

(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS
ESPAÑA



(11) Número de publicación: **2 866 044**

(51) Int. Cl.:

C12Q 1/68

(2008.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Fecha de presentación y número de la solicitud internacional: **13.02.2015 PCT/US2015/015939**

(87) Fecha y número de publicación internacional: **27.08.2015 WO15126766**

(96) Fecha de presentación y número de la solicitud europea: **13.02.2015 E 15708379 (1)**

(97) Fecha y número de publicación de la concesión europea: **24.03.2021 EP 3108009**

(54) Título: **Métodos y composiciones para la elaboración de perfiles de ADN**

(30) Prioridad:

**18.02.2014 US 201461940942 P
28.08.2014 US 201462043060 P
14.01.2015 US 201562103524 P**

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

19.10.2021

(73) Titular/es:

**ILLUMINA, INC. (100.0%)
5200 Illumina Way
San Diego, CA 92122, US**

(72) Inventor/es:

**STEPHENS, KATHRYN M.;
HOLT, CYDNE;
DAVIS, CAREY;
JAGER, ANNE;
WALICHIEWICZ, PAULINA;
HAN, YONMEE;
SILVA, DAVID;
SHEN, MIN-JUI RICHARD;
AMINI, SASAN y
STEEMERS, FRANK**

(74) Agente/Representante:

ELZABURU, S.L.P

ES 2 866 044 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Métodos y composiciones para la elaboración de perfiles de ADN

Campo de la descripción

Las realizaciones proporcionadas en este documento se refieren a métodos y composiciones para la elaboración de perfiles de ADN. Algunas realizaciones se refieren a métodos de amplificación de secuencias diana de tamaños variables en una única reacción, seguida de la secuenciación posterior del banco.

Antecedentes de la descripción

Históricamente, el uso de un subconjunto de marcadores en un genoma humano se ha utilizado para determinar la identidad personal de un individuo, o su huella o perfil de ADN. Estos marcadores incluyen ubicaciones o loci de secuencias repetidas en tandem cortas (STR) y secuencias repetidas en tandem intermedias (ITR) que, en combinación, son útiles para identificar un individuo de otro a nivel genético. El análisis de estos marcadores se ha normalizado en el análisis del ADN encontrado en escenas de crímenes. Por ejemplo, en los Estados Unidos, varias de estas secuencias repetidas se han combinado para crear un sistema de índice de ADN combinado ("Combined DNA Index System", CODIS), que sirve como patrón de laboratorio para la elaboración de perfiles de ADN en crímenes. De manera similar, otros países han adoptado un sistema convencional para la elaboración de perfiles de ADN. Estos sistemas también se han utilizado para determinar la paternidad y las relaciones familiares. Sin embargo, los sistemas actuales se basan todos en la separación por tamaño de estos loci repetidos en un sistema electroforético y, por lo tanto, se limitan al número de loci que puede diferenciarse en tal sistema. Por ejemplo, algunos de los sistemas comerciales actuales para la elaboración de perfiles de ADN con fines forenses diferencian solo 16 marcadores debido a las limitaciones de los métodos de detección electroforética.

El documento WO2006/004659A1 describe métodos para genotipificar una muestra que comprende un ácido nucleico, comprendiendo los métodos analizar los loci STR y SNP. El documento WO2012/155084A1 describe métodos para la amplificación por PCR múltiple de loci STR que pueden usarse para generar rápidamente un perfil de STR específico a partir de ácidos nucleicos diana. Rook *et al.*, Am. J. Pathol., 164 (1), 2004, 23-33., describen un método de amplificación del genoma completo para amplificar el ADN de células cancerosas microdisecionadas por captura con láser para la genotipificación de SNP y STR de alto rendimiento. Fraige *et al.*, Brazilian Archives of Biology and Technology, 56 (2), 2013, 213-221, describen el análisis de siete loci humanos STR para pruebas de paternidad mediante electroforesis de microchip. Teemu *et al.*, Nature Methods, 9 (1), 2012, 72-74, describen un método de secuenciación de ARNm y cariotipificación humana a escala genómica en *Drosophila melanogaster* utilizando identificadores moleculares exclusivos para diferenciar cada molécula de la población.

Resumen de la descripción

Las realizaciones se refieren a métodos que no están limitados por contenido y que reúnen diferentes piezas de información genética sobre un individuo para proporcionar un perfil de ADN más completo y exhaustivo de un individuo. La presente descripción describe métodos y composiciones que permiten obtener este perfil de un individuo, avanzando así en los campos de la genómica personal y forense.

La elaboración de perfiles de ADN utiliza actualmente marcadores biológicos seleccionados para determinar la identidad de una muestra de ADN. Por ejemplo, el análisis más común para determinar un perfil de ADN es determinar el perfil de una serie de secuencias cortas repetidas en tandem (STR) que se encuentran en el genoma de un organismo. El análisis consiste en amplificar secuencias STR definidas que pueden tener hasta 400 pb de longitud que se pueden diferenciar por tamaño en un gel electroforético o mediante electroforesis capilar (CE). La electroforesis se usa para detectar cambios de tamaño debido a diferencias en el número de STR repetidas en un locus dado y, como tal, la longitud de los amplicones de PCR, que para el sistema CE está entre 50-500 pb. Para ayudar a superar los límites impuestos por las metodologías de diferenciación de tamaño (es decir, las STR de tamaño de amplicón solapantes no se pueden diferenciar), los métodos actuales de elaboración de perfiles de ADN utilizan diferentes conjuntos de cebadores marcados, de modo que los amplicones que se solapan en tamaño se pueden marcar con diferentes tintes fluorescentes, por lo cual, tras la excitación, los espectros de emisión difieren, lo que permite diferenciar los amplicones solapantes utilizando diferencias en los espectros de excitación y emisión del tinte. Utilizando el marcaje diferenciado, los métodos actuales permiten la multiplexación de 24 loci STR diferentes utilizando 6 tintes detectables de forma diferente en una prueba de elaboración de perfiles de ADN.

Existen muchas limitaciones para las metodologías actuales de elaboración de perfiles de ADN. Como se mencionó anteriormente, los sistemas diferenciados por tamaño limitan el número de loci que pueden determinarse discretamente en un momento dado. Otra limitación de los métodos establecidos para la elaboración de perfiles de ADN es que el ADN que se analizará a menudo está degradado, y el intervalo de tamaño de algunos de los marcadores no se adapta al ADN degradado, por ejemplo, los amplicones pueden ser más grandes que el tamaño de los fragmentos del ADN degradado. Para el ADN degradado, los amplicones de 400 pb se consideran muy largos y pueden resultar en la pérdida de amplificación de esos loci más largos. Cuando los analistas de ADN amplifican muestras de ADN degradado para identificar su perfil de STR, por ejemplo, una muestra encontrada en la escena de un crimen, a menudo no pueden detectar todos los loci, lo que da como resultado un perfil parcial que puede dificultar o imposibilitar

la correspondencia entre un sospechoso en la escena del crimen con un crimen. Por defecto, con tales muestras, un analista de ADN tiene pocas opciones y si queda alguna muestra, se deben realizar ensayos adicionales para identificar otros marcadores que podrían dar una pista sobre la identidad del individuo, como análisis de polimorfismos de un solo nucleótido (SNP), mini-STR o ADN mitocondrial (ADNm). Sin embargo, se debe gastar una muestra valiosa en cada ensayo sin certeza de éxito en la identificación final de un individuo. La figura 1A demuestra los diferentes caminos potenciales para la identificación del ADN, todos los cuales son flujos de trabajo separados y requieren partes alícuotas de muestras valiosas. Cuando uno o más flujos de trabajo simples deben combinarse y posiblemente repetirse varias veces, el proceso resultante ya no es simple ni se usa de modo eficaz una muestra valiosa.

Las descripciones en la presente solicitud proporcionan métodos, composiciones y sistemas para determinar el perfil de ADN de un individuo u organismo mediante secuenciación de última generación (NGS) proporcionando así una solución a los problemas y limitaciones de las metodologías actuales para la elaboración de perfiles de ADN. La figura 1B muestra un ejemplo de un flujo de trabajo de los métodos descritos en una realización. En este documento se describen métodos y composiciones para combinar una multitud de marcadores pertinentes desde un punto de vista forense en un ensayo que incluyen, entre otros, repeticiones en tandem cortas (STR), repeticiones en tandem intermedias (ITR), polimorfismos de un solo nucleótido informativos de identidad (iSNP), polimorfismos de un solo nucleótido informativos de ascendencia (aSNP) y polimorfismos de un solo nucleótido informativos de fenotipo (pSNP).

La presente descripción describe ensayos que superan las limitaciones de las metodologías actuales para la elaboración de perfiles de ADN. Las realizaciones descritas proporcionan métodos para la amplificación múltiplex, preparación de bancos y secuenciación de STR, ITR, iSNP, aSNP y pSNP combinados procedentes de una muestra de ácido nucleico en una única reacción múltiplex. Los métodos descritos analizan una pluralidad de marcadores en un ensayo experimental con una manipulación mínima de la muestra, utilizando pequeñas cantidades de ADN de muestra, incluido ADN degradado. Algunas realizaciones descritas se pueden utilizar para construir bancos de datos de perfiles de ADN y/o perfiles de ADN que se pueden usar para la investigación de crímenes. Algunas realizaciones proporcionan composiciones y métodos de PCR desarrollados para ser lo suficientemente sensibles como para detectar cantidades de ADN inferiores a nanogramos. Además, los parámetros de diseño de cebadores no convencionales permiten una PCR altamente multiplexada para la identificación de STR, ITR y SNP en una reacción múltiplex. Para la investigación de crímenes, los presentes métodos y composiciones incorporan identificadores de moléculas exclusivos (UMI) que ayudan a eliminar, por ejemplo, errores de PCR y de secuenciación, tartamudeo y similares de los resultados de secuenciación. Ver Kivioja *et al.*, Nat. Meth., 9, 72-74 (2012). Además, los resultados de los métodos y las composiciones descritos en este documento son compatibles con las bases de datos existentes.

Por lo tanto, las realizaciones descritas en este documento proporcionan métodos para construir un perfil de ADN que comprenden: proporcionar una muestra de ácido nucleico, amplificar la muestra de ácido nucleico con una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo de un solo nucleótido (SNP) y al menos una secuencia diana que comprende una repetición en tandem en una reacción múltiplex para generar productos de amplificación, y determinar los genotipos del al menos un SNP y dicha al menos una repetición en tandem en los productos de amplificación, construyendo así el perfil de ADN de la muestra de ácido nucleico.

En algunas realizaciones, los métodos comprenden generar un banco de ácidos nucleicos a partir de los productos de amplificación. En algunas realizaciones, los métodos comprenden determinar las secuencias del banco de ácidos nucleicos. En algunas realizaciones, la muestra de ácido nucleico proviene de un ser humano. En algunas realizaciones, la muestra de ácido nucleico proviene de una muestra ambiental, una planta, un animal no humano, una bacteria, una arquea, un hongo o un virus. En algunas realizaciones, el perfil de ADN se usa para uno o más diagnósticos o pronósticos de enfermedades, identificación de biomarcadores del cáncer, identificación de anomalías genéticas o análisis de diversidad genética. En algunas realizaciones, el perfil de ADN se usa para uno o más de construcción de bancos de datos, análisis forense, investigación de crímenes, paternidad o identificación personal. En algunas realizaciones, dicha al menos un SNP indica la ascendencia o una característica fenotípica de la fuente de la muestra de ácido nucleico. Cada uno de la pluralidad de cebadores tiene una temperatura de fusión baja y una longitud de al menos 24 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una temperatura de fusión inferior a 60 °C. Cada uno de la pluralidad de cebadores tiene una longitud de al menos 24 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una longitud de 24 nucleótidos a aproximadamente 38 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una secuencia de nucleótidos de homopolímero. En algunas realizaciones, la muestra de ácido nucleico se amplifica mediante una reacción en cadena de la polimerasa (PCR). En algunas realizaciones, la muestra de ácido nucleico se amplifica en un tampón de amplificación que tiene una concentración de sales que es mayor, en comparación con la concentración de sales de un tampón de amplificación usado junto con cebadores diseñados de modo convencional. En algunas realizaciones, la sal comprende KCl, LiCl, NaCl o una combinación de las mismas. En algunas realizaciones, la sal comprende KCl. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 100 mM a aproximadamente 200 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es menor que aproximadamente 150 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 145 mM. En algunas realizaciones, el SNP es un SNP de ascendencia, un SNP fenotípico, un SNP de identidad o una combinación de los mismos. En algunas realizaciones, la pluralidad de cebadores se hibrida específicamente con al menos 30 SNP. En algunas realizaciones, la pluralidad de cebadores se hibrida específicamente con al menos 50 SNP. En algunas realizaciones, la repetición en tandem es una repetición en tandem corta (STR), una repetición en tandem intermedia (ITR) o una variante de la misma. En

algunas realizaciones, la pluralidad de cebadores se hibrida específicamente con al menos 24 secuencias repetidas en tandem. En algunas realizaciones, la pluralidad de cebadores se hibrida específicamente con al menos 60 secuencias repetidas en tandem. En algunas realizaciones, la muestra de ácido nucleico comprende de aproximadamente 10 pg a aproximadamente 100 pg de ADN. En algunas realizaciones, la muestra de ácido nucleico comprende de aproximadamente 5 pg a aproximadamente 10 pg de ADN. En algunas realizaciones, la muestra de ácido nucleico comprende ADN genómico. En algunas realizaciones, el ADN genómico proviene de una muestra forense. En algunas realizaciones, el ADN genómico comprende ADN degradado. En algunas realizaciones, se determinan al menos 50 % de los genotipos de dicho al menos un SNP y al menos una repetición en tandem. En algunas realizaciones, se determinan al menos el 80 % de los genotipos de dicho al menos un SNP y al menos una repetición en tandem. En algunas realizaciones, se determinan al menos el 90 % de los genotipos de dicho al menos un SNP y al menos una repetición en tandem. En algunas realizaciones, se determinan al menos el 95 % de los genotipos de dicho al menos un SNP y al menos una repetición en tandem. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una o más secuencias de marcador. En algunas realizaciones, una o más secuencias de marcador comprenden un marcador de cebador, un marcador de captura, un marcador de secuenciación, un marcador de identificador molecular exclusivo o una combinación de los mismos. En algunas realizaciones, dichas una o más secuencias de marcador comprenden un marcador de cebador. En algunas realizaciones, dichas una o más secuencias de marcador comprenden un marcador de identificador molecular exclusivo.

Las presentes descripciones proporcionan métodos para construir un banco de ácidos nucleicos que comprenden: proporcionar una muestra de ácido nucleico y amplificar la muestra de ácido nucleico con una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo de un solo nucleótido (SNP) y al menos una secuencia diana que comprende una secuencia repetida en tandem en una reacción múltiple para generar productos de amplificación.

En algunas realizaciones, la muestra de ácido nucleico no se fragmenta antes de la amplificación. En algunas realizaciones, las secuencias diana no se enriquecen antes de la amplificación. En algunas realizaciones, dicho al menos un SNP indica la ascendencia o una característica fenotípica de la fuente de la muestra de ácido nucleico. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una o más secuencias de marcador. En algunas realizaciones, una o más secuencias de marcador comprenden un marcador de cebador, un marcador de captura, un marcador de secuenciación o un marcador de identificación molecular exclusivo, o una combinación de los mismos. En algunas realizaciones, los métodos incluyen amplificar los productos de amplificación con una segunda pluralidad de cebadores. En algunas realizaciones, cada uno de la segunda pluralidad de cebadores comprende una porción correspondiente al marcador del cebador de la pluralidad de cebadores y una o más secuencias de marcador. En algunas realizaciones, una o más secuencias de marcador de la segunda pluralidad de cebadores comprende un marcador de captura, un marcador de secuenciación o una combinación de los mismos. En algunas realizaciones, los métodos incluyen la adición de proteína de unión monocatenaria (SSB) a los productos de amplificación. En algunas realizaciones, la muestra de ácido nucleico y/o los productos de amplificación se amplifican mediante la reacción en cadena de la polimerasa (PCR). En algunas realizaciones, la muestra de ácido nucleico y/o los productos de amplificación se amplifican en un tampón de amplificación que tiene una concentración de sales más alta, en comparación con la concentración de sales de un tampón de amplificación usado junto con cebadores diseñados de modo convencional. En algunas realizaciones, la sal comprende KCl, LiCl, NaCl o una combinación de las mismas. En algunas realizaciones, la sal comprende KCl. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 100 mM a aproximadamente 200 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es menor que aproximadamente 150 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 145 mM.

Las presentes descripciones proporcionan un banco de ácidos nucleicos que comprende una pluralidad de moléculas de ácidos nucleicos, en el que la pluralidad de moléculas de ácidos nucleicos comprende al menos una secuencia repetida en tandem flanqueada por un primer par de secuencias de marcador y al menos una secuencia de polimorfismo de un solo nucleótido (SNP) flanqueada por un segundo par de secuencias de marcador. Se proporciona además un banco de ácidos nucleicos construido usando los métodos y las composiciones descritos en este documento. En algunas realizaciones, dicho al menos un SNP indica la ascendencia o una característica fenotípica de la fuente de la pluralidad de moléculas de ácidos nucleicos.

Las presentes descripciones proporcionan una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana corta y al menos una secuencia diana larga en una muestra de ácidos nucleicos, en la que la amplificación de la muestra de ácidos nucleicos usando la pluralidad de cebadores en una única reacción múltiple da como resultado en al menos un producto de amplificación corto y al menos un producto de amplificación largo, en la que cada uno de la pluralidad de cebadores comprende una o más secuencias de marcador.

En algunas realizaciones, la secuencia diana corta comprende un polimorfismo de un solo nucleótido (SNP) y la secuencia diana larga comprende una repetición en tandem. En algunas realizaciones, una o más secuencias de marcador comprenden un marcador de cebador, un marcador de captura, un marcador de secuenciación, un marcador de identificación molecular exclusivo o una combinación de los mismos. Cada uno de la pluralidad de cebadores tiene una temperatura de fusión baja de menos de 60 °C y una longitud de al menos 24 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores tiene una longitud de 24 nucleótidos a aproximadamente 38 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una secuencia de

nucleótidos de homopolímero. En algunas realizaciones, la muestra de ácido nucleico se amplifica mediante la reacción en cadena de la polimerasa (PCR). En algunas realizaciones, el SNP es un SNP de ascendencia, un SNP fenotípico, un SNP de identidad o una combinación de los mismos. En algunas realizaciones, la pluralidad de cebadores se hibrida específicamente con al menos 30 SNP. En algunas realizaciones, la pluralidad de cebadores se

5 hibrida específicamente con al menos 50 SNP. En algunas realizaciones, la repetición en tandem es una repetición en tandem corta (STR), una repetición en tandem intermedia (ITR) o una variante de la misma. En algunas realizaciones, la pluralidad de cebadores se hibrida específicamente con al menos 24 secuencias repetidas en tandem. En algunas realizaciones, la pluralidad de cebadores se hibrida específicamente con al menos 60 secuencias repetidas en tandem.

10 Las descripciones de este documento proporcionan kits que comprenden al menos un medio de recipiente, en el que dicho al menos un medio de recipiente comprende una pluralidad de cebadores descritos en el presente documento.

En algunas descripciones, los kits incluyen un reactivo para una reacción de amplificación. En algunas realizaciones, el reactivo es un tampón de amplificación para la reacción en cadena de la polimerasa (PCR). En algunas realizaciones, el tampón de amplificación comprende una concentración de sales que aumenta en comparación con la concentración de sales de un tampón de amplificación usado junto con cebadores diseñados de modo convencional.

15 15 En algunas realizaciones, la sal comprende KCl, LiCl, NaCl o una combinación de las mismas. En algunas realizaciones, la sal comprende KCl. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 100 mM a aproximadamente 200 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es menor que aproximadamente 150 mM. En algunas realizaciones, la concentración de KCl en el tampón de amplificación es de aproximadamente 145 mM.

20 Breve descripción de los dibujos

La figura 1A y la figura 1B muestran las diferencias en el flujo de trabajo actual A) para la elaboración de perfiles de ADN frente a B) el de un ejemplo de una realización de la presente descripción.

La figura 2 muestra un ejemplo de realización de un método para crear un banco útil para la elaboración de perfiles de ADN.

25 25 La figura 3 muestra otro ejemplo de realización de un método para crear un banco útil para la elaboración de perfiles de ADN.

La figura 4A, la figura 4B, la figura 4C y la figura 4D son gráficos lineales que ilustran los resultados del electroferograma de cómo un par de cebadores diseñado por métodos tradicionales y siguiendo los protocolos y limitaciones de diseño de cebadores de PCR establecidos pueden causar una amplificación no específica de dianas genómicas y el oscurecimiento de la detección de amplicones deseados cuando se combinan con cebadores diseñados siguiendo los métodos de la presente descripción; A) 10 parejas de cebadores diseñados por métodos de la presente descripción dirigidos a loci SNP, B) y D) los 10 cebadores más un par de cebadores adicional diseñado por métodos tradicionales que demuestra que el par de cebadores adicional interfiere con los 10 pares de cebadores durante la amplificación, y C) los 10 pares de cebadores más un par de cebadores adicional, en el que el par de cebadores adicional también se diseñó siguiendo los métodos de la presente descripción, dando como resultado una amplificación exitosa de todos los SNP diana. El eje X es el tamaño de los fragmentos del banco (pb) y el eje Y son las unidades de fluorescencia (FU) de los picos amplificados de los fragmentos amplificados.

30 30 La figura 5A, la figura 5B, la figura 5C, la figura 5D y la figura 5E son diagramas de cajas que muestran ejemplos de resultados para un experimento que sigue el flujo de trabajo esbozado en la figura 2, que se utilizó para identificar un panel de 56 STR y una combinación de 75 SNP informativos de identidad (iSNP), SNP informativos de ascendencia (aSNP) y SNP informativos de fenotipo (pSNP) en una reacción de secuenciación y amplificación múltiple de una muestra. Se indican los resultados replicados que demuestran la amplificación y secuenciación exitosa de los loci STR del panel; A) diagrama de cajas que demuestra el equilibrio intralocus para 25 STR heterocigóticos del panel, B) diagrama de cajas que demuestra un bajo tartamudeo para la mayoría de los 56 loci STR, C) diagrama de cajas que demuestra la cobertura de secuenciación para los loci STR, D) diagrama de cajas que demuestra la cobertura de secuenciación para los SNP, y E) diagrama de cajas que demuestra el equilibrio para 22 SNP heterocigóticos del panel. La barra de error inferior indica el valor mínimo, la barra de error superior indica el valor máximo, la caja inferior indica el 25º percentil y la caja superior indica el 75º percentil, siendo la media la intersección entre las cajas superior e inferior.

40 40 La figura 6 muestra una serie de gráficos de barras que muestran ejemplos de gráficos de loci STR del experimento de la figura 5. Los gráficos muestran diferentes llamadas alélicas para los STR en el panel de la figura 5.

45 45 La figura 7A, la figura 7B, la figura 7C, la figura 7D y la figura 7E son diagramas de cajas que muestran ejemplos de resultados para un experimento que sigue el flujo de trabajo esbozado en la figura 3, que se utilizó para identificar un panel de 26 STR y una mezcla de 94 iSNP, aSNP y pSNP en una reacción de secuenciación y amplificación múltiple de una muestra. Se indican los resultados replicados que demuestran la amplificación y secuenciación exitosa de los STR del panel; A) diagrama de cajas que demuestra el equilibrio intralocus para 21 loci STR heterocigóticos del panel, B) diagrama de cajas que demuestra un bajo tartamudeo para los 26 loci STR (39 de los 47 alelos de los 26 loci no mostraron tartamudeo), C) diagrama de cajas que demuestra la cobertura de secuenciación para los loci STR (números de lectura normalizados utilizando los UMI), D) diagrama de cajas que demuestra la cobertura de secuenciación

para los SNP y E) diagrama de cajas que demuestra el equilibrio para 21 iSNP heterocigotos del panel. La barra de error inferior indica el valor mínimo, la barra de error superior indica el valor máximo, la caja inferior indica el 25º percentil y la caja superior indica el 75º percentil, siendo la media la intersección entre las cajas superior e inferior.

5 La figura 8 muestra una serie de gráficos de barras que muestran ejemplos de gráficos de loci STR del experimento de la figura 7. Los gráficos muestran diferentes llamadas alélicas para los STR en el panel de la figura 7.

10 La figura 9 muestra gráficos de barras de muestras analizadas sin UMI y con UMI. El panel izquierdo de cada conjunto representa las muestras analizadas sin UMI y el panel derecho de cada conjunto representa las muestras analizadas con UMI. El eje X designa el número de repeticiones del STR y el eje Y designa el número de recuento del alelo en particular. Las líneas de error dentro de las barras separan el error de secuencia (parte superior de la barra) de la secuencia correcta (parte inferior de la barra) dentro de la secuencia STR.

15 La figura 10A y la figura 10B muestran ejemplos de resultados de un experimento en el que la proporción de ADN era de 90:10 hembra:macho. A) Un subconjunto de resultados de llamadas de loci STR para loci STR cuando se usan los métodos actuales de elaboración de perfiles de ADN de electroforesis capilar, y B) varios resultados de llamadas de loci STR para varios loci STR cuando se usan los métodos de la presente solicitud. Tanto los métodos de CE como los métodos de la presente solicitud detectaron el bajo nivel de contaminación del ADN masculino.

20 La figura 11 muestra gráficos de barras que demuestran que se detectaron loci STR específicos del cromosoma Y en el experimento de la figura 9, lo que demuestra además que la presente solicitud puede detectar ADN masculino contaminante y loci STR específicos de ese ADN masculino, mientras que, con las metodologías actuales de CE, sería necesario realizar dos experimentos.

25 La figura 12 es una tabla que muestra ejemplos de resultados de secuenciación de alto nivel de un experimento con 12 individuos de muestra y un individuo de referencia, lo que demuestra la coherencia de las llamadas STR y SNP entre dos réplicas.

La figura 13 es una tabla que muestra ejemplos de estadísticas de población del experimento mostrado en la figura 12.

30 La figura 14 es una tabla que muestra ejemplos de predicciones de fenotipo basadas en el genotipo de pSNP del experimento mostrado en la figura 12.

La figura 15 es un gráfico que muestra un ejemplo de un cartografiado de ascendencia basado en el genotipo de aSNP del experimento mostrado en la figura 12.

35 La figura 16A, la figura 16B, la figura 16C, la figura 16D y la figura 16E son gráficos de barras que muestran ejemplos de gráficos de loci STR del experimento de la figura 12.

La figura 17A y la figura 17B son gráficos de barras que muestran ejemplos de gráficos de SNP del experimento de la figura 12.

40 La figura 18A y la figura 18B muestran diagramas de cajas que muestran el equilibrio intralocus de ejemplos de loci STR y SNP del experimento de la figura 12.

35 La figura 19A y la figura 19B son gráficos que muestran análisis de tartamudeo de ejemplos de loci STR del experimento de la figura 12.

La figura 20 es una tabla que muestra ejemplos de heterocigotos isométricos en loci STR del experimento de la figura 12.

45 La figura 21 es un diagrama de bloques que muestra un ejemplo de diagrama de herencia basado en variantes en el STR D8S1179 del experimento de la figura 12.

La figura 22 es un diagrama de bloques que muestra un ejemplo de diagrama de herencia basado en variantes en el STR D13S317 del experimento de la figura 12.

La figura 23 es una tabla que muestra ejemplos de resultados de genotipificación utilizando ADN degradado.

La figura 24A y la figura 24B muestran resultados de ejemplos de genotipificación de STR y equilibrio intralocus en diferentes entradas de ADN.

45 La figura 25A y la figura 25B muestran resultados de ejemplos de genotipificación de SNP y equilibrio intralocus en diferentes entradas de ADN.

Descripción detallada

Definiciones

Como se usa en este documento, las formas singulares "un", "una", "el" y "la" incluyen las referencias en plural a menos que se indique lo contrario, expresamente o por contexto. Por ejemplo, "un" dímero incluye uno o más dímeros,

a menos que se indique lo contrario, expresamente o por contexto.

Como se usa en el presente documento, las expresiones "perfil de ADN", "huella genética" y "perfil genotípico" se usan indistintamente en el presente documento para referirse a las variaciones alélicas en una colección de loci polimórficos, tales como una repetición en tandem, un polimorfismo de un solo nucleótido (SNP), etc. Un perfil de ADN es útil en

5 medicina forense para identificar a un individuo basándose en una muestra de ácido nucleico. El perfil de ADN, como se usa en este documento, también se puede usar para otras aplicaciones, como el diagnóstico y pronóstico de enfermedades que incluyen cáncer, identificación de biomarcadores del cáncer, análisis de herencia, análisis de diversidad genética, identificación de anomalías genéticas, construcción de bancos de datos, análisis forense, investigaciones de crímenes, paternidad, identificación personal, etc.

10 Los términos y expresiones "polinucleótido", "oligonucleótido", "ácido nucleico" y "molécula de ácido nucleico" se usan indistintamente en este documento para referirse a una forma polimérica de nucleótidos de cualquier longitud y pueden comprender ribonucleótidos, desoxirribonucleótidos, análogos de los mismos o mezclas de los mismos. Estos términos y expresiones se refieren solo a la estructura primaria de la molécula. Por tanto, los términos y expresiones incluyen ácido desoxirribonucleico ("ADN") de cadena triple, doble y sencilla, así como ácido ribonucleico ("ARN") de cadena triple, doble y sencilla.

15 Como se usa en el presente documento, "identidad de secuencia" o "identidad" u "homología" en el contexto de dos secuencias de nucleótidos incluye la referencia a los restos en las dos secuencias que son iguales cuando se alinean para una correspondencia máxima en una ventana de comparación especificada. La porción de la secuencia de nucleótidos en la ventana de comparación puede comprender adiciones o delecciones (es decir, espacios) en comparación con la secuencia de referencia para el alineamiento óptimo de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que aparece el resto de base de ácido nucleico idéntico en ambas secuencias para obtener el número de posiciones concordantes, dividiendo el número de posiciones concordantes entre el número total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para producir el porcentaje de identidad de secuencia.

20 25 Como se usa en el presente documento, "sustancialmente complementario o sustancialmente concordante" significa que dos secuencias de ácido nucleico tienen al menos un 90 % de identidad de secuencia. Preferiblemente, las dos secuencias de ácido nucleico tienen al menos 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia. Como alternativa, "sustancialmente complementario o sustancialmente concordante" significa que dos secuencias de ácido nucleico se pueden hibridar bajo condiciones de alta rigurosidad.

30 35 Se entiende que los aspectos y las realizaciones de la invención descrita en este documento incluyen aspectos y realizaciones "que consisten" y/o "que consisten esencialmente en".

Otros objetos, ventajas y características de la presente invención resultarán evidentes a partir de la siguiente memoria descriptiva tomada junto con los dibujos adjuntos.

Métodos para construir un perfil de ADN

40 45 50 55 Las metodologías establecidas para determinar el perfil de ADN están limitadas de varias maneras. Por ejemplo, los métodos actuales detectan cambios de tamaño de loci amplificados que difieren debido a cambios en la longitud de las secuencias repetidas en tandem que se encuentran en una muestra de ADN. Para multiplexar amplificaciones de STR para la visualización, las amplificaciones deben diseñarse para espaciar los diferentes tamaños de amplicones dentro de los límites de separación por tamaño del sistema electroforético, que para CE es de aproximadamente 50-500 pb. Como tal, solo se puede visualizar un número limitado de secuencias repetidas en un ensayo. Por ejemplo, el kit de amplificación GLOBALFILER PCR (APPLIED BIOSYSTEMS) es capaz de diferenciar 24 loci STR mediante el uso de 6 tintes diferentes. Además, tales métodos tienen problemas cuando el ADN de la muestra está degradada, tal como es habitual con las muestras de ADN de la escena de un crimen, de modo que no es posible obtener productos de amplificación más largos, lo que da como resultado un perfil de ADN incompleto. Los métodos actuales tampoco suelen ser lo suficientemente sensibles como para detectar pequeñas cantidades de ADN contaminante, por lo que una muestra mixta puede pasar desapercibida y no ser referida, lo cual podría ser crucial para la investigación de crímenes. Como tal, los métodos actuales pueden conducir a unos resultados incompletos que conducen a unos resultados no concluyentes, lo que puede ser perjudicial para la elaboración de perfiles de ADN.

Además, las dianas actuales no incluyen información sobre la ascendencia de la muestra, rasgos fenotípicos como el posible color de ojos y otra información de la muestra individualizada. Algunas metodologías de secuenciación han intentado incluir detección de STR y SNP. Por ejemplo, se ha intentado la preparación de un banco, seguida de un enriquecimiento personalizado para los STR y SNP; aunque no se cubren completamente todos los STR, ya que los métodos de preparación del banco generalmente implican el cizallamiento de muestras que puede borrar la secuencia diana. Además, los métodos y protocolos de diseño de cebadores establecidos pueden proporcionar conjuntos de cebadores para amplificar secuencias largas (por ejemplo, STR) o secuencias cortas (por ejemplo, SNP), pero las combinaciones de ambos en una reacción no han tenido éxito.

La presente descripción describe soluciones a los problemas y limitaciones de los sistemas de perfiles de ADN actuales. Los métodos y las composiciones descritos en este documento permiten la combinación de STR y SNP en

un ensayo usando PCR para amplificar las dianas y generar bancos para la secuenciación. Mientras se desarrollaban los presentes ensayos, se descubrió inesperadamente que, por ejemplo, cuando se utiliza un diseño de cebadores no convencional y contrario a la intuición, tanto los STR como los SNP pueden amplificarse en una reacción que permite determinar la secuencia para todos los loci diana. Sorprendentemente, cuando se diseñaron cebadores de amplificación utilizando parámetros contrarios al dogma actual que rodea al diseño de cebadores, se crearon cebadores que permitieron que las regiones STR más largas se amplificaran y que las regiones SNP cortas se amplificaran de una manera más o menos equilibrada, lo que permitió que ambos STR y SNP pudieran amplificarse juntos en una reacción múltiple.

Los métodos y las composiciones descritos en este documento para determinar el perfil de ADN de un organismo se pueden usar siempre que se deseen conjuntos de amplicones de diferentes tamaños a partir de una reacción de amplificación fuera de la elaboración de perfiles de ADN. Por ejemplo, si las dianas de interés para la PCR incluyen tanto regiones de genes grandes como regiones SNP cortas que pueden dar como resultado amplicones que varían en tamaño de cientos a miles de pares de bases, frente a amplicones de menos de 100 pares de bases, respectivamente, entonces los métodos y las composiciones descritos en el presente documento podrían permitir la amplificación simultánea con éxito de las dianas del gen y de SNP, lo que no habría sido posible sin la práctica de los métodos descritos. Además, los métodos y las composiciones descritos en este documento pueden aplicarse a cualquier organismo, por ejemplo, seres humanos, primates no humanos, animales, plantas, virus, bacterias, hongos y similares. Como tales, los presentes métodos y composiciones no solo son útiles para la elaboración de perfiles de ADN (por ejemplo, análisis forense, paternidad, identificación individual, etc.) y humanos como genoma diana, sino que también podrían usarse para otras dianas tales como marcadores del cáncer y de enfermedades, marcadores de anomalías genéticas y/o cuando el genoma diana no es humano.

Por lo tanto, las realizaciones descritas en este documento proporcionan métodos para construir un perfil de ADN que comprenden: proporcionar una muestra de ácido nucleico, amplificar la muestra de ácido nucleico con una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo de un solo nucleótido (SNP) y al menos una secuencia diana que comprende una repetición en tandem, y determinar los genotipos de dicho al menos un SNP y al menos una repetición en tandem en los productos de amplificación, construyendo así el perfil de ADN de la muestra de ácido nucleico.

Los expertos en la técnica apreciarán que se puede usar cualquier técnica adecuada para determinar los genotipos de las secuencias diana que incluye, pero no se limita a hibridación basada en matrices, secuenciación o similares. Por lo tanto, en algunas realizaciones, los métodos descritos en este documento pueden comprender generar un banco de ácidos nucleicos, tal como un banco de secuenciación, a partir de los productos de amplificación y determinar las secuencias del banco de ácidos nucleicos.

En algunas realizaciones, la presente descripción proporciona métodos y composiciones para la elaboración de perfiles de ADN que comprenden la identificación concurrente de STR e iSNP, por ejemplo, para su uso en la construcción de bancos de datos de población o personales. En tales bancos de datos, los datos personales no son forzosamente necesarios, ya que normalmente se conoce a las personas. Sin embargo, si se desea información adicional, se pueden agregar dianas de información adicionales para la identificación concurrente. Las repeticiones cortas en tandem son bien conocidas en la técnica y consisten en secuencias repetidas de di- o trinucleótidos. Las repeticiones en tandem intermedias se consideran típicamente secuencias repetidas de entre 4 y 7 secuencias de nucleótidos. Los SNP utilizados en este documento pueden ser de cualquier forma que pueda ofrecer información sobre las características físicas de una persona. Los ejemplos en el presente documento son SNP que proporcionan pistas sobre la ascendencia o herencia (aSNP) y aquellos que proporcionan pistas sobre características fenotípicas (SNP informativos de fenotipos). En los métodos descritos en el presente documento, un ensayo de perfil de ADN podría incluir cualquier número de estos SNP en combinación con determinaciones de loci STR e ITR.

Por ejemplo, la presente descripción proporciona métodos y composiciones adicionales en los que, junto con STR y iSNPs, se incluyen dianas adicionales. Si se desea obtener más información sobre un individuo, por ejemplo, cuando una muestra pertenece a un individuo o grupo de individuos desconocido, como puede ser el caso de las investigaciones de crímenes, los otros marcadores de información se pueden agregar al STR e iSNP, tales como SNP relacionados con la ascendencia (aSNP) y SNP relacionados con variantes fenotípicas (SNP informativo de fenotipo). La información adicional se puede utilizar para ayudar a los investigadores, por ejemplo, proporcionando información sobre la herencia, el color de ojos, el color del cabello y similares de un individuo desconocido. Como tal, la adición de toda la información combinada puede proporcionar un perfil de ADN más completo de un individuo que no se conocía anteriormente utilizando los métodos actuales de elaboración de perfiles de ADN.

Los métodos y las composiciones descritos en este documento están diseñados para ser lo suficientemente sensibles como para detectar cantidades inferiores a nanogramos de moléculas de ácidos nucleicos. Además, los métodos y las composiciones descritos en este documento pueden ser útiles para amplificar una muestra de ácido nucleico preparada con moléculas de ácidos nucleicos de baja calidad, como ADN genómico degradado y/o fragmentado de una muestra forense. La muestra de ácido nucleico puede ser una muestra purificada o un lisado que contiene ADN bruto, por ejemplo, derivado de un frotis bucal, papel, tela u otro sustrato que puede estar impregnado con saliva, sangre u otros fluidos corporales. Como tal, en algunas realizaciones, la muestra de ácido nucleico puede comprender pequeñas cantidades o porciones fragmentadas de ADN, tal como ADN genómico. Por ejemplo, la muestra de ácido

nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es, que es aproximadamente o que es menor que 1 pg, 2 pg, 3 pg, 4 pg, 5 pg, 6 pg, 7 pg, 8 pg, 9 pg, 10 pg, 11 pg, 12 pg, 13 pg, 14 pg, 15 pg, 16 pg, 17 pg, 18 pg, 19 pg, 20 pg, 30 pg, 40 pg, 50 pg, 60 pg, 70 pg, 80 pg, 90 pg, 100 pg, 200 pg, 300 pg, 400 pg, 500 pg, 600 pg, 700 pg, 800 pg, 900 pg, 1 ng, 10 ng, 100 ng o está en un intervalo definido por dos de estos valores, por ejemplo, 10 pg a 100 pg, 10 pg a 1 ng, 100 pg a 1 ng, 1 ng a 10 ng, 10 ng a 100 ng, etc. En algunas realizaciones, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) de aproximadamente 100 pg a aproximadamente 1 ng. En algunas realizaciones, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es más de aproximadamente 62,5 pg. En algunas realizaciones, las etapas de fragmentación adicionales, tales como sonicación o digestión con endonucleasas, no se incluyen en los procedimientos de fragmentación.

En algunas realizaciones, los métodos y las composiciones descritos en este documento son capaces de determinar con éxito los genotipos de una o más de las secuencias diana, por ejemplo, SNP, STR, etc., incluso con cantidades inferiores a nanogramos y/o muestras de ácido nucleico degradadas. Por ejemplo, los métodos y las composiciones descritos en este documento son capaces de determinar con éxito el genotipo que es, que es aproximadamente o que es más del 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 100 %, o un intervalo entre dos cualesquiera de los valores anteriores, de las secuencias diana. En algunas realizaciones, los métodos y las composiciones descritos en este documento son capaces de determinar con éxito el genotipo de más de aproximadamente 50 %, 80 %, 90 %, 95 %, 98 % o más de las secuencias diana. En algunas realizaciones, los métodos y las composiciones descritos en este documento son capaces de lograr un equilibrio intralocus de más de aproximadamente 10 %, 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 %, 100 %, o una intervalo entre dos cualesquiera de los valores anteriores, de las secuencias diana.

Para la investigación forense, la pluralidad de cebadores puede incorporar identificadores de moléculas exclusivos (UMI) que ayudan a eliminar, por ejemplo, errores de PCR y de secuenciación, tartamudeo y similares de los resultados de secuenciación. Ver Kivioja *et al.*, *supra*. Como se analiza con más detalle en otra parte de esta descripción, la inclusión de UMI en los cebadores también permite la identificación de variantes dentro de los loci de repetición en tandem, mejorando aún más la utilidad de los métodos y las composiciones actuales para la elaboración de perfiles de ADN y otros fines, tales como el análisis de herencia.

Por consiguiente, en algunas realizaciones, los genotipos de las secuencias de repetición en tandem como se describen en el presente documento pueden incluir variantes de secuencia dentro de los loci de repetición en tandem. Por lo tanto, un homocigoto para una repetición en tandem (por ejemplo, 13, 13 para D9S1122) que usa el método tradicional puede identificarse como un heterocigoto isométrico basado en variantes de secuencia dentro de la repetición en tandem. Como apreciarán los expertos en la técnica, tener en cuenta las variantes de secuencia intralocus mejoraría enormemente la utilidad de los métodos descritos en este documento, por ejemplo, para el análisis de herencia.

Métodos para construir un banco de ácidos nucleicos

Las realizaciones descritas en este documento proporcionan métodos para construir un banco de ácidos nucleicos que comprenden: proporcionar una muestra de ácido nucleico y amplificar la muestra de ácido nucleico con una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo de un solo nucleótido (SNP) y al menos una secuencia diana que comprende una secuencia repetida en tandem.

Los métodos y las composiciones descritos en este documento están diseñados para ser lo suficientemente sensibles como para detectar cantidades inferiores a nanogramos de moléculas de ácidos nucleicos. Además, los métodos y las composiciones descritos en este documento pueden ser útiles para amplificar una muestra de ácido nucleico que consiste en moléculas de ácidos nucleicos de baja calidad, como ADN genómico degradado y/o fragmentado de una muestra forense. La muestra de ácido nucleico puede estar purificada o ser un lisado que contiene ADN bruto, por ejemplo, derivado de un frotis bucal, papel, tela u otro sustrato que puede estar impregnado con saliva, sangre u otros fluidos corporales. Como tal, en algunas realizaciones, la muestra de ácido nucleico puede comprender una cantidad baja de ADN fragmentado, tal como ADN genómico. Por ejemplo, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es, que es aproximadamente o que es menor que 1 pg, 2 pg, 3 pg, 4 pg, 5 pg, 6 pg, 7 pg, 8 pg, 9 pg, 10 pg, 11 pg, 12 pg, 13 pg, 14 pg, 15 pg, 16 pg, 17 pg, 18 pg, 19 pg, 20 pg, 30 pg, 40 pg, 50 pg, 60 pg, 70 pg, 80 pg, 90 pg, 100 pg, 200 pg, 300 pg, 400 pg, 500 pg, 600 pg, 700 pg, 800 pg, 900 pg, 1 ng, 10 ng, 100 ng o está en un intervalo definido por dos de estos valores, por ejemplo, 10 pg a 100 pg, 10 pg a 1 ng, 100 pg a 1 ng, 1 ng a 10 ng, 10 ng a 100 ng, etc. En algunas realizaciones, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) de aproximadamente 100 pg a aproximadamente 1 ng. En algunas realizaciones, la muestra de ácido nucleico puede comprender una cantidad de ácido nucleico (por ejemplo, ADN genómico) que es más de aproximadamente 62,5 pg. En algunas realizaciones, no se incluyen etapas de fragmentación adicionales, como sonicación o digestión con endonucleasas.

En algunas realizaciones, los métodos descritos en el presente documento comprenden la amplificación y la preparación de bancos en previsión de la secuenciación paralela corriente abajo. Un ensayo puede incluir dos mezclas maestras de PCR, dos polimerasas termoestables, dos mezclas de cebadores y adaptadores de bancos. En algunas realizaciones, se puede amplificar una muestra de ADN durante varios ciclos utilizando un primer conjunto de cebadores de amplificación que comprenden regiones específicas de la diana y regiones de marcador no específicas

de diana y una primera mezcla maestra de PCR. La región de marcador puede ser cualquier secuencia, tal como una región de marcador universal, una región de marcador de captura, una región de marcador de amplificación, una región de marcador de secuenciación, una región de marcador UMI y similares. Por ejemplo, una región de marcador puede ser el molde para los cebadores de amplificación utilizados en una segunda ronda de amplificación o posterior, por ejemplo, para la preparación de bancos. En algunas realizaciones, los métodos comprenden la adición de proteína de unión monocatenaria (SSB) a los primeros productos de amplificación. Una parte alícuota de la primera muestra amplificada se puede extraer y amplificar una segunda vez usando un segundo conjunto de cebadores de amplificación que son específicos de la región de marcador por ejemplo, una región de marcador universal o una región de marcador de amplificación, de los primeros cebadores de amplificación, que pueden comprender de una o más secuencias de marcador adicionales, tales como marcadores de secuencia específicas para uno o más flujos de trabajo de secuenciación corriente abajo, y la misma o una segunda mezcla maestra de PCR. Como tal, un banco de la muestra de ADN original estará listo para secuenciar.

Un método alternativo podría comprender que la primera amplificación se realice en un volumen pequeño (por ejemplo, 15 al) y en lugar de transferir una parte alícuota a una nueva ubicación para una segunda ronda de amplificación, se podrían agregar al tubo reactivos adicionales para realizar una segunda ronda de amplificación.

Una vez que se crea el banco, este se puede purificar y cuantificar. En algunos ejemplos, la purificación se puede realizar procesando la muestra a través de un sustrato como AMPURE XP Beads (Beckman Coulter) que actúa para purificar los fragmentos de ADN de los componentes de reacción. Otro método podría ser la incorporación de un resto de purificación, tal como un resto de hapteno, en el segundo conjunto de cebadores de amplificación. Por ejemplo, si se incorporó una biotina en uno de los cebadores del segundo conjunto de cebadores de amplificación, entonces los fragmentos del banco podrían capturarse utilizando un resto de estreptavidina en una esfera, por ejemplo. Utilizando la estrategia de captura, los bancos también podrían normalizarse y cuantificarse mediante la normalización basada en esferas ("Bead Based Normalization", BBN). Sin embargo, los bancos se pueden purificar y cuantificar, o agrupar y cuantificar si se realizan múltiples reacciones, sin el uso de BBN. Por ejemplo, los bancos también podrían cuantificarse mediante métodos electroforéticos en gel, BioAnalyzer, qPCR, métodos espectrofotométricos, kits de cuantificación (por ejemplo, PicoGreen, etc.) y similares como se conoce en la técnica. Después de la cuantificación, el banco puede secuenciarse mediante secuenciación paralela.

En algunas realizaciones, el primer conjunto de cebadores de amplificación utilizados para amplificar un ADN diana se proporciona en una concentración tan limitada que cuando se agrega una parte alícuota de la primera reacción de amplificación a un nuevo tubo y se agregan los reactivos de la segunda reacción de amplificación, se produce una amplificación de arrastre de mínima a indetectable resultante del primer conjunto de cebadores de amplificación, y no es necesaria una etapa de limpieza entre la primera reacción de amplificación y la segunda reacción de amplificación. En algunos ejemplos, la concentración de los cebadores de amplificación para una primera PCR es, es aproximadamente o es menor que 0,5 nM, 0,6 nM, 0,7 nM, 0,8 nM, 0,9 nM, 1,0 nM, 1,5 nM, 2,0 nM, 3,0 nM, 4,0 nM, 5,0 nM, 6,0 nm, 7,0 nm, 8,0 nm, 9,0 nm 10,0 nm, 11,0 nm 12,0 nm, o un intervalo entre cualquiera de estos valores, por ejemplo, 0,5 nM a 1,0 nM, 1,0 nM a 12 nM, 0,8 nM a 1,5 nM, etc. En algunas realizaciones, la concentración de los cebadores de amplificación para una primera PCR es de aproximadamente 0,9 nM a aproximadamente 10 nM.

La figura 2 muestra un ejemplo de un flujo de trabajo de los métodos descritos en la presente en una realización. Una secuencia de ADN genómico diana se amplifica usando un primer conjunto de cebadores que comprende una región que flanquea la secuencia diana y las regiones de marcador de amplificación (que pueden ser iguales o diferentes) dando como resultado amplicones que comprenden la secuencia diana y marcadores en ambos extremos. Una parte alícuota de los amplicones de la primera PCR se amplifica adicionalmente usando un segundo conjunto de cebadores específicos para las primeras secuencias de marcador que comprenden además secuencias de cebadores de secuenciación (secuencias de adaptadores i5 e i7), generando así un banco que comprende la secuencia de ADN diana flanqueada por secuencias utilizadas en una secuenciación paralela, en este caso las secuencias i5 e i7 se utilizan en secuencia mediante métodos de síntesis popularizados por Illumina, Inc.

Un ejemplo de un flujo de trabajo alternativo para determinar un perfil de ADN a partir de una muestra se describe en la figura 3. En este ejemplo, una diana de ADN se amplifica con un primer par de cebadores que comprende secuencias que flanquean la secuencia diana, secuencias de marcador que no son diana (iguales o diferentes) y otras secuencias de identificadores moleculares exclusivos o UMI, que comprenden bases aleatorizadas. Los UMI se pueden usar, por ejemplo, para disminuir o eliminar bioinformáticamente los errores que aparecen durante los procesos de preparación de bancos (por ejemplo, artefactos de PCR o incorporaciones erróneas, etc.). El uso de UMI puede ser importante para la elaboración de perfiles de ADN, pero es de particular importancia para ayudar a eliminar errores cuando las muestras se secuencian para la investigación de crímenes. En este ejemplo, la primera ronda de amplificación se realiza durante 2 ciclos, a lo que sigue la adición de una proteína de unión monocatenaria (SSB) y la incubación a 37 °C durante 15 minutos, seguida de una inactivación a 95 °C/5 minutos que extingue con eficacia la amplificación adicional del primer conjunto de cebadores de amplificación durante la segunda ronda de amplificación. Aunque se desconoce el mecanismo, se contempla que la adición de SSB une irreversiblemente los primeros cebadores de amplificación monocatenarios y evita que participen en reacciones de amplificación posteriores. Despues de la incubación con SSB, se agrega un segundo conjunto de cebadores que comprenden marcadores de secuencia y una segunda mezcla de PCR, lo que da como resultado el banco de secuenciación.

Bancos de ácidos nucleicos

Las presentes descripciones proporcionan bancos de ácidos nucleicos, que pueden usarse para secuenciación. En algunas realizaciones, los bancos de ácidos nucleicos descritos en este documento pueden comprender una pluralidad de moléculas de ácidos nucleicos, en los que la pluralidad de moléculas de ácidos nucleicos comprende al menos una

5 secuencia repetida en tandem flanqueada por un primer par de secuencias de marcador y al menos una secuencia de un polimorfismo de un solo nucleótido (SNP) flanqueada por un segundo par de secuencias de marcador.

Como se describe en este documento, el tamaño de las moléculas de ácidos nucleicos puede variar mucho usando los métodos y las composiciones descritos en este documento. Los expertos en la técnica apreciarán que las moléculas de ácidos nucleicos amplificadas a partir de una secuencia diana que comprende una repetición en tandem (por ejemplo, STR) pueden tener un tamaño grande, mientras que las moléculas de ácidos nucleicos amplificadas a partir de una secuencia diana que comprende un SNP pueden tener un tamaño pequeño. Por ejemplo, las moléculas de ácidos nucleicos pueden comprender desde menos de cien nucleótidos hasta cientos o incluso miles de nucleótidos. Por lo tanto, el tamaño de las moléculas de ácidos nucleicos puede estar en un intervalo entre dos valores cualesquiera de aproximadamente 50 pb, aproximadamente 60 pb, aproximadamente 70 pb, aproximadamente 80 pb, 10

10 aproximadamente 90 pb, aproximadamente 100 pb, aproximadamente 110 pb, aproximadamente 120 pb, aproximadamente 130 pb, aproximadamente 140 pb, aproximadamente 150 pb, aproximadamente 200 pb, aproximadamente 300 pb, aproximadamente 400 pb, aproximadamente 500 pb, aproximadamente 600 pb, aproximadamente 700 pb, aproximadamente 800 pb, aproximadamente 900 pb, aproximadamente 1 kb, o más. En 15 algunas realizaciones, el tamaño mínimo de las moléculas de ácidos nucleicos puede ser una longitud que sea, que sea aproximadamente o que sea menor que 50 pb, 60 pb, 70 pb, 80 pb, 90 pb o 100 pb. En algunas realizaciones, el tamaño máximo de las moléculas de ácidos nucleicos puede ser una longitud que sea, que sea aproximadamente o que sea más de 100 pb, 150 pb, 200 pb, 250 pb, 300 pb, 350 pb, 400 pb, 450 pb , 500 pb o 1 kb.

20

Para la generación de agrupaciones, los fragmentos del banco se inmovilizan sobre un sustrato, por ejemplo, un portaobjetos, que comprende secuencias de oligonucleótidos homólogas para capturar e inmovilizar los fragmentos del banco de ADN. Los fragmentos del banco de ADN inmovilizados se amplifican utilizando metodologías de amplificación de agrupaciones como se ejemplifica en las descripciones de las patentes de EE. UU. n.º 7.985.565 y 7.115.400. Las 25 patentes de EE. UU. n.º 7.985.565 y 7.115.400 describen métodos de amplificación de ácidos nucleicos en fase sólida que permiten inmovilizar productos de amplificación sobre un soporte sólido para formar matrices compuestas por agrupaciones o "colonias" de moléculas de ácidos nucleicos inmovilizadas. Cada grupo o colonia sobre dicha matriz se 30 forma a partir de una pluralidad de cadenas polinucleotídicas inmovilizadas idénticas y una pluralidad de cadenas polinucleotídicas complementarias inmovilizadas idénticas. Las matrices así formadas se denominan generalmente "matrices agrupadas". Los productos de reacciones de amplificación en fase sólida como las descritas en las patentes de EE. UU. n.º 7.985.565 y 7.115.400 son las denominadas estructuras "con puente" formadas por la asociación de pares 35 de cadenas polinucleotídicas inmovilizadas y cadenas complementarias inmovilizadas, estando ambas cadenas inmovilizadas sobre el soporte sólido en el extremo 5', preferiblemente mediante una unión covalente. Las metodologías de amplificación de agrupaciones son ejemplos de métodos en los que se usa un molde de ácido nucleico inmovilizado para producir amplicones inmovilizados. También pueden usarse otras metodologías adecuadas para producir 40 amplicones inmovilizados a partir de fragmentos de ADN inmovilizados producidos de acuerdo con los métodos proporcionados en este documento. Por ejemplo, se pueden formar una o más agrupaciones o colonias mediante PCR en fase sólida, tanto si uno o ambos cebadores de cada pareja de cebadores de amplificación están inmovilizados. Sin embargo, los métodos descritos en este documento no se limitan a ninguna metodología de preparación de secuenciación o plataforma de secuenciación particular y pueden adaptarse a otros métodos de preparación de plataformas de secuenciación paralela y plataformas de secuenciación asociadas.

Cebadores

45 Las realizaciones descritas en este documento proporcionan una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana corta y al menos una secuencia diana larga en una muestra de ácido nucleico, en la que la amplificación de la muestra de ácido nucleico usando la pluralidad de cebadores en una única reacción múltiplex da como resultado al menos un producto de amplificación corto y al menos un producto de amplificación largo, en la que cada uno de la pluralidad de cebadores comprende una o más secuencias de marcador.

50 En el presente documento se describe además una pluralidad de cebadores que tienen las secuencias expuestas en las tablas 1-2.

Para la amplificación múltiplex de una secuencia diana grande (por ejemplo, STR, ITR) y una secuencia diana pequeña (por ejemplo, SNP), se diseñan cebadores que permitan una amplificación equilibrada en todos los tipos de diana. Los 55 métodos y las composiciones descritos en este documento pueden usarse para amplificar múltiples secuencias diana repetidas en tandem en una única reacción múltiplex. Por ejemplo, la pluralidad de cebadores puede hibridarse específicamente con una serie de secuencias repetidas en tandem que son, que son aproximadamente o que son más de 4, 6, 8, 10, 12, 14, 16, 18, 24, 30, 40, 50, , 60, 70, 80, 90, 100, o un intervalo entre cualquiera de los dos valores, como 4 a 12, 10 a 24, 30 a 100, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridarse específicamente a al menos 24 secuencias repetidas en tandem. En algunas realizaciones, la pluralidad de cebadores 60 puede hibridarse específicamente con al menos 60 secuencias repetidas en tandem. Los métodos y las composiciones descritas en este documento pueden usarse para amplificar múltiples secuencias diana SNP en una sola reacción.

Por ejemplo, la pluralidad de cebadores puede hibridarse específicamente con una serie de secuencias SNP que son, que son aproximadamente o que son más de 4, 6, 8, 10, 12, 14, 16, 18, 24, 30, 40, 50, 60, 70, 80, 90, 100 o un intervalo entre cualquiera de los dos valores, como 4 a 12, 10 a 24, 30 a 100, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridarse específicamente con al menos 30 secuencias SNP. En algunas realizaciones, la pluralidad de cebadores puede hibridarse específicamente con al menos 50 secuencias SNP.

Se descubrió durante la experimentación que las secuencias diana de SNP cortas se amplificaban preferentemente frente a las secuencias diana STR más largas cuando se usaban cebadores que se diseñaban siguiendo los criterios establecidos y la sabiduría para el diseño exitoso de cebadores. Además, al menos en el flujo de trabajo de secuencia por síntesis, en el que se generan agrupaciones y las agrupaciones se secuencian en sí mismas (por ejemplo, cuando se utiliza la secuencia por síntesis (SBS, descrita aquí en otro punto) asociada con los secuenciadores de Illumina, Inc.), también aparece una amplificación de agrupación preferencial de los fragmentos de SNP del banco más cortos. Para superar estos dos sesgos, se necesitaba una nueva estrategia para el diseño de cebadores que permitiera una amplificación equilibrada entre las secuencias diana SNP cortas y las secuencias diana STR largas.

Una de las estrategias incluyó el diseño de cebadores para la amplificación de STR. Con las STR, las secuencias repetidas a menudo están incluidas dentro de regiones repetidas más grandes; por lo tanto, el diseño de cebadores específicos para la amplificación de STR puede resultar problemático. Además, las STR y sus regiones flanqueantes a menudo son ricas en AT. En un caso, los cebadores se diseñaron para las regiones problemáticas utilizando una estrategia de diseño contraria a los criterios de diseño de PCR convencionales y bien establecidos. Los criterios establecidos para el diseño de cebadores de PCR indican que, entre otros criterios, 1) la longitud óptima para los cebadores es de 18 a 22 nucleótidos, 2) la Tm debe estar en el intervalo de 55-58 °C, 3) el contenido de GC debe ser de aproximadamente 40-60 %, 4) las regiones de dinucleótidos AT repetidas deben evitarse, siendo <4 repeticiones de dinucleótidos AT el máximo. Se diseñaron cebadores que eran más largos que los cebadores de PCR típicos, por ejemplo, de 23 a 35 nucleótidos de longitud en lugar de 18 a 22 nucleótidos, que tenían bajas temperaturas de fusión (Tm), por ejemplo, aproximadamente 54 °C en lugar de aproximadamente 58 °C, y los cebadores eran ricos en AT, tres parámetros que los criterios de PCR establecidos convencionales indican que deben evitarse para un diseño de cebador óptimo. En efecto, se diseñaron cebadores no óptimos. Sorprendentemente, se descubrió que estos cebadores largos, ricos en AT y con baja Tm en realidad multiplexaban los STR mejor que los cebadores cortos, de alta Tm y bajo contenido en AT. Aunque no se pretenda limitación alguna por la teoría, se contempla que los cebadores más cortos que se diseñaron siguiendo los criterios de diseño de PCR establecidos podrían formar dímeros que tenían altas temperaturas de fusión y, por lo tanto, forman dímeros de manera eficiente en condiciones normales de PCR, mientras que los cebadores más largos y de Tm baja podrían formar dímeros a una Tm realmente baja y, por lo tanto, no serían estables para la formación de dímeros, lo que permite una mayor participación de los cebadores más largos y de Tm baja en condiciones normales de amplificación en comparación con los cebadores cortos de Tm alta (por ejemplo, 18-22 nucleótidos, Tm de 60 °C, 50 % de contenido de GC).

Los cebadores más largos, de baja Tm y ricos en AT para la amplificación de STR se multiplexaron luego con los cebadores más cortos de alta Tm diseñados de modo convencional que se dirigían a los SNP. Sin embargo, las reacciones de amplificación múltiple una vez más no tuvieron éxito en proporcionar una amplificación equilibrada de STR y SNP en una reacción múltiple. Se contempló que quizás la aplicación del diseño de cebador no convencional para amplificar dianas no problemáticas, por ejemplo, para amplificar las dianas SNP, podría producir amplificaciones múltiples exitosas. Como tal, los mismos criterios utilizados para diseñar cebadores no óptimos para STR se aplicaron al diseño de cebadores para SNP (largos, Tm baja, ricos en AT). Sorprendentemente, los nuevos cebadores diseñados dieron como resultado un mejor equilibrio entre la amplificación de STR y SNP en una reacción múltiple.

La figura 4 muestra ejemplos de la interacción entre los cebadores de diseño convencional y no convencional en una reacción múltiple. En la figura 4A, una reacción múltiple de 10 SNP diana muestra una amplificación esperada en el intervalo deseado de aproximadamente 200-350 pb para el banco. Los cebadores utilizados para amplificar los 10 SNP en el múltiple se diseñaron para ser más largos, tener una Tm más baja y ser más ricos en AT, tal como lo aconsejan los criterios de diseño de cebadores de PCR establecidos. Cuando se diseña un 11º pareja de cebadores utilizando los criterios de diseño de PCR establecidos, es decir, los cebadores son cortos, tienen Tm alta y no son ricos en AT, y sumados a los 10 pares, el múltiple resultante muestra una amplificación no específica del ADN diana. Como se ve en la figura 4B y 4D, la adición de un 11º pareja de cebadores diseñado convencionalmente interfiere con el 10-plex de parejas de cebadores no convencionales y da como resultado una amplificación múltiple infructuosa de los SNP diana. Sin embargo, la adición de un 11º pareja de cebadores que también se diseña de manera no convencional siguiendo los mismos criterios que el 10-plex de parejas de cebadores da como resultado la amplificación exitosa de las dianas de SNP (figura 4C).

Por consiguiente, cada uno de la pluralidad de cebadores tiene una temperatura de fusión de menos de 60 °C y tiene una longitud de al menos 24 nucleótidos, por ejemplo, de 24 nucleótidos a aproximadamente 38 nucleótidos. En algunas realizaciones, cada uno de la pluralidad de cebadores comprende una secuencia de nucleótidos de homopolímero.

En algunos ejemplos, los cebadores diseñados de manera no convencional comprenden secuencias que flanquean los STR y SNP diana y secuencias adicionales sin molde. Las secuencias adicionales pueden ser, por ejemplo, secuencias de marcador que sirven para un propósito durante la preparación de bancos o metodologías de

secuenciación. Por ejemplo, una secuencia de marcador puede ser una secuencia de captura, tal como un resto hapteno que puede ser capturado por un resto compañero inmovilizado para purificar fragmentos de banco. Un ejemplo de un resto hapteno es la biotina que puede ser capturada por estreptavidina para fragmentos de banco aislados de componentes de reacción y similares. Una secuencia de marcador también podría ser una secuencia de amplificación, 5 por ejemplo, que sea complementaria a un cebador de amplificación y se use en una o más reacciones de amplificación. Las figuras 2 y 3 muestran ejemplos de secuencias de marcador que se utilizan en una segunda ronda de amplificación después de una primera ronda de amplificación. Una secuencia de marcador también podría ser un marcador de secuencia. Las figuras 2 y 3 también muestran ejemplos de marcadores de secuencia, el adaptador i5 y el adaptador i7 10 que se utilizan en la secuenciación como cebadores de hibridación, de generación de agrupaciones y de secuenciación durante las reacciones de secuencia por síntesis como se describe en el presente documento. Otro ejemplo de una secuencia de marcador es un identificador molecular exclusivo, o UMI, como se muestra en la figura 3.

Un UMI comprende un tramo aleatorio de nucleótidos que se puede usar durante la secuenciación para corregir errores de secuenciación y PCR, agregando así una capa adicional de corrección de errores a los resultados de secuenciación. Los UMI pueden tener una longitud, por ejemplo, de 3-10 nucleótidos, aunque el número dependerá de la cantidad de 15 ADN de entrada. Por ejemplo, si se utiliza 1 ng de ADN para dirigirse a aproximadamente 250 sitios, se prevé que se necesitarían aproximadamente 350 copias x 250 dianas, por lo que serían aproximadamente 90.000 UMI diferentes. Si se utiliza más ADN, por ejemplo, 10 ng, se podrían necesitar aproximadamente 1 millón de UMI diferentes. Todas las réplicas de PCR de la misma reacción de PCR tendrían la misma secuencia de UMI, por lo que las réplicas se 20 pueden comparar y cualquier error en la secuencia, tales como sustituciones, delecciones, inserciones de una sola base (es decir, tartamudeo en PCR), puede excluirse bioinformáticamente de los resultados de la secuenciación. Los identificadores moleculares exclusivo también se pueden utilizar en el análisis de una muestra mixta. Las muestras mixtas, por ejemplo, una muestra de ADN femenino que está contaminada con ADN masculino, se pueden desconvolucionar para indicar las contribuciones de ADN femenino y masculino utilizando secuencias UMI. Por 25 ejemplo, podría haber un total de cuatro números repetidos diferentes para dos ADN mixtos; sin embargo, podría haber menos de cuatro si la mezcla de dos muestras comparte alelos en un locus particular. Estos alelos compartidos se pueden distinguir y se pueden determinar sus porcentajes aproximados utilizando las UMI para determinar el número de alelos diferentes en la población inicial de moléculas de ADN. Por ejemplo, las moléculas iniciales podrían contarse y si un contribuyente menor estuviera presente, por ejemplo, al 5 %, entonces 5 % de los UMI identificaría un genotipo y el 95 % identificaría un segundo genotipo. Después de la PCR, si uno de los alelos (o quizás más) 30 estuviera sesgado tras la amplificación, no se observaría esa proporción de 5:95. Sin embargo, utilizando UMI, podría corregirse una proporción sesgada después de condensar las réplicas de PCR mediante detección y corrección de UMI. Esto es importante cuando se trata de diferenciar un artefacto de tartamudeo de la PCR y un verdadero contribuyente menor.

Un cebador de los presentes métodos puede comprender una o más secuencias de marcador. Las secuencias de 35 marcador pueden ser una o más de las secuencias de cebadores que no son homólogas con la secuencia diana, pero, por ejemplo, pueden usarse como moldes para una o más reacciones de amplificación. La secuencia de marcador puede ser una secuencia de captura, por ejemplo, una secuencia de hapteno, tal como biotina, que puede usarse para purificar amplicones de los componentes de la reacción. Las secuencias de marcador pueden ser secuencias, tales 40 como secuencias de adaptador, que son ventajosas para capturar los amplicones del banco sobre un sustrato, por ejemplo, para la amplificación en puente en anticipación de las tecnologías de secuencia como se describe en este documento. Además, las secuencias de marcador pueden ser marcadores de identificación molecular exclusivos típicamente de entre, por ejemplo, 3-10 nucleótidos que comprenden un tramo aleatorio de nucleótidos que se pueden usar para la corrección de errores durante la preparación del banco y/o los métodos de secuenciación.

Además, es ventajoso que una reacción de PCR multiplexada contenga cebadores oligonucleotídicos para 45 sustancialmente todas las dianas reunidas en una mezcla. Sin embargo, como se describe en el presente documento, los oligonucleotídicos son inusualmente más largos que los cebadores diseñados usando parámetros tradicionales. La adición adicional de secuencias de marcador a los cebadores, tal como la adición de UMI, que agregan una secuencia específica de un gen diana, crea secuencias de cebadores aún más largas. En algunas realizaciones, se puede añadir glicina betaína (aproximadamente 1,5 M) a la pluralidad de cebadores. Por ejemplo, en algunas realizaciones, los 50 tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales, como se describe en este documento, comprenden una concentración de betaína que es, que es aproximadamente o que es mayor que 100 mM, 200 mM, 300 mM, 400 mM, 500 mM, 600 mM, 700 mM, 800 mM, 900 mM, 1 M, 1,2 M, 1,3 M, 1,4 M, 1,5 M, 1,6 M, 1,7 M, 1,8 M, 1,9 M, 2 M, 3 M, 4 M, 5 M, 6 M, 7 M, 8 M, 9 M, 10 M o un intervalo entre dos de estos valores, por ejemplo, de 500 mM a 2 M, de 1 M a 1,5 M, etc. Como tal, una mezcla de cebadores como se describe en el presente 55 documento suplementada con betaína, por ejemplo, a aproximadamente 1,5 M, sería ventajosa cuando se practican los métodos de la presente descripción. En algunas realizaciones, se puede añadir glicerol a la pluralidad de cebadores. Por ejemplo, en algunas realizaciones, los tampones de amplificación utilizados en reacciones de amplificación con cebadores no convencionales, como se describe en el presente documento, comprenden una concentración de glicerol que es, que es aproximadamente o que es mayor que 100 mM, 200 mM, 300 mM, 400 mM, 500 mM, 600 mM, 700 mM, 800 mM, 900 mM, 1 M, 1,2 M, 1,3 M, 1,4 M, 1,5 M, 1,6 M, 1,7 M, 1,8 M, 1,9 M, 2 M, 3 M, 4 M, 5 M, 6 M, 7 M, 8 M, 9 M, 10 M o un intervalo entre dos de estos valores, por ejemplo, de 500 mM a 2 M, de 1 M a 1,5 M, etc. Como tal, una mezcla 60 de cebadores como se describe en el presente documento suplementada con glicerol, por ejemplo, a aproximadamente 1,5 M, sería ventajosa cuando se practican los métodos de la presente descripción.

En algunas realizaciones, también se pueden modificar los tampones asociados con un diseño de cebadores no convencional usado en los métodos de amplificación de la presente descripción. Por ejemplo, en algunas realizaciones, las concentraciones de sales, tales como KCl, LiCl, NaCl o una combinación de las mismas, del tampón de amplificación aumentan en comparación con la concentración de sales de un tampón de amplificación usado junto con cebadores diseñados de modo convencional. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe en este documento comprenden una concentración de KCl que es, que es aproximadamente o que es mayor que 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, 200 mM, 250 mM, 300 mM, 400 mM, 500 mM o un intervalo entre dos de estos valores, por ejemplo, de 60 mM a 200 mM, de 100 mM a 250 mM, etc. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe en el presente documento comprenden una concentración de KCl que es aproximadamente 145 mM. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe en este documento comprenden una concentración de LiCl que es, que es aproximadamente o que es mayor que 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, 200 mM, 250 mM, 300 mM, 400 mM, 500 mM o un intervalo entre dos de estos valores, por ejemplo, de 60 mM a 200 mM, de 100 mM a 250 mM, etc. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales, como se describe en el presente documento, comprenden una concentración de LiCl que es aproximadamente 145 mM. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe en este documento comprenden una concentración de NaCl que es, que es aproximadamente o que es mayor que 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, 200 mM, 250 mM, 300 mM, 400 mM, 500 mM o un intervalo entre dos de estos valores, por ejemplo, de 60 mM a 200 mM, de 100 mM a 250 mM, etc. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe en el presente documento comprenden una concentración de NaCl que es aproximadamente 145 mM.

En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe en este documento pueden comprender MgSO₄, MgCl₂, o una combinación de los mismos.

Kits

Las realizaciones descritas en el presente documento proporcionan kits que comprenden al menos un medio de recipiente, en el que dicho al menos un medio de recipiente comprende una pluralidad de cebadores como se describe en el presente documento. En algunas realizaciones, el medio de recipiente puede ser un tubo, un pocillo, una placa de microtitulación, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridarse específicamente con una serie de secuencias repetidas en tandem que es, que es aproximadamente o que es mayor que 4 , 6, 8, 10, 12, 14, 16, 18, 24, 30, 40, 50, 60, 70, 80, 90, 100 o un intervalo entre cualquiera de los dos valores, como 4 a 12, 10 a 24, 30 a 100, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridarse específicamente con al menos 24 secuencias repetidas en tandem. En algunas realizaciones, la pluralidad de cebadores puede hibridarse específicamente con al menos 60 secuencias repetidas en tandem. Los métodos y las composiciones descritos en este documento pueden usarse para amplificar múltiples secuencias diana SNP en una sola reacción. Por ejemplo, la pluralidad de cebadores puede hibridarse específicamente con una serie de secuencias SNP que es, que es aproximadamente o que es mayor que 4, 6, 8, 10, 12, 14, 16, 18, 24, 30, 40, 50, 60, 70, 80, 90, 100 o un intervalo entre cualquiera de los dos valores, como 4 a 12, 10 a 24, 30 a 100, etc. En algunas realizaciones, la pluralidad de cebadores puede hibridarse específicamente con al menos 30 secuencias SNP. En algunas realizaciones, la pluralidad de cebadores puede hibridar específicamente con al menos 50 secuencias SNP.

En algunas realizaciones, dicho al menos un medio de recipiente comprende un tampón de amplificación. En algunas realizaciones, también se pueden modificar los tampones asociados con un diseño de cebador no convencional usado en los métodos de amplificación de la presente descripción. Por ejemplo, en algunas realizaciones, las concentraciones de sales, tales como KCl, LiCl, NaCl o una combinación de las mismas, del tampón de amplificación aumentan en comparación con la concentración de sales de un tampón de amplificación usado junto con cebadores diseñados de modo convencional. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe en este documento comprenden una concentración de KCl, NaCl o LiCl que es, que es aproximadamente o que es mayor que 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 110 mM, 120 mM, 130 mM, 140 mM, 150 mM, 160 mM, 170 mM, 180 mM, 190 mM, 200 mM, 250 mM, 300 mM, 400 mM, 500 mM o un intervalo entre dos de estos valores , por ejemplo, de 60 mM a 200 mM, de 100 mM a 250 mM, etc. En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe aquí comprenden una concentración de KCl, NaCl o LiCl que es aproximadamente 145 mM.

En algunas realizaciones, los tampones de amplificación usados en reacciones de amplificación con cebadores no convencionales como se describe en este documento pueden comprender MgSO₄, MgCl₂, o una combinación de los mismos.

Métodos de secuenciación

Los presentes métodos no se limitan a ninguna plataforma de secuenciación particular, sin embargo, se ejemplifican aquí con respecto a SBS, o secuencia por síntesis, un tipo de secuenciación paralela. Las técnicas particularmente aplicables son aquellas en las que los ácidos nucleicos se unen en ubicaciones fijas en una matriz de modo que sus posiciones relativas no cambien y en las que se forman imágenes repetidamente de la matriz. Son particularmente aplicables los ejemplos en los que se obtienen imágenes en diferentes canales de color, por ejemplo, coincidiendo con diferentes marcadores utilizados para distinguir un tipo de base de nucleótidos de otro.

Las técnicas de SBS generalmente implican la extensión enzimática de una hebra de ácido nucleico naciente mediante la adición iterativa de nucleótidos contra una cadena de molde. En los métodos tradicionales de SBS, se puede proporcionar un monómero de un solo nucleótido a un nucleótido diana en presencia de una polimerasa en cada suministro. Sin embargo, en los métodos descritos en el presente documento, se puede proporcionar más de un tipo de monómero nucleotídico a un ácido nucleico diana en presencia de una polimerasa en un suministro.

Las técnicas de SBS pueden utilizar monómeros de nucleótidos que tienen un resto marcador o que carecen de un resto marcador. Por consiguiente, los acontecimientos de incorporación pueden detectarse basándose en una característica del marcador, como la fluorescencia del marcador; una característica del monómero de nucleótido, como el peso molecular o la carga; un subproducto de la incorporación del nucleótido, como la liberación de pirofosfato; o similar. En algunos ejemplos en los que dos o más nucleótidos diferentes están presentes en un reactivo de secuenciación, los diferentes nucleótidos pueden distinguirse entre sí o, como alternativa, los dos o más marcadores diferentes pueden ser indistinguibles según las técnicas de detección que se utilizan. Por ejemplo, los diferentes nucleótidos presentes en un reactivo de secuenciación pueden tener diferentes marcadores y se pueden distinguir usando ópticas apropiadas como se ejemplifica mediante los métodos de secuenciación desarrollados por Solexa (ahora Illumina, Inc.).

Algunos ejemplos incluyen técnicas de pirosecuenciación. La pirosecuenciación detecta la liberación de pirofosfato inorgánico (PPi) a medida que se incorporan nucleótidos particulares en la hebra naciente (Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. y Nyren, P. (1996), "Real-time DNA sequencing using detection of pyrophosphate release ", Analytical Biochemistry, 242 (1), 84-9; Ronaghi, M. (2001), "Pyrosequencing sheds light on DNA sequencing ", Genome Res., 11 (1), 3-11; Ronaghi, M., Uhlen, M. y Nyren, P. (1998), "A sequencing method based on real-time pyrophosphate ", Science, 281 (5375), 363; patente de EE. UU. n.º 6.210.891; patente de EE. UU. n.º 6.258.568 y patente de EE. UU. n.º 6.274.320). En la pirosecuenciación, el PPi liberado puede detectarse por ser inmediatamente convertido en adenosina trifosfato (ATP) por la ATP sulfúrilasa, y el nivel de ATP generado se detecta mediante fotones producidos por luciferasa. Los ácidos nucleicos que se van a secuenciar se pueden unir a características en una matriz y se pueden formar imágenes de la matriz para capturar las señales quimioluminiscentes que se producen debido a la incorporación de nucleótidos en las características de la matriz. Se puede obtener una imagen después de que la matriz se trate con un tipo de nucleótido particular (por ejemplo, A, T, C o G). Las imágenes obtenidas después de la adición de cada tipo de nucleótido diferirán con respecto a las características de la matriz que se detecten. Estas diferencias en la imagen reflejan el contenido de secuencia diferente de las características en la matriz. Sin embargo, las ubicaciones relativas de cada característica permanecerán sin cambios en las imágenes. Las imágenes pueden almacenarse, procesarse y analizarse utilizando los métodos establecidos en este documento. Por ejemplo, las imágenes obtenidas después del tratamiento de la matriz con cada tipo de nucleótido diferente se pueden manipularse de la misma manera que se ejemplifica en el presente documento para imágenes obtenidas de diferentes canales de detección para métodos de secuenciación basados en terminadores reversibles.

En otro ejemplo de SBS, la secuenciación del ciclo se logra mediante la adición discontinua de nucleótidos terminadores reversibles que contienen, por ejemplo, un marcador de colorante escindible o fotoblanqueable como se describe, por ejemplo, en el documento WO 04/018497 y la patente de EE. UU. n.º 7.057.026. Esta estrategia está siendo comercializado por Solexa (ahora Illumina Inc.), y también se describe en los documentos WO 91/06678 y WO 07/123,744. La disponibilidad de terminadores marcados con fluorescencia en los que la terminación puede revertirse y el marcador fluorescente puede escindirse facilita la secuenciación de terminación cíclica reversible (CRT) eficiente. Las polimerasas también pueden diseñarse conjuntamente para incorporar y extenderse eficazmente a partir de estos nucleótidos modificados. Se describen ejemplos de sistemas y métodos de SBS adicionales que se pueden utilizar con los métodos y sistemas descritos en este documento en la publicación de solicitud de patente de EE. UU. n.º 2007/0166705, la publicación de solicitud de patente de EE. UU. n.º 2006/0188901, la publicación de solicitud de patente de EE. UU. n.º 7.057.026, la publicación de solicitud de patente de EE. UU. n.º 2006/0240439, la publicación de solicitud de patente de EE. UU. n.º 2006/0281109, la publicación PCT n.º WO 05/065814, la publicación de solicitud de patente de EE. UU. n.º 2005/0100900, la publicación PCT n.º WO 06/064199, la publicación PCT n.º WO 07/010.251, la publicación de solicitud de patente de EE. UU. n.º 2012/0270305 y la publicación de solicitud de patente de EE. UU. n.º 2013/0260372.

Algunos ejemplos pueden utilizar la detección de cuatro nucleótidos diferentes usando menos de cuatro marcadores diferentes. Por ejemplo, la SBS se puede realizar utilizando métodos y sistemas descritos en los materiales incorporados de la publicación de solicitud de patente de EE. UU. n.º 2013/0079232. Como primer ejemplo, se puede detectar un par de tipos de nucleótidos en la misma longitud de onda, pero distinguirlos en función de la diferencia de intensidad de un miembro del par en comparación con el otro, o en función de un cambio en un miembro del par (por

ejemplo, mediante modificación química, modificación fotoquímica o modificación física) que hace que la señal aparente aparezca o desaparezca en comparación con la señal detectada para el otro miembro del par. Como segundo ejemplo, se pueden detectar tres de cuatro tipos de nucleótidos diferentes en condiciones particulares, mientras que un cuarto tipo de nucleótidos carece de un marcador que sea detectable en esas condiciones, o se detecta mínimamente en esas condiciones (por ejemplo, una detección mínima debido a la fluorescencia de fondo, etc.). La incorporación de los primeros tres tipos de nucleótidos en un ácido nucleico se puede determinar basándose en la presencia de sus respectivas señales, y la incorporación del cuarto tipo de nucleótido en el ácido nucleico se puede determinar basándose en la ausencia o la detección mínima de cualquier señal. Como tercer ejemplo, un tipo de nucleótido puede incluir uno o más marcadores que se detectan en dos canales diferentes, mientras que otros tipos de nucleótidos se detectan en no más de uno de los canales. Los tres ejemplos de configuraciones mencionados anteriormente no se consideran mutuamente excluyentes y pueden usarse en varias combinaciones. Un ejemplo de realización que combina los tres ejemplos es un método SBS basado en fluorescencia que usa un primer tipo de nucleótido que se detecta en un primer canal (por ejemplo, dATP que contiene un marcador que se detecta en el primer canal cuando se excita con una primera longitud de onda de excitación), un segundo tipo de nucleótido que se detecta en un segundo canal (por ejemplo, dCTP que tiene un marcador que se detecta en el segundo canal cuando es excitado por una segunda longitud de onda de excitación), un tercer tipo de nucleótido que se detecta en el primer y el segundo canal (por ejemplo, dTTP que tiene al menos un marcador que se detecta en ambos canales cuando se excita con la primera y/o segunda longitud de onda de excitación) y un cuarto tipo de nucleótido que carece de un marcador que no se detecta en ninguno de los canales o se detecta mínimamente (por ejemplo, dGTP que no tiene marcador).

Además, como se describe en la publicación de solicitud de patente de EE. UU. n.º 2013/0079232, los datos de secuenciación se pueden obtener utilizando un solo canal. En las llamadas estrategias de secuenciación de un colorante, el primer tipo de nucleótido se marca, pero el marcador se elimina después de que se genere la primera imagen, y el segundo tipo de nucleótido se marca solo después de que se genere una primera imagen. El tercer tipo de nucleótido conserva su marcador tanto en la primera como en la segunda imagen, y el cuarto tipo de nucleótido permanece sin marcar en ambas imágenes.

Algunos ejemplos pueden utilizar la secuenciación mediante técnicas de acoplamiento. Tales técnicas utilizan ADN ligasa para incorporar oligonucleótidos e identificar la incorporación de tales oligonucleótidos. Los oligonucleótidos tienen típicamente diferentes marcadores que se correlacionan con la identidad de un nucleótido particular en una secuencia con la que se hibridan los oligonucleótidos. Al igual que con otros métodos de SBS, se pueden obtener imágenes tras el tratamiento de una matriz de características de ácidos nucleicos con los reactivos de secuenciación marcados. Cada imagen mostrará características de ácidos nucleicos que han incorporado marcadores de un tipo particular. Diferentes características estarán presentes o ausentes en las diferentes imágenes debido al contenido de secuencia diferente de cada característica, pero la posición relativa de las características permanecerá sin cambios en las imágenes. Las imágenes obtenidas a partir de métodos de secuenciación basados en el acoplamiento se pueden almacenar, procesar y analizar como se establece en este documento. Se describen ejemplos de sistemas y métodos de SBS que se pueden utilizar con los métodos y sistemas descritos en este documento en la patente de EE. UU. n.º 6.969.488, la patente de EE. UU. n.º 6.172.218, y la patente de EE. UU. n.º 6.306.597.

Algunos ejemplos pueden utilizar la secuenciación de nanoporos (Deamer, D. W. y Akeson, M., "Nanopores and nucleic acids: prospects for ultrarapid sequencing", Trends Biotechnol., 18, 147-151 (2000); Deamer, D. y D. Branton, "Characterization of individual polynucleotide molecules using a membrane channel", Acc. Chem. Res., 35: 817-825 (2002); Li, J., M. Gershon, D. Stein, E. Brandin y J. A. Golovchenko, "DNA molecules and configurations in a solid-state nanopore microscope", Nat. Mater., 2: 611-615 (2003))). En tales realizaciones, el ácido nucleico diana pasa a través de un nanoporo. El nanoporo puede ser un poro sintético o una proteína de membrana biológica, como la α -hemolisina. A medida que el ácido nucleico diana pasa a través del nanoporo, cada par de bases se puede identificar midiendo las fluctuaciones en la conductancia eléctrica del poro (patente de EE. UU. n.º 7.001.792; Soni, G. V. y Meller, "Progress toward ultrafast DNA sequencing using solid-state nanopores", Clin. Chem., 53, 1996-2001 (2007); Healy, K., "Nanopore-based single-molecule DNA analysis", Nanomed., 2, 459 a 481 (2007); Cockroft, S. L., Chu, J., Amorin, M. y Ghadiri, M. R., "A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution", J. Am. Chem. Soc., 130, 818-820 (2008)). Los datos obtenidos de la secuenciación de nanoporos se pueden almacenar, procesar y analizar como se establece en este documento. En particular, los datos pueden tratarse como una imagen de acuerdo con el ejemplo de tratamiento de imágenes ópticas y otras imágenes que se expone en este documento.

Algunos ejemplos pueden utilizar métodos que implican el seguimiento en tiempo real de la actividad de la ADN polimerasa. Las incorporaciones de nucleótidos se pueden detectar mediante interacciones de transferencia de energía por resonancia de fluorescencia (FRET) entre una polimerasa que porta un fluoróforo y nucleótidos marcados con γ -fosfato como se describe, por ejemplo, en la patente de EE. UU. n.º 7.329.492 y la patente de EE. UU. n.º 7.211.414, o pueden detectarse incorporaciones de nucleótidos con guías de onda de modo cero como se describe, por ejemplo, en la patente de EE. UU. n.º 7.315.019 y usando análogos de nucleótidos fluorescentes y polimerasas modificadas como se describe, por ejemplo, en la patente de EE. UU. n.º 7.405.281 y la publicación de solicitud de patente de EE. UU. n.º 2008/0108082. La iluminación se puede restringir a un volumen de escala de zeptolitros alrededor de una polimerasa unida a la superficie de modo que se pueda observar la incorporación de nucleótidos marcados con fluorescencia con un fondo bajo (Levene, M. J. et al., "Zero-mode waveguides for single-molecule analysis at high concentrations", Science, 299, 682-686 (2003); Lundquist, P. M. et al., "Parallel confocal detection of

single molecules in real time", Opt. Lett., 33, 1026-1028 (2008); Korlach, J. et al., "Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures", Proc. Natl. Acad. Sci. U.S.A., 105, 1176-1181 (2008)). Las imágenes obtenidas de dichos métodos pueden almacenarse, procesarse y analizarse como se establece en este documento.

- 5 Algunas realizaciones de SBS incluyen la detección de un protón liberado tras la incorporación de un nucleótido en un producto de extensión. Por ejemplo, la secuenciación basada en la detección de protones liberados puede utilizar un detector eléctrico y técnicas asociadas que están disponibles comercialmente en Ion Torrent (Guilford, CT, una subdivisión de Life Technologies) o métodos y sistemas de secuenciación descritos en los documentos US 2009/0026082 A1; US 2009/0127589 A1; US 2010/0137143 A1; o US 2010/0282617 A1. Los métodos aquí expuestos para amplificar ácidos nucleicos diana usando la exclusión cinética pueden aplicarse fácilmente a sustratos usados para detectar protones. Más específicamente, los métodos establecidos en el presente documento pueden usarse para producir poblaciones clonales de amplicones que se usan para detectar protones.

Los métodos de SBS anteriores se pueden llevar a cabo ventajosamente en formatos múltiplex de modo que se manipulen simultáneamente múltiples ácidos nucleicos diana diferentes. En realizaciones particulares, se pueden tratar diferentes ácidos nucleicos diana en un recipiente de reacción común o sobre una superficie de un sustrato particular. Esto permite la administración conveniente de reactivos de secuenciación, la eliminación de reactivos que no han reaccionado y la detección de acontecimientos de incorporación de una manera múltiplex. En realizaciones que usan ácidos nucleicos diana unidos a una superficie, los ácidos nucleicos diana pueden estar en un formato de matriz. En un formato de matriz, los ácidos nucleicos diana se pueden unir típicamente a una superficie de una manera espacialmente distingible. Los ácidos nucleicos diana pueden unirse mediante unión covalente directa, unión a una esfera u otra partícula, o unión a una polimerasa u otra molécula que esté unida a la superficie. La matriz puede incluir una única copia de un ácido nucleico diana en cada sitio (también denominado característica) o pueden estar presentes múltiples copias que tienen la misma secuencia en cada sitio o característica. Pueden producirse múltiples copias mediante métodos de amplificación tales como amplificación en puente o PCR en emulsión como se describe con más detalle a continuación.

Los métodos de la presente descripción utilizan la tecnología de Illumina, Inc. para secuenciar los bancos de perfiles de ADN creados mediante la práctica de los métodos descritos en este documento. El instrumento de secuenciación MiSeq se utilizó para agrupar y secuenciar los ejemplos descritos en este documento. Sin embargo, como se indicó anteriormente y como entenderán los expertos en la técnica, los presentes métodos no están limitados por el tipo de plataforma de secuenciación utilizada.

Ejemplos

Los siguientes ejemplos describen varios métodos y materiales para la elaboración de perfiles de ADN. Estos métodos y materiales pueden modificarse manteniendo el alcance de la invención. Tales modificaciones resultarán evidentes para los expertos en la técnica a partir de la consideración de esta descripción o la práctica de los métodos aquí descritos. En consecuencia, no se pretende que estos métodos o materiales se limiten a los ejemplos específicos descritos en este documento, sino que cubran todas las modificaciones y alternativas que caen dentro del alcance de las reivindicaciones.

Ejemplo 1: Diseño de cebadores no convencional

40 El programa de diseño informático DesignStudio de Illumina, Inc. (San Diego, CA) fue modificado y utilizado para el diseño de cebadores. Los expertos en la técnica comprenderán, por supuesto, que también se pueden usar programas de diseño de cebadores alternativos, como Primer3, y que los parámetros predeterminados se restablecen para imitar la intención de los parámetros modificados para el diseño de cebadores. La configuración generalmente se restablece en el archivo config.xml que se incluye en el software; sin embargo, esto puede diferir cuando se usa un software diferente y es una práctica común consultar los materiales específicos para acceder a los parámetros predeterminados de cada software. Los siguientes parámetros se pueden restablecer en el software de diseño de cebadores:

- 45 1) El amplicón de longitud mínima deseada se restablece a >60<
- 2) El amplicón de longitud máxima deseada se restablece a >120<
- 3) El espaciado de candidatos ajustado se restablece a >3< (el valor predeterminado es 30 pb)
- 4) El porcentaje máximo en GC de la sonda se restablece a >60< para permitir un mayor número de tramos repetidos ricos en AT
- 5) La Tm promedio se restablece a >57< (59 °C predeterminado) para reducir la Tm promedio
- 6) La Tm máxima se restablece a >60< (predeterminada 71)
- 7) La Tm mínima se restablece a >51< (predeterminado 55)
- 8) La longitud promedio de la sonda se restablece a >28< (predeterminado 27)

9) La longitud máxima de la sonda se restablece a >38< (predeterminado 30)

10) La longitud mínima de la sonda se restablece a >25< (por defecto 22)

Para diseñar los cebadores de SNP, el intervalo para dirigirse al extremo 3' del cebador se estableció en "pequeño" para mantener los cebadores a aproximadamente 1 pb de distancia para el SNP diana. Una vez que se restablecen todos los parámetros, el programa de diseño de cebadores se puede ejecutar en la secuencia para determinar el pareja de cebadores candidatos que se encuentran dentro de los nuevos parámetros. Por ejemplo, un usuario del software puede generar una lista de dianas que indica al software dónde buscar en el genoma para diseñar los cebadores. En el presente ejemplo, las regiones diana se copiaron y pegaron en la aplicación de interfaz gráfica de usuario que usa el software DesignStudio para orientar y dirigirse al diseño del cebador. Una vez que las regiones diana se ingresan en el programa, el programa se dirige al archivo de creación de un diseño para iniciar la herramienta y crear los diseños de cebadores. En el presente ejemplo, la salida principal es un archivo .txt que incluye las secuencias de cebadores y/o algunas de las regiones que contienen fallos y son "indeseables", y en este momento las secuencias objetivo deben redefinirse y volver a ejecutarse. El software utilizado en este experimento proporcionó los cebadores diseñados que se cartografiaron sobre la secuencia que se especificó como la región diana. Siguiendo los parámetros de restablecimiento, se diseñaron cebadores que no siguieron los criterios convencionales para el diseño de cebadores para la amplificación; sin embargo, estos permitieron la amplificación múltiplex de STR largos y SNP cortos.

Los ejemplos de cebadores dirigidos a STR diseñados ventajosos en los métodos descritos en este documento incluyen los enumerados en la tabla 1. Los ejemplos de cebadores dirigidos a SNP ventajosos en los métodos descritos aquí incluyen los enumerados en la tabla 2.

20 Tabla 1- Cebadores dirigidos a STR sin marcadores y tamaños de amplicón

SEQ ID NO	CEBADOR LOCUS STR	DEL	EJEMPLOS DE CEBADORES DE STR SIN MARCADORES	TAMAÑO DEL AMPLOCÓN
1	AmeIPP_F_T		CCCTGGGCTCTGTAAAGAA	106, 112
2	AmeIPP_R_Sm		ATCAGAGCTAAACTGGGAAGCTG	
3	CSF1PO_F1_T		ACAGTAAC TG CTT CAT AGA TAG	117
4	CSF1PO_R1_Sm		GTGTCAGAC CCT GTT CTA AGTA	
5	D5S818_F2_T		TGATTT CCT CTT GGT AT CCT TAT GTA AT	112
6	D5S818_R2_Sm		ACA ACAT TT GT AT CCT TAT CT GT AT CCT	
7	D8S1179_F1_T		TTT GTAT TT CAT GT GT AC ATT CGT ATC	110
8	D8S1179_R1_Sm		ACCT AT CCT GTAG ATT ATT TC ACT GTG	
9	D18S51_F1_T		CT CT GAGT GACA AAT T GAG AC CT	184
10	D18S51_R1_T		TTA ACTT CT CT GG GT GT GG AG ATG	
11	D19S433_F1_T		TTT GGT GC ACC CATT ACC CG	188
12	D19S433_R1_Sm		AGG AGG TT GAGG CT GC AAAA	
13	D7S820_F2_T		CACCA AA TATT GG TA ATT AA AT GT TT ACT AT AG AC	167
14	D7S820_R2_Sm		TAA AGG GT AT GAT AGA AAC ACT TG TC	
15	D16S539_F2_T		CAA AGG CAG AT CCC AAG CT CT	160
16	D16S539_R2_Sm		TGT GT GT GC AT CT GT AAGC AT	
17	D3S1358_F2_T		TGGT GT GT ATT CC CT GT GCC	170
18	D3S1358_R2_Sm		GCAGT CCA AT CT GG GT GACA	
19	D10S1248_F1_T		CCA AT CT GG TC ACA AAC AT ATT AAT GAA	148
20	D10S1248_R1_Sm		TTT CC CTT GT CT TG TT ATT AA AGG AA AC	
21	TH01_F1_T		TTCCC ATT GG CTC TT C CTC	112
22	TH01_R1_Sm		CT GT AC AC AG GG CT T CCG AG	
23	FGA_F2_T		GCT GAGT GAT TT GT CT GT AATT G	188
24	FGA_R2_Sm		GA ACT CA CAG AT TA ACT GT A ACC AAA AT AAA ATT AG	
25	D61043_F1_T		CA A TAGT GTG CA AGG AT GGG TG	175

SEQ ID NO	CEBADOR LOCUS STR	DEL	EJEMPLOS DE CEBADORES DE STR SIN MARCADORES	TAMAÑO DEL AMPLICÓN
26	D61043_R1_Sm		TCTGTGGTTCTCCAGCTTAC	
27	TPOX_F1_T		CTTAGGGAACCCTCACTGAATG	77
28	TPOX_R1_Sm		GTCCTTGTCAAGCGTTATTG	
29	D13S317_F2_T		TTGGGTTGAGCCATAGGCAG	162
30	D13S317_R2_Sm		GCATCCGTGACTCTGGAC	
31	D21S11_F1_T		GTTATGGGACTTTCTCAGTCTCCAT	226
32	D21S11_R3_Sm		GAGACTAATAGGAGGTAGATAGACTGG	
33	D12S391_F1_T		GAGACTGTATTAGTAAGGCTTCTC	253
34	D12S391_R2_Sm		CCTGGACTGAGCCATGCTCC	
35	D1S1656_F2_T		CAGTCCTGTGTTAGTCAGGATT	173
36	D1S1656_R1_Sm		TCAAGGGTCAACTGTGTGATGT	
37	D9S1122_F3_T		CTTCTGAAAGCTCTAGTTACCT	120
38	D9S1122_R2_Sm		TTGCTTATTGTGGGGTATTCA	
39	PentaE_F1_T		AAGAATTCTCTTATTGGTTATTAATTG	362
40	PentaE_R1_Sm		AAATTGTGGACAGGTGCGGT	
41	D17S1301_F2_T		CCATGAAAAATACATGCATGTGTTATTATAC	142
42	D17S1301_R2_Sm		TGATTAAAAGAACATGAAGGTAAGATGTATAAC	
43	D2S441_F2_T		CCAAATGTTATGATTAATCTTAAATTGGAGC	160
44	D2S441_R3_Sm		GTAACAAGGGCTACAGGAATCATGAG	
45	D4S2408_F3_T		TCATCCACTGAAATGACTGAAAAATAG	102
46	D4S2408_R9_Sm		AGGTACATAACAGTTCAATAGAAAG	
47	D2S1338_F2_T		GAGTTATTAGTAAGTAAAGGATTGCAG	162
48	D2S1338_R2_Sm		GGGAGCCAGTGGATTGGAACACAG	
49	PentaD_F3_T		GCATGGTGAGGCTGAAGTAG	268
50	PentaD_R1_Sm		CTAACCTATGGTCATAACGATT	
51	vWA_F3_T		GATGATAAGAATAATCAGTATGTGACTTGG	160
52	vWA_R3_Sm		ATAGGTTAGATAGAGATAGGACAGATGATA	
53	SE33_F1_T		CCCTACCGCTATAGTAACCTGC	380
54	SE33_R2_Sm		CACGTCTGTAATTCCAGCTCCTA	
55	D20S482_F3_T		GGAAGCGTGTACTAGAGTTCTCAG	145
56	D20S482_R2_Sm		GGACAGCCTCCATATCCACATG	
57	DXS10074_F1_T		TTCTACTGCCACCTTATTG	212
58	DXS10074_R1_sm		TTTATGGTCTCAGTGGCCCTCAGA	
59	DXS10103_F1_sm		TCATAATCACATATCACATGAGC	177
60	DXS10103_R1_T		AAACAGAACCAAGGGGAATGAA	
61	DXS10135_F1_T		TGAAACTAAAGTCAAATGGGCTAC	268
62	DXS10135_R1_sm		TAAGGGGTGACACCTCTGGATA	
63	DXS8377_F2_sm		CCCAGCCTACATCTACCACTTCATG	276
64	DXS8377_R2_T		CTAATGTTGTATGGACCTTGGAAAGC	
65	DXS7423_F1_sm		GTCTCCAGTACCCAGCTAGCTTAG	191

SEQ ID NO	CEBADOR LOCUS STR	DEL	EJEMPLOS DE CEBADORES DE STR SIN MARCADORES	TAMAÑO DEL AMPLICÓN
66	DXS7423_R1_T		TCTCCCAACCTGCCCTTATCA	
67	DXS8378_F1_sm		TTTGGGCTGACACAGTGGCT	442
68	DXS8378_R1_T		TTGATCACACACAGGAGGTTGACC	
69	HPRTB_F1_sm		TATACCACTTGATGTTGACACTAGTTAC	213
70	HPRTB_R1_T		CCTGTCTATGGTCTCGATTCAAT	
71	DXS10148_F3_sm		TGCATGACAGAGGGAGATTCT	256
72	DXS10148_R3_T		AGAGGGGAAATAGTAGAATGAGGATG	
73	DXS7132_F3_sm		GCCAAACTCTATTAGTCACACGTT	204
74	DXS7132_R4_T		CTGGTTCTCTAGCTCACATACAGT	
75	DYF387S1ab_F2_T		TTTAOCCCTAACAGAAAAAAAAGAAGAA	227,231
76	DYF387S1ab_R2_Sm		CAGTGTGAGAAGTGTGAGAAGTGC	
77	DYS385a_b_F1_T		GACACCATGCCAACACAACAC	260,248
78	DYS385a_b_R1_Sm		ATCTATCTATTCCAATTACATAGTCC	
79	DYS389I_II_F3_T		TCATTATACCTACTTCTGTATCCAACCTC	183,303
80	DYS389I_II_R3_Sm		GGAACACAATTATCCCTGAGTAGCAG	
81	DYS390_F2_T		GGTAGCATAATAGAAATTATGAGTGGG	318
82	DYS390_R2_Sm		GAAGACAGACTTCAATATCACAGAACATCG	
83	DYS391_F1_T		GTGTATCTATTCAATTCAATCATACACCC	143
84	DYS391_R1_Sm		CTCCCTGGTTGCAAGCAATTGCC	
85	DYS438_F1_T		CCAAAATTAGTGGGAATAGTTGAAC	149
86	DYS438_R2_Sm		GTCGAGATCACACCATTGCATTTC	
87	DYS439_F1_T		GCCTGGCTTGAATTCTTTACCC	195
88	DYS439_R1_Sm		TTTAAGTCTTAATCTATCTGAATTAATAGATT	
89	DYS481_F1_T		CTTTAAGAGGAGTCTGCTAAAGGAATG	144
90	DYS481_R3_Sm		TCACCAGAAGGTTGCAAGAC	
91	DYS505_F1_T		TCTGGCGAAGTAACCCAAAC	174
92	DYS505_R1_Sm		TCGAGTCAGTTCACCAAGAAGG	
93	DYS522_F2_T		GGAACCAGTGAGAGCCG	306
94	DYS522_R2_Sm		CTCAGAGTGCTGAACCCAG	
95	DYS533_F2_T		GTATTTATTGATGATCAGTTCTTAACCAACC	206
96	DYS533_R2_Sm		CTACCTAATATTTATCTATCATTCTAATTATGTCTCTTC	
97	DYS549_F1_T		CTCTAAAGGTTTTGGTGGCATAAG	222
98	DYS549_R1_Sm		GATTAATACAACAAAAATTGGTAATCTGAAA	
99	DYS570_F1_T		CAACCTAAGCTGAAATGCAGATATT	170
100	DYS570_R1_Sm		GTTATGAAACGTAACATGACTAG	
101	DYS576_F2_T		GCAGTCTCATTCTGGAGATGAAGG	191
102	DYS576_R1_Sm		CTTGGGCTGAGGAGTTCAATC	
103	DYS612_F2_T		GCCAGTAAGAATAAAATTACAGCATGAAG	287
104	DYS612_R2_Sm		GAATAATCTACCAGCAACAAATGGCT	
105	DYS635_F4_T		TGCCCAATGGAATGCTCT	274

SEQ ID NO	CEBADOR LOCUS STR	DEL	EJEMPLOS DE CEBADORES DE STR SIN MARCADORES	TAMAÑO DEL AMPLICÓN
106	DYS635_R2_Sm		GCTCCATCTCAAACAAACAAAAACACAAAAATG	
107	DYS643_F2_T		GGGTCAATTGAACCTCATGCTTG	170
108	DYS643_R1_Sm		CCCCCCTAAATTCTACTGAAGTAAA	
109	Y_GATAH4_F2_T		TAACAGGATAAATCACCTATCTATGTAT	175
110	Y_GATAH4_R2_Sm		GCTGAGGAGAATTCCAATTTA	

Tabla 2 - Cebadores dirigidos a SNP

SEQ ID NO	CEBADOR DE SNP	EJEMPLOS DE CEBADORES DE SNP SIN MARCADORES
111	rs10092491_iSNPI_T_F2	CCCGCAAACACTAACTAGGATAAAATCTCA
112	rs1015250_iSNPI_T_F	CGACATGGAAATGTCAGATCATAAGAC
113	rs1024116_iSNPI_T_F2	CCAGGGAGTGAAAATCCTTTATCATC
114	rs1028528_iSNPI_T_F2	GAGGATGAAGGTTAGAGCCAGACCT
115	rs1029047_iSNPI_T_F2	TGTGGAATAAACTGAAGGCTAAAGAAAA
116	rs1031825_iSNPI_T_F2	CAAGCCCTATGCCAAGGATAACAATG
117	rs10488710_iSNPI_T_F	GAGGTTTACTGTATTAGGAGTCCCAC
118	rs10495407_iSNPI_T_F	CAGATGTGAGATGATAATTGCGTCTCC
119	rs1058083_iSNPI_T_F	TTGTTCTCTCCATCCCATTTCACCC
120	rs10773760_iSNPI_T_F	CTTGTACATTCCCTTATCTGCTATGTGG
121	rs1294331_iSNPI_T_F2	CTCTCTTGGAGTTTATGTGTTGCTAC
122	rs12997453_iSNPI_T_F	CTCTGATGATGTGCAAGAAAGGTAGGTA
123	rs13182883_iSNPI_T_F	TCAGACTATGTTAAGGAGACTATGAGG
124	rs13218440_iSNPI_T_F	CTAAGTATCTACCAATGTGCTACGTACC
125	rs1335873_iSNPI_T_F	CACGTGGATGATATGGTTCTCAAGG
126	rs1336071_iSNPI_T_F2	AGCACCTATATATTACCTGAAAGCAT
127	rs1355366_iSNPI_T_F	CCCATGATTTCTGTGGTGAGAATTTC
128	rs1357617_iSNPI_T_F	CACCCCTCTGTACTTTAATTGACTTCCC
129	rs1382387_iSNPI_T_F	GTTTTCTTCATTCCATGTTGACTTCCC
130	rs1413212_iSNPI_T_F	CACTCTCTGAATCCTGGTCAACAAAC
131	rs1454361_iSNPI_T_F	CAAGTTATATCATAGAGTCTACGACCCC
132	rs1463729_iSNPI_T_F	CTGCAACTATCAGTCTCTGCCCTTATTTC
133	rs1493232_iSNPI_T_F	GATGTGCTCAAACGTGTTATTGTGAGG
134	rs1498553_iSNPI_T_F	GAACTCATTTATCCAGAGACCTGTTCTC
135	rs1523537_iSNPI_T_F	CATAATACAACCTGTCTTGGAGTTACT
136	rs1528460_iSNPI_T_F	GTGACCAGTAGTTCTATGAGCAAGTATG
137	rs159606_iSNPI_T_F	CCACATTGTATGGTTTTAGGCACCATG
138	rs1736442_iSNPI_T_F	CTAATAAGTGGGACAGTTAAGAGAAGGC
139	rs1821380_iSNPI_T_F	CAAGACAAGCGATTGAAAGAAGTGGAT
140	rs1886510_iSNPI_T_F	CCTTGTCAATCTTCTACCAGAGGGTAA
141	rs1979255_iSNPI_T_F	GAATCATAGCTTGTGTTGGTCAGGG

SEQ ID NO	CEBADOR DE SNP	EJEMPLOS DE CEBADORES DE SNP SIN MARCADORES
142	rs2016276_iSNPI_T_F	GAATTACAAGTATTCGCATCCCAGCCT
143	rs2040411_iSNPI_T_F	GACCAACTGGCTTAACAGATGCAAAT
144	rs2046361_iSNPI_T_F2	TCCTTACCTTAAGACTTTCCATTG
145	rs2056277_iSNPI_T_F2	CATTATCTCGTCATACTTCCCTGTCTTG
146	rs2076848_iSNPI_T_F	GCATCAAATTACCCAGTGAAATTATTGA
147	rs2107612_iSNPI_T_F	ATGAGTACATTATCAACTGTTTGGAG
148	rs2111980_iSNPI_T_F	CAGCCATGTTGAAACATTTTACGGTC
149	rs214955_iSNPI_T_F	GCACATTCTAAGAACCTGGTGATTCTATC
150	rs221956_iSNPI_T_F	GCTAGAAAAAGCTGAGATAGCTGTGAAG
151	rs2342747_iSNPI_T_F	CCTTGAAGCTCATTCTTGTGTC
152	rs2399332_iSNPI_T_F	CTGGACACCAGACCAAAACAAATAACC
153	rs251934_iSNPI_T_F	GTAATTAGAGGGCAGTGAGGCTTTAA
154	rs279844_iSNPI_T_F	CTCCAGAACGACTGGATATTAATTAG
155	rs2830795_iSNPI_T_F	TGAGCCAATCAGCAATATAATAGGACT
156	rs2831700_iSNPI_T_F	CCTAGAACCAATTATCTGCTTTGGC
157	rs2920816_iSNPI_T_F2	CCATTGATTCTCTACAGTCTGCAGGTA
158	rs321198_iSNPI_T_F	CTCCACACTTATACAGGTGAAATCTGA
159	rs338882_iSNPI_T_F	CATTTTCTCTCCTCTGTCTCACCTTC
160	rs354439_iSNPI_T_F	GCTTCTCTTCCCTTATGTATCTCTC
161	rs3780962_iSNPI_T_F	GGCTTTGAAGAAAACACTAACCTGTC
162	rs430046_iSNPI_T_F	CACCTATGGCTCTTATTCTCC
163	rs4364205_iSNPI_T_F	CATTGATGCCATTGGTTGTTCCA
164	rs445251_iSNPI_T_F	CCATCACACTATCCTGACATGAACAAAT
165	rs4606077_iSNPI_T_F	GAAGATTGCATCCCAGTGAAAGCAC
166	rs560681_iSNPI_T_F	GCACTTCATAAAGAACAGTCAGGATGC
167	rs6444724_iSNPI_T_F	GGAGAACATCAGGAAATAGTCACCTCCTAC
168	rs6811238_iSNPI_T_F	CATTGACCTCTAGCCAAATGAAGTAC
169	rs7041158_iSNPI_T_F	GGAATTCTGAGAATAACATTGCCTCTC
170	rs717302_iSNPI_T_F	CATATGTTGGGGAGCTAACCTAATGA
171	rs719366_iSNPI_T_F	CACTGTGACCACAGCATTTAACTC
172	rs722098_iSNPI_T_F2	GGGTAAAGAAATATTCAAGCACATCCAAA
173	rs722290_iSNPI_T_F	GAGTATCCTTATCTAAATGCTGGTCC
174	rs727811_iSNPI_T_F	CTTTCTCTTACCGGAACCTCAACGAC
175	rs729172_iSNPI_T_F	CCTCATTATGACCAAGGCTCCTCTG
176	rs733164_iSNPI_T_F	TGACTCTAATTGGGGATGTGGAATTAG
177	rs735155_iSNPI_T_F	GACCTAACCTGGAGAAAACGGAGA
178	rs740598_iSNPI_T_F	GTTTCTCTCTGTGAAACCTTGTCTCAG
179	rs740910_iSNPI_T_F	GCAAACACACAAAGATAGGTTGAGTTT
180	rs763869_iSNPI_T_F	CATATCAAGTGCTTCTGTTGACATTG
181	rs8037429_iSNPI_T_F	CTGAAAAGTGCTACGTAAGAGGTCATTG

SEQ ID NO	CEBADOR DE SNP	EJEMPLOS DE CEBADORES DE SNP SIN MARCADORES
182	rs8078417_iSNPI_T_F	CATCTGAGTGTGAGAAGAGCCTCAA
183	rs826472_iSNPI_T_F2	CCCAGCAAAAACCTCTTTCTCCAGTAA
184	rs873196_iSNPI_T_F	GCTAGGAAAGTTCTCTGGTTACA
185	rs876724_iSNPI_T_F	GAATATCTATGAGCAGGCAGTAGCAG
186	rs891700_iSNPI_T_F2	CTAACAGTGTCACTATGTGAGCTAT
187	rs901398_iSNPI_T_F	CATCATACAGACTCAAGGAGCTAGCTG
188	rs907100_iSNPI_T_F	CTTCCAAGCCTGGAAAACACAGAAAA
189	rs914165_iSNPI_T_F	GTACCTTATAAATCACGGAGTGCAGAC
190	rs917118_iSNPI_T_F	CAAGTGGTAAGAGATGACTGAGGTCAA
191	rs938283_iSNPI_T_F	CTTCTCTCTTAGAAGGACACTGGTCAG
192	rs964681_iSNPI_T_F	GTTATGGAGGATTGGTAAGAACCCAGAG
193	rs987640_iSNPI_T_F	GAGCTGTTAAGGGAAAGGGTAGTTA
194	rs9905977_iSNPI_T_F	GCAGACAAAACCATGACAATGATCTTAG
195	rs993934_iSNPI_T_F	CCCATGATGAAACAGTTGCACTAAATG
196	rs9951171_iSNPI_T_F	CTCAATTCTGTCCCTGCTTCATG
197	rs10092491_iSNPI_S_R2	TTAGAAATTCCAGATAGAGCTAAACTG
198	rs1015250_iSNPI_S_R	GTTAGGAAAAGAACCCAGGTGTTT
199	rs1024116_iSNPI_S_R2	GCAAAAGTAAATACAAGGCATACTTT
200	rs1028528_iSNPI_S_R2	CAATGAAAAGAAAGGTCTTACTCGAC
201	rs1029047_iSNPI_S_R2	CATTCTAAACTCTAAACAAACATTG
202	rs1031825_iSNPI_S_R2	GGTCCTAACCTATTAAATTAAATGAG
203	rs10488710_iSNPI_S_R	GACTTCAATTATGTCAGCATTAAAA
204	rs10495407_iSNPI_S_R	CCTCTGGTTGCATTGGATTCTCATTG
205	rs1058083_iSNPI_S_R	TCTCCATGAAACTGGGTAATTTC
206	rs10773760_iSNPI_S_R	TGTCTGGAAGTTCGTCAAATTGCAG
207	rs1294331_iSNPI_S_R2	GTAGCATAAAACATTCCAAAATTCAAT
208	rs12997453_iSNPI_S_R	TGCTTAAAGATACAGGTTATCTGTATTAC
209	rs13182883_iSNPI_S_R	CTCTCCGTTACTTTCTCCTGCCCTT
210	rs13218440_iSNPI_S_R	GATCCTGAGATTCACCTCTAGCCCT
211	rs1335873_iSNPI_S_R	CCGTACCAGGTACCTAGCTATGTACT
212	rs1336071_iSNPI_S_R2	CTTTCTGTTTGTCATCTGAAATTCT
213	rs1355366_iSNPI_S_R	CAAAGTTAAGTATCACCATCCAGCTGG
214	rs1357617_iSNPI_S_R	ATAGGGATAGCTGATAAGAACATGACC
215	rs1382387_iSNPI_S_R	CTTAATAAGACGCTGCATCTGCCCA
216	rs1413212_iSNPI_S_R	TCCAGGAGACATTGTCATATAAGTGA
217	rs1454361_iSNPI_S_R	AGACACTTTCACTATCCATTAGAAC
218	rs1463729_iSNPI_S_R	GTTTCACATGTGCATGCTTTGGGT
219	rs1493232_iSNPI_S_R	CCAAAGCTATTCTCTTTGGGTGC
220	rs1498553_iSNPI_S_R	GAAAGTCACTTCAGATGTTCAAAGCC
221	rs1523537_iSNPI_S_R	GGGTTTCAGTCTGCAACAAGATCTG

SEQ ID NO	CEBADOR DE SNP	EJEMPLOS DE CEBADORES DE SNP SIN MARCADORES
222	rs1528460_iSNPI_S_R	TGGAGATCAATATTAGCCTAACATAT
223	rs159606_iSNPI_S_R	GACTGTTCTCATCCTGTTATTAGTGT
224	rs1736442_iSNPI_S_R	AACACACAGAAACATCAAGCTGAGC
225	rs1821380_iSNPI_S_R	TTCCTGACATTCTCCTCTTCTATCTG
226	rs1886510_iSNPI_S_R	TATGACGCCGGATTTACAACAAAC
227	rs1979255_iSNPI_S_R	CAGAGACTATGGATGGTATTAGGTCAA
228	rs2016276_iSNPI_S_R	ACTTTGTGTGGCTGAGAGAGAGAAA
229	rs2040411_iSNPI_S_R	TGAGTGTTCTGTATTCTACTCTAAG
230	rs2046361_iSNPI_S_R2	ATTTTGTCATTGTTGACACTTCACC
231	rs2056277_iSNPI_S_R2	GGTGTTAGGGAGACAGGCATGAATG
232	rs2076848_iSNPI_S_R	TGAAACTTTCAACTCTCCTACCGCC
233	rs2107612_iSNPI_S_R	GTTAAAATTGCCACTAATTATGTGTTT
234	rs2111980_iSNPI_S_R	AACTGATCCTATGCAGCAAGATCTTG
235	rs214955_iSNPI_S_R	GATGCTTGCACAAAGACTGAAAAGG
236	rs221956_iSNPI_S_R	GTCTGTGTGTCCTCTGAGATGATGAATG
237	rs2342747_iSNPI_S_R	GGGAGGAAGAAAACAGAGAGTCTGA
238	rs2399332_iSNPI_S_R	AGTTTGTGGCTTCTTTGAGAAGTATC
239	rs251934_iSNPI_S_R	GGCAGATGAAGTAGATAGATCTGGCTG
240	rs279844_iSNPI_S_R	GTTCAGTGTCAATTGACCAGATATT
241	rs2830795_iSNPI_S_R	AGACATAGGACACACCATTATTGTCT
242	rs2831700_iSNPI_S_R	TCAAAATATTGGCTAAACTATTGCCGG
243	rs2920816_iSNPI_S_R2	CTGGAGTTATAATAAAATTGGATTATAGC
244	rs321198_iSNPI_S_R	TTACCTGTTCCCTTTGTGATTCCAC
245	rs338882_iSNPI_S_R	ACCAAGTCAAGAGCTCTGAGAGACAT
246	rs354439_iSNPI_S_R	ACAGTGAATGATATTCAAATATTGTGC
247	rs3780962_iSNPI_S_R	GAACAAGGTCAAGATATCAGCTTCACC
248	rs430046_iSNPI_S_R	AGGTCAACATGAATGGTGTGATGT
249	rs4364205_iSNPI_S_R	ATCCACCCATGAGAAATATCCACAA
250	rs445251_iSNPI_S_R	ACAATTCAAATTAAATGTAAAACTGCAAGTG
251	rs4606077_iSNPI_S_R	TAGTTCTAGTGTGGATCTGACTCC
252	rs560681_iSNPI_S_R	GAACATCTGTCAGGTTCTCTCCATC
253	rs6444724_iSNPI_S_R	GAAAGGACTAAATTGTTAACACTGGT
254	rs6811238_iSNPI_S_R	TGTGTGTTTAAAGCCAGGTTGTT
255	rs7041158_iSNPI_S_R	GATGGACTGGAAC TGAGGATTTC
256	rs717302_iSNPI_S_R	AGCTTAGAAAGGCATATCGTATTAACTG
257	rs719366_iSNPI_S_R	TTATAGTGAGTAAGGACAGGCC
258	rs722098_iSNPI_S_R2	ACACATCTGTTGACAGTAATGAAATATCC
259	rs722290_iSNPI_S_R	GTTTAAACTGGATACCATCCCCAAGAC
260	rs727811_iSNPI_S_R	ATGAGATTGCTGGAGATGCAGATG
261	rs729172_iSNPI_S_R	CACATTCCCTCTGCGGTTACATAC

SEQ ID NO	CEBADOR DE SNP	EJEMPLOS DE CEBADORES DE SNP SIN MARCADORES
262	rs733164_iSNPI_S_R	GACAAGCCTCGCTTGAGTTTCTT
263	rs735155_iSNPI_S_R	TGTGAGAGTGTACCGAATTCAACG
264	rs740598_iSNPI_S_R	AAATAGCAATGGCTCGTCTATGGTTAG
265	rs740910_iSNPI_S_R	TGCTAAGTAAGGTGAGTGGTATAATCA
266	rs763869_iSNPI_S_R	ATAAAATATGATGTGGCTACTCCCTCAT
267	rs8037429_iSNPI_S_R	GCTACACCTCCATAGTAATAATGTAAGAG
268	rs8078417_iSNPI_S_R	TGAAGCAGCTAGAGAACTCTGTACGT
269	rs826472_iSNPI_S_R2	TTTTGTCTCTGTTATATTAGTCACCTATCTC
270	rs873196_iSNPI_S_R	ATAGCCCTGCATTCAAATCCCAAGTG
271	rs876724_iSNPI_S_R	TCCATTTTATACCAC TGCACTGAAG
272	rs891700_iSNPI_S_R2	GCAGTAAAACATTTCATCAAATTCCA
273	rs901398_iSNPI_S_R	TCTGGGTGCAAACTAGCTGAATATCAG
274	rs907100_iSNPI_S_R	GAAAATCTGGAGGCAATT CATGATGCC
275	rs914165_iSNPI_S_R	ATACAATGATGATCACACGGGACCC
276	rs917118_iSNPI_S_R	CCATGAAGATGGAGTCAACATT TACA
277	rs938283_iSNPI_S_R	TCCTAACCCCTAGTACGTTAGATGTG
278	rs964681_iSNPI_S_R	GAGGTGATTCTGTGAGGAACGTCG
279	rs987640_iSNPI_S_R	GTACATTCACTAACAGGCTCTTTCC
280	rs9905977_iSNPI_S_R	AATT CATGAGCTGGTGTCCAAGGAG
281	rs993934_iSNPI_S_R	ATAACAGTCTCCAGAGTATATTAGCTTAG
282	rs9951171_iSNPI_S_R	GTT CCTCTGGATGCAACATGAGAG
283	rsI0497191_aSNPI_T_F	GAAAGGATGAAGAGGGTGGATATTGGAG
284	rs1079597_aSNPI_T_F	CCAAACCTCATCATCTTACCTGGATT
285	rs11652805_aSNPI_T_F	GTCCAAAGTCAAGTGCAAGTATAGTTGG
286	rs1229984_aSNPI_T_F	ACAATCTTCTGAATCTGAACAGCTTC
287	rs12439433_aSNPI_T_F	CAAAGGAAGGCATT CCTAATGATCTTC
288	rs12498138_aSNPI_T_F	CTTGCTTGCTTTCTTCTTCAGGGAA
289	rs12913832_pSNPI_NU_T_F	CTGCTTCAAGTGTATATAAAACTCACAGT
290	rs1426654_aSNPI_T_F	CCTAGGAAAGCAGTA ACTAATT CAGGAG
291	rs1462906_aSNPI_T_F	GCAATTGTTCACTTTAGTTCTGTAGC
292	rs1572018_aSNPI_T_F	GGCCTAATATGCATGTGTTCATGTCTCT
293	rs16891982_aSNPI_T_F	CAGAGTTCTCATCTACGAAAGAGGAGT
294	rs174570_aSNPI_T_F	ATCCTAGACCTCCAGGTGGAATGATC
295	rs17642714_aSNPI_T_F	CTTGGCTGCTCAATATTTGGAGTAAG
296	rs1S00414_aSNPI_T_F	GAGTAAATGAGCTGTGGTTCTCTTAA
297	rs1834619_aSNPI_T_F	CTTCCATGTGGACCC TTAACATT CAG
298	rs1876482_aSNPI_T_F	GCATAGT GAGCT GTT GATAGAGCTTTG
299	rs1919550_aSNPI_T_F	CTAGAACAAATCATTGGCTCCTAGT
300	rs192655_aSNPI_T_F	GTCTGGT GAGTACTGGCTGAATGTAAA
301	rs200354_aSNPI_T_F	CCAGAGGATGCTGCTAACATTCTACAA

SEQ ID NO	CEBADOR DE SNP	EJEMPLOS DE CEBADORES DE SNP SIN MARCADORES
302	rs2024566_aSNPI_T_F	GCTCATGCCTGGAATTCACCTTATTT
303	rs2042762_aSNPI_T_F	CTAACTAGACATTGGGCCACCTACTT
304	rs2166624_aSNPI_T_F	GTCTATGGTGCCTATAGAATGTACAGGT
305	rs2196051_aSNPI_T_F	CCCTCTCAAGTTGTGAGCAAATATCAC
306	rs2238151_aSNPI_T_F	CTCTATCTGCTGCAATGGACTTCC
307	rs260690_aSNPI_T_F	CCTAGAACAGATTGAAGGGCTTTG
308	rs2814778_aSNPI_T_F	AAATGAGGGGCATAGGGATAAGGGA
309	rs310644_aSNPI_T_F	CCTAGAAATCTGATACGTTATCCTATGA
310	rs3737576_aSNPI_T_F	AGGAGAGATATTCACATGAACCCAA
311	rs3811801_aSNPI_T_F	GAACATCTGACCAGAAATTCCAGTA
312	rs3823159_aSNPI_T_F	GTGTAGTGAAATCCTTAGACTTAGGTAA
313	rs3916235_aSNPI_T_F	AATACATGAAAAAGTAATACATGGGGCA
314	rs4471745_aSNPI_T_F	ATTAAATGTTACTTCTATCTACAAGGA
315	rs4833103_aSNPI_T_F	CATTTGTGAAATGCAAAGGGCAAATCT
316	rs4891825_aSNPI_NU_T_F	GCTGAGAGGCTTAATTCCATCAAGATGA
317	rs4918664_aSNPI_NU_T_F	CCCATCCTAAACTTAGTTATGGCAG
318	rs6754311_aSNPI_T_F	GTAACACATTCTCTTGGGAAGCTAGC
319	rs6990312_aSNPI_NU_T_F	CTTAGCTTCAGTGAAAATGGTCCTCTC
320	rs7226659_aSNPI_NU_T_F	CTTTCTTAGCTCCTCTCCATTCTCTTC
321	rs7326934_aSNPI_NU_T_F	GTCTATGCAGTGCTTCAGTGAGGATTAT
322	rs735480_aSNPI_NU_T_F	CTCTATCTGCTCAGAGCCTGCTTAAAG
323	rs7554936_aSNPI_NU_T_F	GGAAAGGATACAGTGTGAGCAAGATAG
324	rs7657799_aSNPI_NU_T_F	GCCAACTTGATTCTCTTCAAATGCTTG
325	rs7722456_aSNPI_T_F	AGATGGGTTTACCATGTTCCCAG
326	rs798443_aSNPI_T_F	GTACAGTAGTTAGTTCCAGACTGATGA
327	rs7997709_aSNPI_T_F	GTAAATATCTAACTGTGTTCCCTCAGT
328	rs870347_aSNPI_T_F	GAACCAAAAGGAATTAGAGACTAGGGG
329	rs917115_aSNPI_T_F	CTGCTTTACGGCTCTCCTTCTTC
330	rs10497191_aSNPI_S_R	CCCACATCCTCCCATTATAGGCAA
331	rs1079597_aSNPI_S_R	TACATGATCCTAAGGGCAGCAGGAA
332	rs11652805_aSNPI_S_R	GTTTGGTGCATCCTCTTCTCTC
333	rs1229984_aSNPI_S_R	GACTGTAGTCACCCCTCTCCAACA
334	rs12439433_aSNPI_S_R	AGAGTGAATACATAGAAAAGAAACTAAAG
335	rs12498138_aSNPI_S_R	ATTGCGAGAACAGATAATATTGAAG
336	rs12913832_pSNPI_NU_S_R	ACAGGAACAAAGAATTGTTCTCATGG
337	rs1426654_aSNPI_S_R	CCTTGGATTGTCTCAGGATGTTGCA
338	rs1462906_aSNPI_S_R	CTGGGATGTTGGCTTGTG
339	rs1572018_aSNPI_S_R	ATTGGTAGTACACTAATGGATATATGTGAG
340	rs16891982_aSNPI_S_R	GAATAAAGTGAGGAAAACACGGAGTTG
341	rs174570_aSNPI_S_R	GAGAGAGGCAGAAAGGAGGGATGAA

SEQ ID NO	CEBADOR DE SNP	EJEMPLOS DE CEBADORES DE SNP SIN MARCADORES
342	rs17642714_aSNPI_S_R	TACTCTGTCTTCAGTAGCPGTTCTTGG
343	rs1S00414_aSNPI_S_R	TTAGACTACCAAGATCAAGATGAATGC
344	rs1834619_aSNPI_S_R	ATCTCAATAAGCTGTTAAAACAGAAAG
345	rs1876482_aSNPI_S_R	TAAAGAAAATGCCATGGGCTGTACCC
346	rs1919550_aSNPI_S_R	ATTGTGCAGCAGAACAGAGTAGTG
347	rs192655_aSNPI_S_R	ATTCTTGACATAGCTCACGAAATTCCC
348	rs200354_aSNPI_S_R	AAAATGAGACCTCGTATCTTGCAGC
349	rs2024566_aSNPI_S_R	AAATGCAGAACTGCCAAAAGAAACCC
350	rs2042762_aSNPI_S_R	GAGAATCTGTGAATGCCAGGGTCTG
351	rs2166624_aSNPI_S_R	ATGGATTCTATGTTTCAGACATCTAATT
352	rs2196051_aSNPI_S_R	ATCACTAGAAAGAAAAGAGTCCCTATT
353	rs2238151_aSNPI_S_R	GAAGTTAAAAGAGTGGAACATGGGG
354	rs260690_aSNPI_S_R	CTACGTAAGCAAAATGATCACGCAC
355	rs2814778_aSNPI_S_R	AACCTGATGCCCTCATTAGTCCTT
356	rs310644_aSNPI_S_R	CACCAGATTCTAGGAATAGCATGTGAG
357	rs3737576_aSNPI_S_R	AAGAGCATAGTGAGGGGTTAGACCT
358	rs3811801_aSNPI_S_R	CTTTATATTAGTGTAGAGATCAGTCTCC
359	rs3823159_aSNPI_S_R	TGAGTCCTTACCTAACCTGGTTGTC
360	rs3916235_aSNPI_S_R	AATCCAAGCAACTCTCTTTGACCAC
361	rs4471745_aSNPI_S_R	TTTACTGGAACCCTGATTTGTTGGA
362	rs4833103_aSNPI_S_R	TGCCACTGATATATCAGTACCTGAGT
363	rs4891825_aSNPI_NU_S_R	ACAATCTCAATCCCCCTTAATGTTTC
364	rs4918664_aSNPI_NU_S_R	GTGGGCAGAGAGAGTAAGAGAACCT
365	rs6754311_aSNPI_S_R	CAAACCAGATTCTGGCAGAACATAGTTAGC
366	rs6990312_aSNPI_NU_S_R	CTTCTCTCCATCCTCCTCTCCAC
367	rs7226659_aSNPI_NU_S_R	AGATCAAGGGATCTGTGGGACAATAAC
368	rs7326934_aSNPI_NU_S_R	GGGGAGTGATTCAAGCATCCTGATT
369	rs735480_aSNPI_NU_S_R	CATGAGTTGAGGTAAGATGAAGGAGA
370	rs7554936_aSNPI_NU_S_R	TCTCTCTCATCCTAGTGAATGCCATC
371	rs7657799_aSNPI_NU_S_R	GGGTGATGATCTACCTGCAGGTATA
372	rs7722456_aSNPI_S_R	CTCAAGGCCCTGGTCTGAATTAC
373	rs798443_aSNPI_S_R	ACATCTCCAGTTAATAATTCCACTAAC
374	rs7997709_aSNPI_S_R	TGGATTGCTCAACAAATAGTGCTAAAA
375	rs870347_aSNPI_S_R	CATGCGACATCCAGGTAGCTAAATAC
376	rs917115_aSNPI_S_R	ATGGATAAAAATGGAACCTTCAAGAGAA
377	rs12203592_pSNPI_T_F	GTTTTATGAAAGCTCGTCATATGGCT
378	rs12821256_pSNPI_T_F	GTTCCAACCTAGTCATAAGTCCCTGG
379	rs12896399_pSNPI_T_F	GGGTCTTGATGTTGATTGATGAGGAAG
380	rs1393350_pSNPI_T_F	CCTAACAGAAAGTCACTGTTGTATCTG
381	rs1800407_pSNPI_T_F	TCACTCTGGCTTGACTCTCTGTG

SEQ ID NO	CEBADOR DE SNP	EJEMPLOS DE CEBADORES DE SNP SIN MARCADORES
382	rs2378249_pSNPI_T_F	GGCTGGTTCAGTCTGGAGACTTATTT
383	rs2402130_pSNPI_T_F	CTTCACCTCGATGACGATGATGATGAT
384	rs4959270_pSNPI_T_F	GACAATAACAGCACAAAGGATGGAAAAG
385	rs1805009_pSNPI_T_F	GAACCAGACCACACAATATCACCAC
386	rs28777_pSNPI_T_F	TCTACCTCTTGATGTCCCCTCGATAG
387	rs16891982_pSNPI_T_F	CAGAGTTCTCATCTACGAAAGAGGGAGT
388	rs683_pSNPI_T_F	CCCAGCTTGAAAAGTATGCCTAGAACT
389	rs12913832_pSNPI_T_F	CTGCTTCAAGTGTATATAAAGTCACAGT
390	rs12203592_pSNPI_S_R	TTGTTTCATCCACTTGGTGGTAAAAG
391	rs12821256_pSNPI_S_R	TAATTAAAGCTCTGTGTTAGGGTTTT
392	rs12896399_pSNPI_S_R	CAATTCTTGTCTTAGGTCAGTATAT
393	rs1393350_pSNPI_S_R	TACTCTCCTCAGTCCCTCTGC
394	rs1800407_pSNPI_S_R	TGAGACAGAGCATGATGATCATGGC
395	rs2378249_pSNPI_S_R	GCACAAGTCTAGGAACACTTTGCAC
396	rs2402130_pSNPI_S_R	GAAGTATTGAAACCATAACGGAGCCC
397	rs4959270_SNPI_S_R	TGAGGAACACATCCAAACTATGACAC
398	rs1805009_SNPI_S_R	TTTCTCGCCCTCATCATCTGCAATG
399	rs28777_pSNPI_S_R	TCAGTTGATTCATGTGATCCTCACAG
400	rs16891982_pSNPI_S_R	GAATAAAGTGGAGAAAACACGGAGTTG
401	rs683_pSNPI_S_R	ATTACCTCTTCTAATACAAGCATATG
402	rs12913832_pSNPI_S_R	ACAGGAACAAAGAATTGTTCTCATGG

Ejemplo 2: Elaboración de perfiles de ADN para un banco de datos

Este ejemplo describe un experimento que sigue el flujo de trabajo de la figura 2. Este ejemplo no utiliza UMI, ya que se considera que las muestras obtenidas son de individuos cuya identidad ya se conoce.

- 5 Para este experimento, las STR se multiplexan con iSNP como se muestra en la tabla 3.

Tabla 3 - SNP y STR informativos de identidad

SNP informativos de identidad				
rs1005533	rs1357617	rs2076848	rs4530059	rs763869
rs10092491	rs1360288	rs2107612	rs4606077	rs8037429
rs1015250	rs1382387	rs2111980	rs560681	rs8078417
rs1024116	rs1413212	rs214955	rs576261	rs826472
rs1028528	rs1454361	rs221956	rs6444724	rs873196
rs1029047	rs1463729	rs2269355	rs6811238	rs876724
rs1031825	rs1490413	rs2342747	rs6955448	rs891700
rs10488710	rs1493232	rs2399332	rs7041158	rs901398
rs10495407	rs1498553	rs251934	rs717302	rs907100
rs1058083	rs1523537	rs279844	rs719366	rs914165
rs10773760	rs1528460	rs2830795	rs722098	rs917118
rs10776839	rs159606	rs2831700	rs722290	rs938283

ES 2 866 044 T3

SNP informativos de identidad				
rs1109037	rs1736442	rs2920816	rs727811	rs964681
rs1294331	rs1821380	rs321198	rs729172	rs987640
rs12997453	rs1886510	rs338882	rs733164	rs9905977
rs13182883	rs1979255	rs354439	rs735155	rs993934
rs13218440	rs2016276	rs3780962	rs737681	rs9951171
rs1335873	rs2040411	rs430046	rs740598	
rs1336071	rs2046361	rs4364205	rs740910	
rs1355366	rs2056277	rs445251	rs7520386	
STR autosómicos				
D1S1656	CSF1PO	vWA	D21S11	D4S2408
D2S441	D7S820	D13S317	TPOX	D17S1301
D2S1338	D8S1179	Penta E	SE33	D9S1122
D3S1358	D10S1248	D16S539	Penta D	D6S1043
FGA	TH01	D18S51	D22S1045	Amelogenina
D5S818	D12S391	D19S433	D20S482	
STR de X				
DXS8378	DXS8377	DXS10101	DXS10148	DXS10146
DXS7132	DXS10135	DXS10134	DXS10079	
HPRTB	DXS10074	DXS7423	DXS10103	
STR de Y				
DYS456	DYS393	DYS437	DYS533	DYS449
DYS389I/II	DYS391	DYS438	DYS518	DYS522
DYS390	DYS439	DYS448	DYS570	DYS505
DYS458	DYS635	DYS576	DYS643	DYS627
DYS19	DYS392	DYS481	DYS460	DYF387S1a/b
DYS385a/b	YGATAH4	DYS549	DYS612	

Por supuesto, se podrían agregar SNP y STR adicionales a la lista anterior. Los ejemplos de otras dianas posibles incluyen, entre otros, los marcadores que se encuentran en la tabla 4.

5 Tabla 4 - Ejemplos de STR y SNP adicionales para la multiplexación

SNP informativos de identidad				
rs1004357	rs1554472	rs2567608	rs521861	rs9606186
rs1019029	rs1872575	rs2811231	rs5746846	rs985492
rs1027895	rs2073383	rs2833736	rs590162	rs9866013
rs10500617	rs2175957	rs315791	rs6591147	
rs10768550	rs2255301	rs3744163	rs689512	
rs12480506	rs2270529	rs4288409	rs7205345	
rs13134862	rs2272998	rs464663	rs7229946	
rs1358856	rs2291395	rs4789798	rs7704770	
rs1410059	rs2292972	rs4796362	rs8070085	

SNP informativos de identidad			
rs1478829	rs2503107	rs4847034	rs9546538
STR autosómicos			
D1S1677	D3S4529	D18S853	D10S1435
D11S4463	D6S1017	D14S1434	D5S2500
D1S1627	D1GATA113	D2S1776	

- Los cebadores se diseñaron para contener una secuencia de cebador de PCR específica de gen en el extremo 3' y una secuencia de marcador adaptador en el extremo 5'. En este experimento, los cebadores directos contienen la secuencia de marcador para los adaptadores TruSeq Custom Amplicon i5 y los cebadores inversos contienen la secuencia de marcador para los adaptadores TruSeq Small RNA kit i7. Los marcadores se pueden usar como sitios de cebadores de amplificación, así como sitios de cebadores de secuenciación.

Secuencia de marcadores del adaptador i5	5'TACACGACGCTCTCCGATCT3 '	(SEQ ID NO:403)
Secuencia de marcadores del adaptador i7	5'CTTGGCACCCGAGAATTCCA3 '	(SEQ ID NO:404)

- Para equilibrar la amplificación entre los STR y los SNP en el múltiplex, se modificaron los parámetros de diseño de cebadores para los SNP como se describe en el ejemplo 1. El conjunto original de cebadores de SNP diseñados con Design Studio de Illumina eran cebadores de PCR clásicos: secuencias cortas con altas temperaturas de fusión. y poca o ninguna estructura secundaria. Design Studio se utilizó para diseñar las sondas de amplicón personalizadas TruSeq y para crear el complemento inverso de la sonda cadena abajo para preparar el cebador de PCR inverso. Sin embargo, estos cebadores no se multiplexaron bien y un cebador malo podía convertir el ensayo de bueno a malo (por ejemplo, todo dímeros de cebadores y ningún producto) (figura 4). En un intento por crear mejores cebadores para multiplexar, se usó Primer3 (shareware) que contiene una función de banco de cebado incorrecto. Se descubrió que los cebadores diseñados con Primer3 funcionaron incluso peor en el ensayo múltiplex que los cebadores de Design Studio. Sorprendentemente, se estaban generando datos que mostraban que los cebadores de STR se multiplexaban bien. Se observó que las parejas de cebadores mal diseñados dirigidos a dianas STR no causaron fallos de multiplexación, como sí lo hicieron los cebadores de SNP. Los cebadores de STR son largos, ricos en AT y tienen bajas temperaturas de fusión, al contrario de lo que se conoce como un cebador "bueno".

- Los cebadores de SNP se rediseñaron siguiendo los parámetros del ejemplo 1. Los cebadores se mezclaron para todas las dianas. Para este ejemplo, se mezclaron parejas de cebadores para 56 STR con parejas de cebadores para 75 iSNP, aSNP y SNP informativos de fenotipo. Se añadió polimerasa (Hot-start Phusion II en este ejemplo) a una mezcla maestra de todos los componentes necesarios para la PCR y se añadieron los cebadores. La mezcla se pipeteó hacia los pocillos de una placa de PCR, pero la amplificación también se pudo realizar en tubos, etc. Se añadió ADN a la placa como ADN purificado en un volumen de 15 microlitros; sin embargo, también se pueden utilizar extractos lisados de muestras de sangre o bucales de hisopos o papel de filtro no tratado, o directamente de muestras de sangre o bucales en tarjetas FTA, etc. Para este experimento, se utilizó ADN de control 2800M purificado a 1 ng y 100 pg. Las reacciones se sometieron a PCR durante un número determinado de ciclos (en el caso del ejemplo, 25 ciclos) siguiendo el protocolo:

95 °C	-	3 min	
96 °C	-	1 min	
54 °C	-	2 min *	* pendiente 0,5 °C/segundo
68 °C	-	45 seg **	**pendiente 0,2 °C/segundo
60 °C	-	30 min	
4 °C	-	mantenimiento	

- Después de ciclar, las placas se retiraron del termociclador. La reacción se llevó a 50 microlitros con polimerasa (Kapa HiFi, Kapa Biosystems), mezcla maestra de PCR que contenía todos los componentes necesarios para la PCR y un par de adaptadores (un adaptador i7 y un adaptador i5). Se realizó una segunda ronda de PCR para un número determinado de ciclos (10 ciclos en el caso del ejemplo) para generar los bancos de secuenciación, siguiendo el protocolo:

98 °C	-	30 segundos
98 °C	-	15 segundos
66 °C	-	30 segundos
72 °C	-	1 min
72 °C	-	5 min
10 °C	-	mantenimiento

Después de ciclar, la placa que contenía los bancos completados se retiró del termociclador. En este momento, las muestras pueden agruparse por volumen y purificarse como una sola muestra utilizando, por ejemplo, esferas magnéticas (SPRI). Las muestras también se pueden purificar individualmente. La reunión o los bancos individuales

5 se pueden cuantificar utilizando un método basado en qPCR, utilizando un analizador de fragmentos o un bioanalizador, o utilizando PicoGreen y un lector de placas (como en el caso del ejemplo). Los expertos en la técnica conocerán la gran cantidad de opciones para la cuantificación de bancos. Si los bancos se purifican individualmente, cada concentración puede normalizarse a 2 nM y reunirse por volumen.

10 Las reuniones de bancos purificados se desnaturizaron, se diluyeron, se agruparon y se secuenciaron en el instrumento de secuenciación MiSeq con una ejecución de secuenciación de 350 ciclos y las dos lecturas de índice. Después de la secuenciación, las muestras se desmultiplexaron de acuerdo con las secuencias de adaptador y se analizaron a través del canal de Forensics Genomics (Illumina, Inc.). Las lecturas de STR se separaron de las lecturas de SNP y se analizaron de forma independiente. Las STR se analizaron utilizando el algoritmo descrito en una solicitud de patente anterior (PCT/US2013/30867). Los números de repeticiones y cualquier variación de secuencia se indicaron 15 junto con los números leídos. Los SNP se analizaron mediante un manifiesto y las llamadas se indicaron junto con los números leídos. Se calculó el equilibrio relativo entre los alelos (% mínimo/máximo), el equilibrio entre los loci (% CV), las tasas de error y las tasas de tartamudeo para los loci STR. Los resultados de las STR en el múltiplex del banco de datos inicial se muestran en la figura 5A-C. El equilibrio (equilibrio promedio 80 %), el tartamudeo (aproximadamente 3 %) y las tasas de error (menos del 5 %) cumplen los requisitos de entrada del diseño para los loci incluidos en este 20 ejemplo. El porcentaje de CV (~142%) se calculó utilizando los 56 loci. Aunque los cebadores utilizados muestran equilibrio entre locus (interlocus), se prevé una mayor optimización del cebador para mejorar el equilibrio entre locus. Las llamadas para los loci conocidos coinciden con los resultados publicados para 2800M. Los resultados de los SNP 25 se muestran en la figura 5D-E. La cobertura, llamadas de alelos, tartamudeo y otros artefactos para los 56 loci STR en el múltiplex grande se muestran en la figura 6. Estos gráficos imitan los electroferogramas generados por la tecnología CE. Las barras son análogas a los picos del alelo definido (eje X), y los recuentos leídos (eje Y) son análogos a RFU. La cobertura de los SNP oscilaba entre 10 y 2500X, según el SNP, sin embargo, cada SNP que se multiplexó se contó y proporcionó llamadas precisas.

Ejemplo 3: Elaboración de perfiles de ADN para crímenes

30 Este ejemplo describe un experimento que sigue el flujo de trabajo de la figura 3. Este ejemplo incorpora UMI en los cebadores, ya que se considera que las muestras obtenidas son de individuos cuya identidad aún no se conoce.

Para este experimento, las STR se multiplexaron con iSNP, aSNP y SNP informativos de fenotipo como se muestra en la tabla 5.

Tabla 5 - Investigaciones con STR y SNP

SNP informativos de identidad				
rs1005533	rs1357617	rs2076848	rs4530059	rs763869
rs10092491	rs1360288	rs2107612	rs4606077	rs8037429
rs1015250	rs1382387	rs2111980	rs560681	rs8078417
rs1024116	rs1413212	rs214955	rs576261	rs826472
rs1028528	rs1454361	rs221956	rs6444724	rs873196
rs1029047	rs1463729	rs2269355	rs6811238	rs876724
rs1031825	rs1490413	rs2342747	rs6955448	rs891700
rs10488710	rs1493232	rs2399332	rs7041158	rs901398
rs10495407	rs1498553	rs251934	rs717302	rs907100
rs10S8083	rs1523537	rs279844	rs719366	rs914165
rs10773760	rs1528460	rs2830795	rs722098	rs917118

ES 2 866 044 T3

SNP informativos de identidad				
rs10776839	rs159606	rs2831700	rs722290	rs938283
rs1109037	rs1736442	rs2920816	rs727811	rs964681
rs1294331	rs1821380	rs321198	rs729172	rs987640
rs12997453	rs1886510	rs338882	rs733164	rs9905977
rs13182883	rs1979255	rs354439	rs735155	rs993934
rs13218440	rs2016276	rs3780962	rs737681	rs9951171
rs1335873	rs2040411	rs430046	rs740598	
rs1336071	rs2046361	rs4364205	rs740910	
rs1355366	rs2056277	rs445251	rs7520386	
STR autosómicos				
D1S1656	CSF1PO	vWA	D21S11	D4S2408
D2S441	D7S820	D13S317	TPOX	D17S1301
D2S1338	D8S1179	Penta E	SE33	D9S1122
D3S1358	D10S1248	D16S539	Penta D	D6S1043
FGA	TH01	D18S51	D22S1045	Amelogenina
D5S818	D12S391	D19S433	D20S482	
STR de X				
DXS8378	DXS8377	DXS10101	DXS10148	DXS10146
DXS7132	DXS10135	DXS10134	DXS10079	
HPRTB	DXS10074	DXS7423	DXS10103	
STR de Y				
DYS456	DYS393	DYS437	DYS533	DYS449
DYS389I/II	DYS391	DYS438	DYS518	DYS522
DYS390	DYS439	DYS448	DYS570	DYS505
DYS458	DYS635	DYS576	DYS643	DYS627
DYS19	DYS392	DYS481	DYS460	DYF387S1a/b
DYS385a/b	YGATAH4	DYS549	DYS612	
SNP informativos de fenotipo				
N29insA	rs1805006	rs1110400	rs12203592	rs2378249
rs11547464	rs1805007	rs28777	rs1042602	rs12896399
rs885479	rs1805009	rs16891982	rs1800407	rs1393350
rs1805008	Y1520CH	rs12821256	rs2402130	rs683
rs1805005	rs2228479	rs4959270	rs12913832	
SNP informativos de ascendencia				
rs10497191	rs17642714	rs2238151	rs4471745	rs7554936
rs1079597	rs1800414	rs2593595	rs459920	rs7657799
rs11652805	rs1834619	rs260690	rs4833103	rs7722456
rs1229984	rs1871534	rs2814778	rs4891825	rs798443
rs12439433	rs1876482	rs310644	rs4918664	rs7997709
rs12498138	rs1919550	rs3737576	rs671	rs870347

SNP informativos de ascendencia				
rs12913832	rs192655	rs3811801	rs6754311	rs917115
rs1426654	rs200354	rs3814134	rs6990312	rs9522149
rs1462906	rs2024566	rs3823159	rs7226659	
rs1572018	rs2042762	rs3827760	rs7251928	
rs16891982	rs2166624	rs3916235	rs7326934	
rs174570	rs2196051	rs4411548	rs735480	

Este ejemplo incluye UMI para los cebadores de STR. Para estos ejemplos, solo los cebadores de STR contienen UMI, sin embargo, tanto los cebadores de STR como de SNP podrían incluir UMI si se desea y esta opción no está excluida de la práctica. Sin embargo, para este ejemplo, solo los cebadores de STR incorporan UMI con fines de demostración. Se introdujeron identificadores de moléculas exclusivos durante dos ciclos de PCR (figura 3). En primer lugar, como en el ejemplo 2, los cebadores de PCR contienen una secuencia de cebador de PCR específica de gen en el extremo 3' y una secuencia de marcador adaptador en el extremo 5', lo mismo que para las secuencias de marcador utilizadas en el ejemplo 2 para las secuencias i5 e i7. En este experimento, los UMI se colocan entre la secuencia del cebador específico del gen y la secuencia de marcador. En el caso de este ejemplo, se utilizaron cinco bases aleatorias para el UMI en los cebadores directo e inverso. Los cebadores se mezclaron para todas las dianas. La mezcla de cebadores constaba de 26 parejas de cebadores de STR autosómicos y 86 parejas de cebadores de SNP (92 SNP cubiertos). Se añadió polimerasa (Hot-start Phusion II en este ejemplo) a una mezcla maestra de todos los componentes necesarios para la PCR y se añadieron los cebadores. La mezcla se pipeteó hacia los pocillos de una placa de PCR. Se añadió ADN a la placa como ADN purificado, óptimamente 1 ng. Como en el ejemplo 2, el ADN purificado del control 2800M se ensayó a 1 ng. La mezcla de reacción múltiple se sometió a dos ciclos de PCR siguiendo el protocolo:

98 °C	3 min		
98 °C	2 min		
54 °C	12 min	pendiente 0,2 °C/seg	
72 °C	4 min	2 ciclos	
4 °C	mantenimiento		

Después de ciclar, las muestras se retiraron del termociclador y se añadió a la reacción proteína de unión a ADN monocatenario (SSB) de *E. coli*. Se contempló que la SSB reduce los dímeros de cebadores mediante los cebadores específicos de genes marcados no utilizados y evita cualquier amplificación posterior de estos cebadores. La SSB se incubó con la muestra en hielo, y como alternativa también se podría usar incubación a TA o 37 °C. Después de esta incubación, se agregó polimerasa (Hot-start Phusion II en este ejemplo) a una mezcla maestra de todos los componentes necesarios para la PCR, y la mezcla maestra se agregó a la muestra con un par de adaptadores (adaptadores i7 e i5) y se cicló un número determinado de ciclos (en este experimento 34 ciclos), siguiendo el protocolo:

95 °C	3 min		
95 °C	30 segundos		
66 °C	30 segundos		
72 °C	1 min		
72 °C	5 min		
10 °C	mantenimiento		

Las muestras se purificaron con esferas SPRI y los bancos individuales se pudieron cuantificar utilizando un método basado en qPCR, utilizando un analizador de fragmentos (como en el caso del ejemplo) o BioAnalyzer, o utilizando PicoGreen y un lector de placas. Cada concentración de los bancos se normalizó a 2 nM y estos se combinaron por volumen.

Los grupos de bancos purificados se desnaturalizaron, se diluyeron, se agruparon y se secuenciaron utilizando MiSeq con una ejecución de secuenciación de 350x100 ciclos y las dos lecturas de índice. Después de la secuenciación, los

datos se determinaron como se informa en el ejemplo 2. Sin embargo, dado que los cebadores contienen UMI, los UMI se usaron para colapsar los datos empleando réplicas de PCR para eliminar los errores y artefactos de secuenciación y de PCR. Los SNP se analizaron mediante un manifiesto, y las llamadas se indicaron junto con los números leídos. Se calculó el equilibrio relativo entre los alelos (% mínimo/máximo), el equilibrio entre loci (% CV), las 5 tasas de error y las tasas de tartamudeo (solo STR). Los resultados del múltiplex de investigaciones inicial se muestran en la figura 7A-E. La cobertura, las llamadas de alelos, el tartamudeo y otros artefactos para los 26 loci STR en el múltiplex grande se muestran en la figura 8. Estos gráficos imitan los electroferogramas generados por CE. Las barras son análogas a los picos y los recuentos de lectura son análogos a RFU. La cobertura de los SNP oscila entre 10 y 5500X, dependiendo del SNP, sin embargo, cada SNP que se multiplexó se contó y proporcionó resultados útiles.

- 10 Un resultado que se generó con estos estudios fue que se demostró que el tartamudeo era un artefacto de la PCR. Muchos investigadores han planteado la hipótesis de esto (y el deslizamiento de la polimerasa se ha indicado en los cánceres de colon humanos), pero esto no se ha demostrado en los ensayos forenses. Los UMI se pueden utilizar para demostrar que el tartamudeo es de hecho un artefacto de PCR. Los productos con n+1 o n-1 repeticiones tienen los mismos UMI que los productos con el número correcto de repeticiones (figura 9). En la figura 9A, cada locus 15 muestra los resultados sin corrección de UMI en comparación con la figura 9B, en la que se realiza la corrección de UMI. Como se demostró, sin la corrección de UMI, el equilibrio entre los alelos no es tan bueno como cuando se realiza la corrección de UMI. Además, hay considerablemente más tartamudeo, que es evidente sin la corrección de UMI. La parte de la barra sobre la línea entre barras representa el error de secuenciación, mientras que debajo de la línea representa la secuencia correcta dentro de la secuencia STR. El error se reduce considerablemente con la corrección 20 de UMI. Por ejemplo, el locus SE33 tiene un error que se elimina con la corrección de UMI. La corrección de errores puede ser extremadamente importante para que la investigación de crímenes proporcione el perfil de ADN más preciso posible.

Ejemplo 4: Elaboración de perfiles de ADN usando 12 individuos de muestra

Métodos y materiales

- 25 Se ensayó el ADN de 12 individuos de muestra (muestra n.º: 1, 3, 4, 5, 6, 7, 10, 13, 14, 15, 16, 17) y un genoma de referencia (2800M) siguiendo el flujo de trabajo de la figura 3. Este experimento incorpora UMI en los cebadores de STR como se describe en el ejemplo 3. Se analizaron dos réplicas de cada muestra con el kit de preparación de bancos de firmas de ADN ForenSeq de Illumina® en el secuenciador MiSeq. Se usó 1 ng de ADN para cada réplica utilizando la mezcla de cebadores de ADN B: mezcla de muestras recolectadas, que contiene cebadores para 61 STR más amelogenina, 95 SNP informativos de identidad, 56 SNP informativos de ascendencia, 22 SNP informativos de fenotipo (también se emplearon 2 SNP de ascendencia para la predicción del fenotipo).

Configuración por defecto

STR: umbral analítico = 6,5 %; umbral de interpretación = 15 %. SNP: umbral analítico = 3 %; umbral de interpretación = 15 %.

- 35 Las llamadas de secuenciación de alto nivel, como la cobertura y los loci llamados, para la elaboración de perfiles de ADN de los 12 individuos de muestra se muestran en la figura 12. Como puede verse, cada locus fue cubierto por al menos 100.000 lecturas en ambas repeticiones. Solo dos muestras dieron como resultado una llamada de STR fallida (1 de 61). Todos los 173 SNP fueron llamados con éxito en todos los individuos en ambas réplicas. Las llamadas de STR de muestra de dos individuos de la muestra se muestran en la figura 16. Las llamadas de SNP de muestra de 40 dos individuos de muestra se muestran en la figura 17. La figura 13 muestra las estadísticas de población, como la probabilidad de apareamiento aleatorio (RMP) de los auto-STR del National Institute of Standards and Technology (NIST), la frecuencia de haplotipos con confianza del 95 % de los NIST Y-STR, el RMP de los iSNP de dbSNP y el RMP de los STR de la base de datos de US Y-STR.

- 45 Los fenotipos, como el color de los ojos y el color del cabello, de los 12 individuos de la muestra y el individuo de referencia se predijeron basándose en el genotipo de pSNP en el experimento, y se compararon con los fenotipos autoinformados (figura 14). Se observó un alto grado de correlación entre los fenotipos predichos e informados.

- La ascendencia de los 12 individuos de la muestra se predijo utilizando el genotipo de 56 aSNP en el experimento. Se calcularon las puntuaciones de PCA1 y PCA3 de cada individuo de la muestra y se trazaron frente a muestras de referencia en una gráfica de ascendencia. Como se muestra en la figura 15, la ascendencia de los individuos de la muestra se puede predecir en función de la ubicación en la gráfica de ascendencia. Se incluyeron catorce puntos centroides en la gráfica de ascendencia (círculos). Sobre la base del punto centroide más cercano, se predijo la ascendencia de cada individuo de la muestra.

- 55 El experimento de elaboración de perfiles de ADN también mostró un alto nivel de equilibrio intralocus tanto en los loci STR como en los loci SNP, como puede verse en la figura 18, y bajo nivel de tartamudeo, como puede verse en la figura 19.

Seis de los 12 individuos más 2800M tienen al menos un locus heterocigótico isométrico, que se muestra en la figura 20. Un locus heterocigótico isométrico se define como un STR que tiene el mismo número de repeticiones, dos

secuencias diferentes, que están igualmente equilibradas. Usando la información sobre los variantes en el STR D8S1179, el alelo 13 de la muestra 15 se rastreó hasta la abuela, muestra 17 (figura 21). Se usó una información de variantes similar en el STR D13S317 para rastrear los alelos de la muestra 15. Sin embargo, en este caso no se puede determinar el origen de ninguno de los alelos (figura 22).

5 Ejemplo 5: Elaboración de perfiles de ADN para su uso en investigación, análisis forense o de paternidad

Este ejemplo se basa en el flujo de trabajo descrito en ForenSeq™ DNA Signature Prep Guide (Illumina, San Diego, CA).

Para este ejemplo se puede usar ADN purificado o lisado bruto. Para el ADN purificado, cada muestra de 1 ng se diluye hasta 0,2 ng/μl con agua sin nucleasas. Para el lisado bruto, cada muestra de 2 μl se diluye con 3 μl de agua sin nucleasas. Se configura una mezcla maestra para ocho o más reacciones. Para cada reacción, se añaden 5,4 μl

10 de ForenSeq PCR1 Reaction Mix, 0,4 μl de ForenSeq Enzyme Mix y 5,8 μl de mezcla de cebadores de ADN (A o B) en un tubo de microcentrífuga de 1,5 ml. Se transfieren 10 μl de Master Mix a cada pocillo de la placa de PCR y se agrega el ADN o el lisado. La mezcla de reacción múltiple se somete a PCR siguiendo el protocolo:

98 °C durante 3 min.

8 ciclos de:

15 96 °C durante 45 seg.

80 °C durante 30 seg.

54 °C durante 2 min., con modo de pendiente especificado

68 °C durante 2 min., con modo de pendiente especificado

10 ciclos de:

20 96 °C durante 30 seg.

68 °C durante 3 min., con modo de pendiente especificado

68 °C durante 10 min.

Mantenimiento a 10 °C.

25 Después de ciclar, las muestras se retiran del termociclador. La mezcla de reacción ForenSeq PCR2 se agrega a las muestras con un par de adaptadores (adaptadores i7 e i5) y se cicla durante 15 ciclos, siguiendo el protocolo:

98 °C durante 30 seg.

15 ciclos de:

98 °C durante 20 seg.

66 °C durante 30 seg.

30 68 °C durante 90 seg.

68 °C durante 10 min.

Mantenimiento a 10 °C.

35 Las muestras se purifican con esferas de purificación de muestras y los bancos se normalizan y se agrupan por volumen. Los bancos agrupados se diluyen en tampón de hibridación (HT1), se añaden con control de secuenciación humana (HSC) y se desnaturalizan por calor en preparación para la secuenciación.

Ejemplo 6: Genotipificación con ADN degradado

40 La figura 23 muestra los resultados de la genotipificación utilizando ADN cortado y/o tratado con ADNasa que representa ADN degradado. Como se muestra, más del 50 % de los loci de STR y SNP fueron llamados correctamente con el ADN cortado. Se logró una probabilidad de coincidencia aleatoria (RMP) de 10^{-19} con ADN de menos de 100 pb. También se predijo la ascendencia correcta utilizando ADN degradado.

Ejemplo 7: Sensibilidad de la genotipificación

Las figuras 24 y 25 muestran los resultados de la sensibilidad de la genotipificación a niveles de entrada de ADN inferiores a nanogramos de 7,82 pg a 1 ng. Como se muestra, el 100 % de los alelos fueron llamados con éxito a 125 ng de ADN de entrada tanto para STR como para SNP. Más del 50 % de los alelos fueron llamados con éxito a tan

solo 7,82 pg de ADN de entrada. El equilibrio intralocus fue superior al 70 % para la mayoría de los loci a 1 ng de ADN de entrada.

Cada parámetro numérico debe interpretarse a la luz del número de dígitos significativos y los métodos de redondeo habituales.

5 En la medida en que las publicaciones y patentes o solicitudes de patente a las que se hace referencia en este documento contradigan la descripción contenida en la memoria descriptiva, se pretende que la memoria descriptiva reemplace y/o tenga prioridad sobre cualquier material contradictorio de este tipo.

La citación de las publicaciones o documentos anteriores no pretende ser una admisión de que cualquiera de los anteriores sea técnica anterior pertinente, ni constituye una admisión en cuanto al contenido o la fecha de estas 10 publicaciones o documentos.

Aunque la presente invención se ha descrito completamente en conexión con realizaciones de la misma con referencia a los dibujos adjuntos, debe observarse que varios cambios y modificaciones dentro del alcance de las reivindicaciones resultarán evidentes para los expertos en la técnica. Debe entenderse que las diversas realizaciones de la invención 15 se han presentado a modo de ejemplo únicamente, y no a modo de limitación. Asimismo, los diversos diagramas pueden representar un ejemplo de configuración arquitectónica o de otro tipo para la invención, lo cual se ha realizado para ayudar a comprender las características y la funcionalidad que se pueden incluir en la invención. La invención no se limita a las arquitecturas o configuraciones de ejemplo ilustradas, sino que se puede ejecutar usando una variedad de arquitecturas y configuraciones alternativas. Además, aunque la invención se describe anteriormente en términos 20 de varios ejemplos de realizaciones e implementaciones, debe entenderse que las diversas características y funcionalidades descritas en una o más de las realizaciones individuales no están limitadas en su aplicabilidad a la realización particular con la que están describiendo. Por el contrario, pueden aplicarse, solas o en alguna combinación, a una o más de las otras realizaciones de la invención, se describan o no tales realizaciones, y se presenten o no tales características como parte de una realización descrita. Por tanto, la amplitud y el alcance de la invención no deberían estar limitados por ninguna de las realizaciones exemplares descritas anteriormente, sino por las reivindicaciones.

25 Los términos y frases utilizados en este documento, y sus realizaciones, a menos que se indique expresamente lo contrario, deben interpretarse como abiertos en lugar de limitativos. Como ejemplos de lo anterior: el término "que incluye" debe entenderse en el sentido de "que incluye, sin limitación" o similar; el término "ejemplo" se usa para proporcionar ejemplos de casos del objeto analizado, no como una lista exhaustiva o limitante del mismo; y los adjetivos como "convencional", "tradicional", "normal", "conocido" y términos de significado similar, no deben 30 interpretarse como una limitación del objeto descrito a un período de tiempo determinado, ni a un objeto disponible. a partir de un momento dado. Además, aunque los objetos, elementos o componentes de la invención pueden describirse o reivindicarse en singular, se contempla que el plural esté dentro del alcance de la misma, a menos que se indique explícitamente una limitación al singular. Por ejemplo, "al menos uno" puede referirse al singular o al plural 35 y no se limita a ninguno de los dos. La presencia de palabras y frases ampliadas como "uno o más", "al menos", "pero no se limita a", u otras frases similares en algunos casos, no se interpretará como que el caso más restringido está previsto o requerido en algunos casos en los que puedan faltar tales frases ampliadoras.

Lista de secuencias

<110> Illumina, Inc.

Stephens, Kathryn M.

40 Holt, Cydne

Davis, Carey

Jager, Anne

Walichiewicz, Paulina

Han, Yonmee

45 Silva, David

Shen, Min-Jui Richard

Amini, Sasan

Steemers, Frank J.

<120> MÉTODOS Y COMPOSICIONES PARA LA ELABORACIÓN DE PERFILES DE ADN

50 <130> ILLINC.276WO/IP-1192

<150> 62/103,524

<151> 14-01-2015

<150> 62/043,060

<151> 28-08-2014

55 <150> 61/940,942

<151> 18-02-2014

<160> 404

<170> FastSEQ para Windows Versión 4.0

5 <210> 1
 <211> 19
 <212> ADN
 <213> *Homo sapiens*

10 <400> 1
 ccctgggctc tgtaaagaa 19

10 <210> 2
 <211> 24
 <212> ADN
 <213> *Homo sapiens*

15 <400> 2
 atcagagctt aaactggaa gctg 24

15 <210> 3
 <211> 23
 <212> ADN
 <213> *Homo sapiens*

20 <400> 3
 marcador 23 acagtaactg cttcataga

20 <210> 4
 <211> 22
 <212> ADN
 <213> *Homo sapiens*

25 <400> 4
 gtgtcagacc ctgttctaag ta 22

25 <210> 5
 <211> 30
 <212> ADN

30 <213> *Homo sapiens*

30 <400> 5
 tgatttcctt ctttgtatc ttatgtaaat 30

35 <210> 6
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

35 <400> 6
 acaacatttg tttttttatc ttatgtaaat 28

40 <210> 7
 <211> 27
 <212> ADN
 <213> *Homo sapiens*

40 <400> 7
 ttgttatttc atgtgtatcat tcgtatc 27

45 <210> 8
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

45 <400> 8
 acctatccctg tagattttt tcactgtg 28

50 <210> 9

<211> 24
 <212> ADN
 <213> *Homo sapiens*
 5 <400> 9
 ctctgagtga caaattgaga cctt 24
 <210> 10
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 10 <400> 10
 ttaacttc tgggtgtgg agatg 25
 <210> 11
 <211> 20
 <212> ADN
 15 <213> *Homo sapiens*
 <400> 11
 ttgggtcac ccattacccg 20
 <210> 12
 <211> 20
 20 <212> ADN
 <213> *Homo sapiens*
 <400> 12
 aggaggttga ggctgcaaaa 20
 <210> 13
 25 <211> 35
 <212> ADN
 <213> *Homo sapiens*
 <400> 13
 caccaaatat tggtaattaa atgtttacta tagac 35
 30 <210> 14
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 14
 35 taaagggtat gatagaacac ttgtc 25
 <210> 15
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 40 <400> 15
 caaaggcaga tcccaagctc t 21
 <210> 16
 <211> 21
 <212> ADN
 45 <213> *Homo sapiens*
 <400> 16
 tgtgtgtca tctgttaagca t 21
 <210> 17
 <211> 20
 50 <212> ADN
 <213> *Homo sapiens*
 <400> 17

tggtgttat tccctgtgcc 20
 <210> 18
 <211> 20
 <212> ADN
 5 <213> *Homo sapiens*
 <400> 18
 gcagtccaat ctgggtgaca 20
 <210> 19
 <211> 28
 10 <212> ADN
 <213> *Homo sapiens*
 <400> 19
 ccaatctggt cacaaacata ttaatgaa 28
 <210> 20
 15 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 20
 ttccccitgt ctgttattaa aaggaac 27
 20 <210> 21
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 21
 25 ttcccatgg cctgttcctc 20
 <210> 22
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 30 <400> 22
 ctgtacacag ggctccgag 20
 <210> 23
 <211> 23
 <212> ADN
 35 <213> *Homo sapiens*
 <400> 23
 gctgagtat ttgtctgtaa ttg 23
 <210> 24
 <211> 37
 40 <212> ADN
 <213> *Homo sapiens*
 <400> 24
 gaactcacag attaaaactgt aacccaaata aaatttag 37
 <210> 25
 45 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 25
 caatagtgtg caaggatggg tg 22
 50 <210> 26
 <211> 20
 <212> ADN

<213> *Homo sapiens*
 <400> 26
 tctgtggttc tccagcttac 20
 5 <210> 27
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 27
 ctttaggaaac cctcaactgaa tg 22
 10 <210> 28
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 <400> 28
 15 gtcccttgica gcgtttattt gc 22
 <210> 29
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 20 <400> 29
 ttgggttgag ccataggcag 20
 <210> 30
 <211> 20
 <212> ADN
 25 <213> *Homo sapiens*
 <400> 30
 gcatccgtga ctctctggac 20
 <210> 31
 <211> 26
 30 <212> ADN
 <213> *Homo sapiens*
 <400> 31
 gttatgggac ttttctcagt ctccat 26
 <210> 32
 35 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 32
 gagactaata ggaggttagat agactgg 27
 40 <210> 33
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 33
 45 gagactgtat tagtaaggct tcct 24
 <210> 34
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 50 <400> 34
 cctggactga gccatgtcc 20

<210> 35
 <211> 23
 <212> ADN
 <213> *Homo sapiens*

5 <400> 35
 cagtcctgtg ttagtcagga ttc 23
 <210> 36
 <211> 22
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 36
 tcaagggtca actgtgtat gt 22
 <210> 37
 <211> 24
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 37
 ctctgaaag ctcttagttt acct 24
 <210> 38
 20 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 38
 ttgcttattt gtggggtat ttca 24

25 <210> 39
 <211> 29
 <212> ADN
 <213> *Homo sapiens*
 <400> 39
 30 aagaattctc ttatttgggt tattaattt 29
 <210> 40
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 40
 35 aaattgtgga caggtgcgg 20
 <210> 41
 <211> 34
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 41
 ccatgtaaaa atacatgcat gtgtttttt atac 34
 <210> 42
 <211> 35
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 42
 tgataaaaaa gaatgaaggt aaaaatgtgt ataac 35
 <210> 43
 <211> 34
 50 <212> ADN
 <213> *Homo sapiens*

<400> 43
 ccaaatgtt atgattaatc ttttaattt gagg 34

5 <210> 44
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

10 <400> 44
 gtaacaaggg ctacaggaat catgag 26

15 <210> 45
 <211> 27
 <212> ADN
 <213> *Homo sapiens*

20 <400> 45
 tcatccactg aaatgactga aaaatag 27

25 <210> 46
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

30 <400> 46
 aggtacataa cagtcaata gaaag 25

35 <210> 47
 <211> 29
 <212> ADN
 <213> *Homo sapiens*

40 <400> 47
 gagttattca gtaagttaaa ggattgcag 29

45 <210> 48
 <211> 24
 <212> ADN
 <213> *Homo sapiens*

50 <400> 48
 gggagccagt ggatttgaa acag 24

55 <210> 49
 <211> 20
 <212> ADN
 <213> *Homo sapiens*

60 <400> 49
 gcatggtag gctgaagtag 20

65 <210> 50
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

70 <400> 50
 ctaacctatg gtcataacga tttt 25

75 <210> 51
 <211> 30
 <212> ADN
 <213> *Homo sapiens*

80 <400> 51
 gatgataaga ataatcgta tgtgacttgg 30

85 <210> 52
 <211> 30

<212> ADN
<213> *Homo sapiens*

<400> 52
ataggttaga tagagatagg acagatgata 30

5 <210> 53
<211> 22
<212> ADN
<213> *Homo sapiens*

<400> 53
10 ccctaccgct atagtaactt gc 22

<210> 54
<211> 23
<212> ADN
<213> *Homo sapiens*

15 <400> 54
cacgtctgta attccagtc cta 23

<210> 55
<211> 25
<212> ADN
20 <213> *Homo sapiens*

<400> 55
ggaagcgtgt actagagttc ttca 25

<210> 56
<211> 22
25 <212> ADN
<213> *Homo sapiens*

<400> 56
ggacagcctc catatccaca tg 22

<210> 57
30 <211> 23
<212> ADN
<213> *Homo sapiens*

<400> 57
ttcctactgc cccacccta ttg 23

35 <210> 58
<211> 24
<212> ADN
<213> *Homo sapiens*

<400> 58
40 ttatggtct cagtccccct caga 24

<210> 59
<211> 23
<212> ADN
<213> *Homo sapiens*

45 <400> 59
tcataatcac atatcacatg agc 23

<210> 60
<211> 21
<212> ADN
50 <213> *Homo sapiens*

<400> 60
aaacaqaacc agggaaatga a 21

<210> 61
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

5 <400> 61
 tgaaaactaaa gtcaaatggg gctac **25**
 <210> 62
 <211> 24
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 62
 taaggggtga cacctctcg gata 24
 <210> 63
 <211> 25
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 63
 cccagcctac atctaccact tcatt 25
 <210> 64
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 64
 ctaatgttcg tatggacctt tggaaagc 28
 25 <210> 65
 <211> 24
 <212> ADN
 <213> *Homo sapiens*
 <400> 65
 30 gtcctccagta cccagctagc ttat 24
 <210> 66
 <211> 22
 <212> ADN
 <213> *Homo sapiens*
 35 <400> 66
 tctcccaacc tgcccttat ca 22
 <210> 67
 <211> 20
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 67
 ttggggctga cacagtggct 20
 <210> 68
 <211> 24
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 68
 ttatcaaca caggaggttt gacc 24
 <210> 69
 <211> 30
 50 <212> ADN
 <213> *Homo sapiens*

<400> 69
 tataccactt ttagtggac actagttac 30

5 <210> 70
 <211> 23
 <212> ADN
 <213> *Homo sapiens*

10 <400> 70
 cctgtctatg gtctcgattc aat 23

15 <210> 71
 <211> 21
 <212> ADN
 <213> *Homo sapiens*

20 <400> 71
 tgcatgacag agggagattc t 21

25 <210> 72
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

30 <400> 72
 agaggggaaa tagtagaaatg aggatg 26

35 <210> 73
 <211> 24
 <212> ADN
 <213> *Homo sapiens*

40 <400> 73
 gccaaactct attagtcaac gttc 24

45 <210> 74
 <211> 24
 <212> ADN
 <213> *Homo sapiens*

50 <400> 74
 ctggttctct agtcacata cagt 24

55 <210> 75
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

60 <400> 75
 ttacccccata acaagaaaaa aagaagaa 28

65 <210> 76
 <211> 24
 <212> ADN
 <213> *Homo sapiens*

70 <400> 76
 cagtgtgaga agtgtgagaa gtgc 24

75 <210> 77
 <211> 21
 <212> ADN
 <213> *Homo sapiens*

80 <400> 77
 gacaccatgc caaacaacaa c 21

85 <210> 78
 <211> 26

<212> ADN
 <213> *Homo sapiens*

<400> 78
 atctatctat tccaattaca tagtcc 26

5 <210> 79
 <211> 30
 <212> ADN
 <213> *Homo sapiens*

<400> 79
 10 tcattataacc tacttctgta tccaactctc 30

<210> 80
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

<400> 80
 15 ggaacacaat tatccctgag tagcag 26

<210> 81
 <211> 29
 <212> ADN
 20 <213> *Homo sapiens*

<400> 81
 ggttagcataa tagaaatttt atgagtggg 29

<210> 82
 <211> 30
 25 <212> ADN
 <213> *Homo sapiens*

<400> 82
 gaagacagac ttcaatatca cagaacatcg 30

<210> 83
 30 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 83
 gtgtatctat tcattcaatc atacaccc 28

35 <210> 84
 <211> 23
 <212> ADN
 <213> *Homo sapiens*

<400> 84
 40 ctccctgggtt gcaagcaatt gcc 23

<210> 85
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

<400> 85
 45 caaaaattag tgggaaatag ttgaac 26

<210> 86
 <211> 24
 <212> ADN
 50 <213> *Homo sapiens*

<400> 86
 gtcgagatca caccattgca ttcc 24

<210> 87
 <211> 24
 <212> ADN
 <213> *Homo sapiens*

5 <400> 87
 gcctggcttg gaattcttt accc 24
 <210> 88
 <211> 35
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 88
 ttaagtott taatctatct tgaattaata gattc 35
 <210> 89
 <211> 28
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 89
 cttaagagg agtctgctaa aaggaatg 28
 <210> 90
 20 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 90
 tcaccagaag gttgcaagac 20
 25 <210> 91
 <211> 20
 <212> ADN
 <213> *Homo sapiens*
 <400> 91
 30 tctggcgaag taacccaaac 20
 <210> 92
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 35 <400> 92
 tcgagtcaagt tcaccagaag g 21
 <210> 93
 <211> 17
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 93
 ggaaccagtg agagccg 17
 <210> 94
 <211> 19
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 94
 ctcagagtgc tgaacccag 19
 <210> 95
 50 <211> 32
 <212> ADN
 <213> *Homo sapiens*

<400> 95
 gtatttattc atgatcagg tctaactcaa cc 32
 5 <210> 96
 <211> 41
 <212> ADN
 <213> *Homo sapiens*
 <400> 96
 ctacctaata ttatctata tcattctaat tatgtcttt c 41
 10 <210> 97
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 97
 ctctaaaggt ttttttggt ggcataag 28
 15 <210> 98
 <211> 32
 <212> ADN
 <213> *Homo sapiens*
 <400> 98
 20 gattaataca acaaaaattt ggtaatctga aa 32
 <210> 99
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 99
 caacctaagc taaaaatgcag atattc 26
 <210> 100
 <211> 30
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 100
 gttatgaaac gtaaaatgaa tggacttag 30
 <210> 101
 <211> 26
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 101
 gcagtctcat ttccggaga tggagg 26
 40 <210> 102
 <211> 21
 <212> ADN
 <213> *Homo sapiens*
 <400> 102
 ctggggctga ggaggtaat c 21
 45 <210> 103
 <211> 29
 <212> ADN
 <213> *Homo sapiens*
 <400> 103
 50 gccagtaaga ataaaattac agcatgaag 29
 <210> 104
 <211> 25

<212> ADN
 <213> *Homo sapiens*

<400> 104
 gaataatcta ccagcaacaa tggct 25

5 <210> 105
 <211> 20
 <212> ADN
 <213> *Homo sapiens*

<400> 105
 10 tgcccaatgg aatgcgtct 20

<210> 106
 <211> 33
 <212> ADN
 <213> *Homo sapiens*

<400> 106
 15 gctccatctc aaacaacaaa aacacaaaaa atg 33

<210> 107
 <211> 23
 <212> ADN
 20 <213> *Homo sapiens*

<400> 107
 gggtcattga accatcatgct ctg 23

<210> 108
 <211> 25
 25 <212> ADN
 <213> *Homo sapiens*

<400> 108
 ccccccaaaa ttctactgaa gtaaa 25

<210> 109
 30 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 109
 taacaggata aatcacctat ctatgtat 28

35 <210> 110
 <211> 23
 <212> ADN
 <213> *Homo sapiens*

<400> 110
 40 gctgaggaga atttccaaat tta 23

<210> 111
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 111
 45 cccgcaaact aactaggata aatctcta 28

<210> 112
 <211> 28
 <212> ADN
 50 <213> *Homo sapiens*

<400> 112
 cgacatggga aatgtcagat cataagac 28

<210> 113
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

5 <400> 113
 ccagggagtg aaaaatcctt ttatcatc 28
 <210> 114
 <211> 25
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 114
 gaggatgaag gtttagagcca gacct 25
 <210> 115
 <211> 28
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 115
 tgtgaaataa actgaaggct aaagaaaa 28
 <210> 116
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 116
 caagccctat gccaggata taacaatg 28
 25 <210> 117
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 117
 30 gaggtttac tgtatttagga gttccac 28
 <210> 118
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 35 <400> 118
 cagatgtgag atgataattt cgttctcc 28
 <210> 119
 <211> 26
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 119
 ttgtctctt ccatcccatt tcaccc 26
 <210> 120
 <211> 28
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 120
 ctgtacatt cccttatctg ctatgtgg 28
 <210> 121
 <211> 28
 50 <212> ADN
 <213> *Homo sapiens*

<400> 121
 ctctttgg agtttatgt gttgctac 28
 <210> 122
 <211> 28
 5 <212> ADN
 <213> *Homo sapiens*
 <400> 122
 ctctgatgat gtgcaagaaa ggttagta 28
 <210> 123
 10 <211> 29
 <212> ADN
 <213> *Homo sapiens*
 <400> 123
 tcagactatg ttttaaggag actatgagg 29
 15 <210> 124
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 124
 20 ctaagtatct accaatgtgc tacgtacc 28
 <210> 125
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 125
 cacgtggatg atatggttc tcaagg 26
 <210> 126
 <211> 28
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 126
 agcacctata tattataacct gaaagcat 28
 <210> 127
 <211> 28
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 127
 cccatgattt tcttgtggtg agaatttc 28
 <210> 128
 40 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 128
 caccctctgt acttaattt gactcccc 28
 45 <210> 129
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 129
 50 gttttcttc attcccatgt tgtgtac 27
 <210> 130
 <211> 26

<212> ADN
 <213> *Homo sapiens*

<400> 130
 cactcttcg aatcctggc aacaac 26

5 <210> 131
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 131
 10 caagttatat catagagtct acqacc 28

<210> 132
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 132
 15 ctgcaactat cagtctctgc ccttattc 28

<210> 133
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 133
 20 gatgtgtcic aaactgtta ttgtgagg 28

<210> 134
 <211> 28
 25 <212> ADN
 <213> *Homo sapiens*

<400> 134
 gaactcattt atccagagac ctgtctc 28

<210> 135
 30 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 135
 cataatacaa cctgttttg gagttact 28

<210> 136
 35 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 136
 40 gtgaccagta gtttatgag caagtatg 28

<210> 137
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 137
 45 ccacattgtt tggtttttag gcaccatg 28

<210> 138
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 138
 50 ctaataagtggacagttaa gagaaggc 28

<210> 139
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 5 <400> 139
 caagacaaggattgaaaga agtggat 27
 <210> 140
 <211> 28
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 140
 ccttgtcaat ctttctacca gaggtaa 28
 <210> 141
 <211> 25
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 141
 gaatcatagc ttgtgttgtt caggg 25
 <210> 142
 20 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 142
 gaattacaag tatttgcattcc 27
 25 <210> 143
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 143
 30 gaccaacttg gccttaacag atgcaaat 28
 <210> 144
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 35 <400> 144
 tccttacctt taagactttt ccttattt 28
 <210> 145
 <211> 28
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 145
 cattatctcg tcatacttcc ctgtcttg 28
 <210> 146
 <211> 28
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 146
 gcatcaaatt caccagtcaa attattga 28
 <210> 147
 <211> 28
 50 <212> ADN
 <213> *Homo sapiens*

<400> 147
 atgagtagat tattcaactg ttttggag 28
 <210> 148
 <211> 28
 5 <212> ADN
 <213> *Homo sapiens*
 <400> 148
 cagccatgtt gtaaacattt ttacggtc 28
 <210> 149
 10 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 149
 gcacattcta agaactgggt attctatc 28
 15 <210> 150
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 150
 20 gctagaaaaaa gctgagatag ctgtgaag 28
 <210> 151
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 151
 ctttgaagct cattctttgt tgtccc 26
 <210> 152
 <211> 28
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 152
 ctggacaccca gaccaaaaac aaataacc 28
 <210> 153
 <211> 27
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 153
 gtaatttagag ggcaatgggg cttttaa 27
 <210> 154
 40 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 154
 ctccagaagc tactgggata ttaatttag 28
 45 <210> 155
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 155
 50 tgagccaaat cagcaatata ataggact 28
 <210> 156
 <211> 28

<212> ADN
 <213> *Homo sapiens*

<400> 156
 cctagaacca caattatctg tctttggc 28

5 <210> 157
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 157
 10 ccattgattc tctacagttc tgccaggta 28

<210> 158
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

15 <400> 158
 ctcccacactt tatacaggtg aaatctga 28

<210> 159
 <211> 28
 <212> ADN
 20 <213> *Homo sapiens*

<400> 159
 catttttc tccttgc tcaccctc 28

<210> 160
 <211> 28
 25 <212> ADN
 <213> *Homo sapiens*

<400> 160
 gcttctttt cccttatgtt tctctctc 28

<210> 161
 30 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 161
 ggctttgaa gaaaaacact aaccgttc 28

35 <210> 162
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

<400> 162
 40 cacctatggg ctcttcttat ttctcc 26

<210> 163
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

45 <400> 163
 catttgatag ccatgggt tgttcca 28

<210> 164
 <211> 28
 <212> ADN
 50 <213> *Homo sapiens*

<400> 164
 ccatcacact atcctgacat gaacaaat 28

<210> 165
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

5 <400> 165
 gaagattgc atccaggta aagcac 26
 <210> 166
 <211> 28
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 166
 gcacttcata aagaatcagt caggatgc 28
 <210> 167
 <211> 28
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 167
 ggagaatcag gaaatagtca ctccctac 28
 <210> 168
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 168
 cattgacct tctagccaaa tgaagtac 28
 25 <210> 169
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 169
 30 ggaatttctg agaataacat tgccctc 28
 <210> 170
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 35 <400> 170
 catatgtgg gggagctaaa cctaata 28
 <210> 171
 <211> 27
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 171
 cactgtgacc acagcatctt ttaactc 27
 <210> 172
 <211> 28
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 172
 gggtaaagaa atattcagca catccaaa 28
 <210> 173
 <211> 28
 50 <212> ADN
 <213> *Homo sapiens*

<400> 173
 gagtatccct tatctaaaat gctgggcc 28
 5 <210> 174
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 174
 cttttctct taccggaact tcaacgac 28
 10 <210> 175
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 175
 cctcattaat atgaccaagg ctccctcg 28
 15 <210> 176
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 176
 20 tgactctaatttggggatgtg gtaattag 28
 <210> 177
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 177
 gacctaacctt ggagaaaaacc ggaga 25
 <210> 178
 <211> 28
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 178
 gtttctcttc tctgaaccctt tgtctcag 28
 <210> 179
 <211> 28
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 179
 gcaaacaacac aaagataggt tcgagttt 28
 40 <210> 180
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 180
 catatcaagt gctttctgtt gacatttg 28
 45 <210> 181
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 181
 50 ctgaaaagt ctacgtaaga ggtcattg 28
 <210> 182
 <211> 25

<212> ADN
<213> *Homo sapiens*

<400> 182
catctgagtg tgagaagagc ctcaa 25

5 <210> 183
<211> 28
<212> ADN
<213> *Homo sapiens*

<400> 183
10 cccagcaaaa acttctttc tccagtaa 28

<210> 184
<211> 28
<212> ADN
<213> *Homo sapiens*

15 <400> 184
gcttagaaag tttctctct ggttcaca 28

<210> 185
<211> 27
<212> ADN
20 <213> *Homo sapiens*

<400> 185
gaataatctat gagcaggcag ttagcag 27

<210> 186
<211> 28
25 <212> ADN
<213> *Homo sapiens*

<400> 186
ctaactcgatg tcactatgtg tgagctat 28

<210> 187
30 <211> 28
<212> ADN
<213> *Homo sapiens*

<400> 187
catcatacag actcaaggag cttagctg 28

35 <210> 188
<211> 28
<212> ADN
<213> *Homo sapiens*

<400> 188
40 ctttccaagc ctggaaaaac acagaaaa 28

<210> 189
<211> 27
<212> ADN
<213> *Homo sapiens*

45 <400> 189
gtaccattata aatcacggag tgcagac 27

<210> 190
<211> 27
<212> ADN
50 <213> *Homo sapiens*

<400> 190
caagtggtaa gagatgactg aggtcaa 27

<210> 191
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

5 <400> 191
 ctcttcctct tagaaggaca ctggtcag 28
 <210> 192
 <211> 27
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 192
 gttatggagg attggtaaga accagag 27
 <210> 193
 <211> 28
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 193
 gagctgttta agggtaaagg ggttagtta 28
 <210> 194
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 194
 gcagacaaaa ccatgacaat gatcttag 28
 25 <210> 195
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 195
 30 cccatgatga aacagttgc actaaatg 28
 <210> 196
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 196
 35 ctcaatttc ttgtccctgc tttcatg 27
 <210> 197
 <211> 28
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 197
 ttagaaattc cagatagagc taaaactg 28
 <210> 198
 <211> 25
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 198
 gtttagaaaaa gaaccagggt gtttt 25
 <210> 199
 <211> 27
 50 <212> ADN
 <213> *Homo sapiens*

<400> 199
 gcaaaagtaa atacaaaggc atactt 27
 5 <210> 200
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 200
 caatgcaaaa gaaaggct tactcgac 28
 10 <210> 201
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 201
 catttcataa ctctaaaaca aacattt 28
 15 <210> 202
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 202
 20 ggtccttaac ctataaatt ttaatgag 28
 <210> 203
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 203
 25 gactttcaat ttatgtcagc atttaaaa 28
 <210> 204
 <211> 27
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 204
 cctctgggtt gcattggatt ctcatgg 27
 <210> 205
 <211> 27
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 205
 tctccatgaa acttgggtt attttgc 27
 <210> 206
 40 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 206
 tgtctggaaat ttcttcaaat tgcaat 25
 45 <210> 207
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 207
 50 gtagcataaaa acattccaaa aattcaat 28
 <210> 208
 <211> 30

<212> ADN
<213> *Homo sapiens*

<400> 208
tgcttaaag atacaggta tctgttac 30

5 <210> 209
<211> 26
<212> ADN
<213> *Homo sapiens*

<400> 209
10 ctctccgtta ctttcctcgcctt 26

<210> 210
<211> 26
<212> ADN
<213> *Homo sapiens*

<400> 210
15 gatcctgaga ttcaccccta gtccct 26

<210> 211
<211> 26
<212> ADN
20 <213> *Homo sapiens*

<400> 211
ccgtaccagg taccttagta tgtact 26

<210> 212
<211> 27
25 <212> ADN
<213> *Homo sapiens*

<400> 212
ctttctgttt tgccatctg aaattct 27

<210> 213
30 <211> 27
<212> ADN
<213> *Homo sapiens*

<400> 213
caaagttaag tatcaccatc cagctgg 27

35 <210> 214
<211> 28
<212> ADN
<213> *Homo sapiens*

<400> 214
40 atagggatag ctgataagaa acatgacc 28

<210> 215
<211> 25
<212> ADN
<213> *Homo sapiens*

<400> 215
45 cttataaga cgctgcatact gccca 25

<210> 216
<211> 28
<212> ADN
50 <213> *Homo sapiens*

<400> 216
tccaggagac atttgtcat ataagtga 28

<210> 217
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

5 <400> 217
 agacacttt cagtatccat ttagaaac 28
 <210> 218
 <211> 25
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 218
 gtttcacatg tgcatgcctt tgggt 25
 <210> 219
 <211> 26
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 219
 ccaaagctat tctctctttt ggggtc 26
 <210> 220
 20 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 220
 gaaaggcac ttcaagatgtt caaagcc 27
 25 <210> 221
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 221
 30 gggtttcagt ctgcaacaag atcttg 26
 <210> 222
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 35 <400> 222
 tggagatcaa tatttagcct taacatat 28
 <210> 223
 <211> 28
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 223
 gactgttct catcctgtta ttatttgt 28
 <210> 224
 <211> 25
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 224
 aacacacaga aacatcaago tgagc 25
 <210> 225
 <211> 27
 50 <212> ADN
 <213> *Homo sapiens*

<400> 225
 ttccgtacat ttccttctt ctatctg 27
 5 <210> 226
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 226
 tatgacgcct ggatttcac aacaac 26
 10 <210> 227
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 227
 cagagactat ggatggatt tagtcaa 28
 15 <210> 228
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 228
 20 actttgttg gctgagagag agaaa 25
 <210> 229
 <211> 30
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 229
 tgagtgtct ctgtatttc ttactctaag 30
 <210> 230
 <211> 27
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 230
 attttggtc attgtgaca ctccacc 27
 <210> 231
 <211> 25
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 231
 ggtgttaggg agacaggcat gaatg 25
 <210> 232
 40 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 232
 tgaactttt caactcct accqcc 26
 45 <210> 233
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 233
 50 gttaaaattt ccactaatta tgtgtttt 28
 <210> 234
 <211> 27

<212> ADN
 <213> *Homo sapiens*
 <400> 234
 aactgatcct atgcagcaag atcttg 27
 5 <210> 235
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 235
 10 gatgcttgca aacaaagact gaaaagg 27
 <210> 236
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 15 <400> 236
 gtctgtgt cctctgagat gatgaatg 28
 <210> 237
 <211> 26
 <212> ADN
 20 <213> *Homo sapiens*
 <400> 237
 gggaggaaga aaacagagag tcttga 26
 <210> 238
 <211> 28
 25 <212> ADN
 <213> *Homo sapiens*
 <400> 238
 agtttgtgg ctcttttga gaagtatc 28
 <210> 239
 30 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 239
 ggcagatgaa gtagtagata tctggctg 28
 35 <210> 240
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 240
 40 gttcagtgtc aattttgacc agatatt 27
 <210> 241
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 45 <400> 241
 agacatagga cacaccattt tattgtct 28
 <210> 242
 <211> 28
 <212> ADN
 50 <213> *Homo sapiens*
 <400> 242
 tcaaaatatt tggctaaact attgccgg 28

<210> 243
 <211> 31
 <212> ADN
 <213> *Homo sapiens*

5 <400> 243
 ctggagttat taataaattg gattatatag c 31
 <210> 244
 <211> 27
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 244
 ttacctgttt tccttttgtt attccac 27
 <210> 245
 <211> 26
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 245
 accaagtcaa gagctctgag agacat 26
 <210> 246
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 246
 acagtgaatg atattcagaa tatttgac 28
 25 <210> 247
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 247
 30 gaacaaggc aagatatcag ctttcacc 28
 <210> 248
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 248
 35 aggtcataca atgaatggtg tcatgt 26
 <210> 249
 <211> 27
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 249
 atccacccat gagaaatata tccacaa 27
 <210> 250
 <211> 31
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 250
 acaattcaaa ttaatgtaaa aactgcaagt g 31
 <210> 251
 <211> 25
 50 <212> ADN
 <213> *Homo sapiens*

<400> 251
 tagttctagt gtgggatctg actcc 25

5 <210> 252
 <211> 27
 <212> ADN
 <213> *Homo sapiens*

10 <400> 252
 gaacatctgt tcaggttct ctccatc 27
 <210> 253
 <211> 27
 <212> ADN
 <213> *Homo sapiens*

15 <400> 253
 gaaaggacta aatttgtgaa cactus 27
 <210> 254
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

20 <400> 254
 tgtgtgttt aaagccaggt ttgtt 25
 <210> 255
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

25 <400> 255
 gatggactgg aactgaggat ttca 25
 <210> 256
 <211> 29
 <212> ADN
 <213> *Homo sapiens*

30 <400> 256
 agcttttagaa aggcatatcg tattaactg 29
 <210> 257
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

35 <400> 257
 ttatagttag taaaggacag gcccc 25
 <210> 258
 <211> 29
 <212> ADN
 <213> *Homo sapiens*

40 <400> 258
 acacatctgt tgacagtaat gaaatatcc 29
 <210> 259
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

45 <400> 259
 gtttaaactt ggataccatc cccaagac 28
 <210> 260
 <211> 25

<212> ADN
 <213> *Homo sapiens*
 <400> 260
 atgagattgc tggagatgc agatg 25

5 <210> 261
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 261
 10 cacattccc tctgcgggtt acatac 26
 <210> 262
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

15 <400> 262
 gacaaggcctc gcttgatgtt tcattt 25
 <210> 263
 <211> 25
 <212> ADN
 20 <213> *Homo sapiens*
 <400> 263
 tgtgagatgt tcaccgaatt caacg 25
 <210> 264
 <211> 27
 25 <212> ADN
 <213> *Homo sapiens*
 <400> 264
 aaatagcaat ggctcgatca tggtag 27

30 <210> 265
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 265
 tgctaagtaa ggtgatgtt ataatca 27

35 <210> 266
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 266

40 ataaatatga tggctact ccctcat 27
 <210> 267
 <211> 29
 <212> ADN
 <213> *Homo sapiens*

45 <400> 267
 gctacaccctc catatgtata atgttttt 29
 <210> 268
 <211> 26
 <212> ADN
 50 <213> *Homo sapiens*
 <400> 268
 tgaaggatgtt agagaactct gtacgt 26

<210> 269
 <211> 31
 <212> ADN
 <213> *Homo sapiens*

5 <400> 269
 ttttgtctct gttatattag tcacctatct c 31
 <210> 270
 <211> 26
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 270
 atagccctgc attcaaatcc caagtg 26
 <210> 271
 <211> 26
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 271
 tccattttta taccactgca ctgaag 26
 <210> 272
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 272
 gcagtaaaac atttcatca aatttcca 28
 25 <210> 273
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 273
 30 tctgggtgca aactagctga atatcag 27
 <210> 274
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 274
 35 gaaaatctgg aggcaattca tgatgcc 27
 <210> 275
 <211> 26
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 275
 atacaatgat gtcacacagg gaccct 26
 <210> 276
 <211> 27
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 276
 ccatgaagat ggagtcaaca ttttaca 27
 <210> 277
 50 <211> 26
 <212> ADN
 <213> *Homo sapiens*

<400> 277
 tcctaaccgg tagtacgtta gatgtg 26
 <210> 278
 <211> 25
 5 <212> ADN
 <213> *Homo sapiens*
 <400> 278
 gaggtgattt ctgtgaggaa cgtcg 25
 <210> 279
 10 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 279
 gtacattcac ttaacaggct ctctttcc 28
 15 <210> 280
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 280
 20 aattcatgag ctgggttcca aggag 25
 <210> 281
 <211> 29
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 281
 ataacagtct ccagagtata ttagcttag 29
 <210> 282
 <211> 25
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 282
 gttcctctgg gatgcaacat gagag 25
 <210> 283
 <211> 28
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 283
 gaaaggatga agagggtgga tattggag 28
 <210> 284
 40 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 284
 ccaaacctca tcatcttta cctggatt 28
 45 <210> 285
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 285
 50 gtccaaagtc aagtgcagaat atagttgg 28
 <210> 286
 <211> 28

<212> ADN
 <213> *Homo sapiens*

<400> 286
 acaatcttt ctgaatctga acagctc 28

5 <210> 287
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 287
 10 caaaggaagg catttcctaa tgatctc 28

<210> 288
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

15 <400> 288
 ctttgcttig ctttcttct tcagggaa 28

<210> 289
 <211> 28
 <212> ADN
 20 <213> *Homo sapiens*

<400> 289
 ctgcitcaag tgtatataaa ctcacagt 28

<210> 290
 <211> 28
 25 <212> ADN
 <213> *Homo sapiens*

<400> 290
 cctaggaaag cagtaactaa ttcaggag 28

<210> 291
 <211> 28
 30 <212> ADN
 <213> *Homo sapiens*

<400> 291
 gcaatttgtt cacttttagt ttcgtagc 28

35 <210> 292
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 292
 40 ggcctaataat gcatgtgttc atgtctct 28

<210> 293
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

45 <400> 293
 cagagttct catctacqaa agaggagt 28

<210> 294
 <211> 26
 <212> ADN
 50 <213> *Homo sapiens*

<400> 294
 atcttagacc tccaggtgga atgatc 26

<210> 295
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

5 <400> 295
 ctggctgtc tcaatattt ggagtaag 28
 <210> 296
 <211> 28
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 296
 gagtaaatga gctgtggtt ctctcta 28
 <210> 297
 <211> 28
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 297
 ctttccatgt ggaccctta acattcag 28
 <210> 298
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 298
 gcatagttag ctgttgatag agctttg 28
 25 <210> 299
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 299
 30 ctagaacaaa atcattggct ctcctagt 28
 <210> 300
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 300
 35 gtctggtag tactggctga atgtaaa 27
 <210> 301
 <211> 28
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 301
 ccagaggatg ctgctaaaca ttctacaa 28
 <210> 302
 <211> 28
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 302
 gctcatgcct ggaattcacc ttatattt 28
 <210> 303
 50 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 303
 ctaactagac attgggcca ccttactt 28

5 <210> 304
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

10 <400> 304
 qtctatqqtq cctatagaat gtacaggt 28
 <210> 305
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

15 <400> 305
 ccctctcaag ttgtgagca aatatcac 28
 <210> 306
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

20 <400> 306
 ctctatctg ctgcaatgga ctttcc 26
 <210> 307
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

25 <400> 307
 cctagaaaca gatttgaag ggctcttg 28
 <210> 308
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

30 <400> 308
 aaatqaqqqq cataggata aggga 25
 <210> 309
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

35 <400> 309
 cctagaaatc tgatacgta tcctatga 28
 <210> 310
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

40 <400> 310
 aggagagata tattcaacat gaacccaa 28
 <210> 311
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

45 <400> 311
 gaacatctc gaccagaaat ttccagta 28
 <210> 312
 <211> 28

<212> ADN
 <213> *Homo sapiens*

<400> 312
 qtqtaqtqaa atccttagac ttagttaa 28

5 <210> 313
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 313
 aatacatgaa aaagtaatac atggggca 28

10 <210> 314
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

15 <400> 314
 attaaatgtt tacttctatc tacaagga 28

<210> 315
 <211> 28
 <212> ADN

20 <213> *Homo sapiens*

<400> 315
 cattttgtga aatgcaaagg gcaaatact 28

<210> 316
 <211> 28

25 <212> ADN
 <213> *Homo sapiens*

<400> 316
 gctgagaggc ttaattccat caagatga 28

<210> 317
 <211> 28

30 <212> ADN
 <213> *Homo sapiens*

<400> 317
 cccatcctaa acttagttt atggcag 28

35 <210> 318
 <211> 27
 <212> ADN
 <213> *Homo sapiens*

<400> 318
 gtaaacacatt ctctttggga agcttagc 27

<210> 319
 <211> 28
 <212> ADN

40 <213> *Homo sapiens*

<400> 319
 cttagcttca gtaaaaatgg ttcccttc 28

<210> 320
 <211> 28

45 <212> ADN
 <213> *Homo sapiens*

<400> 320
 ctttcttagc tcctctccat ttctcttc 28

<210> 321
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

5 <400> 321
 gtcatgcag tgcttcactg aggattat 28
 <210> 322
 <211> 28
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 322
 ctctatctgc tcagagcctg cttaaaag 28
 <210> 323
 <211> 28
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 323
 gggaaaggata cagtgttgag caagatag 28
 <210> 324
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 324
 gccaaccttga ttctcttca aatgcttg 28
 25 <210> 325
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 325
 30 agatggggtt taccatgttt cccag 25
 <210> 326
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 35 <400> 326
 gtacagtagt tagttccag actgtatga 28
 <210> 327
 <211> 28
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 327
 gtaaatatct aactgtgttt ccctcagt 28
 <210> 328
 <211> 28
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 328
 gaacccaaaag gaattaagag actagggg 28
 <210> 329
 <211> 27
 50 <212> ADN
 <213> *Homo sapiens*

<400> 329
 ctgctttac ggcttctcc ttcttc 27
 5 <210> 330
 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 330
 cccacatcct tcccatattt aggcaa 26
 10 <210> 331
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 331
 tacatgatcc taagggcagc aggaa 25
 15 <210> 332
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 332
 20 gtttggtgca tcctcttctt ctctc 25
 <210> 333
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 333
 gactgttagtc accccttctc caaca 25
 <210> 334
 <211> 31
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 334
 agagtgaaat acatagaaaa gaaacttaaa g 31
 <210> 335
 <211> 28
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 335
 atttgcgaga aacagataaa tattgaag 28
 40 <210> 336
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 336
 acaggaacaa agaatttgtt ctcatgg 28
 45 <210> 337
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 337
 50 ccttggattt tctcaggatg ttgca 25
 <210> 338
 <211> 25

<212> ADN
 <213> *Homo sapiens*

<400> 338
 ctggatgtt tgtttgct ttgtg 25

5 <210> 339
 <211> 30
 <212> ADN
 <213> *Homo sapiens*

<400> 339
 10 attggtagta cactaatgga tataatgtgag 30

<210> 340
 <211> 27
 <212> ADN
 <213> *Homo sapiens*

15 <400> 340
 gaataaaagtg aggaaaacac ggagttg 27

<210> 341
 <211> 25
 <212> ADN
 20 <213> *Homo sapiens*

<400> 341
 gagagaggca gaaaggaggg atgaa 25

<210> 342
 <211> 28
 25 <212> ADN
 <213> *Homo sapiens*

<400> 342
 tactctgtct tcagtagctg tttcttg 28

<210> 343
 <211> 28
 30 <212> ADN
 <213> *Homo sapiens*

<400> 343
 ttagactcac caagatcaag atgaatgc 28

35 <210> 344
 <211> 29
 <212> ADN
 <213> *Homo sapiens*

<400> 344
 40 atctcaataa agctgtcaa aacagaaag 29

<210> 345
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

45 <400> 345
 taaagaaaaat gccatgggct gtaccc 26

<210> 346
 <211> 26
 <212> ADN
 50 <213> *Homo sapiens*

<400> 346
 attgtgcagc agaacagagt gtatgtg 26

<210> 347
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

5 <400> 347
 attcttgca tagtcacga aatttccc 28
 <210> 348
 <211> 26
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 348
 aaaatgagac ctgcgtatctt tgccgc 26
 <210> 349
 <211> 26
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 349
 aaatgcagaa ctgccaaaag aaaccc 26
 <210> 350
 20 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 350
 gagaatctgt gaatgccagg gtctg 25
 25 <210> 351
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 351
 30 atggattcat gttcagaca tctaatt 27
 <210> 352
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 35 <400> 352
 atcactagaa agaaaagagt tcctattc 28
 <210> 353
 <211> 27
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 353
 gaagttaaa agagtggaa catgggg 27
 <210> 354
 <211> 26
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 354
 ctacgttaaggc aaaaatgtac acgcac 26
 <210> 355
 <211> 25
 50 <212> ADN
 <213> *Homo sapiens*

<400> 355
 aacctgatgg ccctcattag tcctt 25
 <210> 356
 <211> 28
 5 <212> ADN
 <213> *Homo sapiens*
 <400> 356
 caccagattt cttagaatag catgtgag 28
 <210> 357
 10 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 <400> 357
 aagagcatag tgaggggta gacct 25
 15 <210> 358
 <211> 29
 <212> ADN
 <213> *Homo sapiens*
 <400> 358
 20 ctttatattt agtgtagaga tcagtctcc 29
 <210> 359
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 359
 tgagtccctt acctaattt ggttgtc 27
 <210> 360
 <211> 27
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 360
 aatccaaagc aactcttt tcaccac 27
 <210> 361
 <211> 26
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 361
 ttactggaa ccctgattt gtggaa 26
 <210> 362
 40 <211> 26
 <212> ADN
 <213> *Homo sapiens*
 <400> 362
 tgccactgat atatcagtac ctgagt 26
 45 <210> 363
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 363
 50 acaatctcaa tccccctaa tgtttc 27
 <210> 364
 <211> 25

<212> ADN
 <213> *Homo sapiens*

<400> 364
 gtggcagag agagtaagag aacct 25

5 <210> 365
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 365
 10 caaaccagat tctggcagaa tagttgc 28

<210> 366
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

15 <400> 366
 cttcctcccc atccccc tccac 25

<210> 367
 <211> 27
 <212> ADN

20 <213> *Homo sapiens*

<400> 367
 agatcaaggg atctgtggga caataac 27

<210> 368
 <211> 26

25 <212> ADN
 <213> *Homo sapiens*

<400> 368
 ggggagtgtat ttcaaggcatc ctgatt 26

<210> 369
 <211> 27

30 <212> ADN
 <213> *Homo sapiens*

<400> 369
 catgagtttg aggtaagatg aaggaga 27

35 <210> 370
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

<400> 370
 40 tctctctcat cctagtgaat gccatc 26

<210> 371
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

<400> 371
 45 gggtgatgtat ctacccatc ggtata 26

<210> 372
 <211> 25

50 <212> ADN
 <213> *Homo sapiens*

<400> 372
 ctcaaggccc tgggtctgaa attac 25

<210> 373
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

5 <400> 373
 acatctccag ttaataattt ccactaac 28
 <210> 374
 <211> 27
 <212> ADN
 10 <213> *Homo sapiens*
 <400> 374
 tggattgctc aacaaatagt gctaaaa 27
 <210> 375
 <211> 27
 15 <212> ADN
 <213> *Homo sapiens*
 <400> 375
 catgcgacat ccaggttagct aaaatac 27
 <210> 376
 20 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 376
 atggataaaa atggaacttt caagagaa 28
 25 <210> 377
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 377
 30 gttttatgt aagcttcgtc atatggct 28
 <210> 378
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 378
 35 gttccaacctt agtcataaag ttccctgg 28
 <210> 379
 <211> 28
 <212> ADN
 40 <213> *Homo sapiens*
 <400> 379
 gggcttgat gtttattgt tgaggaag 28
 <210> 380
 <211> 28
 45 <212> ADN
 <213> *Homo sapiens*
 <400> 380
 cctaacagaa agtcaactgtt tgtatctg 28
 <210> 381
 <211> 26
 50 <212> ADN
 <213> *Homo sapiens*

<400> 381
 tcactctggc ttgtactctc tctgtg 26
 5 <210> 382
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 382
 ggctggttc agtctggaga ctttattt 28
 10 <210> 383
 <211> 27
 <212> ADN
 <213> *Homo sapiens*
 <400> 383
 cttcacctcg atgacgatga tgatgat 27
 15 <210> 384
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 384
 20 gacaataaca gcacaaagga tggaaaag 28
 <210> 385
 <211> 25
 <212> ADN
 <213> *Homo sapiens*
 25 <400> 385
 gaaccagacc acacaatatc accac 25
 <210> 386
 <211> 28
 <212> ADN
 30 <213> *Homo sapiens*
 <400> 386
 tctacctt tgatgtcccc ttcgatag 28
 <210> 387
 <211> 28
 35 <212> ADN
 <213> *Homo sapiens*
 <400> 387
 cagagttct catctacgaa agaggagt 28
 40 <210> 388
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 388
 cccagcttg aaaagtatgc ctagaact 28
 45 <210> 389
 <211> 28
 <212> ADN
 <213> *Homo sapiens*
 <400> 389
 50 ctgcttcaag tgtatataaa ctcacagt 28
 <210> 390
 <211> 28

<212> ADN
 <213> *Homo sapiens*

<400> 390
 ttgttcatc cactttggtg ggtaaaag 28

5 <210> 391
 <211> 27
 <212> ADN
 <213> *Homo sapiens*

<400> 391
 10 taattaagct ctgtgttagt ggaaaa 27

<210> 392
 <211> 28
 <212> ADN
 <213> *Homo sapiens*

<400> 392
 15 caattcttg ttcttttagt cagtata 28

<210> 393
 <211> 25
 <212> ADN
 20 <213> *Homo sapiens*

<400> 393
 tactcttcct cagccccctc tctgc 25

<210> 394
 <211> 25
 25 <212> ADN
 <213> *Homo sapiens*

<400> 394
 tgagacagag catgatgatc atggc 25

<210> 395
 30 <211> 26
 <212> ADN
 <213> *Homo sapiens*

<400> 395
 gcacaagtct aggaactact ttgcac 26

35 <210> 396
 <211> 25
 <212> ADN
 <213> *Homo sapiens*

<400> 396
 40 gaagtattt aaccatacgg agccc 25

<210> 397
 <211> 26
 <212> ADN
 <213> *Homo sapiens*

<400> 397
 45 tgaggaacac atccaaacta tgacac 26

<210> 398
 <211> 25
 <212> ADN
 50 <213> *Homo sapiens*

<400> 398
 ttctcgccc teatcatctg caatg 25

<210> 399
<211> 27
<212> ADN
<213> *Homo sapiens*

5 <400> 399
tcagttgatt tcatgtgatc ctcacag 27
<210> 400
<211> 27
<212> ADN
10 <213> *Homo sapiens*
<400> 400
gaataaaagtg aggaaaacac ggagttg 27
<210> 401
<211> 28
15 <212> ADN
<213> *Homo sapiens*
<400> 401
attacc ttcttataca agcatatg 28
<210> 402
20 <211> 28
<212> ADN
<213> *Homo sapiens*
<400> 402
acaggaacaa agaatttgtt ctccatgg 28
25 <210> 403
<211> 21
<212> ADN
<213> Secuencia artificial
<220>
30 <223> Secuencia de marcador del adaptador i5
<400> 403
tacacgacgc tcttccgatc t 21
<210> 404
<211> 20
35 <212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia de marcador del adaptador i7
<400> 404
40 ctggcaccc gagaattcca 20

REIVINDICACIONES

1. Un método para construir un perfil de ADN que comprende: proporcionar una muestra de ácido nucleico, amplificar la muestra de ácido nucleico con una mezcla de cebadores que comprende una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo de un solo nucleótido (SNP) y al menos una secuencia diana que comprende una repetición en tandem en una reacción múltiplex para generar productos de amplificación, en el que cada uno de la pluralidad de cebadores tiene una temperatura de fusión que es menor que 60 °C, tiene una longitud de al menos 24 nucleótidos que se hibridan con la secuencia diana, y es rico en AT con un contenido en AT mayor del 60 %, opcionalmente en el que cada uno de la pluralidad de cebadores comprende una secuencia de nucleótidos de homopolímero; y
- 5 10 determinar los genotipos de dicho al menos un SNP y dicha al menos una repetición en tandem en los productos de amplificación, construyendo así el perfil de ADN de la muestra de ácido nucleico.
2. El método según la reivindicación 1, en el que el método comprende amplificar los productos de amplificación con una segunda pluralidad de cebadores.
- 15 3.- El método de la reivindicación 1 o la reivindicación 2, que comprende además generar un banco de ácidos nucleicos a partir de los productos de amplificación, en el que el método comprende opcionalmente determinar las secuencias del banco de ácidos nucleicos.
4. El método de una cualquiera de las reivindicaciones 1-3, en el que la muestra de ácido nucleico es de un ser humano, una muestra ambiental, una planta, un animal no humano, una bacteria, una arquea, un hongo o un virus.
- 20 5. El método de una cualquiera de las reivindicaciones 1-4, en el que dicho al menos un SNP indica la ascendencia o una característica fenotípica de la fuente de la muestra de ácido nucleico.
6. El método de una cualquiera de las reivindicaciones 1-5, en el que el perfil de ADN es adecuado para su uso en uno o más diagnósticos o pronósticos de enfermedades, identificación de biomarcadores del cáncer, identificación de anomalías genéticas o análisis de diversidad genética, construcción de banco de datos, análisis forense, investigación de crímenes, paternidad, o identificación personal.
- 25 7. El método de una cualquiera de las reivindicaciones 1-6, en el que la muestra de ácido nucleico se amplifica mediante la reacción en cadena de la polimerasa (PCR).
- 8.- El método de la reivindicación 7, en el que la muestra de ácido nucleico se amplifica en un tampón de amplificación, en el que el tampón de amplificación comprende KCl en una concentración en el intervalo de 60 mM a 500 mM, LiCl en una concentración en el intervalo de 60 mM a 500 mM, NaCl en una concentración en el intervalo de 60 mM a 500 mM, o una combinación de los mismos.
- 30 35 9. El método de una cualquiera de las reivindicaciones 1-8, en el que el SNP es un SNP de ascendencia, un SNP fenotípico, un SNP de identidad o una combinación de los mismos; y/o en el que la pluralidad de cebadores se hibrida específicamente con al menos 30 SNP o con al menos 50 SNP; y/o en el que la repetición en tandem es una repetición en tandem corta (STR), una repetición en tandem intermedia (ITR) o una variante de las mismas; y/o en el que la pluralidad de cebadores se hibrida específicamente con al menos 24 secuencias repetidas en tandem o con al menos 60 secuencias repetidas en tandem; y/o en el que la muestra de ácido nucleico comprende de 100 pg a 100 ng de ADN, o de 10 pg a 100 pg de ADN, o de 5 pg a 10 pg de ADN; y/o en el que la muestra de ácido nucleico comprende ADN genómico; y/o en el que se determinan al menos el 50 %, 80 %, 90 % o 95 % de los genotipos de dicho al menos un SNP y al menos una repetición en tandem.
- 40 45 10. El método de una cualquiera de las reivindicaciones 1-9, en el que cada uno de la pluralidad de cebadores comprende una o más secuencias de marcador que no son homólogas con la secuencia diana, en el que opcionalmente dichas una o más secuencias de marcador comprenden un marcador de cebador, un marcador de captura, un marcador de secuenciación, un marcador de identificación molecular exclusivo o una combinación de los mismos.
11. Un método para construir un banco de ácidos nucleicos que comprende:
- proporcionar una muestra de ácido nucleico, y
- amplificar la muestra de ácido nucleico con una mezcla de cebadores que comprende una pluralidad de cebadores que se hibridan específicamente con al menos una secuencia diana que comprende un polimorfismo de un solo nucleótido (SNP) y al menos una secuencia diana que comprende una secuencia repetida en tandem en una reacción múltiplex para generar productos de amplificación , en el que cada uno de la pluralidad de cebadores tiene una temperatura de fusión inferior a 60 °C, tiene una longitud de al menos 24 nucleótidos que se hibridan con la secuencia diana y es rico en AT con un contenido en AT superior al 60 %, opcionalmente en el que cada uno de la pluralidad de cebadores comprende una secuencia de nucleótidos de homopolímero.

- 5 12. El método de la reivindicación 11, en el que la muestra de ácido nucleico no se fragmenta antes de la amplificación; y/o en el que las secuencias diana no se enriquecen antes de la amplificación; y/o en el que dicho al menos un SNP indica la ascendencia o una característica fenotípica de la fuente de la muestra de ácido nucleico; y/o en el que cada uno de la pluralidad de cebadores comprende una o más secuencias de marcador que no son homólogas con la secuencia diana; y/o en el que el método comprende además amplificar los productos de amplificación con una segunda pluralidad de cebadores.
- 10 13. El método de la reivindicación 11 o la reivindicación 12, en el que la muestra de ácido nucleico y/o los productos de amplificación se amplifican mediante la reacción en cadena de la polimerasa (PCR).
- 10 14. El método de la reivindicación 13, en el que la muestra de ácido nucleico y/o los productos de amplificación se amplifican en un tampón de amplificación, en el que el tampón de amplificación comprende KCl en una concentración en el intervalo de 60 mM a 500 mM, LiCl en una concentración en el intervalo de 60 mM a 500 mM, NaCl a una concentración en el intervalo de 60 mM a 500 mM, o una combinación de los mismos.

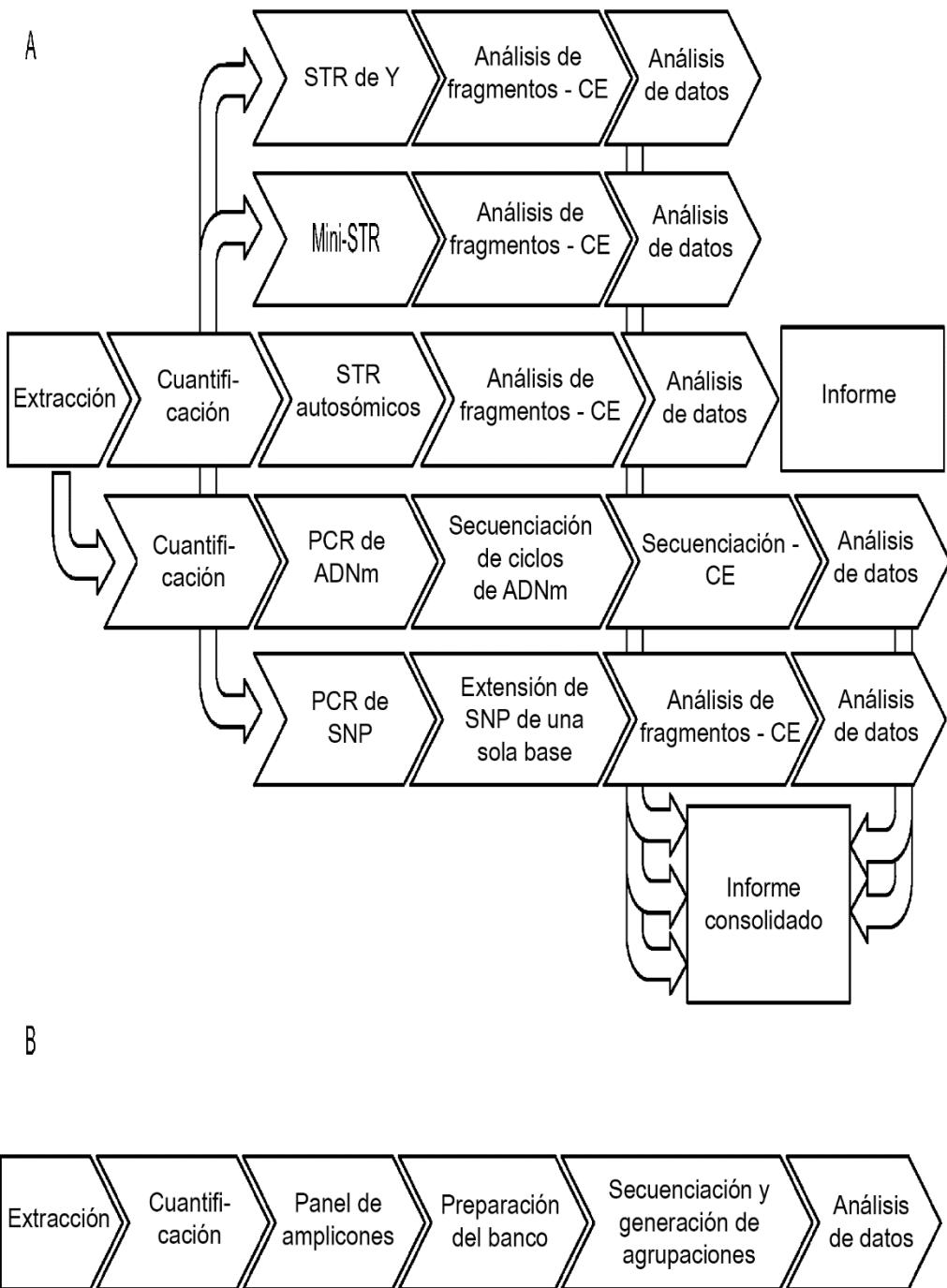


Fig. 1

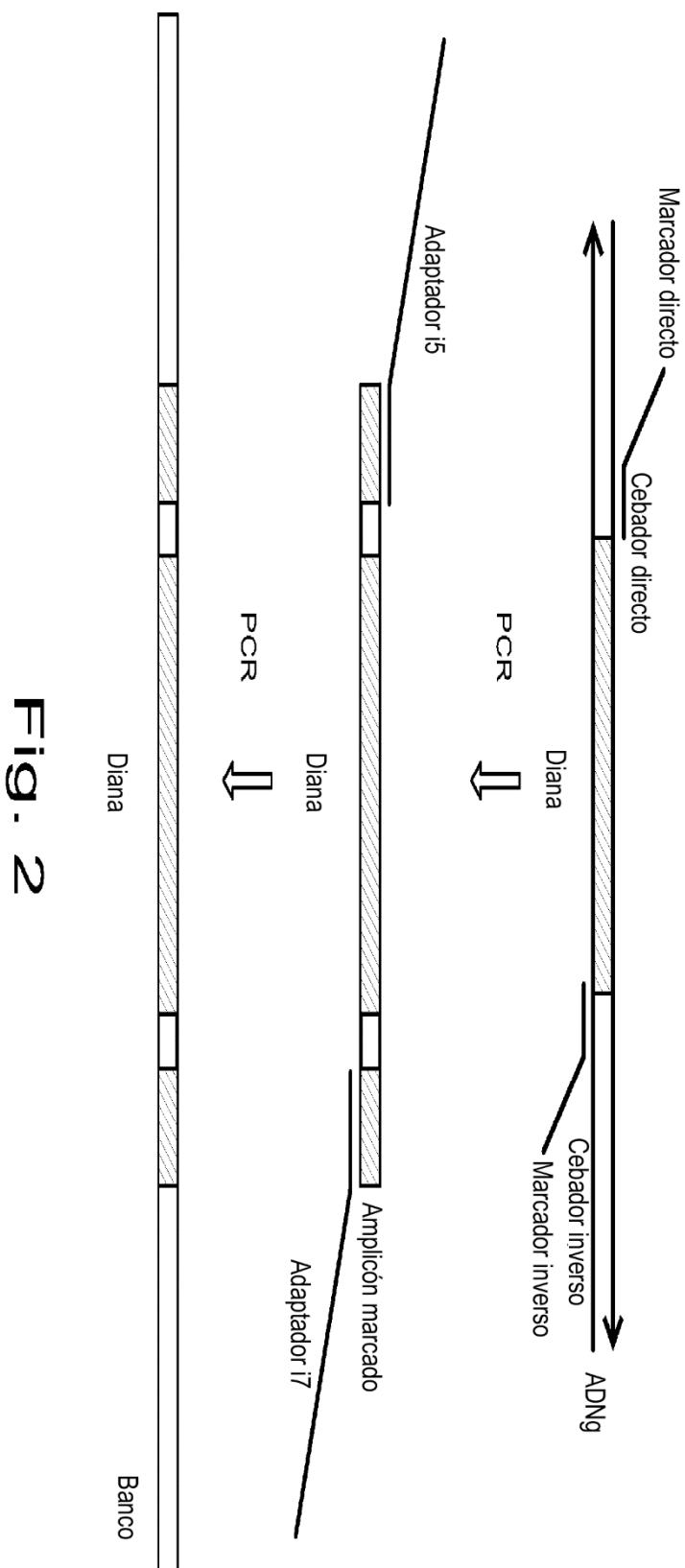


Fig. 2

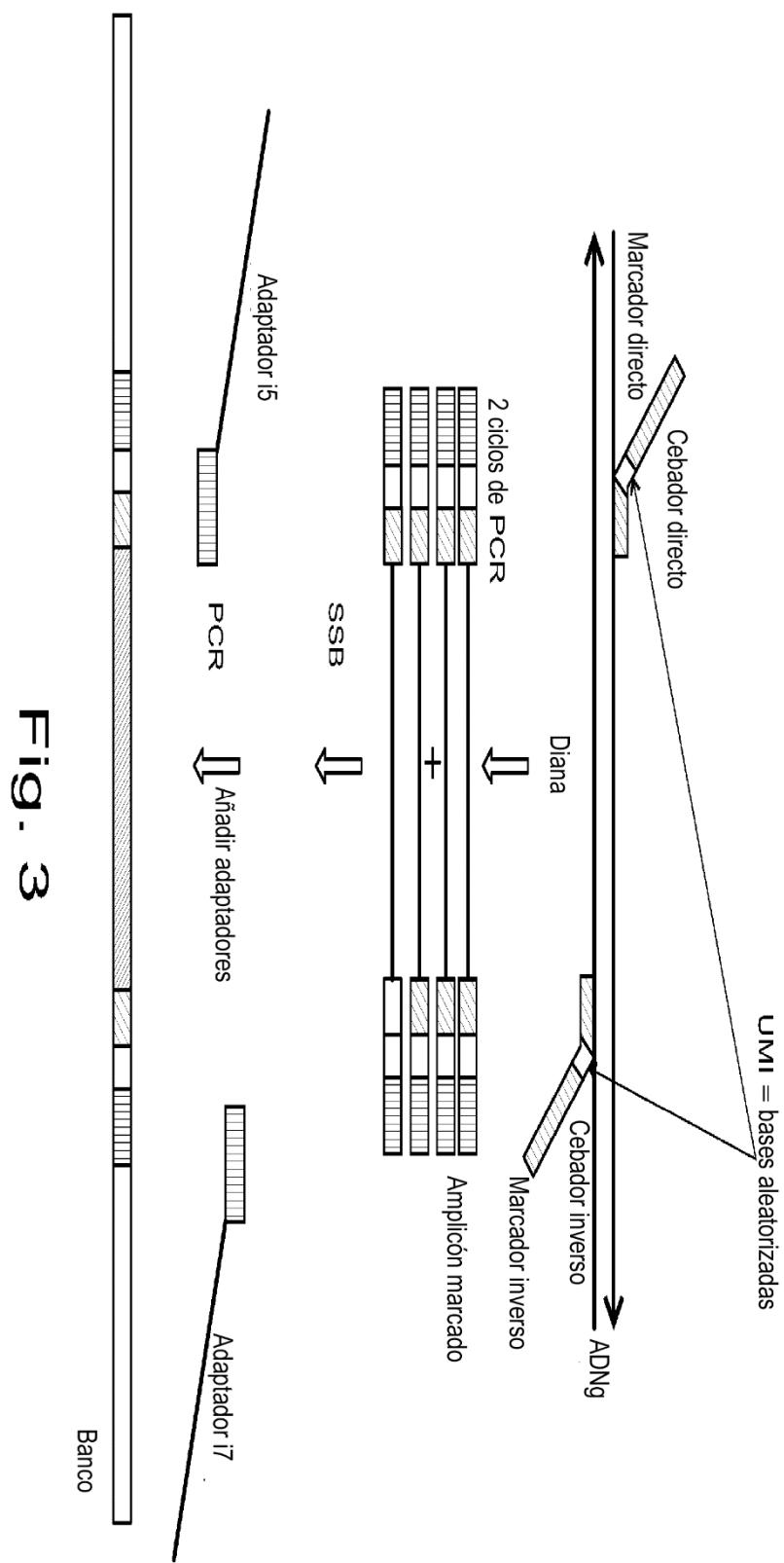
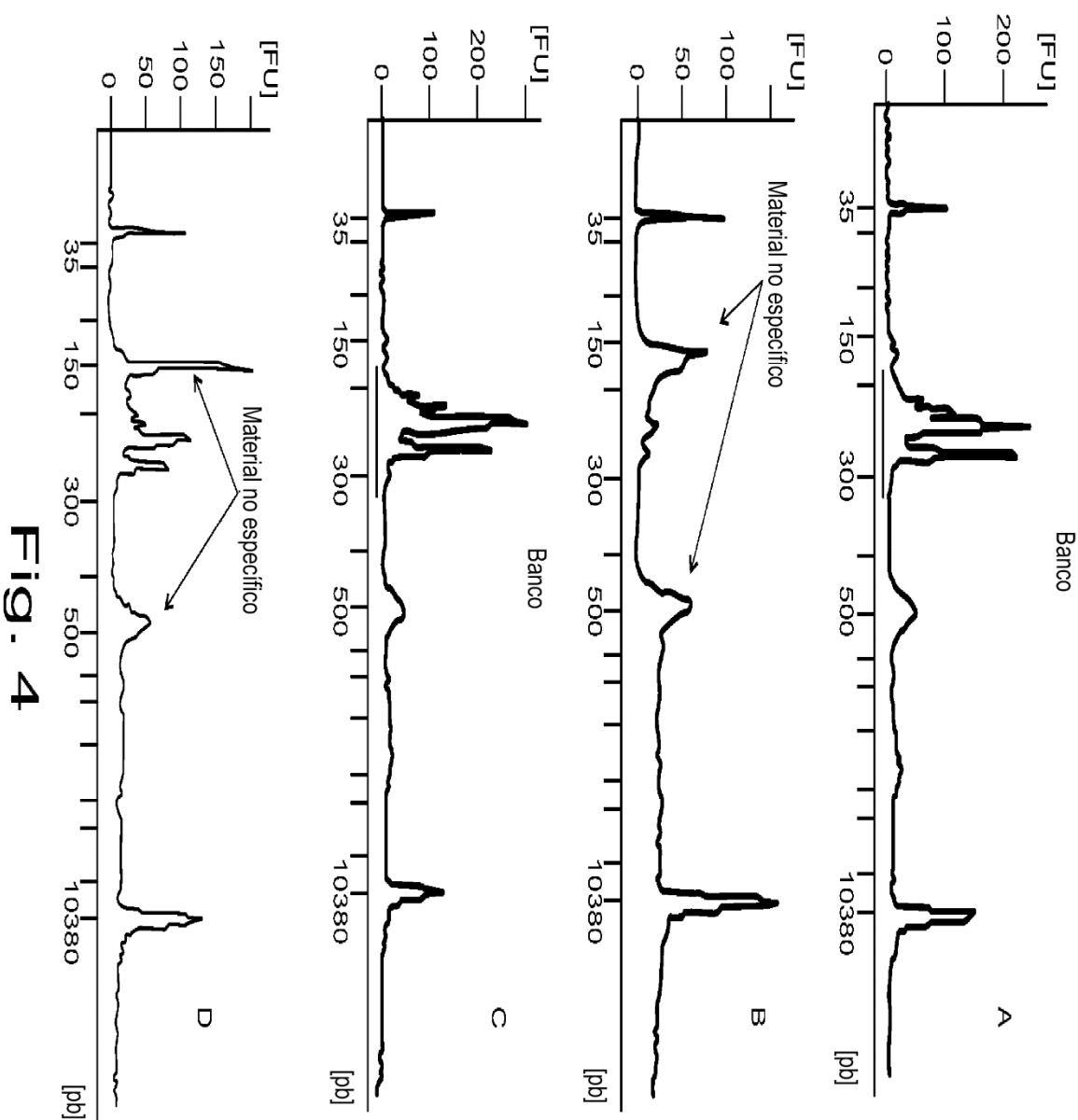


Fig. 3

**Fig. 4**

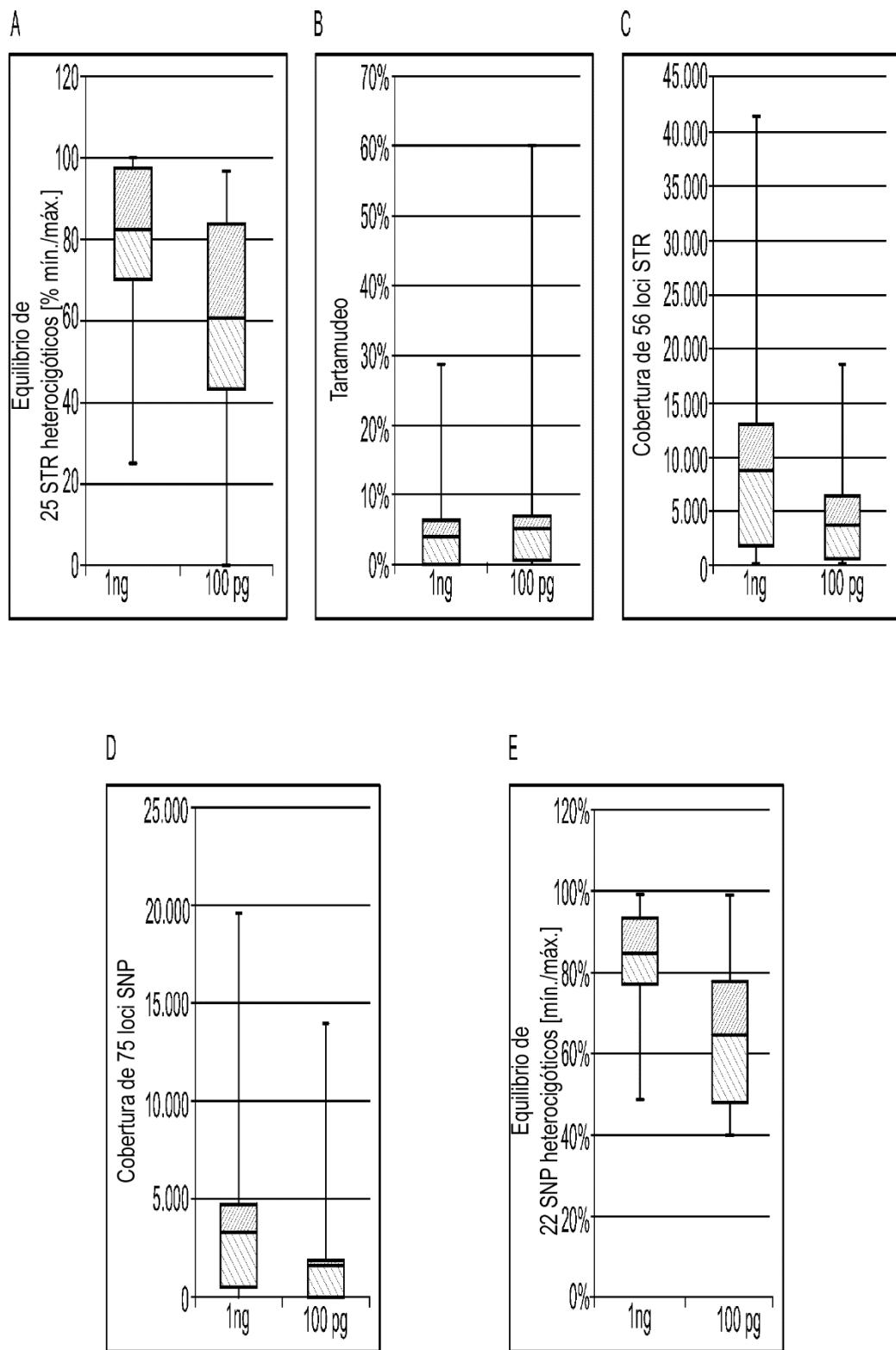


Fig. 5

ES 2 866 044 T3

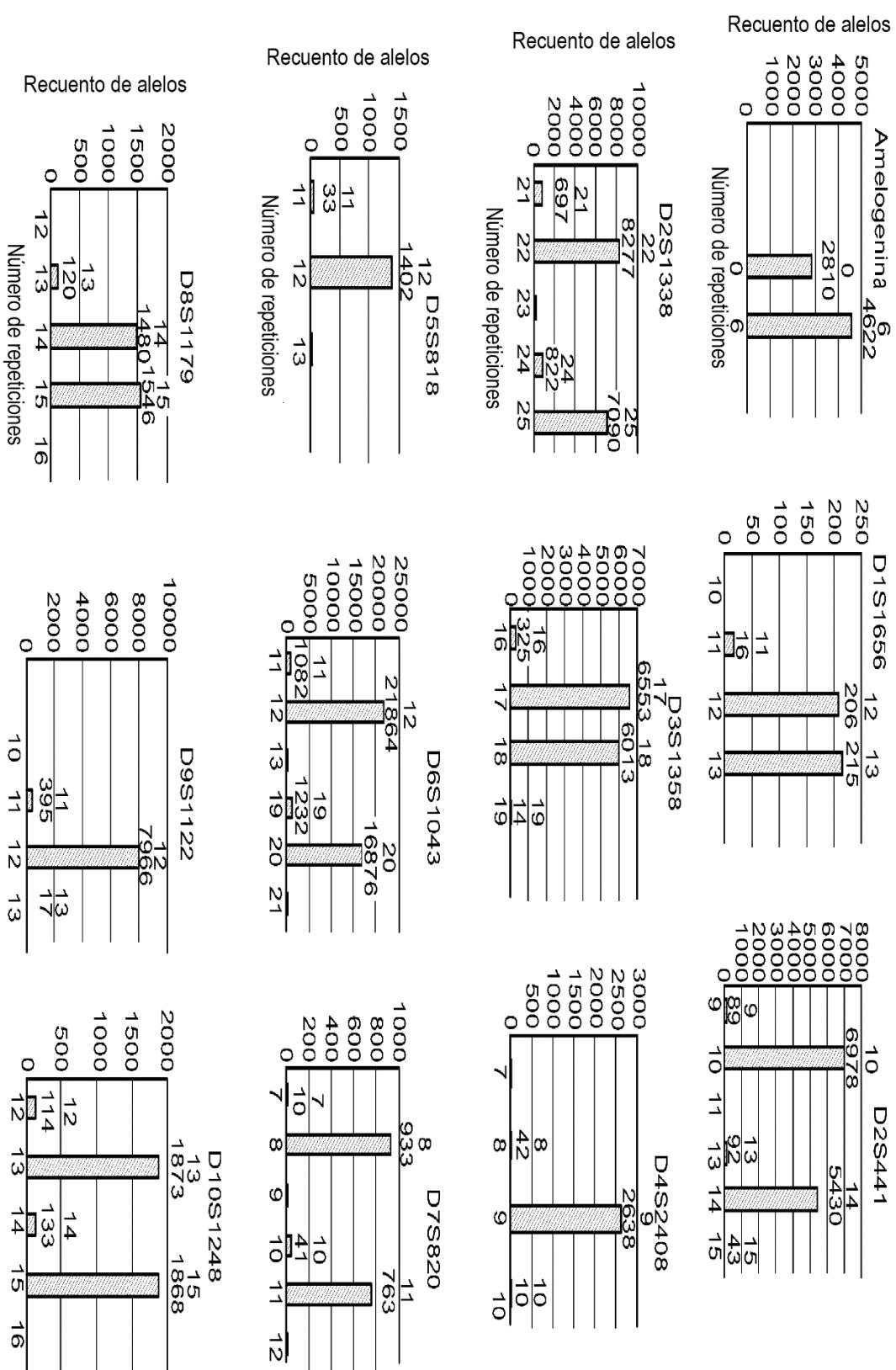
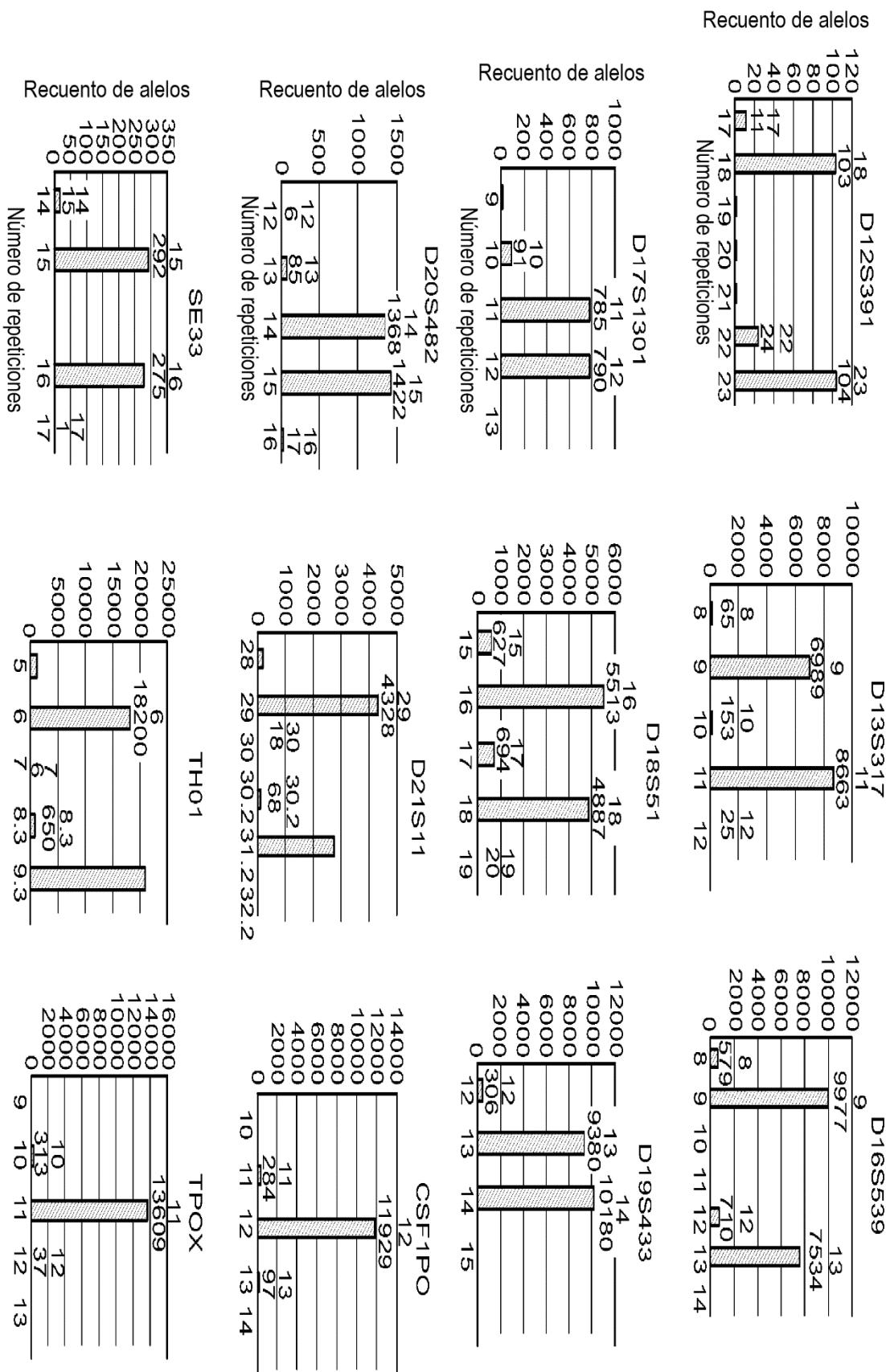
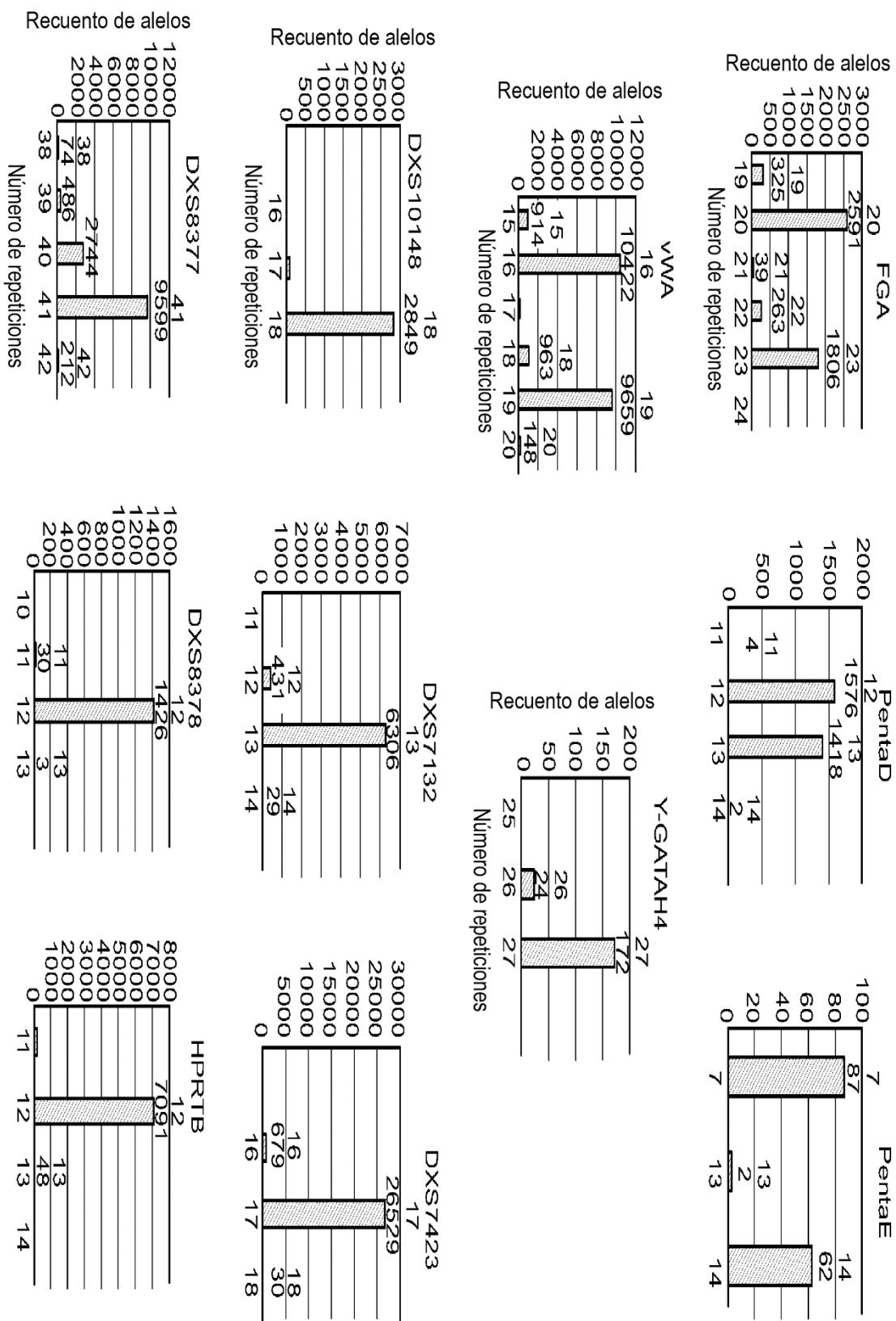
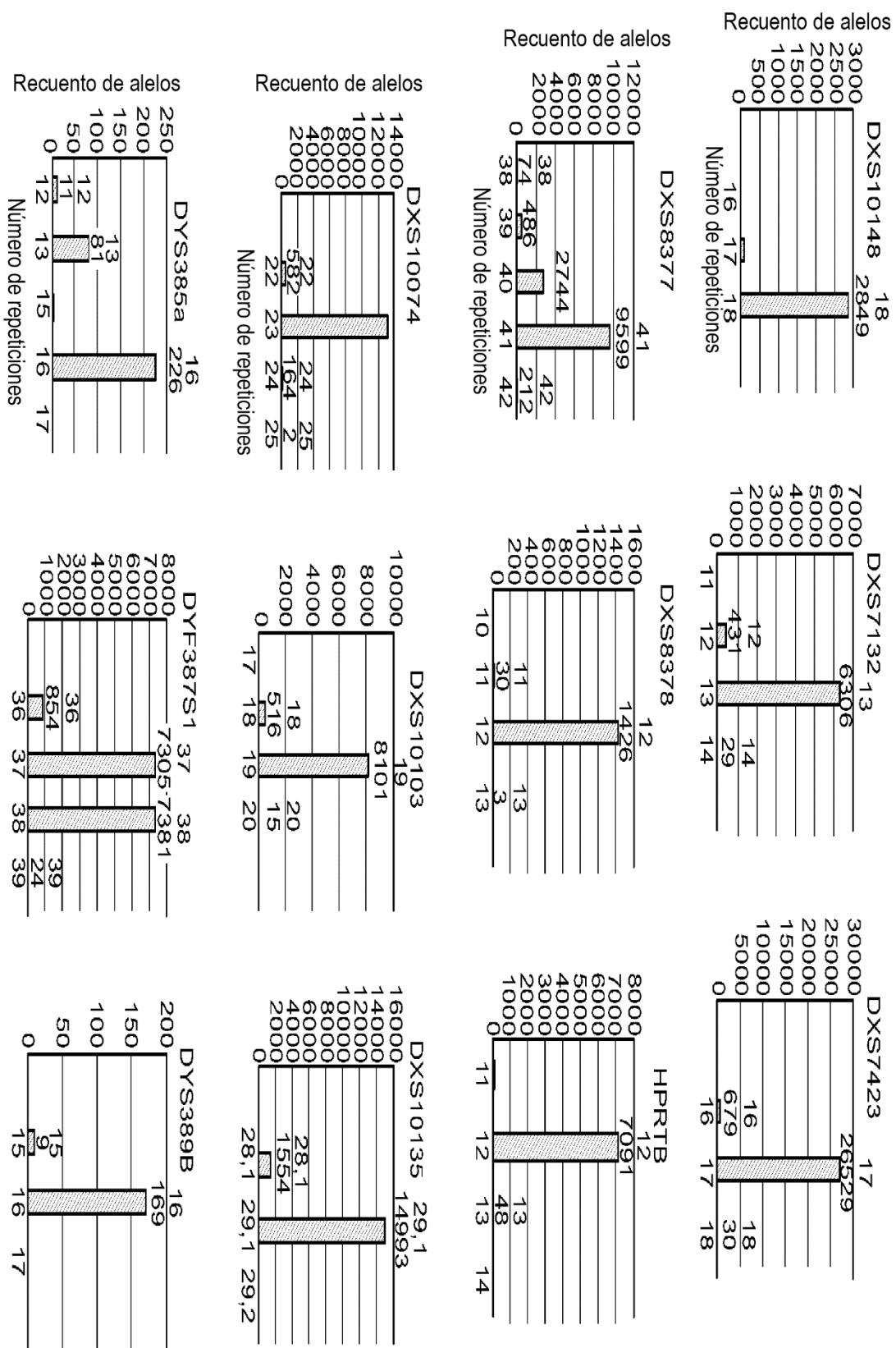


Fig. 6

**Fig. 6 (cont.)**

**Fig. 6 (cont.)**

**Fig. 6 (cont.)**

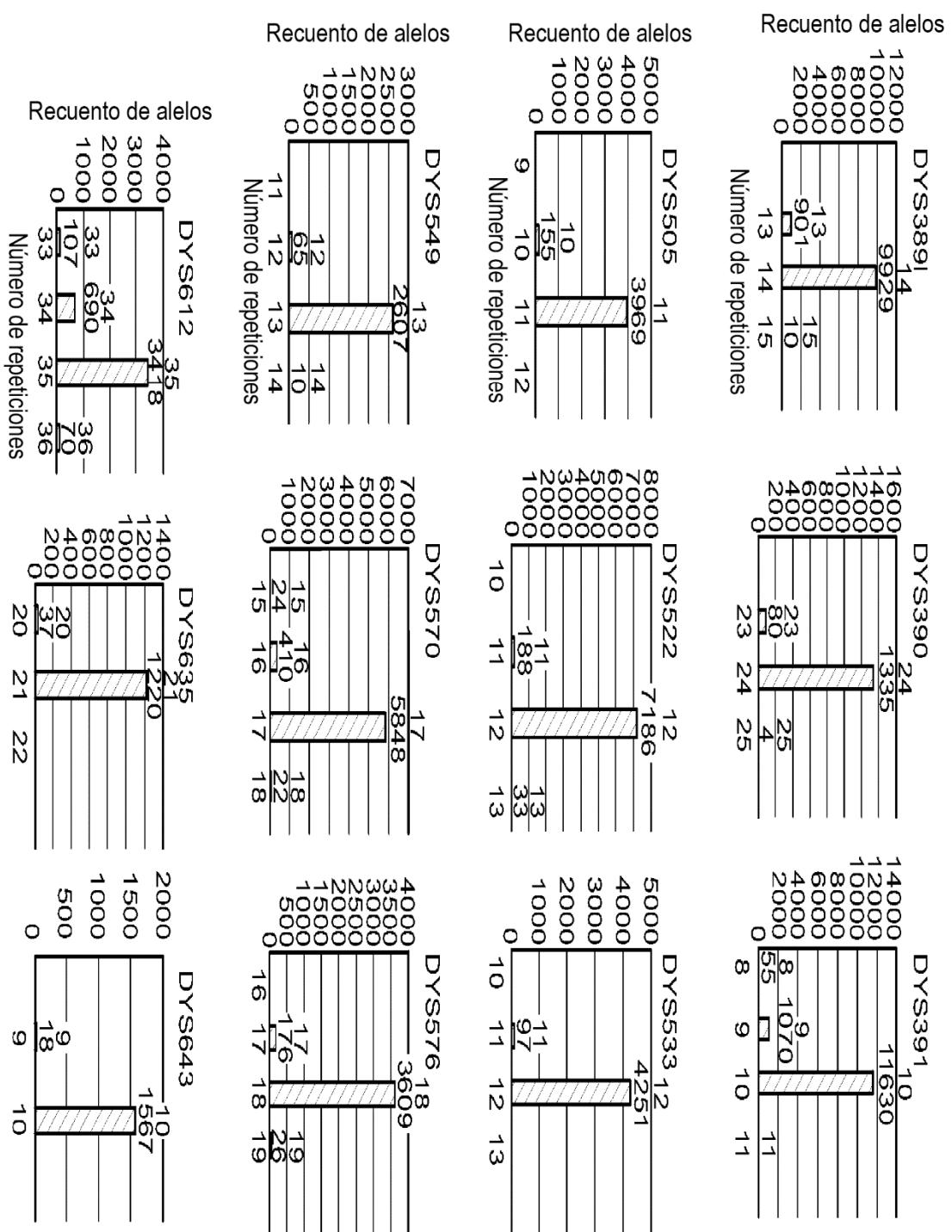
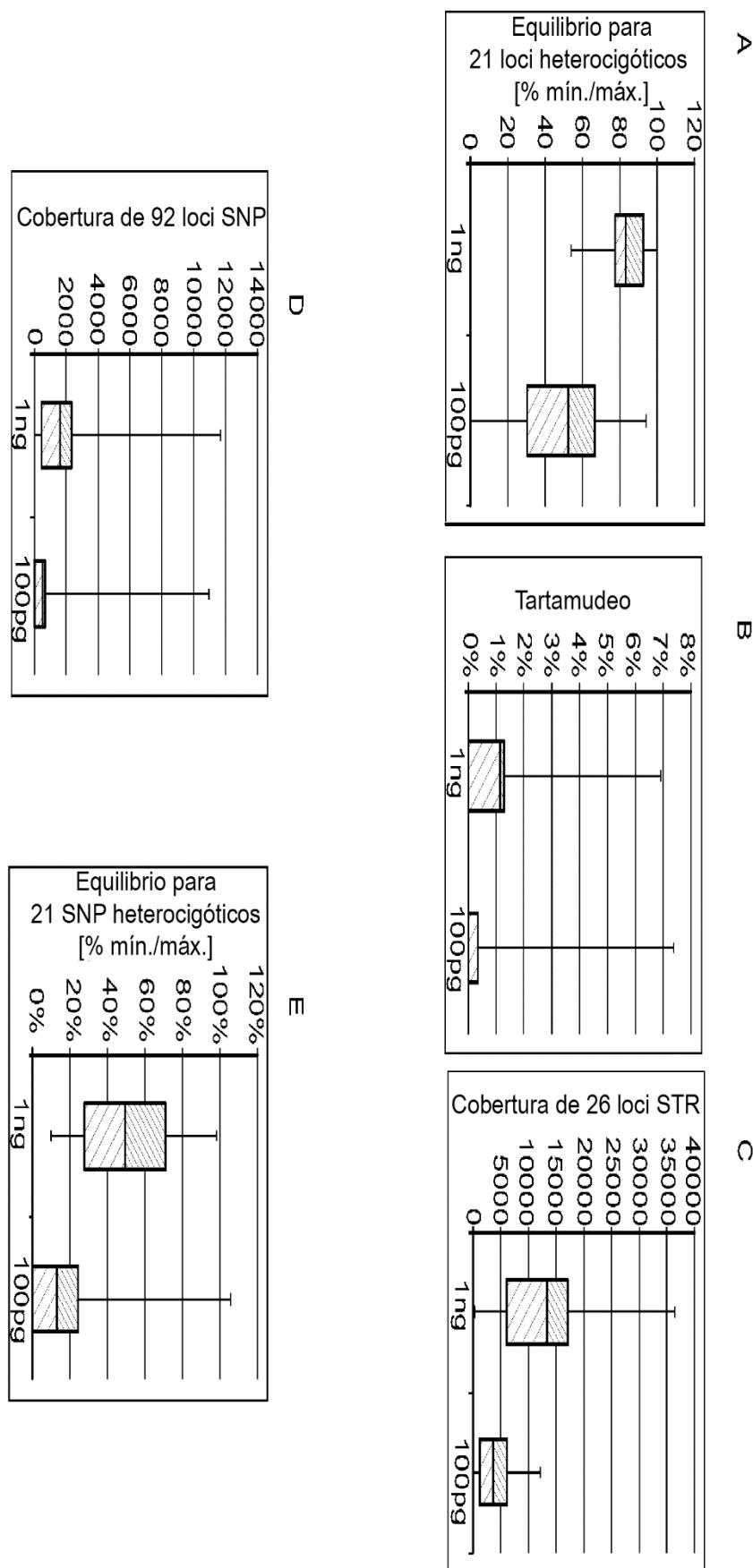
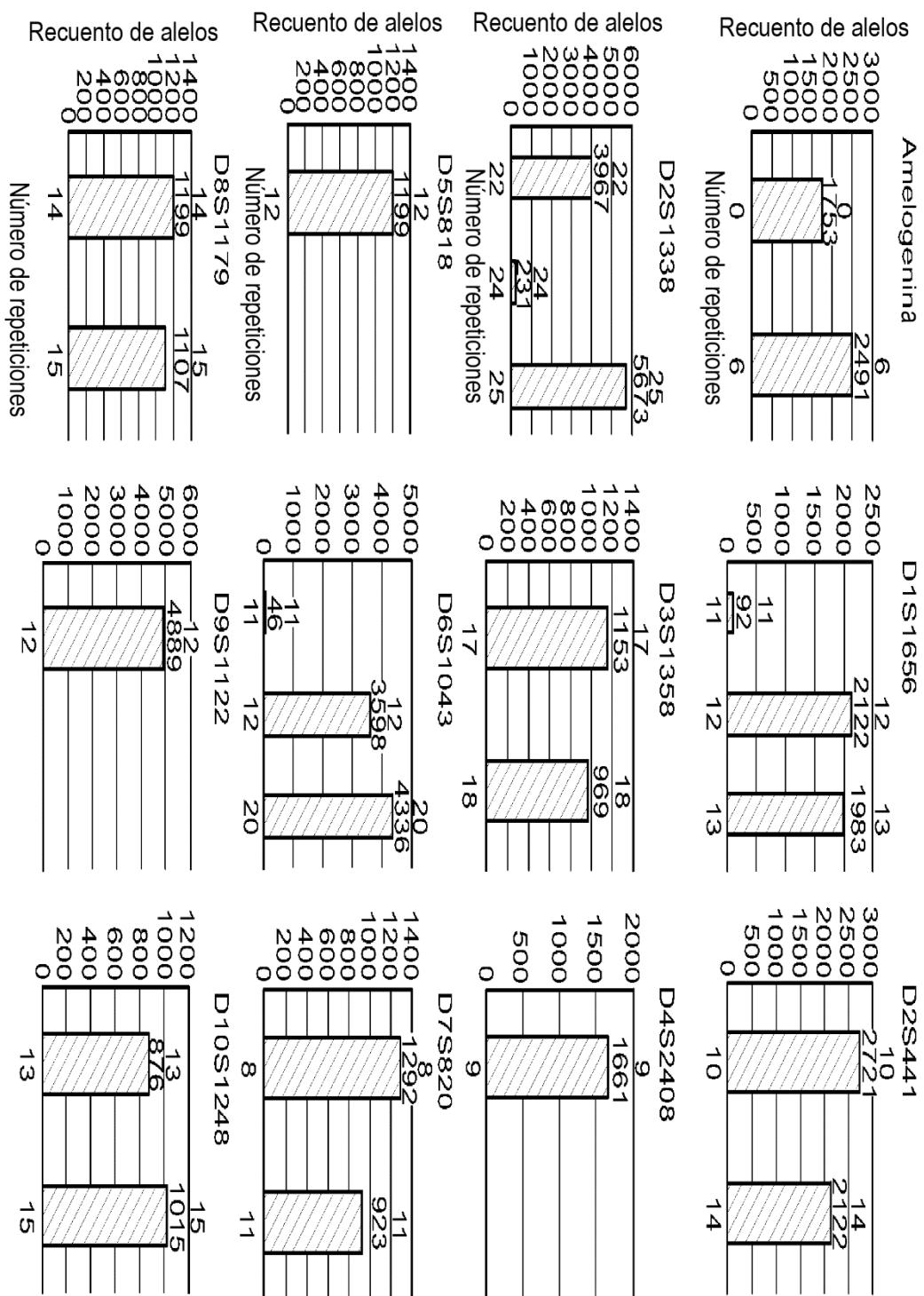


Fig. 6 (cont.)

Fig. 7



**Fig. 8**

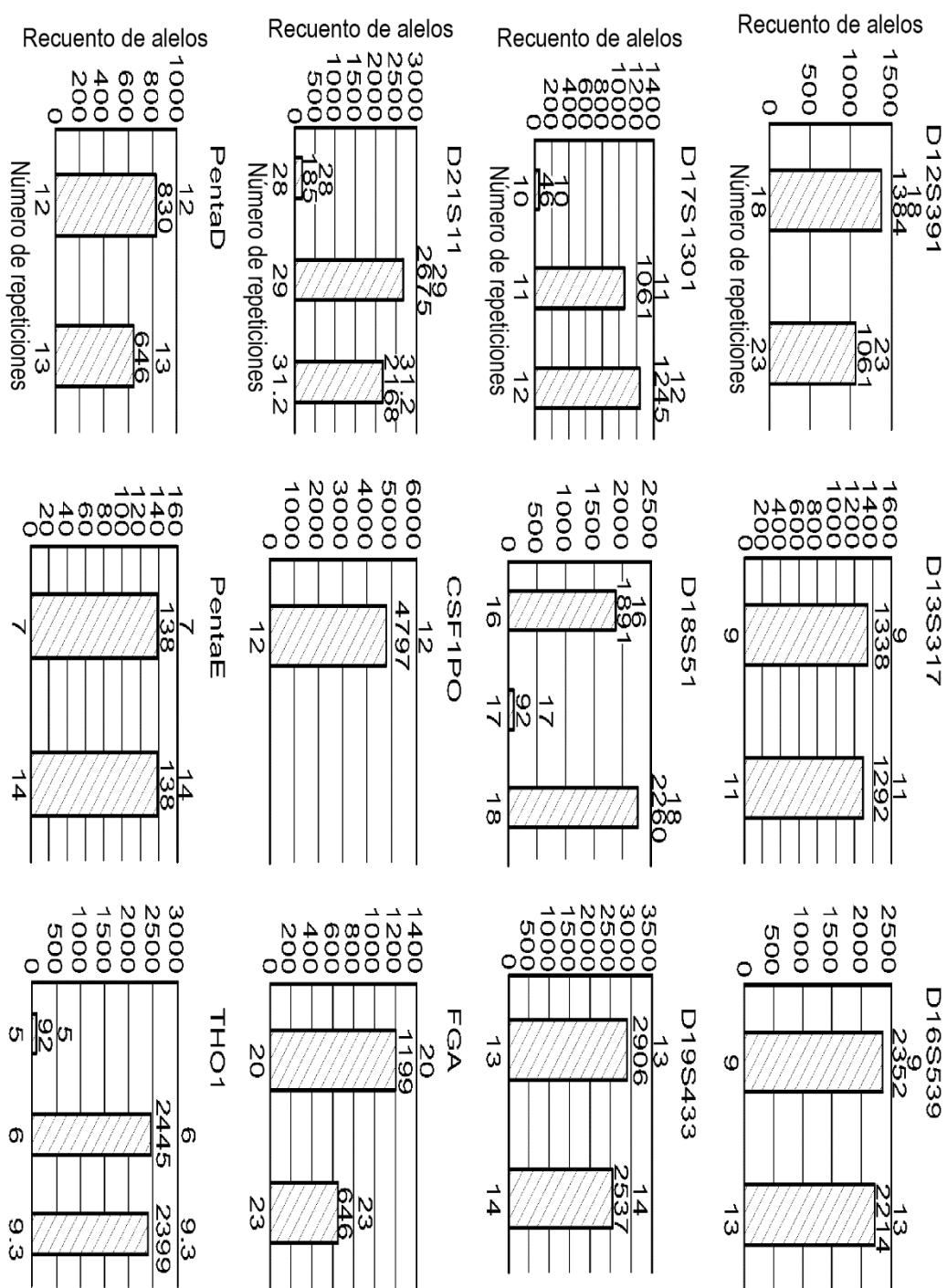
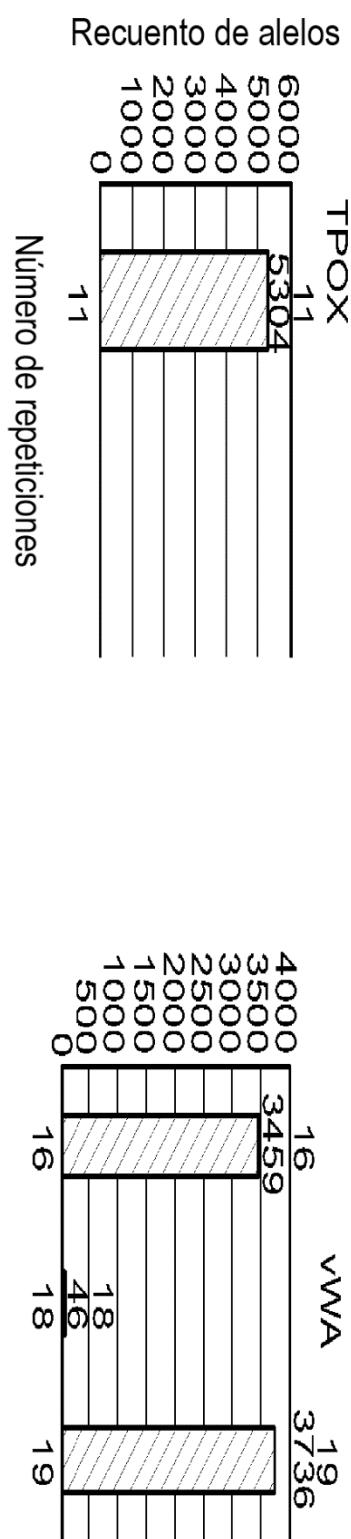
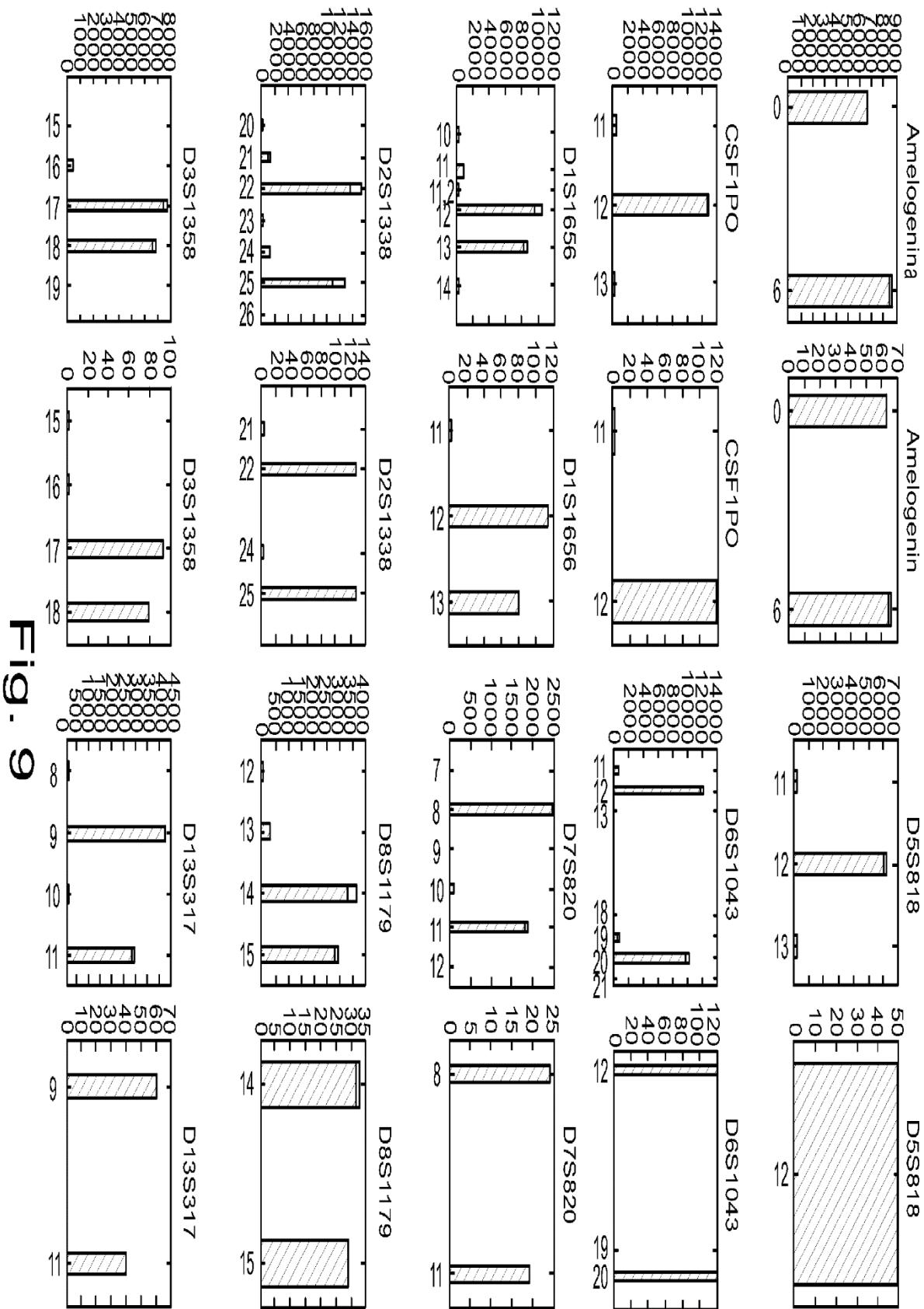
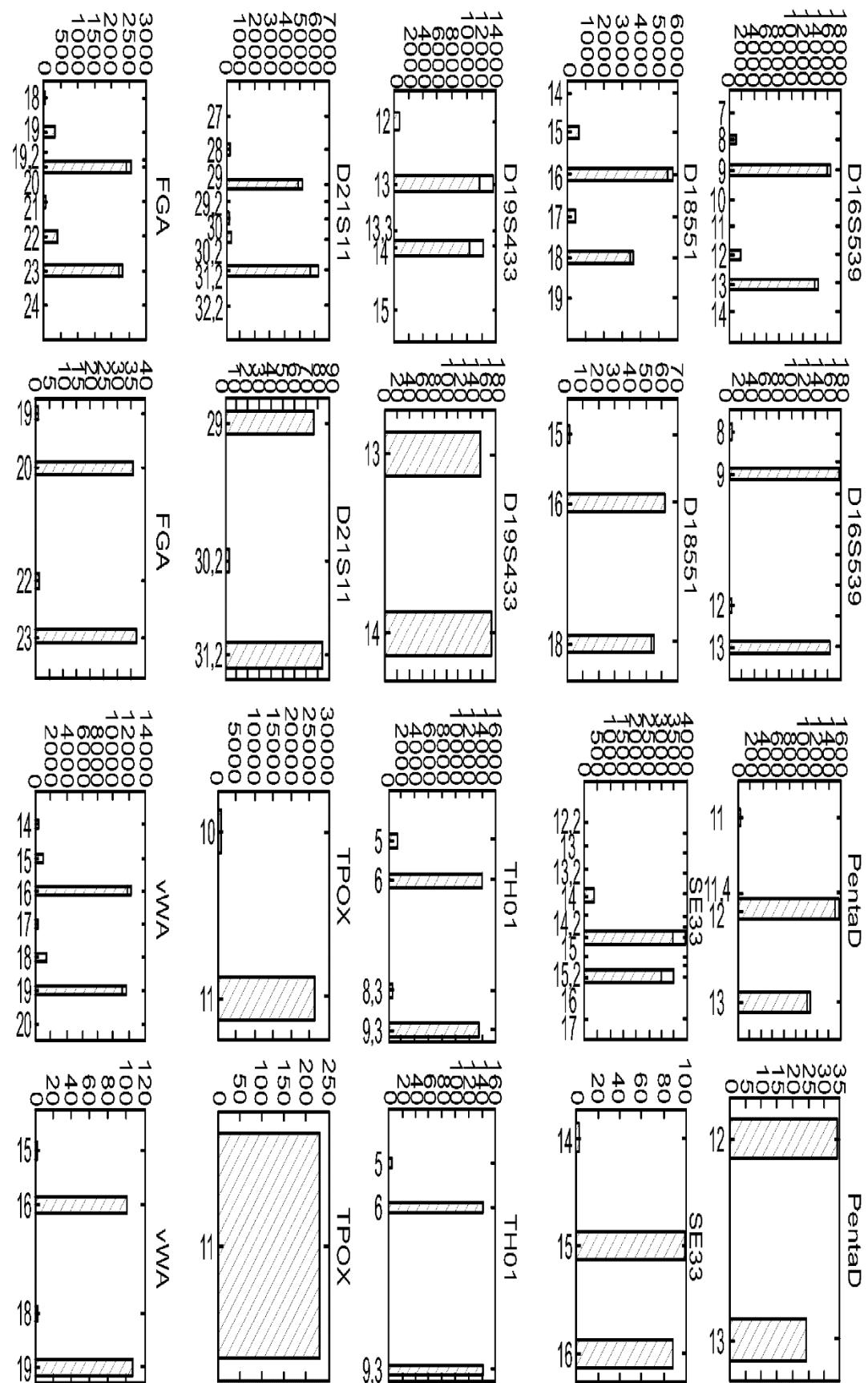


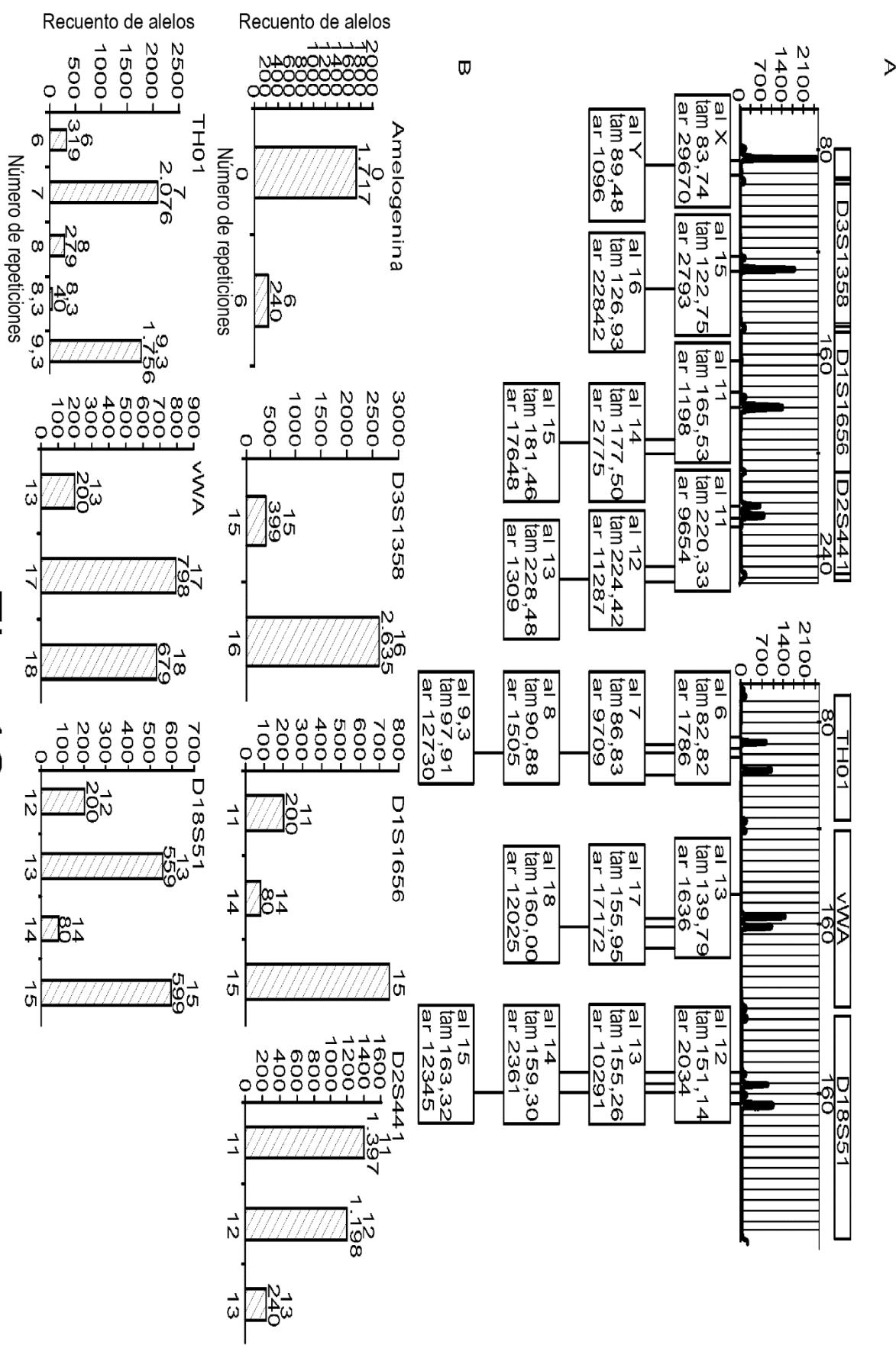
Fig. 8 (cont.)

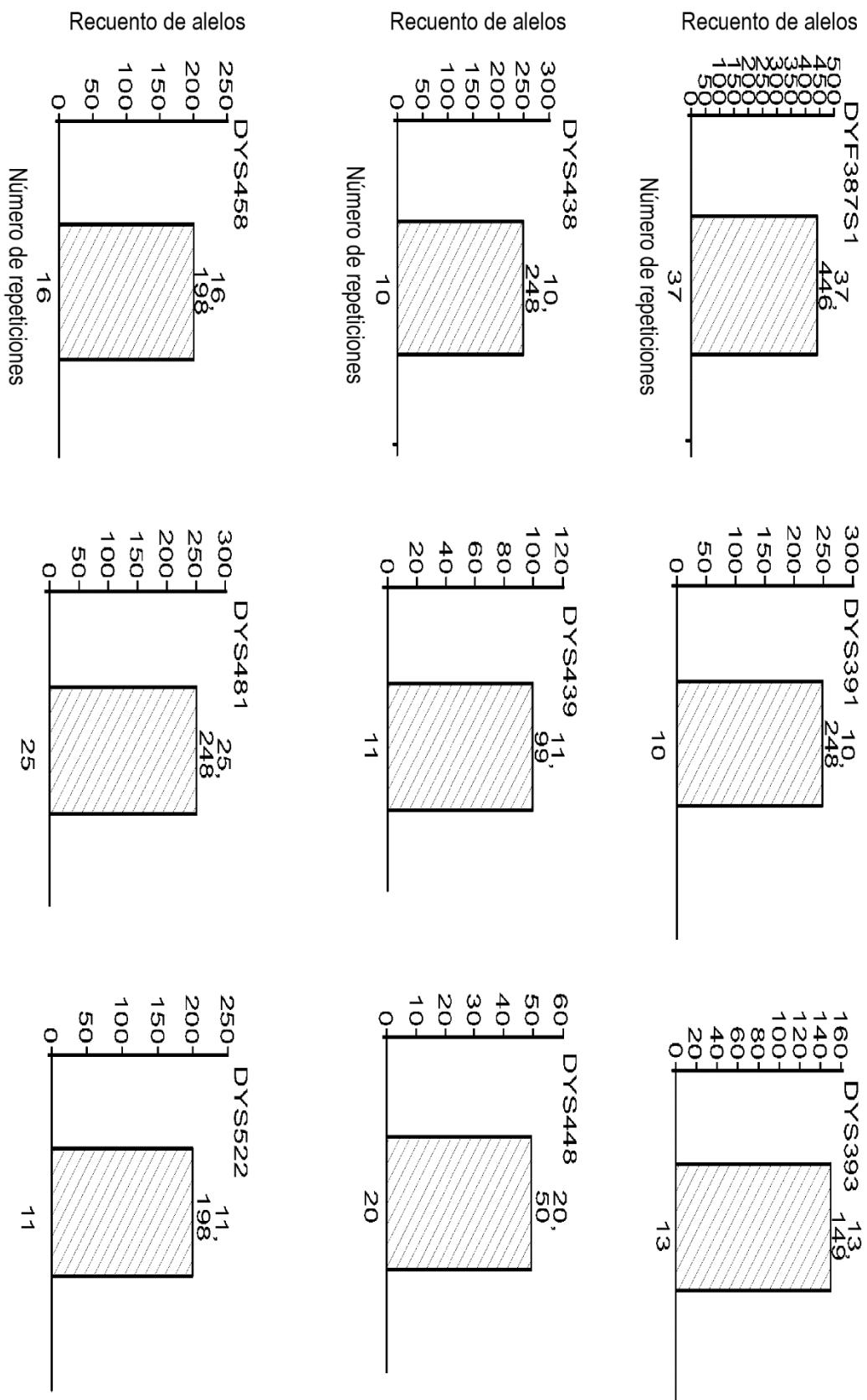
Fig.- 8 (cont.)





**Fig. 9**

**Fig. 10**

**Fig. 11**

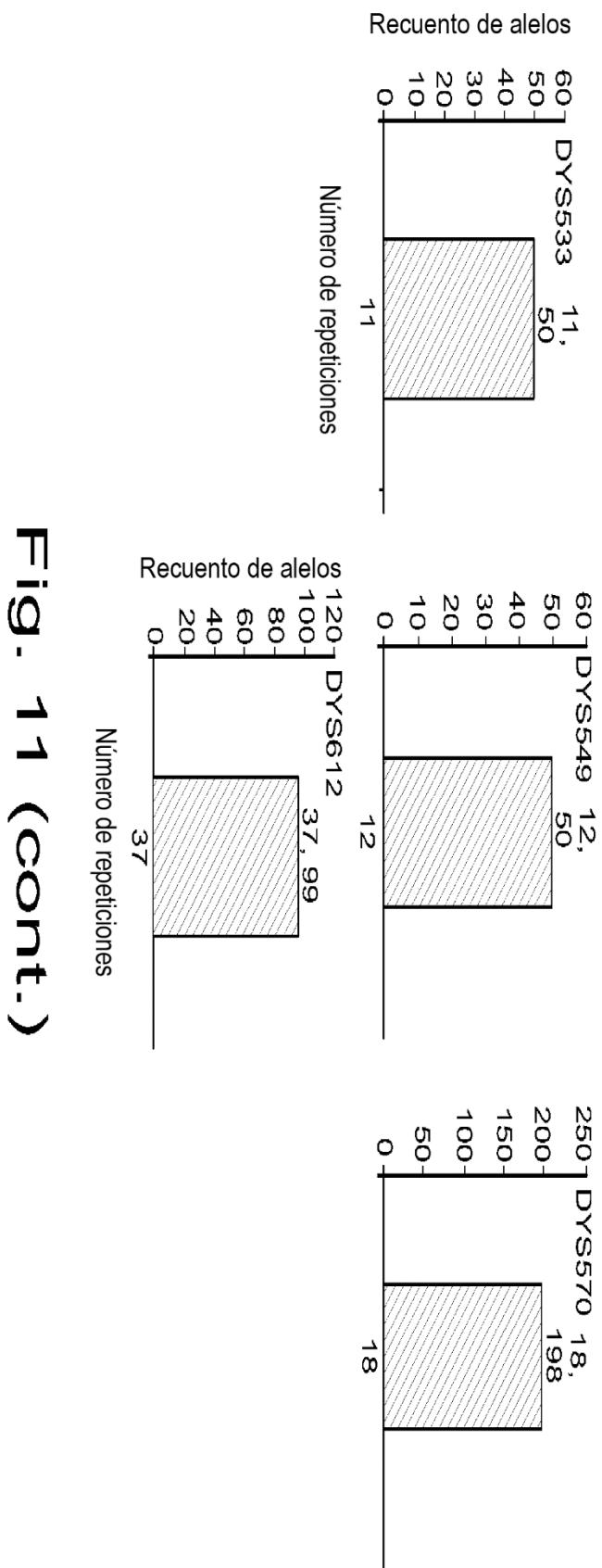


Fig. 11 (cont.)

	Sexo (amelogenina)		Lecturas alineadas con STR		Lecturas alineadas con SNP		n.º de loci STR llamados		n.º de loci SNP llamados	
	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2
Muestra 1	X, Y	X, Y	203068	248442	149466	181778	61	61	173	173
Muestra 3	X	X	216055	257742	212633	240954	37	37	173	173
Muestra 4	X, Y	X, Y	207488	205907	178985	169305	60	60	173	173
Muestra 5	X	X	199073	248681	205600	252732	37	37	173	173
Muestra 6	X, Y	X, Y	257850	281531	108997	129134	61	61	173	173
Muestra 7	X, Y	X, Y	312646	273206	156677	120709	61	61	173	173
Muestra 10	X	X	182882	321643	130252	235847	37	37	173	173
Muestra 13	X	X	266863	319394	183128	207556	37	37	173	173
Muestra 14	X	X	257545	180804	221688	155998	37	37	173	173
Muestra 15	X, Y	X, Y	232712	251490	150815	184715	37	37	173	173
Muestra 16	X, Y	X, Y	275702	213553	176862	126154	61	60	173	173
Muestra 17	X	X	159708	235314	106760	193307	37	37	173	173
2800M	X, Y	X, Y	327295		125254		61		173	
NTC	NA	NA	0	0	0	0	0	0	0	0

Fig. 12

Muestra	NIST Auto-STR RMP	NIST Y-STRs 95% de confianza frecuencia de haplotipos	dbSNP iSNP RMP	US Y-STR DB (probabilidad de apareamiento)
Muestra 1	1,05E-36	8,32E-03	3,98E-45	-2,58
Muestra 3	4,90E-33		2,51E-46	
Muestra 4	7,41E-37	8,32E-03	7,94E-41	-2,58
Muestra 5	2,82E-37		1,00E-43	
Muestra 6	1,45E-34	8,32E-03	6,31E-45	-2,58
Muestra 7	4,17E-36	8,32E-03	1,58E-42	-2,58
Muestra 10	3,55E-38		3,98E-44	
Muestra 13	1,58E-37		7,94E-43	
Muestra 14	2,00E-37		1,26E-44	
Muestra 15	8,71E-38		1,58E-40	
Muestra 16	6,03E-42	8,32E-03	1,26E-44	-2,58
Muestra 17	4,07E-34		7,94E-46	
2800M	5,75E-33	8,32E-03	1,58E-44	-2,58

Fig. 13

	Predicciones								Características del donante		
	Color de ojos			Color de pelo				Color de ojos	Color de pelo		
	azul	intermedio	marrón	marrón	rojo	negro	rubio	claro	oscuro		
Muestra 1	8%	14%	79%	47%	0%	12%	41%	74%	26%	Marrón	Marrón
Muestra 3	0%	1%	99%	35%	0%	63%	3%	3%	97%	Marrón	Marrón oscuro
Muestra 4	1%	3%	97%	30%	0%	68%	2%	3%	97%	Marrón	Negro
Muestra 5	2%	23%	75%	79%	4%	8%	10%	60%	40%	Castaño	Marrón
Muestra 6	3%	6%	91%	66%	0%	18%	16%	54%	46%	Marrón	Marrón
Muestra 7	0%	0%	100%	6%	0%	94%	0%	0%	100%	Marrón oscuro	Negro
Muestra 10	49%	22%	29%	36%	0%	14%	50%	82%	18%	Azul	Marrón/rubio
Muestra 13	0%	1%	99%	32%	0%	67%	1%	2%	98%	Marrón oscuro	Negro
Muestra 14	2%	9%	88%	60%	1%	34%	4%	10%	90%	Marrón oscuro	Marrón oscuro
Muestra 15	1%	4%	95%	33%	0%	66%	0%	0%	100%	Marrón oscuro	Marrón oscuro
Muestra 16	0%	0%	100%	16%	0%	84%	0%	0%	100%	Marrón oscuro	Negro
Muestra 17	85%	8%	6%	63%	8%	11%	19%	62%	38%	Azul	Marrón claro/rubio
2800M	5%	12%	83%	43%	0%	5%	52%	88%	12%		

Fig. 14

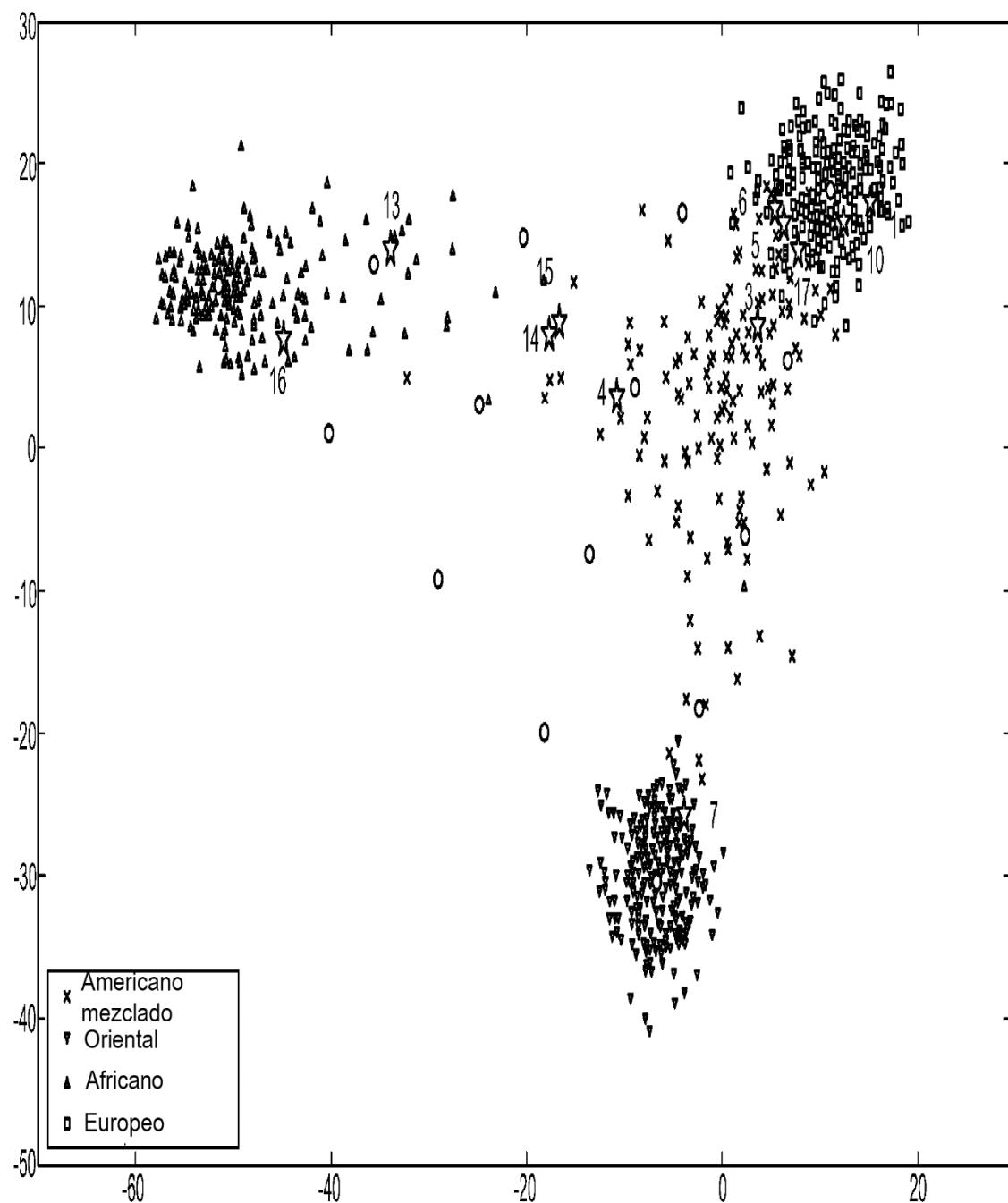


Fig. 15

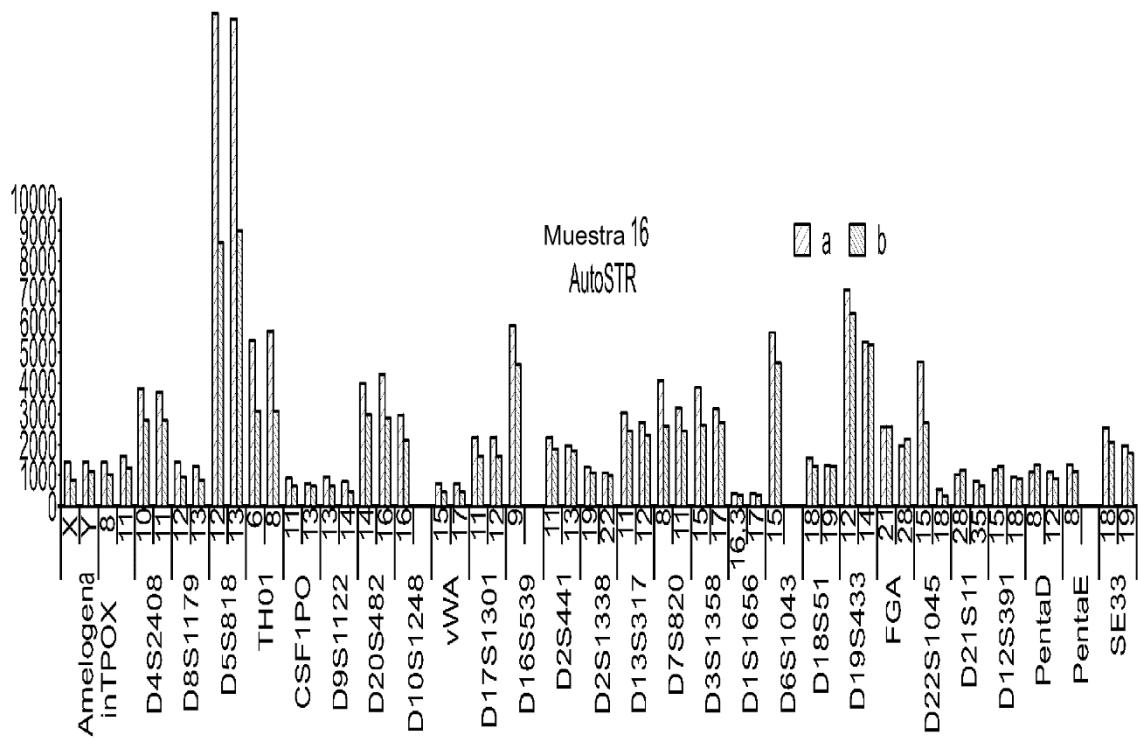


Fig. 16A

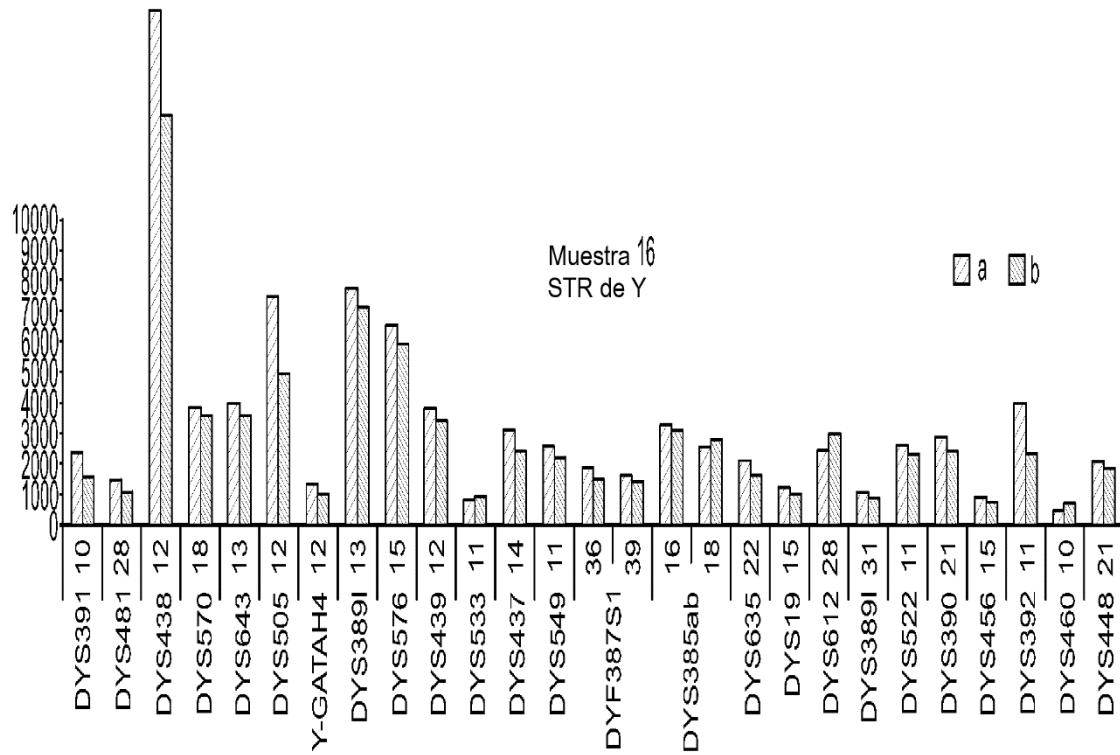


Fig. 16B

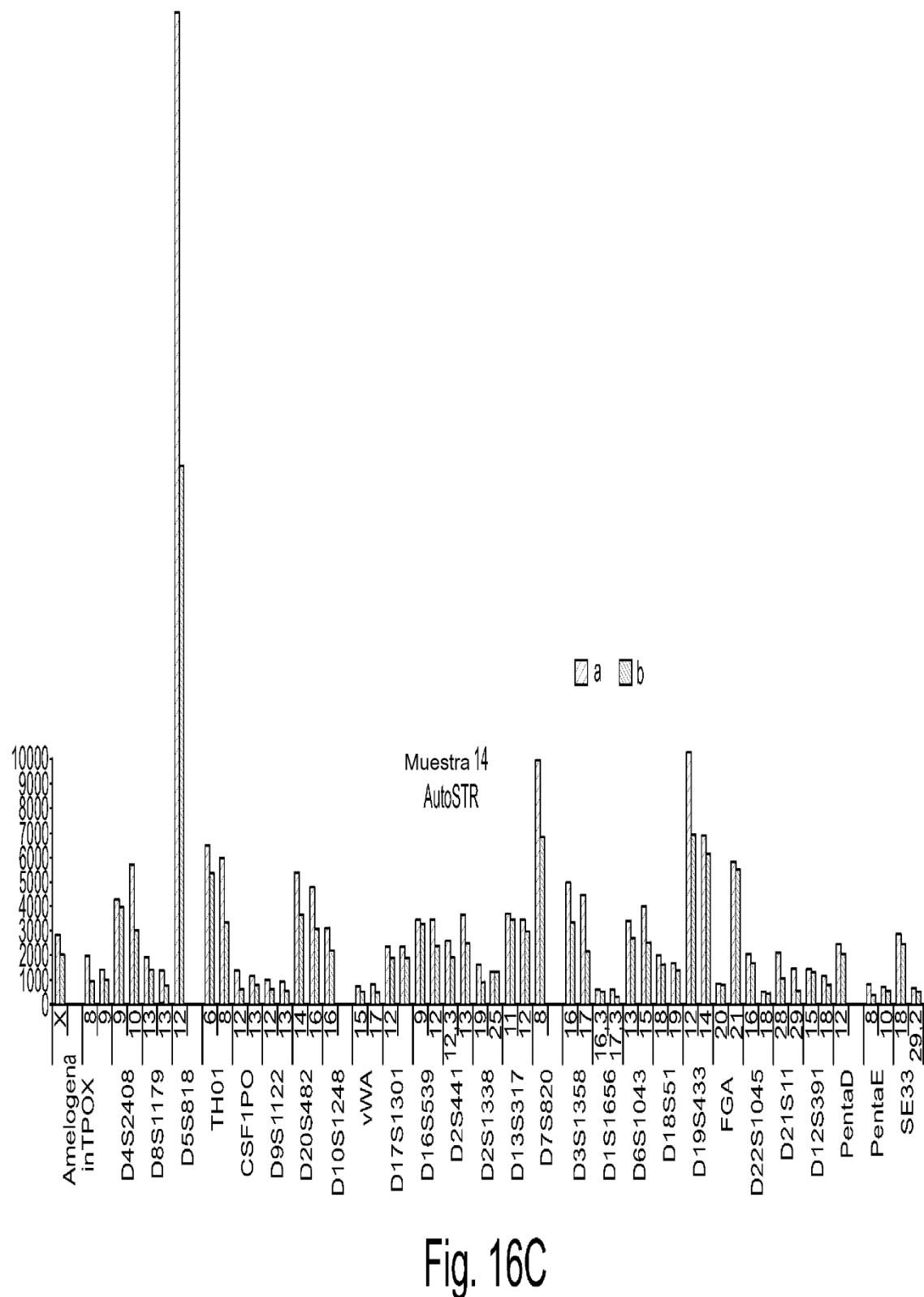


Fig. 16C

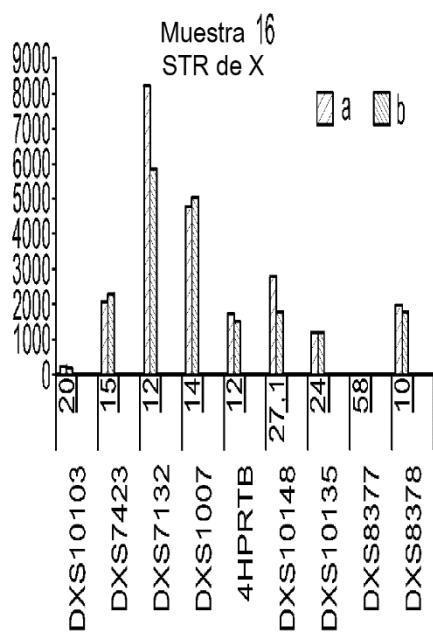


Fig. 16D

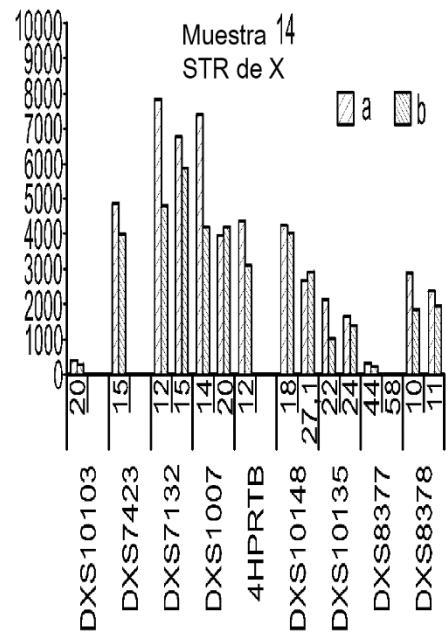


Fig. 16E

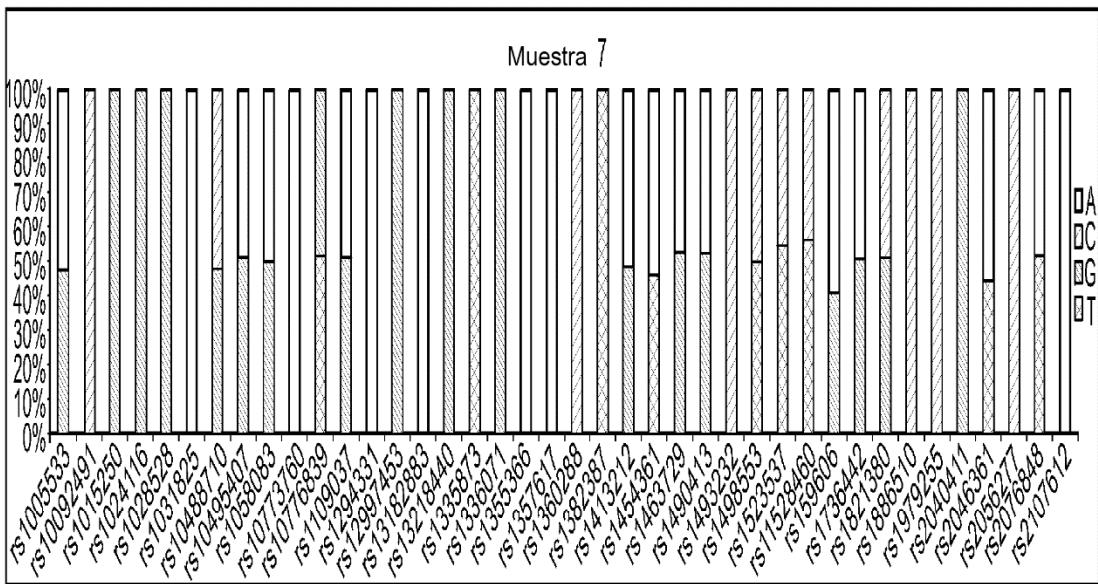


Fig. 17A

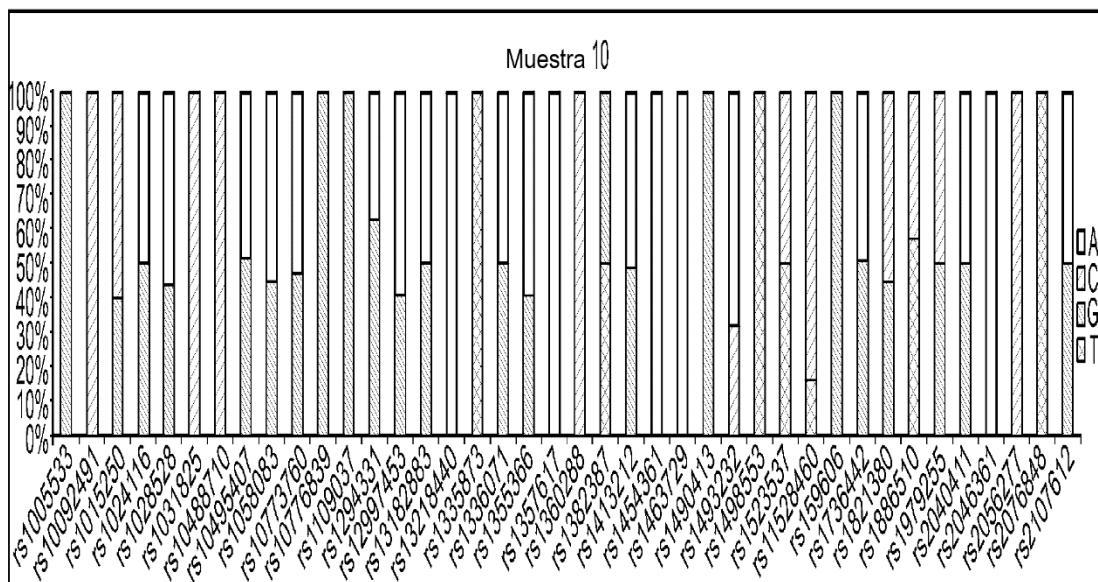


Fig. 17B

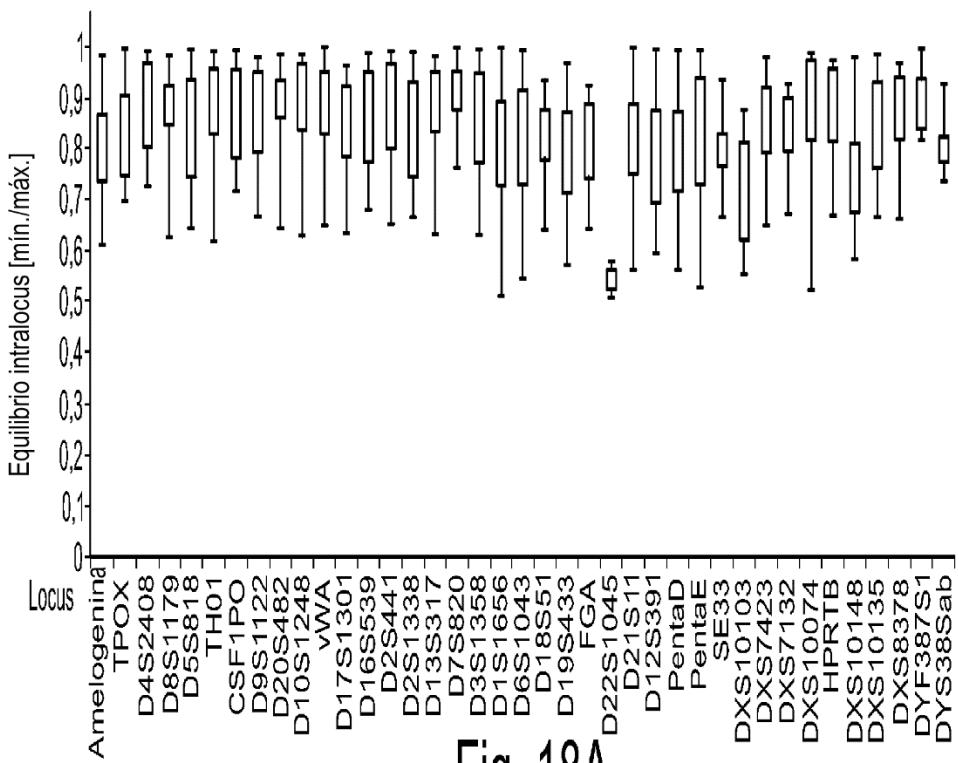


Fig. 18A

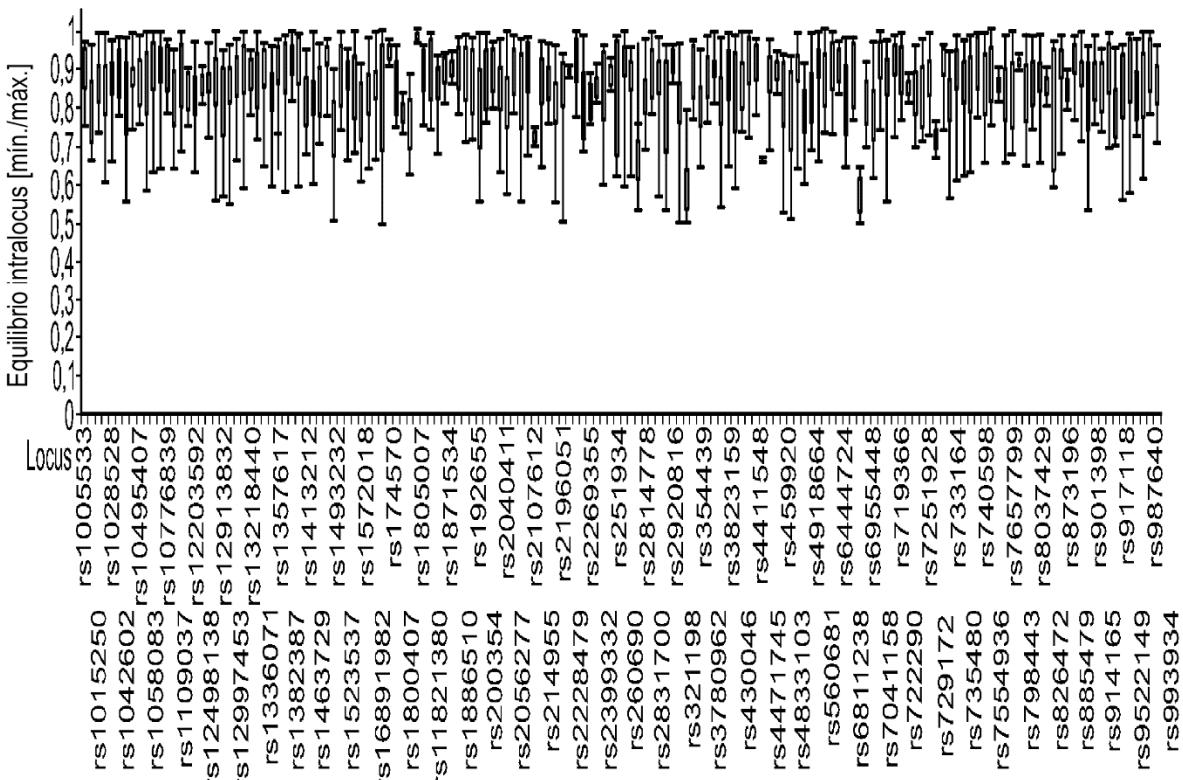


Fig. 18B

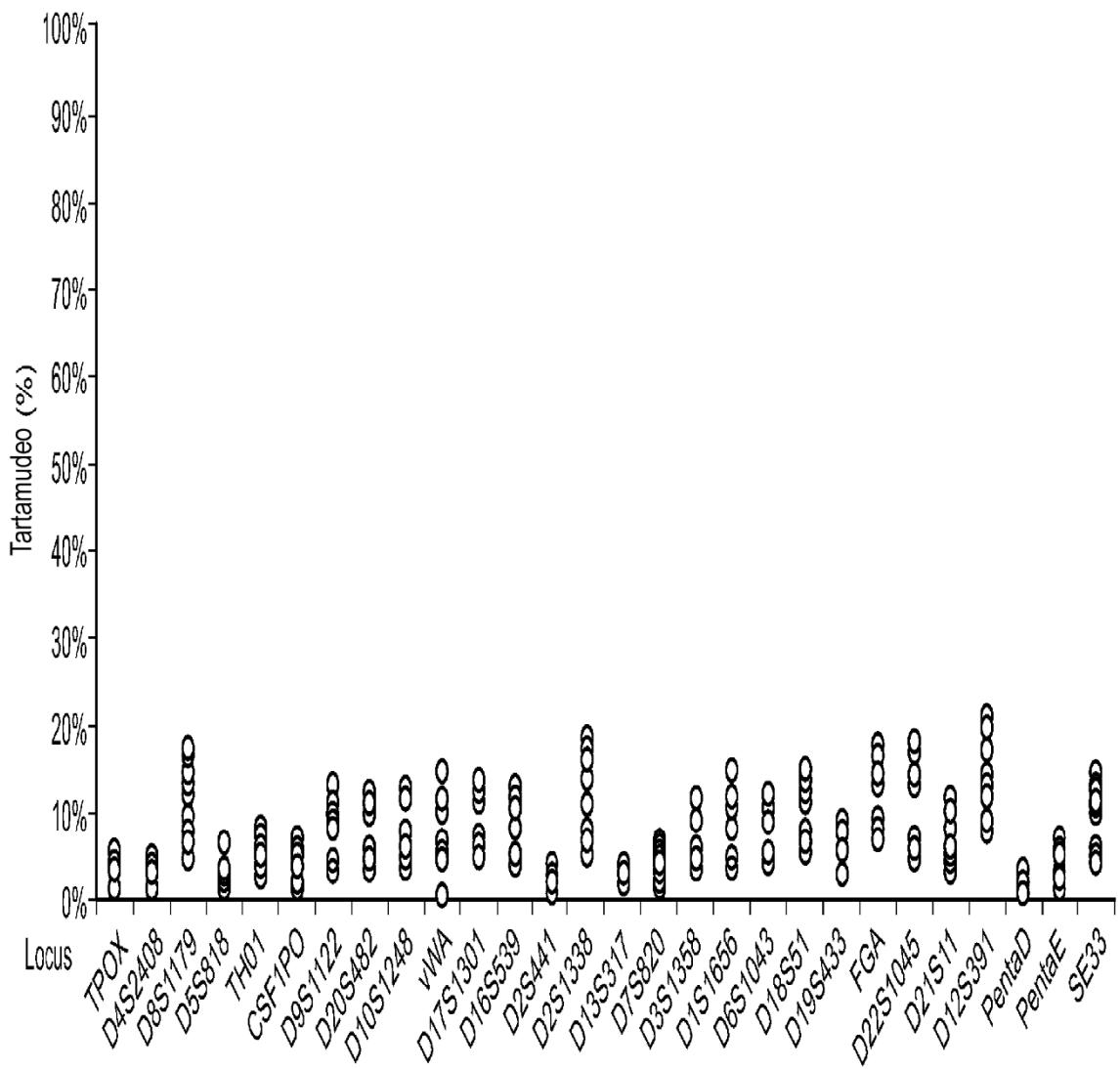


Fig. 19A

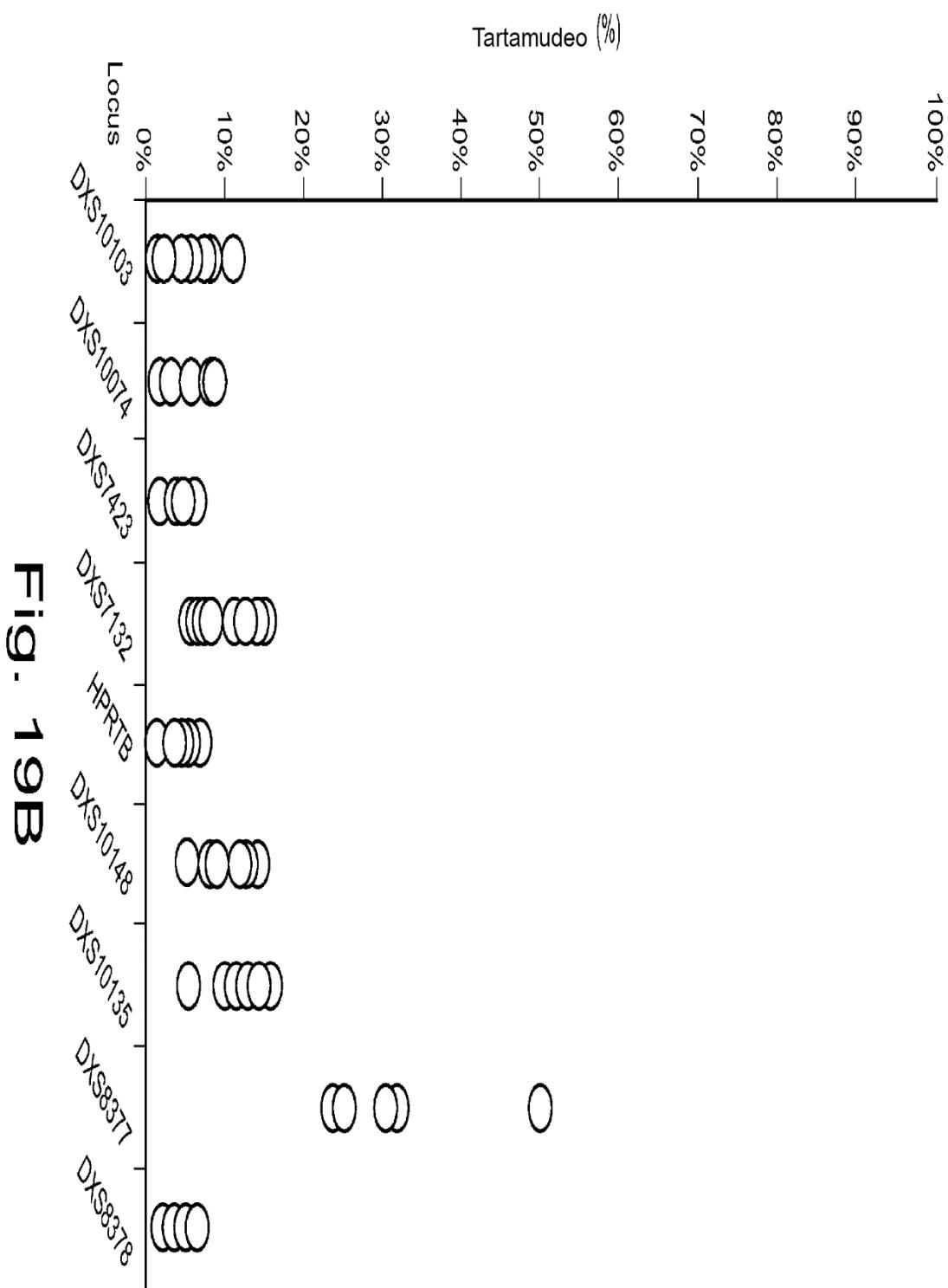


Fig. 20

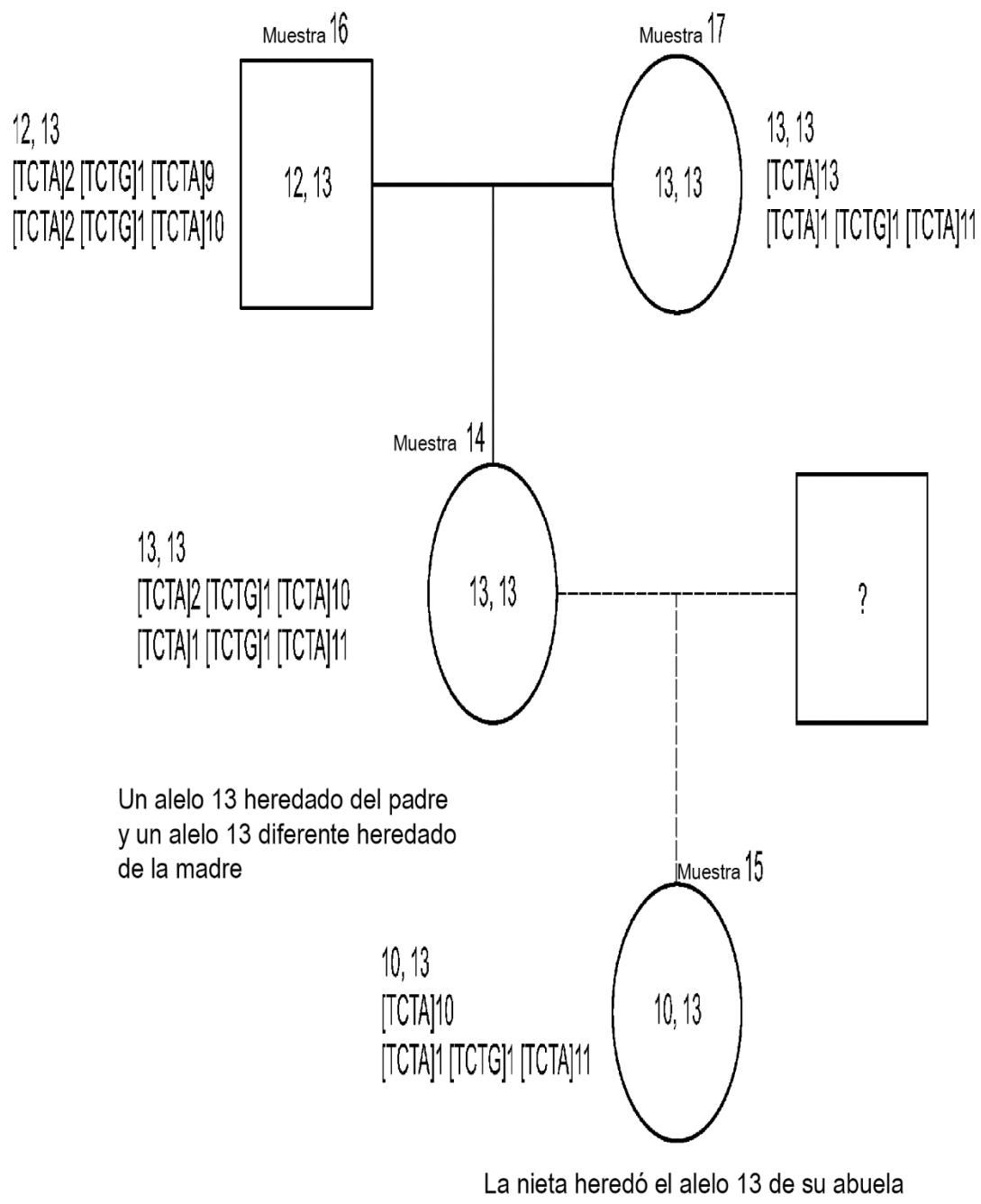


Fig. 21

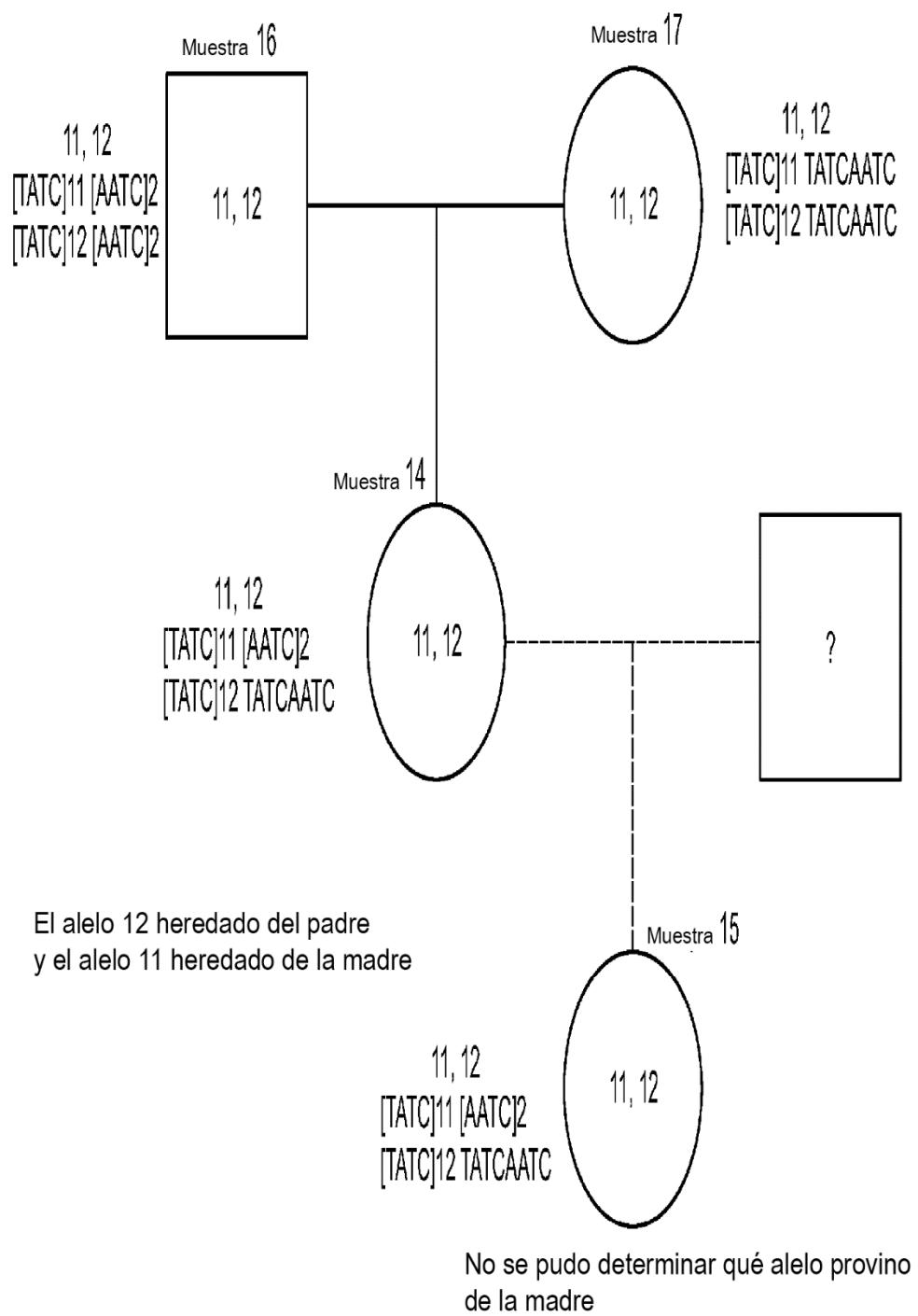


Fig. 22

Muestra	STR llamados	SNP llamados	Alelos STR perdidos	Alelos SNP perdidos	RMP*
Masculino control	60	94	Ninguno	Ninguno	1 en 5,6E+28
Masculino, cortado	45/60	93/94	5	3	1 en 2,3E+28
Masculino, cortado + ADNasa	28/60	81/94	5	8	1 en 5,0E+19
Femenino control	36	94	Ninguno	Ninguno	1 en 1,7E+29
Femenino, cortado	32/36	92/94	7	3	1 en 3,11E+28
Femenino, cortado	21/36	77/94	11	17	1 en 1,41E+16

Fig. 23

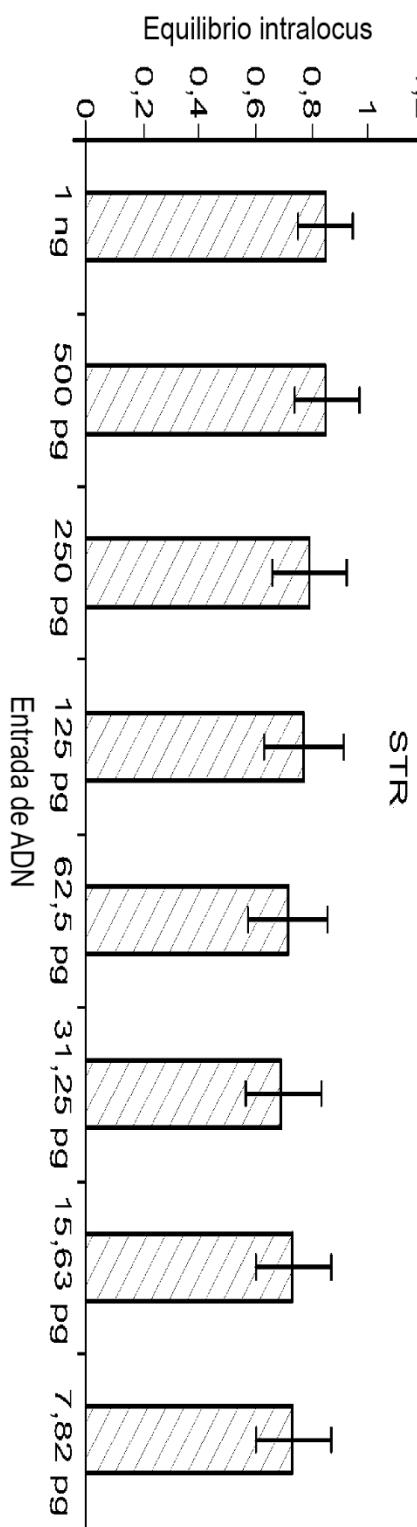


Fig. 24A

	Total de alelos STR llamados			
	Replica			
pg ADN	1	2	3	4
7,82	51	49	59	52
15,6	66	74	67	76
31,3	85	88	86	83
62,5	88	89	89	89
125	89	89	89	89
250	89	89	89	89
500	89	89	89	89
1000	89	89	89	89

Fig. 24B

pg ADN	Total de alelos SNP llamados				
	Réplica	1	2	3	4
7,82	121	129	120	129	129
15,6	175	172	166	173	173
31,3	210	209	215	210	210
62,5	222	226	225	222	222
125	226	226	226	226	226
250	226	226	226	226	226
500	226	226	226	226	226
1000	226	226	226	226	226

Fig. 25A

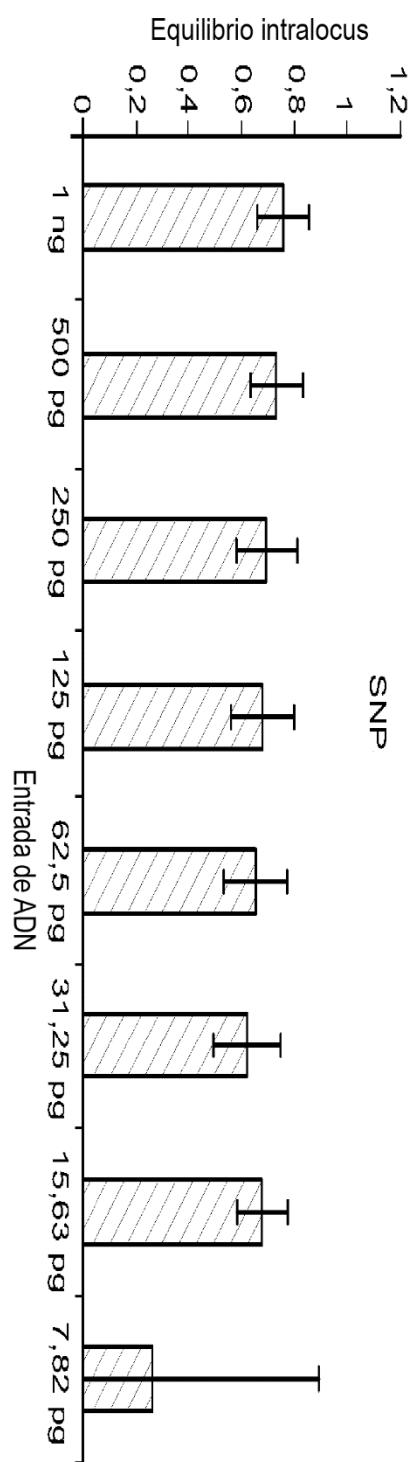


Fig. 25B