargo, la naturaleza estadística
 de los métodos de reducción de complejidad que se aplican actualmente, implica que estos métodos no revelan *a priori*
 aquella diferencias genéticas que se pueden vincular a fenotipos únicos o se mapean cercanamente a genes que
 contribuyen a un genotipo particular.

55 Las tecnologías actuales se enfocan fuertemente en el descubrimiento de polimorfismos de un único nucleótido (SNPs) por
 varias razones: SNPs se presentan más frecuentemente en los genomas que cualquier otro tipo de polimorfismo, SNPs
 permiten la detección precisa de alelos homocigotos y heterocigotos, SNPs se pueden aplicar en aplicaciones de alto
 rendimiento y están disponibles muchas plataformas industriales que hacen la detección de SNP a cualquier escala de

aplicación deseada rentable. Aunque el descubrimiento de SNP sería el método de elección en situaciones donde ocurren bajos niveles de polimorfismo, como áreas de codificación de genes conservados y genomas de individuos estrechamente relacionados, el uso del banco EST para el descubrimiento de SNP en individuos que están estrechamente relacionados puede ser menos efectivo debido a un nivel inherentemente bajo de polimorfismo.

5 En conclusión, un método de descubrimiento de SNP debería idealmente revelar todos los SNPs presentes que están físicamente vinculados a rasgos de interés, pero no debe obstaculizarse por los bajos niveles de polimorfismo que se producen en las áreas de codificación de genes del genoma u obstaculizarse por ningún requisito de conocimiento de las secuencias del genoma. Así se necesita un método que pueda determinar reproduciblemente secuencias concomitantes en áreas de ADN genómico que representan la mayoría de las regiones de codificación de genes y las regiones de su entorno, esto sin previo conocimiento de las secuencias del genoma o el transcriptoma.

10 En la técnica anterior se conoce de Estados Unidos 2003/099962 que se pueden preparar pequeños fragmentos de ADNc monocatenario y pequeños fragmentos de ADN genómico monocatenario, los que después pueden hibridarse, amplificarse y los productos amplificados de los mismos secuenciarse. Sin embargo, en este documento de la técnica anterior los fragmentos genómicos se obtienen mediante el uso de una etapa de clonación consumidora de tiempo.

15 Espelund y otros, (Biotechniques 13(1):74-81, 1992) indica que una etapa de clonación de este tipo puede reemplazarse por una amplificación por PCR usando adaptadores.

Sin embargo, se necesita un método (mejorado) para determinar reproduciblemente las secuencias genómicas en base al transcriptoma, sin conocimiento previo de la información de la secuencia en la misma.

Breve descripción de la invención

20 Los inventores actuales han encontrado ahora un método para analizar una región genómica de un organismo, que comprende cuatro partes principales.

La primera parte implica el aislamiento de ARNm de un organismo seleccionado que se usa para la preparación de fragmentos pequeños de ADN monocatenario con un adaptador que contiene un marcador de biotina y al menos un adaptador que contiene un sitio de endonucleasa de restricción tipo IIs raro. Estos fragmentos de ADN se usan en la tercera parte. En la segunda parte, se aísla el ADN genómico del mismo organismo o de uno relacionado. Este ADN genómico se fragmenta y se liga a las moléculas adaptadoras. En esta tercera parte estos fragmentos genómicos se hibridan con los fragmentos de ADN monocatenario a partir de la parte uno y los híbridos formados en este proceso se usan para la síntesis de fragmentos de ADN. Estos fragmentos se usarán en la cuarta parte, que implica la secuenciación de estos fragmentos usando uno de los métodos de secuenciación de alto rendimiento disponibles.

30 Dicho método de identificación de ADN genómico en una muestra por lo tanto puede comprender las etapas de:

- a) aislamiento y purificación de ARNm de las muestras de tejido de un organismo;
- 35 b) síntesis de ADNc usando dicho ARNm como una plantilla;
- c) opcionalmente reducción de la complejidad de dicho ADNc;
- d) fragmentación de dicho ADNc;
- e) opcionalmente selección de tamaño de dichos fragmentos;
- f) opcionalmente eliminar los fragmentos que contienen polyA mediante la unión a perlas de afinidad recubiertas con estreptavidina;
- 40 g) pulido de dichos fragmentos de ADNc;
- h) ligación de dichos fragmentos con un adaptador que comprende un sitio de reconocimiento para una enzima de restricción rara y otro adaptador que contiene un marcador de biotina;
- i) opcionalmente selección de tamaño de dichos fragmentos;
- j) reparación de mellas de dichos fragmentos;
- 45 k) selección de dichos fragmentos que contienen secuencias adaptadoras;
- l) amplificación de dichos fragmentos usando iniciadores de hibridación para las secuencias adaptadoras descritas en la etapa h, en donde un iniciador es complementario al adaptador con un sitio de restricción raro y el otro iniciador contiene un marcador de biotina.
- m) unir dichos fragmentos a perlas de afinidad recubiertas con estreptavidina;
- 50 n) eliminar los adaptadores que contienen el sitio de restricción raro usando la enzima de restricción correspondiente a partir de dichos fragmentos;
- o) eliminar las cadenas simples no unidas a perlas de afinidad por una interacción biotina-estreptavidina a partir de fragmentos de ADN bicatenarios unidos a las perlas de afinidad;
- p) aislamiento y purificación de ADN genómico por ejemplo a partir del organismo de la etapa a
- 55 q) fragmentación de dicho ADN genómico;
- r) opcionalmente pulir dicho ADN genómico;
- s) ligación de dicho ADN genómico con un solo tipo de adaptador o con dos tipos diferentes de adaptadores;

- t) fusionar dicho ADN genómico en un ADN monocatenario;
- u) hibridar el ADN genómico de la etapa t) con ADN en las perlas de la etapa o);
- v) eliminar el ADN genómico no unido por lavado;
- w) extensión del híbrido del ADNc-ADN genómico por una polimerasa para crear una plantilla bicatenaria;
- x) realizar la PCR sobre dicho híbrido de ADN genómico-ADNc
- y) selección de los fragmentos más grandes de aproximadamente 100 pares de bases a partir de dicha PCR;
- z) opcionalmente purificación de dichos fragmentos y
- aa) secuenciación de alto rendimiento de dichos fragmentos .

10 En otra modalidad, el método se extiende a un método para identificar polimorfismos, que comprende todas las etapas del método de acuerdo con la reivindicación y además comparar los datos de secuencia a partir de dos o más muestras para identificar los polimorfismos.

Definiciones

15 En la siguiente descripción y los ejemplos se utilizan una serie de términos. Con el objetivo de proporcionar una comprensión clara y consistente de la descripción y reivindicaciones, que incluyen el alcance que debe darse a tales términos, se proporcionan las siguientes definiciones. A menos que se defina de otra manera, todos los términos técnicos y científicos que se usan en la presente tienen el mismo significado que el comúnmente entendido por alguien con experiencia en la técnica a la que pertenece esta invención.

20 **Ácido nucleico:** un ácido nucleico de acuerdo con la presente invención puede incluir cualquier polímero u oligómero de bases pirimidina y purina, preferentemente citosina, timina, y uracilo, y adenina y guanina, respectivamente (Ver Albert L. Lehninger, Principes de Biochemistry, en 793-800 (Worth Pub. 1982)). La presente invención contempla cualquier desoxirribonucleótido, ribonucleótido o componente de ácido nucleico péptido, y cualquier variante química de los mismos, tales como las formas metiladas, hidroximetiladas o glicosiladas de estas bases, y similares. Los polímeros u oligómeros pueden ser heterogéneos u homogéneos en su composición, y pueden aislarse de fuentes de origen natural o pueden producirse artificialmente o sintéticamente. Además, los ácidos nucleicos pueden ser ADN o ARN, o una mezcla de los mismos, y pueden existir de manera permanente o transitoria en forma monocatenaria o bicatenaria, incluyendo homodúplex, heterodúplex y estados híbridos.

30 **SNP:** Un polimorfismo de un solo nucleótido es una variación de la secuencia de ADN que ocurre cuando un solo nucleótido - A, T, C, o G - en el genoma difiere entre los miembros de una especie (o entre cromosomas apareados en un individuo) a en un locus específico. Los SNPs son el tipo más común de variación genética. Los SNPs pueden caer dentro de secuencias codificantes de genes, regiones no codificantes de genes, o en regiones intergénicas entre genes. Los SNPs dentro de una secuencia codificante no cambiarán necesariamente la secuencia de amino ácidos de la proteína que se produce, debido a la degeneración del código genético. Un SNP en el que ambas formas llevan a la misma secuencia de polipéptido se denomina sinónimo y si se produce una secuencia de polipéptido diferente se denomina no-sinónimo. Como los SNPs son evolutivamente conservados, ellos pueden usarse como marcadores para un análisis cuantitativo de loci (QTL) y en estudios de asociación.

Intrón: Los intrones son secciones no codificantes de un gen que se eliminan de los pre-ARNms en un proceso denominado empalme para producir ARNms funcionales.

45 **Exón:** Un exón es cualquier región de ADN con un gen que se transcribe a la molécula de ARN mensajero final (ARNm), en lugar de eliminarse por empalme de la molécula de ARN transcrita como los intrones.

ADNc: ADNc es una forma artificial de ADN sintetizada por la enzima transcriptasa inversa usando moléculas de ARN como plantilla.

50 **ADN genómico:** el término ADN genómico indica que el ADN se deriva de la situación 'como es'. Esto significa que el ADN genómico alberga secuencias, como se encontrarán en la naturaleza, por ejemplo que incluyen intrones y secuencias reguladoras. El ADN genómico puede derivarse de diferentes fuentes, como cromosomas, pero además de fuentes extracromosómicas tales como mitocondria, cloroplastos y plásmidos.

55 **ADN Cot-1:** La técnica que se usa para determinar la complejidad de la secuencia de cualquier genoma involucra la desnaturalización y renaturalización del ADN. El ADN se desnaturaliza por calor y esto funde los H-enlaces y rinde el ADN monocatenario. Si el ADN se enfría rápidamente, el ADN permanece de monocatenario. Pero si el ADN se deja enfriar

lentamente, las secuencias que son complementarias se encontrarán una a la otra y eventualmente se aparearán nuevamente. La velocidad a la que el ADN se rehibrida (otro término para renaturaliza) es una función de la especie de la que se aisló el ADN, además identificada como la curva "Cot". El ADN que tiene un alto valor *Cot* es un ADN altamente repetitivo, mientras el ADN con un bajo valor *Cot* está disponible solamente en bajas copias o es único. En el método usamos ADN con un valor *Cot* de 1, una fracción del ADN genómico total que está enriquecida en secuencias de ADN altamente repetitivas.

Anotación: La anotación de las secuencias de ADNc comprende dos etapas. Las secuencias obtenidas se comparan a secuencias de nucleótidos y/o amino ácidos disponibles en las bases de datos (públicas). Los métodos para el alineamiento de secuencias para propósitos de comparación se conocen bien en la técnica. Esta comparación se realiza típicamente con la ayuda de un programa tal como la Herramienta de búsqueda de alineamiento local básica NCBI (BLAST) descrita por Altschul y otros, 1990). Este programa está disponible de varias fuentes, incluyendo El Centro Nacional para la Información Biológica (NCBI, Bethesda, Md.) y en internet ([HTTP://www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). Este programa compara las secuencias ADNc/EST (Etiquetas de secuencia expresada) identificadas con las secuencias presentes en bases de datos y presenta los resultados en base a cierta puntuación y un parámetro de probabilidad. El programa puede seleccionar aquellas secuencias ADNc/EST con un cierto límite inferior predeterminado de dicho parámetro de probabilidad. Las secuencias ADNc/EST seleccionadas después se proporcionan en una segunda etapa con una anotación (es decir un enlace a una secuencia como se presenta en una base de datos). Este tipo de anotación se denomina una "anotación electrónica".

Agrupamiento: el término "agrupamiento" significa la construcción de una colección de secuencias con similitud mediante una comparación pareada de secuencias de dos o más nucleótidos y la selección por la presencia de tramos cortos o largos de nucleótidos idénticos o similares. Varios métodos para el alineamiento de secuencias de nucleótidos se conocen en la técnica, como se explicará adicionalmente más abajo. A veces los términos "ensamble" o "alineación" se usan como sinónimos.

Identificador: una secuencia corta que se puede adicionar a un adaptador o un iniciador o incluida en su secuencia o de otra manera usarse como etiqueta para proporcionar un único identificador. Tal secuencia identificadora puede ser una secuencia de bases única de longitud variada. Por ejemplo etiquetas de 4 bp permiten $4^{(exp4)} = 256$ etiquetas diferentes. Ejemplos típicos son las secuencias ZIP, conocidas en la técnica como etiquetas usadas comúnmente para la detección única por hibridación (Iannone y otros Cytometry 39:131-140, 2000). Usando tal identificador, el origen de una muestra de PCR se puede determinar tras procesamiento adicional. En el caso de la combinación de productos procesados originados a partir de diferentes muestras de ácidos nucleicos, las diferentes muestras de ácidos nucleicos se identifican generalmente usando diferentes identificadores.

Secuenciación: El término secuenciación se refiere a la determinación del orden de los nucleótidos (secuencias de bases) en una muestra de ácidos nucleicos, por ejemplo ADN o ARN. El tamizaje de alto rendimiento, frecuentemente abreviado como HTS, es un método para la experimentación científica especialmente relevante al campo de la biología y la química. Mediante una combinación de robótica moderna y otros hardware de laboratorio especializados, permite a un investigador tamizar eficazmente grandes cantidades de muestras simultáneamente, más específicamente esta es una tecnología de secuenciación tal como la descrita en la presente descripción en otra parte (de 454 Life Sciences, www.454.com e Illumina, www.illumina.com). Por ejemplo el acercamiento de secuenciación Illumina Solexa se basa en la unión de ADN genómico fragmentado aleatoriamente a una superficie plana, ópticamente transparente y la amplificación en fase sólida para crear una celda de flujo de secuenciación de densidad ultra alta con > 10 millones de agrupamientos, cada uno contiene ~1,000 copias de plantilla por sq. cm. Estas plantillas se secuencian usando una robusta tecnología de síntesis por secuenciación de ADN de cuatro colores que emplea terminadores reversibles con fluorescencia removible. Este acercamiento garantiza una alta exactitud y prevención de artefactos con repeticiones homopoliméricas. La detección de fluorescencia de alta sensibilidad se logra usando excitación del láser y ópticas de reflexión total interna.

Endonucleasa de restricción: una endonucleasa de restricción o enzima de restricción es una enzima que reconoce una secuencia de nucleótidos específica (sitio objetivo) en una molécula de ADN bicatenario, y escindirá las dos cadenas de la molécula de ADN en o cerca de cada sitio objetivo.

Fragmentos de restricción: las moléculas de ADN producidas por digestión con una endonucleasa de restricción se denominan fragmentos de restricción. Cualquier genoma determinado (o ácido nucleico, independientemente de su origen) será digerido por una endonucleasa de restricción particular en un conjunto discreto de fragmentos de restricción. Los fragmentos de ADN que resultan de la escisión por endonucleasa de restricción se pueden usar adicionalmente en una variedad de técnicas y pueden, por ejemplo, detectarse por electroforesis en gel.

5 Ligación: la reacción enzimática catalizada por una enzima ligasa en la que dos moléculas de ADN bicatenario se unen covalentemente entre sí se denomina ligación. En general, ambas cadenas de ADN se unen covalentemente, pero es además posible prevenir la ligación de una de las dos cadenas mediante modificaciones químicas o enzimáticas de uno de los extremos de las cadenas. En ese caso la unión covalente ocurrirá solamente en una de las dos cadenas de ADN.

10 Oligonucleótido sintético: las moléculas de ADN monocatenarias que tienen preferentemente de aproximadamente 10 a aproximadamente 50 bases, que se pueden sintetizar químicamente se denominan oligonucleótidos sintéticos. En general, estas moléculas de ADN se diseñan para tener una secuencia de nucleótidos única o, deseada aunque es posible sintetizar familias de moléculas que tienen secuencias relacionadas y que tienen diferentes composiciones de nucleótidos en posiciones específicas dentro de la secuencia de nucleótidos. El término oligonucleótido sintético se usará para denominar moléculas de ADN que tienen una secuencia de nucleótidos diseñada o deseada.

15 Adaptadores: moléculas de ADN bicatenario cortas con un número de pares de bases limitado, por ejemplo aproximadamente 10 a aproximadamente 30 pares de bases en longitud, que se diseñan de forma tal que pueden ligarse a los extremos de los fragmentos de restricción. Los adaptadores generalmente están compuestos de dos oligonucleótidos sintéticos que tienen secuencias de nucleótidos que son parcialmente complementarias entre sí. Cuando se mezclan dos oligonucleótidos sintéticos en solución bajo las condiciones adecuadas, ellas hibridan entre sí formando una estructura bicatenaria. Después de la hibridación, un extremo de la molécula adaptadora se diseña de forma tal que sea compatible con el extremo de un fragmento de restricción y se puede ligar a él; el otro extremo del adaptador se puede diseñar de forma que no puede ligarse, pero no es necesariamente el caso (adaptadores ligados dobles). Específicamente, en la presente invención un adaptador es una molécula de ADN bicatenario, usualmente entre 15 y 60 pares de bases, que se puede preparar por hibridación de dos oligonucleótidos (parcialmente) complementarios. Los adaptadores usados en la presente descripción pueden estar romos en los extremos, o tener un saliente específico para la ligación a moléculas de ADN con una proyección complementaria, tal como aquella creada por la digestión por la endonucleasa de restricción. Los adaptadores pueden tener un saliente adicional, no compatible en el otro extremo, que no se puede ligar a un fragmento de ADN de extremo romo y no a ADN con un saliente específico creado por la endonucleasa de restricción de tipo II, previniendo así la ligación de múltiples adaptadores entre sí.

20 Los adaptadores proporcionan sitios de hibridación para los iniciadores durante la PCR para amplificar los diferentes fragmentos de ADN ligados a los adaptadores en una única reacción de PCR.

25 Uno de los adaptadores ligados a las moléculas de ADNc portará la secuencia de reconocimiento de la endonucleasa de restricción de tipo II para escindir el adaptador de la molécula de ADNc una vez unido a la fase sólida.

30 Fragmentos de restricción ligados a adaptadores: fragmentos de restricción rematados por adaptadores.

35 Iniciadores: en general, el término iniciadores se refiere a cadenas de ADN, que pueden iniciar la síntesis de ADN. La ADN-polimerasa no puede sintetizar ADN *de novo* sin iniciadores: solo puede extender una cadena de ADN existente en una reacción en la que la cadena complementaria se usa como una plantilla para dirigir el orden de los nucleótidos que se van a colocar. Nos referiremos a las moléculas de oligonucleótidos sintéticos que se usan en una reacción en cadena de la polimerasa (PCR) como iniciadores.

40 Amplificación del ADN: el término amplificación del ADN se usará típicamente para denotar la síntesis *in vitro* de moléculas de ADN bicatenarias por medio del uso de PCR o un sistema de amplificación comparable. Se nota que existen otros métodos de amplificación y ellos se pueden usar en la presente invención. Los métodos de la invención pueden, en principio, realizarse mediante el uso de cualquier método de amplificación de ácido nucleico, tales como Reacción en cadena de la polimerasa (PCR; Mullis 1987, patente de los EE.UU. núm. 4,683,195, 4,683,202, en 4,800,159) o mediante el uso de reacciones de amplificación tales como Reacción en cadena de ligasa (LCR; Barany 1991, Proc. Natl. Acad. Sci. Estados Unidos 88:189-193; Sol. EP núm., 320,308), Replicación de secuencia autosostenida (3SR; Guatelli y otros, 1990, Proc. Natl. Acad. Sci. Estados Unidos 87:1874-1878), Amplificación por desplazamiento de cadena (SDA; patente de Estados Unidos núms. 5,270,184, en 5,455,166), Sistema de amplificación transcripcional (TAS; Kwok y otros, Proc. Natl. Acad. Sci. Estados Unidos 86:1173-1177), Q-Beta Replicasa (Lizardi y otros, 1988, Bio/Technology 6:1197), Amplificación por círculo rodante (RCA; patente de los EE.UU. núm. 5,871,921), Amplificación basada en la secuencia de ácido nucleico (NASBA), Polimorfismo de longitud de fragmentos de rotura (patente de los EE.UU. núm. 5,719,028), Amplificación iniciada con iniciador isotérmico y quimérico de ácido nucleico (ICAN), Método de amplificación de extensión de la ramificación (RAM; patente de Estados Unidos núms. 5,719,028 y 5,942,391) u otros métodos adecuados para la amplificación de ADN.

Con el objetivo de amplificar el ADN con un pequeño número de desajustes a uno o más de los iniciadores de amplificación,

se puede realizar una reacción de amplificación bajo condiciones de rigurosidad reducida (por ejemplo, una amplificación por PCR usando una temperatura de hibridación de 38 °C, o la presencia de 3,5 mM MgCl₂). El experto en la técnica será capaz de seleccionar las condiciones de rigurosidad adecuadas.

5 El pulido (además denominado reparación de extremos) se refiere a la conversión de ADN con extremos no romos en ADN con extremos romos.

10 La digestión de pulido de plantillas de ADN genómico (ADNg) con ADNsa I en presencia de Mn²⁺ produce fragmentos de ADN que son de extremos romos o tienen un terminal protuberante con uno o dos nucleótidos de longitud. Similarmente, la fragmentación de ADN por medios mecánicos proporciona una combinación de fragmentos con extremos romos o extremos salientes. Estos fragmentos de ADN, ya sea generados enzimáticamente o mecánicamente, pueden "pulirse" usando los procedimientos descritos más abajo.

15 En un método, el pulido se puede realizar por tratamiento de los fragmentos 3'-salientes con una exonucleasa específica de cadena sencilla, tal como la nucleasa BAL32 o la nucleasa Mung Bean. Generalmente, la nucleasa debe calibrarse antes de usarse.

20 En otro método, los extremos romos se crean con la ADN-polimerasa Pfu o con otras ADN-polimerasas tales como ADN-polimerasa T4 o ADN-polimerasa Klenow. El "pulimiento" Pfu o terminación roma se puede usar para aumentar la cantidad de especies de extremos-romos generadas a continuación de la digestión de la plantilla genómica con ADNsa I. La ADN-polimerasa Pfu rellena los salientes 5'. Además, la ADN-polimerasa Pfu exhibe actividad exonucleasa 3' a 5'. Así, la enzima se puede usar para eliminar extensiones de nucleótidos simples y dobles para aumentar aún más la cantidad de fragmentos de ADN de extremos romos disponibles para la ligación del adaptador (ver, por ejemplo, Costa, G. L. y M. P. Weiner, 1994, Protocols for cloning and analysis of blunt-ended PCR-generated DNA fragments. PCR Methods Appl 3 (5): S95; Costa, G. L., A. Grafsky y M. P. Weiner, 1994, Cloning and analysis of PCR-generated DNA fragments. PCR Methods Appl 3 (6): 338; Costa, G. L. y M. P. Weiner, 1994, Polishing con T4 o Pfu polimerasa increases the efficiency de cloning de PCR products. ácido nucleicos Res. 22 (12): 2423).

30 Los presentes inventores encontraron que al proporcionar el ADN genómico y usando fragmentos derivados de ADNc ligados al adaptador para actuar como iniciadores para la amplificación de fragmentos de ADN genómico, parece posible detectar secuencias de genoma fuera de las regiones que están realmente transcritas, es decir el promotor, intrón y secuencias de terminación. Así combina la identificación rápida de fragmentos de ADN genómico en regiones expresadas del genoma, en donde múltiples muestras se pueden secuenciar en una única corrida, con la posibilidad de estudiar fragmentos de ADN genómico del entorno de estas regiones y la posibilidad de detectar variaciones genéticas en fragmentos que codifican genes, fragmentos de intrones y fragmentos que cubren secuencias reguladoras del genoma.

35 Una ventaja adicional es que el método de la invención es generalmente aplicable, es decir para todos los organismos. No se necesita información sobre el genoma o la organización genómica con antelación. Otra ventaja es que no se requieren etapas de clonación adicionales en el presente método. Esto permite la secuenciación de secuencias que codifican sustancias tóxicas o para las proteínas reguladoras, que de otro modo sería imposible debido a que el organismo huésped en el que tales secuencias se clonarían y expresarían no o apenas sobreviviría. En el mismo sentido además las secuencias que crearían problemas para la clonación pueden abordarse ahora y no hay límite en cuanto a la longitud de las secuencias, lo que además puede causar problemas en el programa de clonación.

40 Una ventaja adicional es que el método actual no requiere ADNc de longitud completa, sino que se puede usar con secuencias más cortas. Esto es especialmente útil cuando se analizan genomas grandes (por ejemplo de plantas bulbosas) ya que entonces se puede mantener el foco en las regiones del genoma más interesantes o relevantes.

45 Una siguiente ventaja es que es posible con el método actual generar los iniciadores para todas las secuencias de ADN expresadas, lo que significa que es posible obtener los datos genómicas de las secuencias expresadas

Breve descripción de las figuras

55 Figura 1. Representación esquemática de la creación de moléculas pequeñas de ADN monocatenario a partir de ARNm. C es un adaptador, D es un adaptador diferente, B es biotina, P es un grupo fosfato, bp es pares de bases, ds es bicatenario, ss es monocatenario, RE es una enzima de restricción, SA es estreptavidina, LD-PCR es PCR de larga longitud, NBA₍₃₀₎ y NVT₍₃₀₎ son códigos de una sola letra para nucleótidos de acuerdo a las reglas

sintácticas para nucleótidos del Comité de nomenclatura de la Unión Internacional de Bioquímica (NC-IUB), A₍₃₀₎ y T₍₃₀₎ indican tramos de 30 As y 30 Ts respectivamente.

Figura 2. Representación esquemática de la creación de fragmentos de ADN genómico y la subsecuente hibridación con pequeñas moléculas de ADN monocatenario a partir de ARNm (figura 1). Después de varias etapas, se obtienen fragmentos de ADN que se pueden secuenciar usando métodos de secuenciación de alto rendimiento. C y D son adaptadores además representados en la figura 1. E y F son diferentes adaptadores, B es biotina, bp es pares de bases, ds es bicatenario, ss es monocatenario y SA es estreptavidina.

Figura 3. Los adaptadores A y B no-fosforilados se ligan a los extremos de los fragmentos de ADN bicatenario genómico fosforilados, pulidos. Los adaptadores A y B se diferencian en la secuencia de nucleótidos y en la presencia de una etiqueta de biotina 5' en el adaptador B. Las mellas están presentes en las uniones 3' de cada uno de los adaptadores y los fragmentos se rellenan por la actividad de desplazamiento de cadena de la ADN-polimerasa Bst. Las interacciones estreptavidina-biotina se pueden usar para eliminar fragmentos flanqueados por juegos de adaptadores homocigotos (A/A y B/B) y para generar plantillas de genoteca monocatenaria. Los fragmentos se unen a las perlas de estreptavidina; el material no unido (compuesto de los juegos de adaptadores homocigotos A/A, que carecen de biotina) se eliminan por lavado. Los fragmentos inmovilizados se desnaturalizan, ambas cadenas de los fragmentos B/B permanecen inmovilizadas mediante el adaptador B biotinilado, mientras los fragmentos A/B se lavan libres y se usan en las etapas subsecuentes.

Figura 4 casos 1 y 2 y 3 de los esquemas ELGD. Panel A. Caso 1. Representación esquemática de la alineación de la secuencia genómica de la manzana FRA8S6E02IN5QW, contiguo EST manzana cg8984st_846_7 y el gen AT1G70160 de *Arabidopsis thaliana* (no dibujados a escala). Se muestra solamente la parte relevante 3' de AT1G70160. Los intrones se representan como recuadros blancos en FRA8S6E02IN5QW y AT1G70160. Los intrones empalmados se muestran como líneas finas en cg8984st_846_7. Las secuencias codificantes de proteína (CDS) se muestran en negro. Las regiones 3'-no traducidas se muestran en gris. La cola de poliA en la secuencia de ADNc se indica con (A)n. Panel B. Caso 2. Representación esquemática del alineamiento de la secuencia genómica de manzana FRA8SE02HOH39 y manzana EST contiguo cg12357st_1428_21 (no dibujados a escala). La secuencia codificante de proteína (CDS) se muestra en negro. Las regiones 5' y 3' no traducidas se muestran en gris. La cola de poliA en la secuencia de ADNc se indica con (A)n. Panel C. Caso 3. Representación esquemática del alineamiento de la secuencia genómica de manzana 02-H03 y manzana EST 91044590 (no dibujados a escala). La secuencia codificante de la proteína (CDS) se muestra en negro. Las regiones 5' y 3' no traducidas se muestran en gris. La región promotora no-transcrita en la secuencia genómica se muestra en blanco.

Descripción detallada de la invención

En un aspecto, la invención se refiere a un método para la identificación de ADN genómico en una muestra que comprende 4 partes diferentes.

Parte 1. En esta parte las secuencias pequeñas se generan a partir de ADNc que se usarán como secuencias iniciadoras en la parte 3. La parte consta de las siguientes etapas:

- a) aislamiento y purificación de ARNm de las muestras de tejido de un organismo;
- b) síntesis de ADNc usando dicho ARNm como una plantilla;
- c) opcionalmente reducción de la complejidad de dicho ADNc;
- d) fragmentación de dicho ADNc;
- e) opcionalmente selección de tamaño de dichos fragmentos;
- f) opcionalmente eliminar los fragmentos que contienen polyA mediante la unión a perlas de afinidad (recubiertas con estreptavidina);
- g) pulido de dichos fragmentos de ADNc;
- h) ligación de dichos fragmentos con un iniciador adaptador que comprende un sitio de reconocimiento para una enzima de restricción rara y un segundo adaptador que contiene un marcador de biotina;
- i) opcionalmente selección de tamaño de dichos fragmentos;
- j) reparación de mellas de dichos fragmentos ;
- k) selección de dichos fragmentos que contienen ambas secuencias adaptadoras;
- l) amplificación de dichos fragmentos usando iniciadores de hibridación para las secuencias adaptadoras descritas en la etapa h; un iniciador complementario al adaptador con el sitio de restricción raro y el otro iniciador contiene un marcador de biotina.
- m) unir dichos fragmentos a perlas de afinidad recubiertas con estreptavidina;
- n) eliminar los adaptadores que contienen el sitio de restricción raros de dichos fragmentos usando la enzima de restricción correspondiente;

o) eliminar las cadenas simples no unidas a perlas de afinidad por una interacción biotina-estreptavidina de los fragmentos de ADN bicatenarios unidos a las perlas de afinidad.

Esta etapa resulta en cadenas simples de ADN unidas a perlas de afinidad de estreptavidina;

Parte 2:

- p) aislamiento y purificación de ADN genómico por ejemplo a partir del organismo de la etapa a);
- q) fragmentación de dicho ADN genómico;
- r) opcionalmente pulir dicho ADN genómico;
- s) ligación de dicho ADN genómico con un solo tipo de adaptador o con dos tipos diferentes de adaptadores (preferido)
- t) fusionar dicho ADN genómico en un ADN monocatenario

Parte 3:

- u) hibridar el ADN genómico de la etapa t) con ADN en las perlas de la etapa o);
- v) eliminar el ADN genómico no unido por lavado
- w) extensión del híbrido de ADNc-ADN genómico por una polimerasa para crear una plantilla bicatenaria
- x) realizar la PCR sobre dicho híbrido de de ADN genómico-ADNc
- y) selección de los fragmentos mayores de aproximadamente 100 pares de base a partir de dicha PCR por fraccionamiento por tamaño
- z) opcionalmente purificación de dichos fragmentos

Parte 4

- aa) Uso de los fragmentos obtenidos en la etapa z) para secuenciación de alto rendimiento de acuerdo con el fabricante.

Además, cuando la secuencia del ADN genómico se determinó puede seguir una etapa para identificar cambios en las secuencias de dos o más muestras: ab) Comparación de los datos de dos o más muestras para identificar polimorfismos.

Al tratar una muestra ácido nucleico en esta forma es posible analizar reproduciblemente regiones genómicas de un organismo cubriendo regiones que codifican genes y conectadas, sin tener ninguna información acerca de la estructura o contenido del genoma del organismo. Cuando el protocolo de secuenciación permite adaptadores con identificadores, se pueden combinar múltiples muestras en una única corrida de secuenciación.

El método comienza con el aislamiento y purificación de una muestra de ARNm a partir de un organismo. Obtener una muestra de ARNm de esta forma es un procedimiento de rutina hoy en día.

Similarmente para la siguiente etapa, en donde - con la ayuda de la enzima transcriptasa inversa - una copia de ADN, el llamado ADNc se elabora a partir del ARN en la muestra. Este ADNc comprende el transcriptoma del cual se derivó el ARN, lo que representa el total de la información genética que se transcribió en el momento que se tomó la muestra. Así, en dependencia del tipo celular, el tejido del que se derivó la célula, la edad de la célula, la fase de desarrollo de la célula y las condiciones ambientales, el transcriptoma de cada célula del mismo organismo será diferente, y aun de la misma célula, cuando se toman las muestras en diferentes momentos y/o diferentes condiciones, se obtendrán diferentes transcriptomas. El ácido nucleico en la muestra original estará usualmente en la forma de ARNm. Sin embargo, además los ARN o ADN derivados a partir de otras fuentes pueden ser útiles, tal como ARN o ADN derivados de librerías de genes. El ácido nucleico en la muestra puede ser bicatenario, monocatenario, y ADN bicatenario desnaturalizado en ADN monocatenario. La muestra puede ser de cualquier organismo, ya sea vegetal, animal, sintético o humano. Se debe entender que si se obtiene una muestra de ADN, no es necesaria la reacción de la transcriptasa inversa.

Ya que la muestra, si se deriva de ARNm de células enteras, contiene el transcriptoma total, a veces sería deseable recuperar sólo un subconjunto de él. Esto se puede alcanzar de varias maneras: una manera posible es discriminar por el tamaño del ADNc, por ejemplo por centrifugación ultrarrápida. El principio que subyace en este método se conoce como reducción de la complejidad. Otras opciones para la reducción de la complejidad son por ejemplo métodos de hibridación que seleccionan por transcritos abundantes o no abundantes o por ejemplo métodos para capturar transcritos específicos ya sea para eliminarlos del conjunto de moléculas de ADNc o para seleccionarlos para análisis adicional o por ejemplo métodos como ADNc-AFLP que crea un subconjunto del conjunto de moléculas de ADNc por digestión de restricción.

Una vez que se obtuvo la muestra de ADNc deseada, el ADNc se fragmenta, lo que se puede hacer enzimáticamente o mecánicamente. La muestra de ácido nucleico se digiere con al menos una endonucleasa de restricción para proporcionar un conjunto de fragmentos de restricción. En ciertas modalidades, se pueden usar dos o más endonucleasas para obtener los fragmentos de restricción. La endonucleasa puede ser una cortadora frecuente (una secuencia de reconocimiento de 3-5 bp, tal como MseI) o una cortadora rara (secuencia de reconocimiento de >5 bp, tal como EcoRI). En ciertas modalidades preferidas, se prefiere una combinación de una cortadora rara y una frecuente. En ciertas modalidades, en particular cuando la muestra contiene o se deriva a partir de un genoma relativamente grande, se puede preferir usar una tercera enzima (cortadora rara o frecuente) para obtener un conjunto grande de fragmentos de restricción de tamaño más corto. Como una endonucleasa de restricción, cualquier endonucleasa será suficiente. Típicamente, se prefieren las endonucleasas Tipo II tales como EcoRI, MseI, PstI etc. En ciertas modalidades una endonucleasa tipo IIs puede usarse, es decir una endonucleasa de la que la secuencia de reconocimiento se localiza distante del sitio de restricción, por ejemplo tales como AcellI, BbvI, BbvII, BbsI, BedI, Bce83I, Bcefl, Bcgl, BlnI, BsaI, BsgI, BsmAI, BsmFI, BspMI, Esp3I, FaulI, FokI, GsuI, HgaI, MboII, MmeI, MnlI, SapI, SfaNI, TaqI y ZthII III.

La fragmentación mecánica puede realizarse por cizallamiento del ADNc, en donde la severidad y duración del cizallamiento determina la cantidad de fragmentación. Uno de tales métodos de cizallamiento es nebulización. Un nebulizador es un pequeño dispositivo plástico que usa aire comprimido para atomizar líquidos. Se adapta fácilmente para el cizallamiento de ADN, es extremadamente eficiente y fácil de usar (Surzycki, S., 2000, Basic Methods in Molecular Biology, NY - Springer Verlag). Están comercialmente disponibles a partir de diferentes fuentes (por ejemplo, Invitrogen Corporation).

Una vez más la reducción de la complejidad puede realizarse opcionalmente en este ADN fragmentado por la selección de tamaño. Opcionalmente la reducción de complejidad se puede obtener por selección de tamaño de los fragmentos, por ejemplo por electroforesis. Además, o alternativamente, los fragmentos poli-A se pueden eliminar por la unión a columnas de afinidad recubiertas con estreptavidina.

En la siguiente etapa del método, los fragmentos de ADNc se hacen romos, un proceso además indicado como "pulido". El pulido del inserto se usa para eliminar los 3 primeros nucleótido(s) salientes o rellenar los 5 primeros nucleótidos salientes de fragmentos de ADN generados por enzimas de restricción, generados por PCR o fragmentos de ADN cizallados. Los estuches para el pulido están disponibles comercialmente (por ejemplo estuche Quick blunt, New England Biolabs Inc.)

En la siguiente etapa del método de la invención, los fragmentos de ADNc se proporcionan con adaptadores mediante una reacción de ligación. En esta reacción dos tipos diferente de adaptadores están ligados a una mezcla de fragmentos de ADNc. Un adaptador porta un marcador de biotina. El otro adaptador contiene un sitio de restricción para una enzima de restricción tipo II que corta afuera de su secuencia de reconocimiento. Un ejemplo de esto es SapI que reconoce la secuencia GCTCTC NNNN. La enzima corta en la secuencia NNNN y deja un saliente NNN sobre el extremo 5'. De esta manera la secuencia completa del adaptador que contiene la secuencia de reconocimiento de restricción se puede eliminar del fragmento. Es importante usar una enzima que es una cortadora rara para prevenir el acortamiento de los fragmentos por una mayor frecuencia de ocurrencia de la secuencia de reconocimientos en el fragmento de ADN.

Nuevamente, en esta etapa, opcionalmente los fragmentos se pueden seleccionar en dependencia de su tamaño mediante por ejemplo electroforesis.

Después de la ligación de los adaptadores a los fragmentos, los fragmentos se someten a la reparación de mellas, para llenar cualquier brecha en el esqueleto de ADN que se pudiera crear por la ligación de los adaptadores. Los fragmentos que sólo contienen ambos adaptadores se pueden seleccionar como se describe en la Figura 1 complementaria de la publicación por M. Margulies y otros en Nature 437, páginas 376-380, 2005 (figura 3). Los fragmentos que contienen ambos adaptadores pueden después amplificarse. En esta etapa se realiza una amplificación por PCR con un iniciador que hibrida a una de las secuencias adaptadoras complementarias en combinación con otro iniciador que porta una molécula de biotina que hibrida a la otra secuencia adaptadora complementaria. Después de realizar una etapa de amplificación usando este conjunto de iniciadores, los fragmentos amplificados de ADN bicatenario se capturan sobre perlas de afinidad con estreptavidina mientras otros productos de la reacción se eliminan de la mezcla de reacción. En la siguiente etapa, los fragmentos de ADN unidos, que tienen diferente secuencias adaptadoras en los terminales 5' y 3', se tratan con una endonucleasa asimétrica que corta el fragmento de forma tal que remueve la secuencia del adaptador completa de un extremo, como se describió antes, proporcionando así un fragmento terminal que es perfectamente complementario al ADN genómico. Después de esta etapa de restricción, las formas monocatenarias libres formadas de los fragmentos se eliminan por elución de las perlas bajo condiciones alcalinas, lo que resulta en un conjunto de fragmentos de ácidos nucleicos monocatenarios unidos a las perlas de estreptavidina. Un método alternativo para generar fragmentos monocatenarios a partir de fragmentos bicatenario es el tratamiento con la enzima exonucleasa lambda. La exonucleasa Lambda degrada partir de una molécula de ADN bicatenario aquellas cadenas que poseen el terminal 5' fosforilado, y deja las cadenas simple con el terminal 5'OH intacto. Ya que uno de los extremos 5' del fragmento se protege por el marcador de biotina que se une a la estreptavidina, el tratamiento con la exonucleasa lambda además crea fragmentos de ADN

monocatenarios unidos a las perlas de estreptavidina. Estos fragmentos de ADN de monocatenarios se usarán en etapas posteriores.

5 A continuación de las etapas descritas anteriormente, en la segunda parte el ADN genómico se aísla del organismo. Este organismo puede ser el mismo del que se aisló el ARNm (cultivado bajo las mismas o diferentes condiciones) o puede ser un organismo diferente (diferente cepa, diferentes especies) y puede incluso ser una colección de ADNs genómicos (por ejemplo una genoteca de clones BAC). Los procedimientos para aislar el ADN genómico son estándar en el campo y son por ejemplo descritos en Ausubel y otros (Preparation of genomic DNA from plant tissue. págs. 2.3.1-2.3.7 en Ausubel y otros, eds. Current protocols in molecular biology. John Wiley & Sons, Inc. Budelier. 1993). Después del aislamiento, el ADN genómico se fragmentó por procedimientos de fragmentación enzimáticos o mecánicos como se describió anteriormente. Si se usa la digestión por enzima de restricción, los fragmentos resultantes se proporcionan con adaptadores que coinciden con los diferentes sitios de restricción salientes en los terminales de los fragmentos. Las moléculas adaptadoras unidas a los fragmentos de ADN genómico tienen secuencias que son diferentes de las moléculas adaptadoras unidas a los fragmentos derivados de ADNc de la parte 1.

10 Si se usa la fragmentación mecánica, los diferentes adaptadores están ligados a los fragmentos de ADN genómico y los fragmentos que tienen diferente adaptadores en los terminales 5' y 3' se seleccionan de acuerdo con el procedimiento previamente descrito para fragmentos de ADNc (por ejemplo se puede aplicar una etapa de pulido). Alternativamente, además un único adaptador se puede ligar a los fragmentos genómicos. En ese caso secuencias adaptadoras con extensiones se pueden usar para prevenir la amplificación no específica en etapas posteriores (D H Jones y S C Winistorfer, PCR Methods Appl. 1993 2: 197-203).

15 Los fragmentos de ADN genómico fragmentados pueden opcionalmente seleccionarse por el tamaño como se describió anteriormente. Para obtener fragmentos de ADN genómico monocatenario se proporciona una etapa de fusión. Después de así proporcionar una muestra apropiada de ADN monocatenario genómico fragmentado, los fragmentos de ADNc monocatenarios de la parte 1, que portan una única molécula adaptadora con un marcador de biotina que opcionalmente se puede unir a perlas de afinidad de estreptavidina, se mezclan en preferentemente en un exceso molar, para la hibridación con el ADN genómico fragmentado que porta diferentes moléculas adaptadoras (parte 3). Después de una etapa de desnaturalización, se aplican las condiciones de hibridación para permitir la formación de moléculas de ADN genómico - ADNc bicatenarias híbridadas. El proceso de hibridación puede incluir un corto período de pre-hibridación con ADNc no-fragmentado (sin adaptadores) para normalizar la reacción de hibridación por la reducción de los efectos de las altas concentraciones de fragmentos de ADNc derivados de los abundantes transcritos. Como una modificación opcional, el ADNCot-1 se puede usar en una etapa de pre-hibridación para reducir las posibles anomalías causadas por las repeticiones de secuencias en el ADN genómico.

20 En esta etapa los fragmentos de ADNc hibridarán al ADN genómico en lugares que son homólogos o idénticos al ARNm muestreado. Después de la extensión del híbrido de ADNc-ADN genómico por una polimerasa se elabora una plantilla bicatenaria. Los fragmentos bicatenarios hibridados ahora se pueden amplificar por una reacción de PCR usando un iniciador complementario con el ADNc adaptador y otro iniciador complementario a uno de los adaptadores del ADN genómico. Opcionalmente el material hibridado se puede separar en dos fracciones separadas para además usar el adaptador de ADNc y el otro iniciador complementario al adaptador de ADN genómico. La amplificación proporciona fragmentos de PCR que no solo contienen copias de la parte correspondiente al ADN genómico del ácido nucleico originalmente muestreado, sino además contiene secuencias no-transcritas, tal como secuencias reguladoras e intrones.

25 Después de la amplificación por PCR, los fragmentos de más de 100 pares de bases, con mayor preferencia más de aproximadamente 200 pares de bases, aun con mayor preferencia más de aproximadamente 300 pares de bases y con la máxima preferencia de aproximadamente 400 pares de bases o más se seleccionan mediante fraccionamiento por tamaño. Opcionalmente, estos fragmentos se purifican de acuerdo con los requisitos a cumplir para la siguiente parte del método de la invención.

30 En la siguiente parte (parte 4) dichos fragmentos se secuencian. La secuenciación de los fragmentos ligados al adaptador amplificados proporciona información de la secuencia en al menos parte de los fragmentos ligados al adaptador y secuencias genómicas que flanquean 3'. La información contenida en la parte derivada del adaptador contiene información acerca de la muestra de la que se obtiene el fragmento si los adaptadores portan etiquetas específicas de la muestra, mientras que la información de la secuencia del propio fragmento (una secuencia identificadora) proporciona información acerca del fragmento y permite la identificación del fragmento. Esta información de secuencia sobre el fragmento se usa para identificar el fragmento con una exactitud que depende del número de nucleótidos que se determina y el número de fragmentos en el conjunto de fragmentos ligados al adaptador amplificados.

35 Para proporcionar una solución al problema de la variación del muestreo en la frecuencia de transcritos entre muestras, que afecta la exactitud de la identificación de marcadores moleculares por secuenciación contenidos en un conjunto de múltiples fragmentos, los presentes inventores encontraron además que la detección de marcadores vía la secuenciación se

realiza preferentemente con suficiente redundancia (profundidad) para muestrear todos los fragmentos al menos una vez y se acompaña de medios estadísticos que se dirigen al problema de la variación del muestreo en relación con la exactitud de los llamados genotipos. Con el objetivo de aumentar la exactitud, preferentemente una etapa de amplificación procede a la etapa de secuenciación. Después de suficientes ciclos de amplificación, la redundancia de los fragmentos de restricción ligados al adaptador es al menos 6, preferentemente al menos 7, con mayor preferencia al menos 8 y con la máxima preferencia al menos 9. En las modalidades preferidas así las secuencias de cada fragmento de restricción ligado al adaptador se determina al menos 6, preferentemente al menos 7, con mayor preferencia al menos 8 y con la máxima preferencia al menos 9 veces. En ciertas modalidades, la redundancia se selecciona tal que, asumiendo una oportunidad total de 50/50 de identificar el locus correctamente como homocigótico, la oportunidad de la identificación correcta del locus es más que 95 %, 96 %, 97 %, 98 %, 99 %, 99.5 %.

La amplificación de los fragmentos de restricción ligados al adaptador lleva a un conjunto amplificado de fragmentos de restricción ligados al adaptador, algunas veces denominados como amplicones. Los amplicones (o al menos parte de los mismos) se someten a una etapa que comprende al menos la determinación de la secuencia del identificador específico de la muestra para determinar el origen del fragmento y de parte de la secuencia del fragmento de restricción. En la práctica esta cantidad equivale además a la determinación de las secciones localizadas en el medio tales como los restos de las secuencias de reconocimiento de la endonucleasa de restricción. Al secuenciar el identificador específico de la muestra en combinación con parte del fragmento localizado adyacente a la secuencia derivada del adaptador, es posible identificar singularmente fragmentos de restricción y sus secuencias genómicas flanqueantes 3'. A partir de esta información es posible restaurar la información genética genómica de un gen completo.

La secuenciación de alto rendimiento usada en la presente invención es un método para la experimentación científica especialmente relevante para el campo de la biología y la química.

Se prefiere que la secuenciación se realice usando métodos de secuenciación de alto rendimiento, tales como los métodos descritos en 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 y otros, (2004) Proc . Natl. Acad. Sci. Estados Unidos 101:5488- 93, y tecnologías de Helios, Solexa, Estados Unidos genómicos, etcetera, las que se incorporan en la presente descripción como referencia.

La tecnología descrita permite la secuenciación de 40 millones de bases en una única corrida y es 100 veces más rápida y barata que la tecnología competidora. La tecnología de secuenciación aproximadamente consiste de 5 etapas: 1) fragmentación de ADN y ligación de adaptadores específicos para crear una genoteca de ADN monocatenario (ADNss), 2) hibridación de ADNss a perlas, emulsificación de las perlas en microrreactores de agua en aceite y realizar la PCR de la emulsión para amplificar las moléculas de ADNss individuales sobre las perlas, 3) selección o enriquecimiento para perlas que contienen las moléculas de ADNss amplificadas en su superficie, 4) deposición de las perlas que portan el ADN en una placa de picotitulación (TM); y 5) secuenciación simultánea en 100,000 pozos por la generación de una señal de luz de pirofosfato. El método se explicará en más detalles más abajo.

Con respecto a esto el siguiente cálculo puede ser ilustrativo: La tecnología de secuenciación de Illumina Solexa como se describe en la presente descripción en otra parte, proporciona 40.000.000 lecturas de aproximadamente 25 bp cada una, totalizando un asombroso 1 billón bp en una única corrida. Asumiendo una redundancia en el muestreo de 10 veces, 4.000.000 fragmentos únicos se pueden evaluar en una corrida. La combinación de 100 muestras permite que se secuencien 40.000 fragmentos para cada muestra.

En una modalidad preferida, la secuenciación comprende las etapas de: (a) hibridar los fragmentos adaptados a perlas, cada perla se hibrida con un fragmento adaptado único.

(b) emulsificar las perlas en microrreactores de agua en aceite, cada microrreactor de agua en aceite comprende una única perla;

(c) cargar las perlas en pozos, cada pozo comprende una perla única; y generar una señal de pirofosfato.

En una primera etapa (a), los adaptadores de secuenciación están ligados a fragmentos dentro de la genoteca de combinación. Dicho adaptador de secuenciación incluye al menos una región "clave" para la hibridación a una perla, una región iniciadora de la secuenciación y una región iniciadora de la PCR. Así, se obtienen los fragmentos adaptados. En una primera etapa, los fragmentos adaptados se hibridan a perlas, cada perla se hibrida con un único fragmento adaptado. Al conjunto de fragmentos adaptados, se añaden las perlas en exceso como para garantizar la hibridación de un único fragmento adaptado por perla para la mayoría de las perlas (distribución de Poisson). En una siguiente etapa, las perlas se emulsifican en microrreactores de agua en aceite, cada microrreactor de agua en aceite comprende una única perla. Los reactivos de PCR están presentes en los microrreactores de agua en aceite lo que permite que una reacción de PCR tenga

lugar dentro del microrreactor. Subsecuentemente, los microrreactores se rompen, y las perlas que comprenden ADN (perlas positivas a ADN) están enriquecidas.

5 En una siguiente etapa, las perlas se cargan en pozos, cada pozo comprende una única perla. Los pozos son preferentemente parte de una placa de micro titulación (TM) lo que permite la secuenciación simultánea de una gran cantidad de fragmentos. Después de la adición de las perlas que portan la enzima, la secuencia de los fragmentos se determina por medio del uso de pirosecuenciación. En etapas sucesivas, la placa de picotitulación (TM) y las perlas así como las perlas de enzimas en la misma se someten a diferentes desoxirribonucleótidos en presencia de reactivos de secuenciación convencionales, y a partir de la incorporación de un desoxirribonucleótido se genera una señal de luz que se registra. La incorporación del nucleótido correcto generará una señal de pirosecuenciación que se puede detectar. 10 La pirosecuenciación por sí misma se conoce en la técnica y se describe entre otros en www.bioetiquetaebio.com; www.pirosecuenciación.com / section technology. La tecnología se aplica adicionalmente en por ejemplo 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). En la presente invención, las perlas se equipan preferentemente con secuencias iniciadoras (unión) o partes de los mismos que son capaces de unirse a amplicones, como debe ser el caso. En otras modalidades, los iniciadores usados en la amplificación están equipados con secuencias, por ejemplo en su extremo 5', que permiten la unión de amplicones a las perlas con el objetivo de permitir la subsecuente polimerización emulsión seguida por secuenciación. Alternativamente los amplicones se pueden ligar con adaptadores de secuenciación antes de la ligación a las perlas o la superficie. Los amplicones secuenciados revelarán la identidad del identificador y así la presencia o ausencia del fragmento de restricción en la muestra. 20

Tecnologías de Illumina-Solexa

25 Uno de los métodos para la secuenciación de alto rendimiento se encuentra disponible de Illumina, Reino Unido (www.illumina.co.uk) y descrito entre otros en WO0006770, WO0027521, WO0058507, WO0123610, WO0157248, WO0157249, WO02061127, WO03016565, WO03048387, WO2004018497, WO2004018493, WO2004050915, WO2004076692, WO2005021786, WO2005047301, WO2005065814, WO2005068656, WO2005068089 y WO2005078130. En esencia, el método empieza con los fragmentos de ADN genómico ligados al adaptador. El ADN ligado al adaptador se une aleatoriamente a un denso césped de iniciadores que se unen a una superficie sólida, típicamente en una celda de flujo. 30 El otro extremo del fragmento ligado al adaptador hibrida a un iniciador complementario en la superficie. Los iniciadores se extienden en presencia de nucleótidos y polimerasas en la llamada amplificación de puente en fase sólida para proporcionar fragmentos bicatenarios. Esta amplificación de puente en fase sólida puede ser una amplificación selectiva. La desnaturalización y repetición de la amplificación de puente en fase sólida resulta en un agrupamiento denso de fragmentos amplificados distribuidos sobre la superficie. La secuenciación se inicia por la adición de cuatro nucleótidos 35 terminadores reversibles etiquetados diferentemente, iniciadores y polimerasa a la celda de flujo. Después de la primera ronda de extensión del iniciador, se detectan las etiquetas, se registra la identidad de las primeras bases incorporadas y el terminal 3' bloqueado y el fluoróforo se eliminan de la base incorporada. Después la identidad de la segunda base se determina de la misma manera y así la secuenciación continúa.

40 En la presente invención, los fragmentos de restricción ligados al adaptador los amplicones están unidos a la superficie vía la secuencia de unión al iniciador o la secuencia de iniciación. La secuencia se determina como se delineó, que incluye la secuencia identificadora y el fragmento de restricción. La tecnología Solexa disponible actualmente permite la secuenciación de fragmentos de aproximadamente 30 pares de bases. Por el diseño inteligente de los adaptadores y los iniciadores unidos a la superficie, la etapa de secuenciación lee a través del identificador de la muestra y los restos de la secuencia de reconocimiento de la endonucleasa de restricción usada. Por ejemplo cuando se usa un identificador de la muestra de 3 bp y los restos de la cortadora rara EcoRI (GAACCT) están presentes, una secuencia interna del fragmento de restricción de 7 bp puede usarse para identificar singularmente el fragmento de restricción en la muestra. 45

50 En una modalidad preferida basada en la tecnología de secuenciación Illumina-Soexa anteriormente, la amplificación de los fragmentos de restricción ligados al adaptador se realiza con un iniciador que contiene cuanto más un nucleótido selectivo en su extremo 3', preferentemente ningún nucleótido selectivo en su extremo 3', es decir el iniciador es solamente complementario al adaptador (a +0 iniciador).

55 En modalidades alternativas dirigidas a los métodos de secuenciación descritos en la presente descripción, los iniciadores usados en la amplificación pueden contener secciones específicas (como alternativa al iniciador descrito en la presente descripción o secuencias de unión a los iniciadores) que se usan en la etapa de secuenciación subsecuente para unir los fragmentos de restricción rematados por adaptador o amplicones a la superficie. Estos generalmente se describen como la región clave o la secuencia compatible al iniciador de 5'.

En una modalidad de la invención, el ácido nucleico muestra se digiere con al menos una enzima de restricción y al menos se liga un adaptador que comprende una secuencia de reconocimiento para una endonucleasa de restricción tipo II. La

subsecuente digestión de los fragmento de restricción ligados al adaptador con una endonucleasa de restricción tipo II rinde, como la distancia entre el sitio de reconocimiento y el de restricción de una endonucleasa de restricción tipo II es relativamente corta (hasta aproximadamente 30 nucleótidos), un fragmento de restricción más corto y más largo, a los que se puede ligar un adaptador compatible con un sitio de restricción IIs. Típicamente, el saliente del sitio restringido-IIs es desconocido de forma tal que se puede usar un conjunto de adaptadores que son degenerados en el saliente. Después de la amplificación (selectiva), los amplicones se pueden secuenciar. La secuencia adaptadora en esta modalidad generalmente se puede describir como: 5' - sitio de unión al iniciador- secuencia identificadora de la muestra - secuencia del extremo cohesivo tipo IIs degenerada - 3'. El iniciador de PCR asociado generalmente sigue: secuencia iniciadora - secuencia identificadora de la muestra - secuencia del extremo cohesivo tipo IIs degenerada - nucleótidos selectivos-3'. El iniciador usado para iniciar la secuenciación-por-síntesis después generalmente tiene la estructura: 5 '-sitio de unión al iniciador-3'. Una etapa de selección por tamaño puede preferirse después de la digestión con la enzima IIs para eliminar los fragmentos más pequeños. Como en esta modalidad los restos del sitio de restricción están para este tipo de enzima típicamente en el orden de 2-4 bp, esto resulta en combinación con un identificador de la muestra de 6 bp en la secuenciación de 15-17 bp de un fragmento de restricción.

Así el método de la invención es perfectamente adecuado para identificar secuencias genómicas regulatorias de genes que pertenecen al transcriptoma de una célula u organismo, sin ninguna información inicial de la secuencia o conocimiento genético anterior de la célula y/o el organismo del que la célula se deriva. En consecuencia, la región promotora, secuencias líder y otras regiones UTR 5', intrones y exones, las secuencias UTR 3' y terminadores de un gen expresado se pueden identificar de acuerdo con el presente método. Ya que no está involucrada una etapa de clonación, es además factible determinar las secuencias genómicas de genes que causan problemas en las etapas de clonación, por ejemplo genes que son tóxicos al organismo huésped, genes que codifican para proteínas regulatorias y/o genes que causan de otra manera problemas en la clonación.

Más aun, es posible en base a esta información directamente analizar todos los polimorfismos (que incluyen SNPs) asociados con los alelos de un gen expresado, si estos polimorfismos ocurren en la secuencia codificante o en la secuencia no-codificante del gen. Así, será posible detectar aberraciones en las secuencias promotoras que causan una regulación de la expresión del gen, será posible detectar mutantes con polimorfismos en los intrones que pueden causar diferentes variantes de empalme, y así sucesivamente.

Con el objetivo de aumentar una correcta interpretación de las secuencias de ácido nucleico secuenciados y las diferencias encontradas en las mismas, se puede realizar anotación automática en los fragmentos secuenciados o cóntigos. Similarmente, la información de secuencia obtenida se puede usar para comparar las secuencias con secuencias a partir de la genoteca EST. De esa forma las secuencias de intrones o las secuencias no-codificantes internas del gen, así como secuencias promotoras y UTRs 3' y 5' se pueden identificar. La genoteca EST se puede tomar del mismo organismo o de una especie relacionada.

En un aspecto adicional, la presente descripción se refiere a estuches con los cuales se puede realizar el método de la invención. Tales estuches comprenderán uno o más adaptadores y opcionalmente uno o más iniciadores complementarios a dichos adaptadores, ligasa, y/o enzimas de restricción que son específicas para cortar los adaptadores, aparte de componentes convencionales para estuches de amplificación per se, como dNTPs, polimerasa etc. Más aun, el estuche debe proporcionar instrucciones para el uso, en donde hay un esquema de instrucciones para realizar el método de la invención.

Además, la presente invención encuentra aplicación en, entre otras, el uso del método para la identificación de marcadores moleculares, para genotipado, análisis de segregantes masivos, mapeo genético, cruzamiento hacia atrás asistido por el marcador, mapeo del loci de rasgo cuantitativo, mapeo del desequilibrio de unión, y determinación de los patrones de metilación.

Ejemplo

Procedimiento de ADNc

Aislamiento de ARN y síntesis de ADNc

El ARN total se aisló a partir de la piel del fruto del manzano, *Malus x domestica*, cultivar Kanzi, siguiendo el método de Chang y otros (1993). Las manzanas se derivaron de 4 huertos diferentes y se escogieron en 5 puntos de tiempo diferentes en 2007 desde principios de agosto hasta finales de septiembre. Iguales cantidades de ARN total de estas 20 muestras se

combinaron en una muestra y se purificaron con el kit RNeasy Plus Micro (QIAGEN, Hilden, Alemania, 74034) para eliminar la contaminación de ADN genómico de acuerdo con las instrucciones del fabricante.

5 La síntesis de la primera cadena de ADNc se ejecutó con 2 µg de ARN total de entrada mediante el uso del kit Mint de síntesis de ADNc (Evrogen, Moscú, Rusia, SK001) de acuerdo con las instrucciones del fabricante. La síntesis del ADNc de cadena doble (ds) por amplificación por PCR se ejecutó mediante el uso del kit Mint de síntesis de ADNc para un óptimo de 18 ciclos. El ADNc ds obtenido se purificó mediante el uso de columnas de purificación por PCR QIAquick® (QIAGEN, 28104) y la concentración se midió mediante espectrofotómetro. El ADNc se analizó en un gel de agarosa al 1%. El ADNc ds osciló de 200 a 2000 pb.

10 *Embotamiento, fosforilación, concatenación y nebulización del ADNc*

15 El ADNc ds se embotó y fosforiló mediante el uso del kit Quick Blunting™ (New England Biolabs, Ipswich, MA, USA, E1201S). 38 µl de ADNc (8 µg) se mezcló con 5 µl 10 x tampón de embotamiento, 5 µl de mezcla de solución de desoxinucleótido 1 mM y 2 µl de mezcla de enzima de embotamiento y se incubó a temperatura ambiente durante 30 minutos, seguido por 10 minutos a 70 °C.

20 Posteriormente, 48 µl de esta mezcla de embotamiento de ADNc se concentró mediante mezcla con 10 µl 10 x tampón de reacción ADN-ligasa T4, 5 µl ADN-ligasa T4 (ambos de New England Biolabs, M0202S, 400,000 U/ml), 25 µl de polietilenglicol 8000 al 40% (p/v) y 12 µl de agua. La mezcla de ligación se incubó durante 2 horas a temperatura ambiente y la concatenación se confirmó por análisis en gel de agarosa. La ligasa T4 se inactivó para incubación a 65 °C durante 10 minutos.

25 El ADNc concatenado se sometió a cizallamiento por nebulización. La mezcla de ligación de ADNc de 100 µl se mezcló con 650 µl de tampón de nebulización (Tris-HCl 10 mM, EDTA 1 mM, glicerol al 50 %, pH 8.0) y se midió con pipeta en un nebulizador (Invitrogen, Paisley, UK, K7025-05). La nebulización se llevó a cabo siguiendo las instrucciones del fabricante durante 15 minutos a 48 psi mediante el uso de gas nitrógeno 5.0 (Praxair, Danbury, CT, Estados Unidos). Después de la centrifugación corta del nebulizador, el ADNc nebulizado colectado se transfirió a un tubo de microcentrífuga y se precipitó por adición de 2 µl de glucógeno (Sigma-Aldrich, St. Louis, MO, Estados Unidos, 20 mg/ml, G 1767), 0.1 volumen de acetato sódico 3M pH5.2 y 1 volumen de isopropanol, e incubación durante 10 minutos a -80 °C. El ADNc se sedimentó por centrifugación durante 15 minutos a 20,800 g, se lavó con etanol al 70 %, se secó y se disolvió en 50 µl de Tris-HCl 10 mM, EDTA 1 mM, pH 8.0.

30 *Dimensionamiento y embotamiento del ADNc*

35 El ADNc cortado por cizallamiento se incubó a 65 °C durante 10 minutos, se añadió tampón de carga del gel y el ADNc se dividió sobre 5 ranuras de un gel de agarosa al 2 % en tampón Tris-acetato (TAE) (Sambrook y *otros*, 1989). Después de la electroforesis, fragmentos de ADNc de 100-400 pb se aislaron del gel mediante el uso del kit de extracción de gel GenElute (Sigma-Aldrich, NA1111). Una pequeña muestra del ADNc purificado se comprobó en el gel y la concentración se encontró que era baja. Por lo tanto, los procedimientos anteriores de embotar, fosforilar, concatenación, nebulización y purificación de gel del ADNc se repitieron varias veces con un total de 24 µg ADNc ds obtenido con el kit Mint de síntesis de ADNc. Los fragmentos de ADNc de 100-400 pb se concentraron por precipitación con etanol y se disolvieron en 19 µl de agua grado biología molecular.

45 Los extremos deshilachados del ADNc cortado por cizallamiento se embotaron y se fosforilaron mediante mezcla con 2.5 µl 10 x tampón de embotamiento, 2.5 µl mezcla de solución de Desoxinucleótido 1 mM y 1 µl mezcla de enzima de embotamiento del kit Quick Blunting™ (New England Biolabs, E1201S) e incubación a temperatura ambiente durante 30 minutos, seguido por 10 minutos a 70 °C. Posteriormente, el ADNc se purificó mediante el uso del kit MinElute® (QIAGEN, 28004) de purificación por PCR.

Ligación de adaptadores y reparación de mellas

50 El adaptador ELTD-AdC se preparó por hibridación de los oligonucleótidos parcialmente complementarios ELTD-iniciador-C (5'-AGTCCGTCGCATCGCTCTTC-3') y ELTD-AdC2 (5'-GAAGAGCGATGCGACG-3'). Este adaptador es romo en un lado y tiene un extremo saliente de 4 nt (AGTC) 5' en el otro lado para llevar a cabo el ligado direccionalmente del ADNc y prevenir el ligado de múltiples adaptadores al ADNc. El adaptador ELTD-AdC además contiene el sitio de restricción SapI de corte raro:

GCTCTTCN/NNN
CGAGAAGNNNN/

5 Este sitio de restricción facilita la eliminación de ELTD-AdC del ADNc durante una etapa posterior en el protocolo. El adaptador ELTD-AdD se preparó por hibridación de los oligonucleótidos parcialmente complementarios ELTD-iniciador-D (5'-Biotina-TEG-AGTGGGTGTCCTGGGTC AAC-3') y ELTD-AdD2 (5'-GTTGACCCAGGACACC-3'). Este adaptador además tiene un extremo saliente de 4 nt (AGTG) 5' en un lado, que está marcado con biotina por medio de un brazo espaciador de tetraetilenglicol (TEG). El marcador de biotina facilitará la inmovilización del ADNc a perlas recubiertas con estreptavidina durante etapas posteriores en el protocolo. Todos los oligonucleótidos se purificaron por HPLC ordenado de Sigma-Aldrich y se disolvieron en Tris-HCl 1 mM, EDTA 0.1 mM, pH 8.0. Los adaptadores se prepararon mediante mezcla de 50 µl de cada oligonucleótido adecuado (400 µM) con 100 µl de 2 x tampón de hibridación (Tris-HCl 20 mM, NaCl 100 mM, EDTA 2 mM, pH 7.6), incubar la mezcla durante 5 minutos a 95 °C en una unidad térmica (Thermomixer Compact, Eppendorf, Hamburg, Alemania), después apagar la unidad térmica lo que permite que las muestras dentro se enfríen lentamente a menos de 30 °C (lo que toma 3 horas). Esto produjo adaptadores bicatenarios ELTD-AdC y ELTD-AdD en concentraciones de 100 pM.

20 Ambos adaptadores se ligaron al ADNc en la siguiente reacción: 9.2 µl de ADNc a partir de la columna de purificación MinElute, 1.25 µl de agua, 0.4 µl del adaptador ELTD-AdC (100 µM), 0.4 µl del adaptador ELTD-AdD (100 µM), 12.5 µl 2 x tampón de reacción de rápida ligación y 1.25 µl ADN-ligasa T4 rápida (kit Quick Ligation™, New England Biolabs, M2200S). La mezcla de ligación se incubó en 25 °C durante 20 minutos y se purificó mediante el uso del kit GenElute de limpieza de la PCR (Sigma-Aldrich, NA1020).

25 El ADNc ligado al adaptador se le reparó las mellas en la siguiente reacción: 47 µl de ADNc a partir de la columna de purificación GenElute, 8 µl 10 x tampón de reacción ThermoPol (New England Biolabs), 8 µl BSA 1 mg/ml, 2 µl dNTPs 10 mM, 1 µl ADN-polimerasa Bst 8U/µl, Fragmento grande (New England Biolabs, M0275) y 14 µl de agua. La reacción de reparación de mellas se incubó durante 30 minutos a 65 °C y se purificó mediante el uso de una columna QIAquick de purificación por PCR que rinde 50 µl del ADNc ligado al adaptador de 100-400 pb.

Amplificación del ADNc ligado al adaptador

30 El ADNc se amplificó con una ADN-polimerasa de alta fidelidad en una reacción PCR que contiene: 10 µl de ADNc a partir de la columna de purificación por PCR (QIAquick, 10 µl 5 x tampón HF Phusion™, 1 µl de dNTPs 10 mM, 2.5 µl de ELTD-iniciador-C 10 µM, 2.5 µl de ELTD-iniciador-D 10 µM, 0.5 µl ADN-polimerasa 2U/µl Phusion Hot Start (Finnzymes, Espoo, Finlandia, F-540) y 23.5 µl de agua. Primero, se llevó a cabo un ensayo para determinar el número óptimo de ciclos de PCR para la amplificación del ADNc. La mezcla de reacción se colocó en un ciclador térmico, se desnaturalizó a 98°C durante 30 seg y posteriormente se sometió a 5 ciclos de desnaturalización-hibridación-elongación: 5 seg a 98°C, 10 seg a 60°C, 15 seg a 72°C. Después de esto, se eliminó 5 µl de la mezcla de reacción y se mantuvo en hielo (muestra después de 5 ciclos). La mezcla de reacción se sometió a 3 ciclos más de PCR como anteriormente y se eliminó 5 µl y se mantuvo en hielo (muestra después de 8 ciclos). Lo anterior se repitió 5 veces más hasta que se había alcanzado un total de 23 ciclos. Las muestras de 5 µl de los ciclos 5, 8, 11, 14, 17, 20 y 23 se analizaron en un gel de agarosa 1.5%. El número óptimo de ciclos se determinó que es 17 ciclos, después de eso se alcanzó una meseta ya que más ciclos condujeron a la aparición de una mancha por encima del tamaño esperado del ADNc. Para hacer más ADNc, se prepararon dos mezclas de reacción por PCR cada una con 10 µl de ADNc como se describe anteriormente. Las mezclas de reacción se colocaron en un ciclador térmico, se desnaturalizó a 98°C durante 30 seg y posteriormente se sometieron a 17 ciclos de desnaturalización-hibridación-elongación: 5 seg a 98°C, 10 seg a 60°C, 15 seg a 72°C. Se siguió una etapa de extensión final de 5 min a 72°C. El ADNc amplificado se purificó mediante el uso de una columna QIAquick de purificación por PCR, seguido por una columna GenElute de limpieza de la PCR (Sigma-Aldrich) para eliminar iniciadores y posibles dímeros de iniciadores.

Aislamiento de ADNc monocatenario adaptado con CD

50 A continuación, el ADNc obtenido en la etapa previa se enriqueció con moléculas que portan ELTD-Ad-C en un extremo y ELTD-Ad-D en el otro extremo por unión a perlas recubiertas de estreptavidina, lavado y elución alcalina. Las moléculas de ADNc que portan ELTD-Ad-C en ambos extremos (a partir de ahora llamadas moléculas CC) no se unen a estreptavidina y se desprenden de las perlas. Las moléculas de ADNc que portan ELTD-Ad-D en ambos extremos (a partir de ahora llamadas moléculas DD) permanecerán unidas a las perlas durante la elución alcalina, ya que ambas cadenas están biotiniladas. Las moléculas de ADNc que portan ELTD-Ad-C en un extremo y ELTD-Ad-D en el otro extremo (a partir de

ahora llamadas moléculas CD) permanecerán unidas a las perlas con una cadena biotinilada, la otra cadena no biotinilada eluirá por tratamiento con NaOH.

5 Las perlas Dyna[®] M-270 paramagnéticas recubiertas con estreptavidina (Invitrogen, 653.05) se resuspendieron completamente y 50 μ l (correspondiente a 0.5 mg) de perlas se transfirió a un tubo de microcentrífuga siliconizado (Sigma-Aldrich, T4816). Las perlas se lavaron 3 veces con 100 μ l 1 x B & W tampón (Tris-HCl 5 mM, EDTA 0.5 mM, NaCl 1 M, pH 7.5) mediante el uso de un soporte Dynal magnético (MPC[®]-E-1, Invitrogen) para separación de las perlas siguiendo las instrucciones del fabricante. Las perlas se resuspendieron en 100 μ l 2 x B&W tampón que contiene Tween-20 al 0.02% para disminuir la unión específica. A continuación, 45 μ l de ADNc purificado y amplificado por PCR mezclado con 55 μ l de agua se añadió a la suspensión de perlas. El ADNc se dejó unir durante 15 min a temperatura ambiente con rotación suave del tubo. El tubo se colocó en el soporte magnético para separar las perlas del sobrenadante, el cual se transfirió a un tubo fresco. Esta fracción se llama AB (después de la unión) y contiene ADNc que no se unió a las perlas. Posteriormente, el sedimento de perlas se lavó como sigue: una vez con 200 μ l 2 x B&W tampón que contiene Tween-20 al 0.02%, una vez con 500 μ l 2 x B&W tampón que contiene Tween-20 al 0.02% y dos veces con 500 μ l de agua. Después de la resuspensión de las perlas en agua por segunda vez, las perlas se transfirieron a un tubo siliconizado fresco antes de transferir el magneto. Finalmente, las perlas se resuspendieron en 250 μ l de NaOH 0.1 M preparado en fresco y el tubo se roto suavemente durante 2-3 minutos. El sobrenadante representa la primera elución. Las perlas se resuspendieron de nuevo en 250 μ l de NaOH 0.1 M y se rotaron suavemente durante 2-3 minutos, este sobrenadante representa la segunda elución. Las eluciones se mezclaron cada una con 1250 μ l de tampón PBI (kit QIAquick de purificación por PCR) y 7.2 μ l de ácido acético al 20% y se purificó a través de columnas QIAquick de purificación por PCR. Además, la fracción AB se purificó a través de una columna de QIAquick purificación por PCR. Las perlas restantes se lavaron una vez con 200 μ l de agua, una vez con 200 μ l de Tris-HCl 10 mM, EDTA 1 mM, pH 8.0, una vez de nuevo con 200 μ l de agua y finalmente se resuspendió en 50 μ l de agua y se almacenó a 4 °C.

25 Cinco microlitros de la fracción AB y ambas eluciones alcalinas se comprobaron en un gel de agarosa al 1.5 %. El ADNc se encontró en la fracción AB y la primera elución alcalina, pero no en la segunda elución alcalina, que se descartó. Las reacciones de control de la PCR se llevaron a cabo en 1 μ l de la fracción AB, primera elución alcalina y las perlas (volumen total de 50 μ l cada uno). Cada una se mezcló con 12.5 μ l de REDTaq[®] ReadyMix[™] (Sigma-Aldrich, R2523), 1 μ l de ELTD-iniciador-C 10 μ M o 1 μ l de ELTD-iniciador-D 10 μ M o [1 μ l de ELTD-iniciador-C 10 μ M y 1 μ l de ELTD-iniciador-D 10 μ M] y agua hasta un volumen total de 25 μ l. Las condiciones de la PCR fueron: 1 min@94°C, (30 seg@94°C, 30 seg@50°C, 30 seg@72°C) durante 6, 9, 12 y 15 ciclos, 5 min@72°C. De cada reacción se cargó 5 μ l en un gel de agarosa al 1.5 %. Los resultados mostraron que no hubo más moléculas CC que DD y CD en la fracción AB como se esperaba, porque las moléculas CC no se unen a las perlas. En la elución alcalina y la fracción de perlas la ocurrencia fue: CC < DD < CD. La conclusión fue que hubo enriquecimiento por moléculas CD en el eluato alcalino como se esperaba, pero esas moléculas CC y especialmente DD estuvieron además aún presentes, probablemente debido a una unión específica de CC y unión incompleta de moléculas DD a las perlas. La elución alcalina (que se purificó con QIAquick) se llama el ADNc monocatenario enriquecido con CD.

Amplificación del ADNc enriquecido con CD

40 Después de probar las condiciones de PCR óptimas, el ADNc monocatenario enriquecido con CD se amplificó como sigue. Se programaron dieciséis reacciones de PCR cada una contiene: 0.5 μ l de la primera elución alcalina purificada en columna QIAquick de las perlas Dyna descrito anteriormente, 10 μ l 5 x tampón HF Phusion[™], 1 μ l de dNTPs 10 mM, 2.5 μ l de ELTD-iniciador-C 10 μ M, 2.5 μ l de ELTD-iniciador-D 10 μ M, 0.5 μ l 2 U/ μ l de ADN-polimerasa Phusion Hot Start (Finnzymes, F-540) y 33 μ l agua. Las mezclas de reacción se colocaron en un ciclador térmico, se desnaturalizó en 98°C durante 30 seg y posteriormente se sometió a 11 ciclos de desnaturalización-hibridación-elongación: 5 seg a 98°C, 10 seg a 60°C, 15 seg a 72°C. Se siguió una etapa de extensión final de 5 min a 72°C. El ADNc amplificado se purificó mediante el uso de tres columnas QIAquick de purificación por PCR. El ADNc purificado se analizó en un gel de agarosa al 1.2 % y la concentración se midió mediante espectrofotómetro. Se obtuvo un total de 27.5 μ g de ADNc de doble cadena enriquecido con CD.

Unión del ADNc enriquecido con CD a perlas Dyna

50 Cinco microgramos de ADNc de doble cadena enriquecido con CD de la etapa previa se usó para unirse a perlas Dyna M-270. El procedimiento descrito anteriormente como 'Aislamiento de ADNc monocatenario adaptado con CD' con las siguientes modificaciones. Una cantidad de 27.32 μ l correspondiente a 5 μ g del ADNc enriquecido con CD se mezcló con agua hasta un volumen de 100 μ l en total, y esta mezcla se añadió a las perlas en 100 μ l 2 x B&W tampón que contiene Tween-20 al 0.02%. Después de la unión de ADNc y lavado con 2 x B&W tampón que contiene Tween-20 0.02% y lavado

con agua, las perlas se lavaron 2 veces con 200 μ l 1 x NEBuffer 4 (New England Biolabs). Finalmente, las perlas con el ADNc unido se resuspendieron en 100 μ l 1 x NEBuffer 4 y se transfirieron a un nuevo tubo de microcentrífuga siliconizado.

Digestión con SapI de la preparación de perlas con ADNc enriquecido con CD

5 El ADNc enriquecido con CD sobre las perlas se digirió con SapI para eliminar el adaptador ELTD-AdC de las moléculas de ADNc, mientras el ADNc permaneció unido a las perlas mediante el adaptador biotinilado ELTD-AdD. Cinco microlitros de Sap I (2 U/ μ l, New England Biolabs, R0569) se añadió a la suspensión de perlas con ADNc y se incubó durante 1.5 horas a 37°C. Las perlas se agitaron mediante vortex a 1400 rpm cada 10 minutos para mantenerlas en suspensión durante esta etapa. A continuación, las perlas se colocaron en el soporte magnético durante 1 min para separar las perlas, el sobrenadante se descartó y las perlas se lavaron dos veces con 500 μ l 2 x B& W tampón que contiene Tween-20 al 0.02% y dos veces con 500 μ l de agua.

Elución alcalina para preparar la genoteca de perlas con ADNc monocatenario enriquecido con CD

15 El sedimento de perlas se resuspendió en 250 μ l de NaOH 0.1 M (preparado en fresco) y el tubo se rotó suavemente durante 2-3 minutos. El tubo se puso en el soporte magnético durante 1min y el sobrenadante (= eluato alcalino) se transfirió a un tubo fresco. El eluato alcalino se mezcló con 1250 μ l de tampón PBI (kit QIAquick de purificación por PCR) y 7.2 μ l de ácido acético al 20% y se purificó a través de la columna QIAquick de purificación por PCR. Las perlas restantes se lavaron una vez con 200 μ l de agua, una vez con 200 μ l de Tris-HCl 10 mM, EDTA 1 mM, pH 8.0, una vez más con 200 μ l de agua y finalmente se resuspendieron en 50 μ l de agua y se almacenaron a 4 °C. Esta es la genoteca de perlas con ADNc monocatenario enriquecido con CD, lista para hibridación con el ADN genómico.

20 El eluato alcalino purificado se analizó en gel de agarosa al 1.2% junto con la fracción después de la unión a partir de la 'Unión del ADNc enriquecido con CD a perlas Dyna' y unas series de diluciones del ADNc bicatenario de concentración conocida a partir de la 'Amplificación del ADNc ligado al adaptador'. Aproximadamente 2 μ g del ADNc se encontró en la fracción AB y no estaba unido a las perlas Dyna. El eluato alcalino mostró una mancha del tamaño esperado. La concentración del eluato alcalino se midió mediante espectrofotómetro y se encontró que 410 ng del ADNc monocatenario se habían eluido de las perlas. En teoría, una cantidad igual de las cadenas de ADNc complementarias deben estar en la genoteca de perlas con ADNc monocatenario enriquecido con CD, correspondiente a aproximadamente 4 pmol asumiendo un tamaño promedio de 300 nt.

ADN genómico

30 *Aislamiento de ADN genómico*

ADN genómico (ADNg) se aisló de hojas Kanzi siguiendo el protocolo de Kobayashi y *otros* (1998). Después del tratamiento con ribonucleasa de acuerdo con Kobayashi y *otros* (1998), el ADNg se precipitó con concentrado de sal para eliminar impurezas por adición de dos tercios del volumen de NaCl 5M y dos volúmenes de etanol (p.a.), seguido por centrifugación durante 15 minutos a 20.000 g, lavar el sedimento con etanol al 70%, secar y disolver el sedimento en Tris-HCl 10 mM, EDTA 1 mM, pH 8.0.

Digestión con enzima de restricción del ADNg

40 El ADNg se fragmentó por digestión con enzima de restricción para crear fragmentos no solapados. Los fragmentos solapados pueden interferir con la etapa de hibridación posterior en el protocolo, lo que conduce a redes de fragmentos que hibridan. Los digestos de restricción se seleccionan que rindan principalmente fragmentos de 1-3 kb. El ADNg se digirió con HindIII/BstYI y con EcoRI/BstYI para crear dos conjuntos de fragmentos diferentes. Veinte microgramos de ADNg se digirieron por adición de 10 μ l de NEBuffer 2, 3 μ l de EcoRI (New England Biolabs, 20U/ μ l, R0101) o 1 μ l de HindIII (New England Biolabs, 100 U/ μ l, R0104) y agua hasta un volumen total de 100 μ l, seguido por incubación a 37°C durante 1 hora. Posteriormente, 6 μ l de BstYI (New England Biolabs, 10U/ μ l, R0523) se añadieron a cada tubo, seguido por incubación a 60°C durante 1 hora. Los ADN digeridos se cargaron cada uno en 4 carriles de un gel de agarosa al 1% y se separaron. Los fragmentos entre 1 y 3 kb se cortaron del gel y se purificaron mediante el uso del kit GenElute de extracción de gel (Sigma-Aldrich, NA1111). El procedimiento anterior se repitió una vez para generar suficientes fragmentos de ADN. Los fragmentos genómicos de EcoRI/BstYI (EB) y HindIII/BstYI (HB) se ligaron a adaptadores ELTD-AdE-Eco, ELTD-AdE-Hind y ELTD-AdF-Bst.

Ligación del adaptador al ADNg

5 El adaptador ELTD-AdE-Eco se preparó por hibridación de los oligonucleótidos parcialmente complementarios ELTD-AdE-Eco1 (5'-CTTGTAGGGCACGGGTCGAGAG-3') y ELTD-AdE-Eco2 (5'-AATTCTCTCGACCCGTGCCCTA-3'). Este adaptador tiene un extremo saliente 5'-AATT en un lado, que es compatible con el saliente de EcoRI de los fragmentos de ADNg y un extremo saliente 5'-CTTG en el otro lado. Estos extremos salientes logran direccionalidad de la ligación con el ADNg y previenen la ligación de múltiples adaptadores al ADNg. El adaptador ELTD-AdE-Hind se preparó por hibridación de los oligonucleótidos parcialmente complementarios ELTD-AdE-Hind1 (5'-CTTGTAGGGCACGGGTCGGAGA-3') y ELTD-AdE-Hind2 (5'-AGCTTCTCCGACCCGTGCCCTA-3'). El análogo a ELTD-AdE-Eco, el adaptador de ELTD-AdE-Hind tiene un saliente 5'-AGCT compatible con HindIII en un lado y un saliente 5'-CTTG en el otro lado. El adaptador ELTD-AdF-Bst se preparó por hibridación de los oligonucleótidos parcialmente complementarios ELTD-AdF-Bst1 (5'-GAATGGCTGGGAGAGTGCTGAG-3') y ELTD-AdF-Bst2 (5'-GATCCTCAGCACTCTCCAGCC-3'). El análogo a ELTD-AdE-Eco, el adaptador de ELTD-AdF-Bst tiene un saliente 5'-GATC compatible con BstYI en un lado y un saliente 5'-GAAT en el otro lado. Todos los oligonucleótidos se ordenaron purificados por HPLC de Sigma-Aldrich y se disolvieron en Tris-HCl 1 mM, EDTA 0.1 mM, pH 8.0.

20 Los adaptadores se prepararon mediante mezcla de 15 µl de cada oligonucleótido adecuado (800 µM) con 60 µl de 2 x tampón de hibridación (Tris-HCl 20 mM, NaCl 100 mM, EDTA 2 mM, pH 7.6) y 30 µl de agua e incubar la mezcla durante 5 minutos a 95°C en una unidad térmica, después apagar la unidad térmica lo que facilita que las muestras dentro se enfríen lentamente a menos de 30°C (lo que toma 3 horas). Esto rinde adaptadores bicatenarios ELTD-AdE-Eco, ELTD-AdE-Hind y ELTD-AdF-Bst a concentraciones de 100 µM.

25 Los adaptadores ELTD-AdE-Eco y ELTD-AdF-Bst se ligaron a los fragmentos de ADNg de 1-3 kb EcoRI/BstYI (EB) en la siguiente reacción: 1.3 µg de fragmentos EB, 0.4 µl del adaptador ELTD-AdE-Eco (100 µM), 0.4 µl del adaptador ELTD-AdF-Bst (100 µM), 40 µl de 2 x tampón de reacción de ligación rápida, 4 µl de ADN-ligasaT4 rápida (kit Quick Ligation™, New England Biolabs, M2200S) y agua hasta un volumen total de 80 µl. Los adaptadores ELTD-AdE-Hind y ELTD-AdF-Bst se ligaron a los fragmentos de ADNg de 1-3 kb HindIII/BstYI (HB) en la siguiente reacción: 1.0 µg de fragmentos HB, 0.4 µl del adaptador ELTD-AdE-Hind (100 µM), 0.4 µl del adaptador ELTD-AdF-Bst (100 µM), 40 µl de 2 x tampón de reacción de ligación rápida, 4 µl de ADN-ligasaT4 rápida (kit Quick Ligation™, New England Biolabs, M2200S) y agua hasta un volumen total de 80 µl. Las mezclas de ligación se incubaron a 25°C durante 20 minutos y se purificaron mediante el uso del kit GenElute de limpieza de la PCR (Sigma-Aldrich, NA1020).

Reparación de mellas y purificación del ADNg EB y HB

35 El ADNg ligado al adaptador se le reparó las mellas en la siguiente reacción: 40 µl de ADNg de la columna GenElute de purificación, 8 µl 10 x tampón de reacción ThermoPol, 8 µl de BSA 1 mg/ml, 2 µl de dNTPs 10 mM, 3 µl 8U/µl Bst ADN-polimerasa, fragmento grande (New England Biolabs, M0275) y 19 µl de agua. La reacción de reparación de mellas se incubó durante 30 minutos a 65°C y se purificó mediante el uso del kit GenElute de limpieza de la PCR (Sigma-Aldrich, NA1020). Esto rindió 50 µl del ADNg ligada al adaptador de 0.02 µg/µl (EB) y 0.014 µg/µl (HB), listo para hibridación con la genoteca de perlas de ADNc monocatenario enriquecido con CD.

40 La etapa de ligación del adaptador se comprobó por PCR. Un nanogramo de la preparación de EB ligado al adaptador y con mellas reparadas se usó como molde en reacciones PCR mediante el uso de cualquiera de los iniciadores ELTD-AdE-Eco1 o ELTD-AdF-Bst1 o una combinación de ambos (E, F, EF respectivamente). Análogamente, la preparación de HB ligado al adaptador se usó como molde en reacciones PCR mediante el uso de cualquier de los iniciadores ELTD-AdE-Hind1 o ELTD-AdF-Bst1 o una combinación de ambos (E, F, EF respectivamente). Para EB y HB, las reacciones PCR dieron una mancha en la región de 1-3 kb como se esperaba.

Hibridación y amplificación de fragmentos de ADN genómico

Hibridación de la genoteca de perlas con ADNc con ADNg fragmentos

50 Las condiciones de hibridación se ensayaron primero en fragmentos del gen de luciferasa de luciérnagas (Luc). Brevemente, un fragmento de Luc monocatenario de 200 nt (Luc200) con ELTD-AdC y ELTD-AdD se unió a perlas Dyna M-270 mediante el marcador de biotina de ELTD-AdD. Esta preparación de perlas con Luc se hibridó con un fragmento de Luc de 1600 nt ligado con ELTD-AdE y -AdF mezclado con un fragmento específico de ADN control de 1400 nt. Después de la hibridación y lavado, los fragmentos unidos a la sonda Luc200 se eluyeron por tratamiento alcalino y se amplificaron por PCR. Se

encontró que el fragmento de Luc de 1600 nt se eluyó a mucho más alta concentración que el fragmento control específico, que no hibrida. Hubo una diferencia de 24 ciclos entre la aparición de estos fragmentos durante la PCR ($2^{24} = 1.7 \times 10^7$ veces de enriquecimiento de Luc1600 asumiendo la eficiencia de la PCR del 100%). Las mismas condiciones de hibridación y lavado se usaron para la genoteca de perlas con ADNc y el ADNg.

Las preparaciones de ADNg EB y HB descritas anteriormente, se desnaturalizaron por calentamiento. Primero, 13 μ l de EB y 12 μ l de HB se redujeron a 5 μ l en un concentrador centrífugo a vacío. La cantidad de EB corresponde a 0.26 μ g o aproximadamente 0.2 pmol de fragmentos de 2 kb como promedio. La cantidad de HB corresponde a 0.17 μ g o aproximadamente 0.13 pmol de fragmentos de 2 kb como promedio. Las muestras después se desnaturalizaron en un bloque térmico durante 5 min a 95°C y se colocaron directamente en hielo.

La genoteca de perlas con ADNc monocatenario enriquecido con CD descrita anteriormente se dividió en dos tubos siliconizados (25 μ l de perlas cada uno). Las perlas se lavaron tres veces con 200 μ l 6 x SSC/0. SDS al 1% (pre-calentado a 60°C) y después se resuspendieron en 100 μ l 6 x SSC/SDS al 0.1% y se mantuvieron a 60°C. A un tubo se añadió ADNg EB, a el otro tubo se añadió ADNg HB. Los tubos se incubaron durante 4 horas a 60°C con rotación suave. Las perlas se lavaron dos veces rápidamente con 500 μ l pre-calentados 6 x SSC/SDS al 0.1% (60°C), tres veces 15 min con 500 μ l 6 x SSC/SDS al 0.1% a 60°C con rotación, dos veces 5 min con 500 μ l 6 x SSC a temperatura ambiente con rotación y una vez rápidamente con 500 μ l de agua.

Elongación de híbridos y amplificación por PCR de fragmentos de ADNg seleccionados

Los híbridos ADNc/ADNg formados en las perlas se usaron para la elongación de los extremos 3' del ADNc mediante el uso de ADNg hibridado como molde. Después de la etapa final de lavado con agua, las perlas se resuspendieron directamente en una mezcla de 77 μ l de agua y 20 μ l 5 x tampón HF Phusion™ y se transfirieron a un nuevo tubo siliconizado. A continuación se añadieron: 2 μ l de dNTPs 10 mM y 1 μ l de 2U/ μ l de ADN-polimerasa Phusion Hot Start (Finnzymes, F-540). La elongación fue a 72 °C durante 2 min. Las perlas se mantuvieron durante la noche a 0°C.

Las perlas se usaron para la amplificación por PCR con diferentes combinaciones de iniciadores. Las perlas hibridadas con ADNg EB se amplificaron con los siguientes conjuntos de iniciadores:

- 1) ELTD-AdE-Eco1 y ELTD-AdF-Bst1
- 2) ELTD-AdE-Eco1
- 3) ELTD-AdF-Bst1.

Las perlas hibridadas con ADNg HB se amplificaron con los siguientes conjuntos de iniciadores:

- 4) ELTD-AdE-Hind3 y ELTD-AdF-Bst1
- 5) ELTD-AdE-Hind3
- 6) ELTD-AdF-Bst1.

ELTD-iniciador E-Hind3 tiene la siguiente secuencia:

5'-GTAGGGCACGGGTCGGAGAAGC-3'. Esta es idéntica a la mayor parte de ELTD-AdE-Hind1 y contiene 3 nt extra en el extremo 3' (AGC) para ajustar en el sitio de HindIII entre el adaptador ligado y el ADNg y para conferir unión más fuerte del extremo 3' a la diana durante la hibridación.

Las reacciones PCR incluyeron 1 μ l de perlas EB o HB, 10 μ l de 5 x tampón HF Phusion™, 1 μ l de dNTPs 10 mM, 2.5 μ l de cada iniciador (10 μ M), 0.5 μ l de 2U/ μ l ADN-polimerasa Phusion Hot Start (Finnzymes, F-540) y agua hasta un volumen final de 50 μ l. Para determinar el número óptimo de ciclos, las mezclas de reacción se dividieron entre cinco tubos de PCR de 0.2 ml (10 μ l cada uno), revestidos con aceite mineral, colocados en un ciclador térmico, se desnaturalizaron a 98°C durante 30 seg y posteriormente se sometieron a 11, 14, 17, 20 y 23 ciclos de desnaturalización-hibridación-elongación: 5 seg a 98°C, 10 seg a 60°C, 2 min a 72°C. Se siguió con una etapa de extensión final de 5 min a 72°C. De cada reacción, se analizó 2.5 μ l en un gel de agarosa al 1.2%.

En las reacciones PCR con iniciadores ELTD-AdE y/o ELTD-AdF, se encontraron manchas de fragmentos que varían de 1-3 kb como se esperaba para fragmentos de ADNg. La combinación de ambos iniciadores dio productos PCR para las muestras EB y HB ya en 17 ciclos. Con ELTD-AdE solo, los productos se encontraron después de 23 ciclos, mientras la misma intensidad de productos se encontró con iniciadores ELTD-AdF después de 20 ciclos. Esto es debido a la presencia

de más sitios de restricción con BstYI (sitio de reconocimiento: RGATCY) en el genoma que los sitios con EcoRI o HindIII (GAATTC o AAGCTT).

Amplificación del ADNg híbrido seleccionado para secuenciación de alto rendimiento.

5 La reacción PCR con ADNg HB en combinación con iniciadores ELTD-AdE-Hind3 y ELTD-AdF-Bst1 y 20 ciclos dio el mejor resultado: una mancha homogénea que varía de 1-3 kb. La siguiente mezcla de reacción PCR se preparó veinte veces para aumentar suficiente ADN para la secuenciación de alto rendimiento: 1 µl de perlas HB, 10 µl de 5 x tampón HF Phusion™, 1 µl de dNTPs 10 mM, 2.5 µl de ELTD-AdE-Hind3 (10 µM), 2.5 µl de ELTD-AdF-Bst1 (10 µM), 0.5 µl de 2U/µl ADN-polimerasa Phusion Hot Start y agua hasta un volumen final de 50 µl. La mezcla reacción se colocó en un ciclador térmico, se desnaturalizó a 98°C durante 30 seg y posteriormente se sometió a 20 ciclos de desnaturalización-hibridación-elongación: 10 seg a 98°C, 10 seg a 60°C, 2 min a 72°C. Se siguió con una etapa de extensión final de 5 min a 72°C. Este llamado ADNg HB híbrido seleccionado y amplificado se purificó mediante el uso de dos columnas QIAquick de purificación por PCR paralelas. El rendimiento, que se midió mediante espectrofotómetro, fue 14.8 µg. Parte del ADNg HB híbrido seleccionado se sometió a secuenciación de alto rendimiento. Otra parte se usó para clonación y secuenciación de Sanger.

Amplificación de ADNc para secuenciación de alto rendimiento

15 Para aumentar suficiente ADNc enriquecido con CD para secuenciación de alto rendimiento, la siguiente mezcla de reacción PCR se preparó veinte veces: 0.5 µl de la primera elución alcalina purificada en columna QIAquick de las perlas Dyna descrita en 'Aislamiento del ADNc monocatenario adaptado con CD', 10 µl de 5 x tampón HF Phusion™, 1 µl de dNTPs 10 mM, 2.5 µl de ELTD-iniciador-C 10 µM, 2.5 µl de ELTD-iniciador-D-NB 10 µM, 0.5 µl de 2U/µl ADN-polimerasa Phusion Hot Start (Finnzymes, F-540) y 33 µl de agua. ELTD-iniciador D-NB tiene la misma secuencia como ELTD-iniciador D, pero no contiene un marcador de biotina en el extremo 5'. Las mezclas de reacción se colocaron en un ciclador térmico, se desnaturalizaron a 98°C durante 30 seg y posteriormente se sometieron a 11 ciclos de desnaturalización-hibridación-elongación: 5 seg a 98°C, 10 seg a 60°C, 15 seg a 72°C. Se siguió con una etapa de extensión final de 5 min a 72°C. El ADNc amplificado se purificó mediante el uso de cuatro columnas paralelas QIAquick de purificación por PCR. El ADNc purificado se analizó en un gel de agarosa al 1.2% y la concentración se midió mediante espectrofotómetro. Se obtuvo un total de 29 µg de ADNc bicatenario enriquecido con CD. El ADNc se sometió a secuenciación de alto rendimiento.

Análisis de secuencia de moléculas de ADN genómico capturado.

Después de capturar el ADN genómico con perlas de ADNc, las perlas con híbridos ADNc-ADNg se usaron para amplificación por PCR mediante el uso del conjunto de iniciadores específicos de ADN genómico ELTD-AdE-Hind3 y ELTD-AdF-Bst1. Una parte del ADN genómico se clonó y 800 clones se sometieron a secuenciación de Sanger (longitud de lectura media 781 pb). Otra parte del ADN genómico se sometió a una tecnología de secuenciación de alto rendimiento que produce lecturas de secuencia que tiene longitud media de 337 pb. Ambos conjuntos de datos de secuencia que consisten en 800 secuencias de Sanger individuales y 1370 lecturas de secuencia de alto rendimiento se analizaron en detalle y se compararon con todas las ESTs de manzano disponibles en un banco de datos públicos (NCBI, 262.411 acceso el 9 de marzo, 2009). El programa BLASTN (S. F. Altschul y otros NAR 25:3389-3402, 1997) se usó para el alineamiento global de las secuencias de ADN genómico con secuencias EST de manzano.

A partir de 800 secuencias de Sanger 488 mostraron similitud significativa (valor e 10^{-10}) con secuencias de la colección de EST del manzano en la base de datos pública y del conjunto aleatorio de 1370 secuencias de ADN genómico, 765 exhibieron similitud significativa (valor e 10^{-10}) lo que muestra que la captura específica del ADN genómico relacionado con el ADNc fue exitosa. Adicionalmente, el alineamiento local de estas secuencias de ADN genómico reveló que estas secuencias frecuentemente se extienden desde los límites 5' o 3' del ADNc o regiones de intrones, lo que muestra que la técnica facilita la identificación de áreas no codificantes. Si se toma en cuenta que los fragmentos se han secuenciado solamente a partir de un lado, que el tamaño de los fragmentos varía entre 1 Kb y 3 Kb y que más del 50% de las secuencias tienen similitud identificable con secuencias EST, parece obvio que un porcentaje mucho más alto de apareamiento de secuencias se habría encontrado si los fragmentos se habrían secuenciado completamente y construidos en contiguos antes del alineamiento con datos públicos de secuencias EST. Dado que el tamaño del genoma del manzano es alrededor de 750 Mb y el número de genes en un genoma vegetal es alrededor de 30000 con una longitud promedio de 1500 pb lo que rinde 45 Mb de secuencias codificantes en el genoma,

esto significa que mediante secuenciación aleatoria solo el 6 por ciento de los fragmentos genómicos podrían ser anotados, mientras con el método descrito encontramos por encima del 50 por ciento de anotación.

Tres casos de fragmentos de ADN genómico que abarcan los límites de EST se describen a partir del ejemplo anterior.

5

Descripción de tres casos por ejemplo.

Tres casos de secuencias genómicas de manzano que se encontraron con el método de la invención se muestran en la figura 4 y el alineamiento de secuencias más abajo.

10 Caso 1.

La secuencia genómica de manzano FRA8S6E02IN5QW (secuencia, ver más abajo) mostró 98.9% de identidad con la parte de el cóntigo de manzano cg8984st_846_7, construido a partir de 7 secuencias EST (GenBank Acc: CO899363, CO419003, CO052855, CO752637, CO901846, CN927506 y CO066317).

15

El mejor acierto de BLASTX (S. F. Altschul y otros NAR 25:3389-3402, 1997) con cóntigo cg8984st_846_7 de *Arabidopsis thaliana* fue el gen AT1G70160 (etiqueta del locus del genoma, The Arabidopsis Information Resource, <http://www.arabidopsis.org>) con e-valor $2e^{-104}$. El alineamiento de las tres secuencias mostró que FRA8S6E02IN5QW contiene parte de la secuencia codificante de proteína (CDS) de una proteína con función desconocida, que se solapa parcialmente con cóntigo cg8984st_846_7. Esto se muestra esquemáticamente en la figura 4, panel A y la secuencia de alineamiento de las partes pertinentes de las secuencias se muestran a continuación. La secuencia codificante de proteína de FRA8S6E02IN5QW es además 77.8% idéntica con AT1G70160, pero se interrumpe en el extremo 5' de la secuencia de un intrón en la misma posición que un intrón en AT1G70160 (Figuras 4, panel A y alineamiento de secuencia). La identidad entre FRA8S6E02IN5QW y AT1G70160 en esta región es baja, 37.6%. En AT1G70160 el intrón es menor que en el manzano (el límite 5' de este intrón se indica en negrita en la secuencia de alineamiento). En FRA8S6E02IN5QW, hay codones de parada en los tres marcos de lectura de la secuencia de intrón y las búsquedas BLAST indican que la CDS de FRA8S6E02IN5QW es homóloga a las proteínas similares de otras especies de plantas, pero que la homología se pierde en la región del intrón. Esto indica que FRA8S6E02IN5QW es de hecho una secuencia genómica, que contiene una secuencia de intrón de interrupción. También aquí la conservación mucho menor de secuencias de intrones indica que las secuencias de intrones contienen más SNP que las secuencias de exones.

30

Caso 2.

El clon genómico de manzano FRA8SE02HOH39 (secuencia, ver más abajo) mostró homología con el cóntigo de manzano cg12357st_1428_21, formado a partir de las 21 secuencias EST (GenBank Acc: CN930585, CV525017, CN873920, EB149394, EB121634, EB116211, CN909797, EB115871, EB154300, CN877800, EB121026, CN932122, CN860924, EB110988, CO865849, CN488473, CN497072, CN903918, CN903403, DR996731 y CN894330). El cóntigo cg12357st_1428_21 contiene la secuencia codificante completa para una proteína con similitud a la subunidad proteolítica de la proteasa Clp ATP dependiente (mejor acierto BLASTX EEF49880, *Ricinus communis*, puntuación $3e^{-128}$).

35

FRA8SE02HOH39 se solapa con la región 3' no traducida de cg12357st_1428_21, y se extiende más allá de la cola poliA en el ADN genómico no transcrito como se muestra en la figura 4. Panel B y el alineamiento de secuencias a continuación.

40

Por lo tanto, FRA8SE02HOH39 es un ejemplo de una secuencia genómica que contiene las secuencias 3' flanqueantes de una región transcrita.

Caso 3.

45

La secuencia genómica de manzano 02-H03 (secuencia, ver más abajo) se solapa con la parte 5' de 40 ESTs de manzano, con similitud con el factor de respuesta a etileno (mejor acierto BLASTX AAV66332, *Cucumis sativus*, puntuación $6e^{-34}$). Esto se muestra esquemáticamente en la figura 4, panel C, para EST 91044950 (GenBank Acc: EB155368). Las otras ESTs tenían extremos 5' similares. El alineamiento de una parte relevante de las secuencias se muestra más abajo. La secuencia genómica se extendió 744 pb corriente arriba del codón de iniciación de la traducción (ATG) del gen. No se encontró similitud de secuencia a cualquier gen fuera de la región codificante en 02-H03, lo que confirma que esta es una secuencia genómica. Una secuencia conservada, corriente arriba del inicio de la transcripción es la caja TATA. Una caja TATA putativa (TATAAA) se encontró 39 pb corriente arriba del inicio de la EST.

50

Referencias

Chang, S., Puryear, J. y Cairney J. 1993. A simple and efficient method for isolating ARN from pine trees. *Plant Mol. Biol. Rep.* 11: 113-116.

Kobayashi, N., Horikoshi, T., Katsuyama, H., Handa, T. y Takayanagi, K. 1998. A simple and efficient ADN extraction method for plants, especially woody plants. *Plant Tissue Culture and Biotechnology* 4: 76-80.

Sambrook, J., Fritsch, E.F. y Maniatis T. 1989. *Molecular cloning. A laboratory manual.* Segunda edición. Cold Spring Harbor Laboratory Press. USA.

FRA8S6E02IN5QW

TATGTTGTGATAACCATATGGCTTCCCTGACATGCTCTGAACATACTCCCA
 TGCTGCAGTAGAGTTGAATTTTGCACGCACCTCTGGATGCAAGGGAAGCA
 AGGCTATTTGTGGATTAGAACTATCCTTGAGTGTCAACTCCCACCACTCAT
 CCCATGGAATCACCGCTATAATTTCTTCACCCTGCAATATTAATTTATTAA
 TAAATGTAAAAATCAACCAAAAAGAAAAGAACTAACCACAATAAACTCTAC
 AAAAAAGAAAAGAACTAAAGCAAAGTTTAAAATAATTAAGAAATCTGTGC
 AAGATTGT CATATATTTAATTTTGTCCCTAAACAAACGCTCATCATATGTT
 CATCACTACAATCCTGATTCAACTATTATTCCTAAAGGCAAAGAACCAA
 AACATTTAGCTTAATTTCTATTCCCTAATAAATCCCAAANACATGAAATGAG
 TTGCTTGCATAAGCATATACTCAATTGAAAAT

FRA8S6E02HOH39

CAGCGACCTGTTTACGTGCAAGGTTTGGATGGAAGAGATTGAACAGTGAT
 GCCAAATTGAATTGCCTCCAGAACAAATCTGAAGGGTGCAAAAACATGT
 ACTTTTTGAGAGTTGAAGAATGACGACACTTTCTTATGTTCTATATTATCT
 TGGTTAAGTTTTTTGCAGGACGGAATGAATCCTCGTCTTTTTTTCCCTAT
 CAAAAGAAGAAAGCTGAGTTTTTATGTTTGATGTCTTGATGATGGATGA
 CCTAGTGTTCAAGTGAAAAATTCGACGGACAAAACGCTTGGCAATCCAAT
 TTGTGCCGTGTATAATGTGTACAGTCCAAAACGAGTTTCACATCGAAGAAA

02_H03

TGTTATTGTTTCATTGAAACATAACGTTACATAACAATATAGGNNNCATTI
GGAACAACTTTTAAAATGGCTGAAAACGCATTTTGTGAAAATGATTTTAA
ACAGTTTTGAGTAAAATACAATGAATCATAGAAAAGTACTTGAAATGCTI
TCTACAAATAGCATATAACTAGTGCTTATTTCAAAAAATATTNNNAAAAC#
TAAACAAAATTCTCTAAAATATTTACGGTCATTGTAAAATCATTTTCAA
CGTGATTATAATCAGGCTATGTAAAATATTCTTTAATATTGACTCAACAAT
AAACGGCGAATCGAAACGATACACGGAGTGGAGCGTGGGAGATGGGAGG
AAAGGATCACCGCACGCAATCAAAGAGTGCATTTCGCAGCCGTCAGATGAT
GATAAAAATGATGGGTGTGCTCTCTCGACAACGCACACATGCCACGTAAT
ACGGAAACGAACATTGCACAATTAATAAATTGCCACCGATGGAGAGCCGC
CCCTCCCTAATCCCATCTCAGTCAAATCCCTTGTTGACTGTGCGCCTCTCT
CT
TCATTTCTATATAAAAACCCACAGCCTGCCTCCTAGTCCTCCATCGCCATCT
CCACACCCGTTTCTCTCACATATTTTCTGCAGCCAAACACTCTTTCCACCC
AAACACTACATACACAAAACGCCACCGTTTAGTTATGGCGCCGAGAGAGA
AGACGGCCACCGCCCGGTTAGGATGAACGGTAACGGAAACGTGAAGGA
GGTGCATTTTAGAGGTGTGAGGAAGAGGCCGTGGGGGAGGTACGCCGCC
GAGATCAGA


```

FRA86E02HOH39A      ACCTTTCAGGCTTCAAGAAATCACCGACATTCGGWAGTCCATATAATATCTGGATPAAATTTCTGGACGGAGATGATGCGGTGGTCTTTTTCCTTCCGT
CG12357ST_1428_21  <CTTTTCTGACAGTCCGAGAAACGACGACACATCTTCCTTAAAGTTCGATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT
FRA86E02HOH39A      ATCGAATAAAGACAGACAGTGGATGATTTTTCCTTTGATGCTGCTTGCATGATGGATGACCTAGTGTTCACATCGAAAAATTCGACGGACAAACACGCTTGCCAAATCCA
CG12357ST_1428_21  ATCGAATAAAGACAGACAGTGGATGATTTTTCCTTTGATGCTGCTTGCATGATGGATGACCTAGTGTTCACATCGAAAAATTCGACGGACAAACACGCTTGCCAAATCCA
                                                                polyA
FRA86E02HOH39A      ATTGTGCGGTGTATATGTGTCTCAGTCCGACGAGGTTTCACATCGAAGAAA
CG12357ST_1428_21
    
```

Leyenda (alineación caso 2)
Alineación de la secuencia genómica de la manzana FRA8E02HOH39 contig EST manzana cg12357st_1428_21. Sólo se muestran las partes relevantes de la secuencia. Las coincidencias entre dos secuencias se muestran en gris. La cola de poliA de la secuencia de ADnc se indica. La secuencia genómica se extiende más allá de la región no traducida 3' del ADnc

Reivindicaciones

1. Un método para la identificación de ADN genómico en una muestra, que comprende

- 5 - aislamiento del ARNm a partir de un organismo seleccionado y preparar a partir de dicho ARNm como molde fragmentos pequeños de ADNc monocatenario con un adaptador que contiene un marcador de biotina y al menos un adaptador que contiene un sitio de restricción de la endonucleasa de restricción tipo IIs raro;
- 10 - aislamiento de ADN genómico a partir del mismo o un organismo relacionado y preparar a partir de dicho ADN genómico monocatenario fragmentos de ADN genómico ligados a las moléculas adaptadoras;
- 10 - hibridación de dichos fragmentos de ADN genómico monocatenarios con dichos fragmentos de ADNc monocatenario y amplificación de dichos híbridos; y
- secuenciación de alto rendimiento de dichos híbridos.

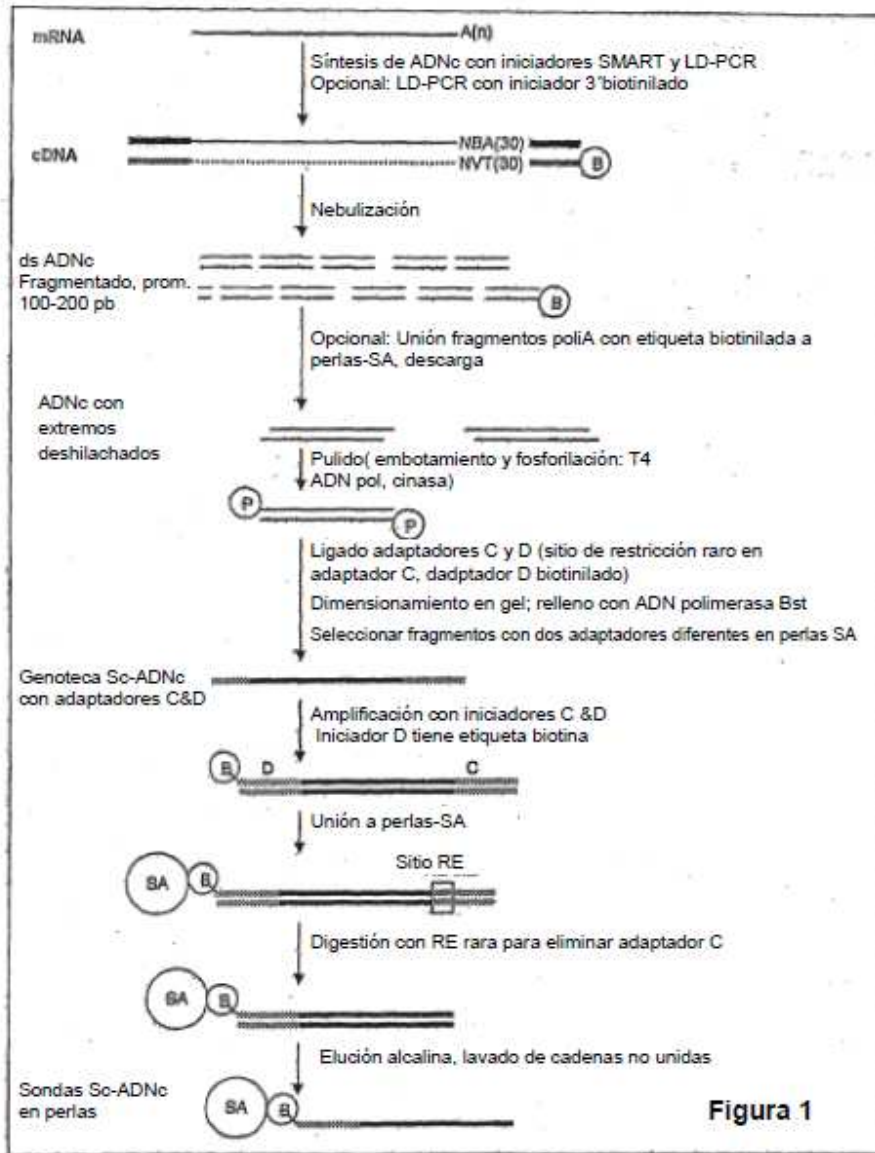
2. Método de acuerdo con la reivindicación 1, que comprende las etapas de

- 15 a) aislamiento y purificación de ARNm de las muestras de tejido de un organismo;
- b) síntesis de ADNc mediante el uso de dicho ARNm como un molde;
- c) opcionalmente reducción de la complejidad de dicho ADNc;
- 20 d) fragmentación de dicho ADNc;
- e) opcionalmente selección de tamaño de dichos fragmentos;
- f) opcionalmente eliminar los fragmentos que contienen poliA mediante la unión a perlas de afinidad recubiertas con estreptavidina;
- g) pulido de dichos fragmentos de ADNc;
- 25 h) ligación de dichos fragmentos con un adaptador que comprende un sitio de reconocimiento para una enzima de restricción rara y otro adaptador que contiene un marcador de biotina;
- i) opcionalmente selección de tamaño de dichos fragmentos;
- j) reparación de mellas de dichos fragmentos;
- k) selección de dichos fragmentos que contienen ambas secuencias adaptadoras;
- 30 l) amplificación de dichos fragmentos mediante el uso de iniciadores de hibridación para las secuencias adaptadoras descritas en la etapa h, en donde un iniciador es complementario al adaptador con un sitio de restricción raro y el otro iniciador contiene un marcador de biotina.
- m) unir de dichos fragmentos a perlas de afinidad recubiertas con estreptavidina;
- n) eliminar los adaptadores que contienen el sitio de restricción raros usando la enzima de restricción correspondiente a partir de dichos fragmentos;
- 35 o) eliminar las cadenas simples no unidas a perlas de afinidad por una interacción biotina-estreptavidina de los fragmentos de ADN bicatenarios unidos a las perlas de afinidad a través de una interacción biotina-estreptavidina que resulta en cadenas sencillas de ADN unidas a las perlas de afinidad con estreptavidina;
- p) aislamiento y purificación del ADN genómico por ejemplo a partir del organismo de la etapa a
- 40 q) fragmentación de dicho ADN genómico;
- r) opcionalmente pulir dicho ADN genómico;
- s) ligación de dicho ADN genómico con un solo tipo de adaptador o con dos tipos diferentes de adaptadores (preferido)
- t) fusionar dicho ADN genómico en un ADN monocatenario
- 45 u) hibridar el ADN genómico de la etapa t) con ADN en las perlas de la etapa o);
- v) eliminar el ADN genómico no unido por lavado
- w) extensión del híbrido del ADN genómico-ADNc por una polimerasa para crear un molde bicatenario
- x) realizar la PCR sobre dicho híbrido de ADN genómico-ADNc
- y) selección de los fragmentos mayores de aproximadamente 100 pares de base a partir de dicha PCR por fraccionamiento de tamaño
- 50 z) opcionalmente purificación de dichos fragmentos
- aa) secuenciación de alto rendimiento de dichos fragmentos .

3. Método para la identificación de polimorfismos, que comprende todas las etapas del método de acuerdo con la reivindicación 1 o 2 y además ab) comparar los datos de la secuencia de dos o más muestras para identificar los polimorfismos.

4. Método de acuerdo con la reivindicación 2 o 3, en donde las secuencias de la etapa aa) se combinan en contiguas de las secuencias solapantes individuales

5. Método de acuerdo con cualquiera de las reivindicaciones 2-4, en donde las secuencias de la etapa ab) o los cóntigos de la reivindicación 3 se anotan por la anotación automática.
- 5 6. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que las secuencias se obtienen a partir de individuos que pertenecen a una especie y se comparan con datos de EST disponibles para revelar secuencias no codificantes, tales como secuencias de intrones y secuencias no codificantes internas genes.
- 10 7. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que las secuencias se obtienen a partir de uno o más individuos pertenecientes a especies relacionadas y comparado con datos de EST disponibles revelan secuencias no codificantes, tales como secuencias de intrones y secuencias no codificantes internas de genes.
- 15 8. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que las secuencias se obtienen a partir de dos o más individuos que pertenecen a la misma especie y se compararon para revelar sitios polimórficos.
- 20 9. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que las secuencias se obtienen a partir de uno o más individuos de diferentes especies y se compararon para revelar sitios polimórficos.
- 25 10. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que las secuencias se obtienen a partir de uno o más individuos de diferentes especies y se compararon para revelar áreas conservadas en el ADN genómico.
11. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el adaptador que comprende un sitio de reconocimiento para una enzima de restricción rara desde la etapa h) comprende el sitio de reconocimiento para la enzima SapI.
- 30 12. Método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la fragmentación de los ácidos nucleicos se logra mediante nebulización.



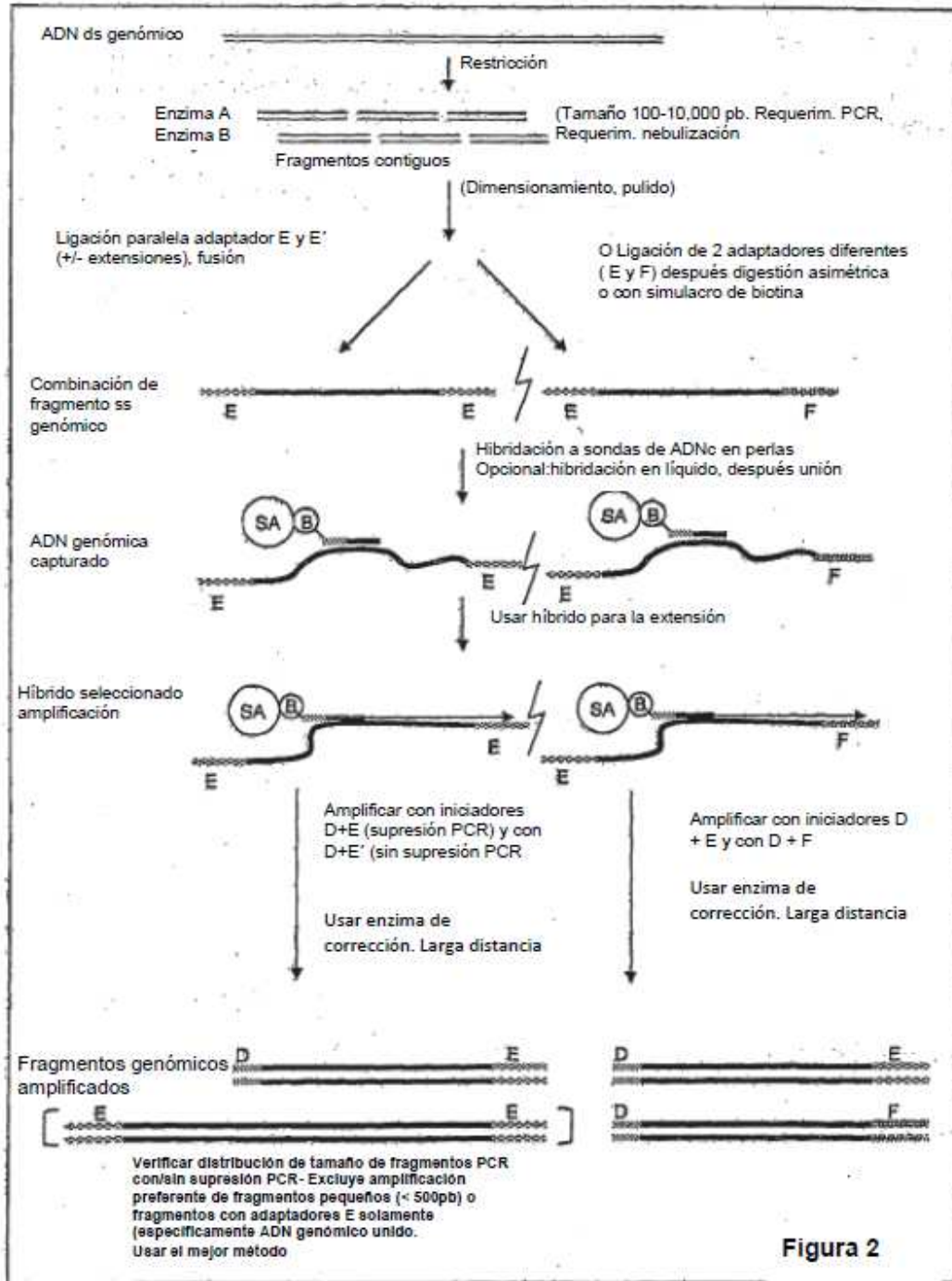


Figura 3

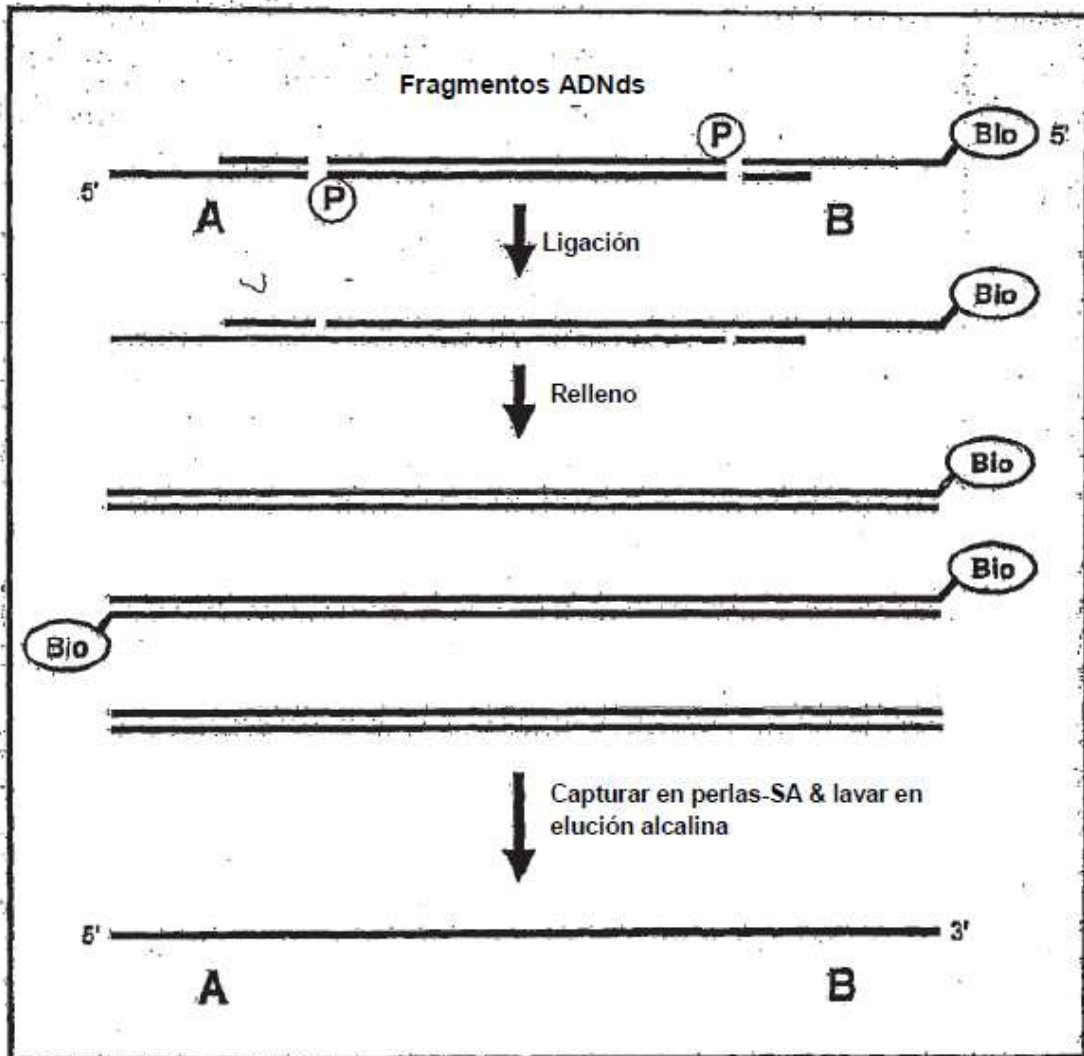


Figura 4

