

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 609 130**

21 Número de solicitud: 201531475

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación:

**14.10.2015**

43 Fecha de publicación de la solicitud:

**18.04.2017**

Fecha de la concesión:

**18.01.2018**

45 Fecha de publicación de la concesión:

**25.01.2018**

73 Titular/es:

**UNIVERSIDAD DE LA LAGUNA (50.0%)  
Pabellón de Gobierno, C/ Molinos de Agua, s/n  
38200 La Laguna (Tenerife) ES y  
SERVICIO CANARIO DE SALUD (50.0%)**

72 Inventor/es:

**FLORES INFANTE, Carlos Alberto;  
MORENO VEGA, José Marcos y  
BASALDÚA LEMARCHAND, Santiago**

74 Agente/Representante:

**ARIAS SANZ, Juan**

54 Título: **Método para determinar la ascendencia geográfica de un sujeto**

57 Resumen:

Método para determinar la ascendencia geográfica de un sujeto.

La presente invención se relaciona con métodos para determinar la ascendencia geográfica de un sujeto basados en la determinación de la secuencia de un panel de 7 polimorfismos nucleotídicos sencillos. La invención también se relaciona con kits para llevar a cabo dichos métodos y con sus usos.

ES 2 609 130 B1

## DESCRIPCIÓN

### Método para determinar la ascendencia geográfica de un sujeto

#### CAMPO DE LA INVENCION

5           La presente invención está relacionada con métodos para determinar el origen geográfico de un sujeto a partir de muestras de ADN, así como con kits para llevar a cabo dicha determinación.

#### ANTECEDENTES DE LA INVENCION

10           La utilización de información genética para la distinción de grupos humanos es un objetivo común muy extendido en muchos estudios biomédicos, sean estos dirigidos a la genética médica, forense o a la caracterización de diferencias interindividuales en la respuesta a fármacos. Las categorizaciones raciales y étnicas, comúnmente utilizadas para designar la ascendencia individual, constituyen  
15           aproximaciones imperfectas pues implican factores genéticos y no genéticos. Así, a pesar de que dichas categorías son importantes para el estudio, el diagnóstico y el tratamiento de enfermedades, tales indicadores de ascendencia (basado en declaraciones del individuo) están sujetas a error pues no describen con detalle la constitución genética de un individuo tal que evidencie el origen (biogeográfico) de sus  
20           ancestros, de quienes ha heredado su genoma. En este escenario, la inferencia de la ascendencia genética (también conocida como ascendencia genética global) ha demostrado ser una alternativa no sesgada extremadamente útil.

          Una multitud de estudios han mostrado que, como media, el 87-96% de la variación genética humana existente a nivel mundial es atribuible a las diferencias  
25           entre dos individuos cualesquiera, mientras que solo el 4-13% será debido a diferencias genéticas existentes entre las dos poblaciones de las que esos individuos son escogidos. Estos estudios también han evidenciado la existencia de hasta siete grupos de poblaciones mundiales identificables genéticamente utilizando varios cientos de miles de polimorfismos.

30           Muchos estudios han intentado identificar pequeños conjuntos o paneles de polimorfismos genéticos que se comporten como marcadores informativos de la ascendencia (conocidos como AIM, del inglés *Ancestry Informative Markers*), para permitir una estimación de los orígenes genéticos recientes de un individuo. Tal y como ocurre en lo concerniente al desarrollo de pruebas forenses basadas en el ADN  
35           (por ejemplo el IrisPlex), los polimorfismos genéticos preferidos para construir dichos

paneles de AIM son los llamados polimorfismos nucleotídicos sencillos (conocidos como SNP, del inglés *Single Nucleotide Polymorphisms*). La utilización de SNP ha recibido un trato preferente en distintos ámbitos científicos para desarrollo de tales pruebas de ADN por involucrar ensayos técnicamente fáciles de diseñar, altamente sensibles, rápidos y económicos, así como por ofrecer gran versatilidad por su capacidad de adaptación a un amplio rango de sistemas de detección. A pesar de la disponibilidad de métodos de alto rendimiento para la determinación de cientos de miles de SNP, la utilización de tecnologías de alto rendimiento puede suponer un coste prohibitivo en algunas situaciones. Además de ello, dichas tecnologías requieren de una cantidad sustancial de ADN no fragmentado de partida que no siempre se encuentra disponible en la muestra biológica (por ejemplo en el contexto de la medicina forense). Puesto que los polimorfismos genéticos que se localizan en los cromosomas sexuales o en regiones del genoma donde no existe recombinación pueden dar lugar a interpretaciones conflictivas, particularmente cuando se trata de inferencias en individuos procedentes de poblaciones que han sufrido sesgos en la proporción de sexos que contribuyen en la mezcla, estos estudios se han focalizado exclusivamente en SNP autosómicos.

La asignación de ascendencia a muestras humanas de ADN mediante la utilización de paneles de AIM es, y será, de gran utilidad en el contexto de estudios y procedimiento donde la obtención de datos genómicos mediante tecnologías de alto rendimiento sea impracticable.

En el ámbito de la medicina forense, donde el material biológico existente en la escena de un crimen o en una catástrofe es normalmente escaso y/o se encuentra degradado, un panel de una docena de SNP con características de AIM resulta de gran utilidad para predecir la ascendencia del sospechoso o sujeto, reduciendo el ámbito de su identificación. Por tanto, el panel de AIM que se propone ayudaría en las investigaciones forenses a concentrar los esfuerzos en la identificación de individuos desconocidos en situaciones donde la obtención de los perfiles de ADN de rutina no daría ninguna información de utilidad práctica.

En el ámbito de la biomedicina, donde el número de muestras biológicas disponibles para los estudios es habitualmente enorme, un panel de una docena de SNP con características de AIM resulta muy económico y útil para el control de calidad rutinario de biobancos, y en el diseño y organización del mismo. Además, específicamente en el ámbito de investigación farmacéutica, la identificación y clasificación económica de las muestras o sujetos de ensayo en categorías poblacionales es esencial por la conocida diversidad en las respuestas a fármacos y

su relación con la ascendencia biogeográfica. En el contexto de estudio de enfermedades complejas, donde el número de sujetos estudiados es de miles y los costes de determinación son elevados, el uso de paneles de AIM de coste reducido en etapas tempranas del proyecto supone reducción de gastos, permitiendo la  
5 identificación temprana de los estratos poblacionales más apropiados para el estudio.

Los métodos previamente utilizados para identificar o priorizar AIM (de los millones de SNP existentes en nuestro genoma) para permitir inferencias eficientes de la ascendencia individual han estado basados en puntuaciones relacionadas con el contenido informativo de ascendencia de cada SNP individualmente. Aunque  
10 intuitivamente fáciles de llevar a cabo, los paneles resultantes de combinar subconjuntos de AIM identificados ofrecen soluciones subóptimas, permitiendo tan solo inferencias poco precisas en poblaciones diversas, en comparación con las que podrían ofrecer estrategias basadas en AIM identificados mediante análisis combinatorios (multi-marcador). Una de las razones que podría explicar que tales  
15 análisis combinatorios no hayan sido implementados podría ser el hecho que el derivado de modelos combinatorios que incluyeran más de 3 AIM sin tener ninguna hipótesis previa que favorezca su combinación constituye un problema computacionalmente complejo. En consecuencia, las medidas de información de ascendencia utilizadas hasta hoy día para priorizar los AIM de uso habitual son  
20 esencialmente la diferencia neta en la frecuencia alélicas entre las poblaciones, distancias genéticas (tipo  $F_{ST}$ ), y el índice  $I_n$  de información de asignación de ascendencia. Utilizando estas medidas, se han derivado diferentes conjuntos de AIM una vez han sido eliminados aquellos que muestran correlaciones potenciales con algún otro del panel. Dichos análisis han sido dirigidos a poblaciones europeas y  
25 americanas, típicamente necesitando de paneles de varios cientos de AIMS para ser capaces de realizar inferencias precisas de la ascendencia. La solicitud de patente internacional WO2006089238A2 propone la utilización de varios cientos de AIM en este contexto.

Además de lo descrito, otros tres estudios han realizado un análisis exhaustivo  
30 en un intento de reducir al mínimo el número de AIM necesario para permitir la asignación de ascendencia genética, al tiempo que se mantiene una asignación precisa de individuos al máximo posible de divisiones poblacionales a nivel mundial. Mediante la aplicación de algoritmos avanzados, dos estudios fueron capaces de construir paneles de menos de 15 AIM con capacidad para asignar con precisión  
35 individuos de diverso origen a tan solo 3-4 de las siete subdivisiones poblacionales reconocidas. Sin embargo, de los paneles de AIM diseñados hasta hoy día, el más

preciso es el deducido por Paschou *et al.* (J Med Genet 2010, 47: 835-47) que, con tan solo 50 AIM seleccionados, permite una asignación precisa de individuos a cinco de las siete subdivisiones poblacionales reconocibles a nivel mundial. Para la asignación a las siete subdivisiones poblacionales proponen la necesidad de 350 AIM seleccionados. Es de destacar que incluso en este último caso, el error de asignación de individuos de origen europeo es elevado (donde el panel tan solo mantiene el 65,4% de acierto en la asignación).

Por lo tanto, sigue siendo necesario el disponer de una metodología basada en un panel de AIM que permita la correcta asignación de individuos a una población de procedencia de sus ancestros a partir de muestras de ADN. Idealmente, dicho panel debe consistir en un número mínimo de polimorfismos genéticos tipo SNP que ofrezca la máxima información sobre la ascendencia, y que permita la inferencia de la población de origen de manera ciega y precisa, utilizando para ello exclusivamente datos genotípicos derivados del panel y sin ningún otro conocimiento previo del sujeto bajo estudio.

### **COMPENDIO DE LA INVENCION**

Los autores de la presente invención han identificado, a partir de datos genómicos derivados del ADN de 938 individuos sin relación familiar de 52 poblaciones diversas del mundo, un panel de 7 SNP que permite la asignación genética precisa de individuos a una de las siete divisiones poblacionales humanas identificables en el mundo (Figura 1), mejorando considerablemente el resto de paneles de AIM que han sido propuestos para el mismo objetivo, incluso aquellos que necesitan de la determinación de varios cientos de SNP.

Este panel constituye una herramienta que:

- A) Ofrece menos error en la clasificación de ascendencia humana a nivel mundial que el resto de paneles desarrollados con un menor número de SNP.
- B) Proporciona una mayor información de ascendencia, por permitir la clasificación en hasta siete subdivisiones mundiales.
- C) Ofrece mayor versatilidad, al tratarse de un panel reducido en cuanto al número de SNP, cuya determinación puede fácilmente adaptarse a casi cualquier tecnología actual de determinación de genotipos.

Por tanto, en un primer aspecto, la invención se relaciona con un método para determinar la ascendencia geográfica de un sujeto que comprende

- (i) determinar en una muestra que comprende material genético de dicho sujeto el genotipo de los dos alelos de los polimorfismos de un solo

nucleótido (SNP) rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622 y, opcionalmente, cualquier otro polimorfismo que se encuentre en desequilibrio de ligamiento con cualquiera de los anteriores, en ambos alelos y

- 5 (ii) determinar la ascendencia geográfica de dicho sujeto en base a los genotipos de dichos SNP.

En un segundo aspecto, la invención se relaciona con un kit que comprende reactivos adecuados para la detección de la secuencia de los polimorfismos de un solo nucleótido (SNP) rs1257016, rs971779, rs238547, rs1482326, rs4760332, 10 rs2104388, y rs2427622, en donde dichos reactivos se seleccionan del grupo formado por

- (i) un conjunto de parejas de cebadores adecuados para amplificar las regiones genómicas que comprenden dichos SNP o fragmentos de las mismas que contienen dichos SNP o las secuencias complementarias 15 de las mismas y
- (ii) un conjunto de sondas que hibridan con las regiones genómicas que comprenden las secuencias de dichos SNP o con las secuencias complementarias de las mismas, en donde dichas sondas hibridan de forma específica con una de las variantes de dichos SNP.

20 En un tercer aspecto, la invención se relaciona con un uso de un kit según el segundo aspecto para determinar la ascendencia geográfica de un sujeto.

En un cuarto aspecto, la invención se relaciona con un medio leíble por ordenador que contiene instrucciones ejecutables que, al ejecutarse, resultan en

- (i) la recepción por al menos un procesador del genotipo de un sujeto en 25 los dos alelos de las posiciones polimórficas rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622 y
- (ii) proporcionar una indicación sobre la probabilidad de que un sujeto tenga una determinada ascendencia geográfica.

### 30 **BREVE DESCRIPCIÓN DE LAS FIGURAS**

**Figura 1.** Matriz de confusión entre la ascendencia declarada por los sujetos de HGDP-CEPH (Proyecto de Diversidad del Genoma Humano del Centre d'Etude du Polymorphisme Humain) (938 sujetos) y la inferida para ellos al utilizar únicamente la 35 información del panel mínimo de 7 AIM desarrollado. La intensidad del entramado es indicativa de la proporción de sujetos que son clasificados correcta (diagonal) e

incorrectamente (resto de sectores) en las siete subdivisiones biogeográficas. En la diagonal también se expresa numéricamente la proporción de aciertos en cada subdivisión. Nótese que la mayoría de las confusiones ocurren en la clasificación de sujetos de origen Eurasiático. Clave: EUR, Europa; AFR, África sub-sahariana; MEA, Oriente Próximo y Norte de África; CSA, Centro y Sur de Asia; EAS, Este de Asia; AME, Américas; OCE, Oceanía.

**Figura 2.** Matriz de confusión entre la ascendencia declarada por los sujetos de HGDP-CEPH (938 sujetos) y la inferida para ellos al utilizar únicamente la información de uno de los paneles de 12 AIM desarrollado (panel mínimo con suplemento nº 1). La intensidad del entramado es indicativa de la proporción de sujetos que son clasificados correcta (diagonal) e incorrectamente (resto de sectores) en las siete subdivisiones biogeográficas. En la diagonal también se expresa numéricamente la proporción de aciertos en cada subdivisión. Nótese que los aciertos en la clasificación de sujetos de origen Eurasiático y del Este Asiático mejoran considerablemente con respecto a las del uso del panel mínimo de 7 AIM. Clave: EUR, Europa; AFR, África sub-sahariana; MEA, Oriente Próximo y Norte de África; CSA, Centro y Sur de Asia; EAS, Este de Asia; AME, Américas; OCE, Oceanía.

**Figura 3.** Matriz de confusión entre la ascendencia declarada por los sujetos del Proyecto 1000 Genomas (2000 sujetos) y la inferida para ellos al utilizar únicamente la información de uno de los paneles de 12 AIM desarrollado (panel mínimo con suplemento nº 1). La intensidad del entramado es indicativa de la proporción de sujetos que son clasificados correcta (diagonal) e incorrectamente (resto de sectores) en las siete subdivisiones biogeográficas. En la diagonal también se expresa numéricamente la proporción de aciertos en cada subdivisión. Clave: EUR, Europa; AFR, África sub-sahariana; CSA, Centro y Sur de Asia; EAS, Este de Asia.

## DESCRIPCIÓN DETALLADA

### Método para determinar la ascendencia geográfica de un sujeto

En un primer aspecto, la invención se relaciona con un método para determinar la ascendencia geográfica de un sujeto, en adelante “método de la invención”, que comprende

- i) determinar en una muestra que comprende material genético de dicho sujeto el genotipo de los dos alelos de los polimorfismos de un solo nucleótido (SNP) rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622 y, opcionalmente, cualquier otro

polimorfismo que se encuentre en desequilibrio de ligamiento con cualquiera de los anteriores, en ambos alelos y

- ii) determinar la ascendencia geográfica de dicho sujeto en base a los genotipos de dichos SNP.

5 El término “método para determinar la ascendencia geográfica de un sujeto”, tal y como aquí se utiliza, se refiere a un método para inferir, con una determinada probabilidad, la región geográfica de la que son originarios los ancestros de un sujeto. Dicha región geográfica puede ser:

- 10 - África sub-sahariana: el término “África sub-sahariana”, tal y como aquí se utiliza, se refiere a la región del continente africano ubicada al sur del desierto del Sáhara, e incluye las regiones correspondientes a los países del continente africano que no limitan con el mar Mediterráneo.
- 15 - Oriente Próximo y Norte de África: el término “Oriente Próximo”, tal y como aquí se utiliza, se refiere a las regiones correspondientes con los países Arabia Saudí, Baréin, Chipre, Emiratos Árabes Unidos, Irak, Irán, Israel, Palestina, Jordania, Kuwait, Líbano, Omán, Catar, Siria, Turquía, Yemen, Georgia, Armenia y Azerbaiyán. El término “Norte de África”, tal y como aquí se utiliza, se refiere a la región geográfica más boreal del continente africano, que limita al norte con el mar Mediterráneo, al este con el mar Rojo y el Sinaí, al oeste con el océano Atlántico y al sur con la región del desierto del Sahara, e incluye las regiones correspondientes a los países Argelia, Egipto, Libia, Marruecos y Túnez.
- 20 - Centro y Sur de Asia: el término “centro de Asia”, tal y como aquí se utiliza, se refiere a una región del continente asiático comprendida entre el mar Caspio, la frontera de China y la de Rusia y el Sur de Asia. Incluye las regiones correspondientes a los países Kazajistán, Kirguistán, Tayikistán, Turkmenistán y Uzbekistán. El término “sur de Asia”, tal y como aquí se utiliza, se refiere a la región de Asia que incluye la meseta iraní y el subcontinente de la India incluyendo las regiones correspondientes a los países Afganistán, Bangladés, Bután, India, Maldivas, Nepal, Pakistán y Sri Lanka.
- 25 - Este de Asia: el término “este de Asia”, tal y como aquí se utiliza, se refiere a las regiones del continente asiático correspondiente a los países China, Corea del Norte, Corea del Sur, Japón, Mongolia, Taiwán y las dependencias de Hong Kong y Macao.
- 30 - América: el término “América”, tal y como aquí se utiliza, se refiere al continente americano.
- 35

- Oceanía: el término "Oceanía", tal y como aquí se utiliza, se refiere al continente del mismo nombre.

El experto en la materia entenderá que la determinación del método de la invención se hace con una determinada probabilidad y, por lo tanto, no tiene que ser  
5 correcta para el 100% de los sujetos evaluados, aunque sí para una parte estadísticamente significativa de los sujetos. Es decir, el método de la invención permite determinar correctamente la ascendencia geográfica en una parte estadísticamente significativa de los sujetos. Si una parte es estadísticamente significativa, se puede determinar sin más trámite por la persona experta en la técnica  
10 utilizando varias herramientas estadísticas bien conocidas de evaluación, por ejemplo, la determinación de intervalos de confianza, determinación de valor de p, validación cruzada con índices de clasificación, etc (más detalles en "Estadística para la investigación" Dowdy y Wearden, John Wiley & Sons, New York, 1983). Los intervalos de confianza preferidos son por lo menos 50%, al menos 60%, al menos 70%, al  
15 menos 80%, al menos 90% o al menos 95%. Los valores de p son, preferiblemente, 0,01, 0.005 o inferior.

El término "sujeto", como se usa aquí, se refiere a un ser humano de cualquier edad o sexo. En una realización particular, el sujeto no pertenece a las poblaciones de Japón, Kalash o Hazaras, ni a poblaciones procedentes de cruzamientos históricos (tal  
20 como, por ejemplo, las del Caribe o afroamericanas).

La primera etapa del método de la invención comprende determinar en una muestra que comprende material genético de dicho sujeto el genotipo de los dos alelos de los polimorfismos de un solo nucleótido (SNP) rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622 y, opcionalmente, cualquier otro  
25 polimorfismo que se encuentre en desequilibrio de ligamiento con cualquiera de los anteriores, en ambos alelos.

El término "polimorfismo de un solo nucleótido" o "SNP" (del inglés "single nucleotide polymorphism") o simplemente "polimorfismo", tal y como se usa aquí, se refiere a una variación en la secuencia de nucleótidos de un ácido nucleico que se  
30 produce en un solo nucleótido (A, C, T o G), donde cada posible secuencia está presente en una proporción igual o mayor que un 1% de la población. Los SNP se denominan habitualmente según el número de registro en la base de datos SNP (dbSNP) en el Centro Nacional de Información Biotecnológica (NCBI, accesible en <http://www.ncbi.nlm.nih.gov/projects/SNP/>). En general, los SNP representan una de las  
35 formas más comunes de variaciones genéticas. Estos polimorfismos aparecen cuando un solo nucleótido en el genoma se altera (por ejemplo, por medio de sustitución,

adición o supresión). Cada versión de la secuencia con respecto al sitio polimórfico se refiere como un alelo del sitio polimórfico. Los SNP tienden a ser estables evolutivamente de generación en generación y, como tal, se puede utilizar para estudiar anomalías genéticas específicas en una población o relaciones entre poblaciones. Si los SNP ocurren en la región codificante de la proteína, puede conducir a la expresión de una variante, a veces defectuosa, de la proteína que puede conducir al desarrollo de una enfermedad genética. Algunos SNP se pueden producir en regiones no codificantes, lo que puede provocar alteraciones en el procesamiento del material genético, o alteraciones en los niveles de expresión de proteínas. Cada versión de la secuencia con respecto al SNP se conoce como un alelo del SNP.

La primera etapa del método de la invención comprende determinar el genotipo de los dos alelos de los siguientes SNP:

- rs1257016: localizado en el gen *FAM178B*, corresponde a la secuencia TCAATATATAATCTTTGTCTCTTGT[A/G]CTTTTTGACTTAAAGTTTATTT, con SEQ ID NO: 1
- rs971779: localizado en el gen *RPSAP52*, corresponde a la secuencia CACTGCCATCACAAAGAAAACATA[C/T]GTAAATTTCTTCAAGGCTACAA, con SEQ ID NO: 2
- rs238547: localizado en el gen *SCNN1B*, corresponde a la secuencia TAGGCTTCAAGACCATGGACTTCCC[C/T]GCCGTCACCATCTGCAATGCTA, con SEQ ID NO: 3.
- rs1482326: está localizado en el gen *ADGRB3* (también conocido como *BAI3*), corresponde a la secuencia GCTATGTTTCAAAGGAATTGAAGA[A/G]CAACATCTCCATAGAATACAGT, con SEQ ID NO: 4.
- rs4760332: localizado en el gen *CTDSP2*, corresponde a la secuencia TACCCACAAAATCCCAAATGCCTTG[A/C]ATGAGTATGGGAGAGAACTGCT, con SEQ ID NO: 5.
- rs2104388: localizado en el gen *KLF12*, corresponde a la secuencia ACAGCTGGGAAATATGTGTATGTCT[A/G]TGTGGGGAAGTTTAACTTTCC, con SEQ ID NO: 6.
- rs2427622: localizado en el gen *MYT1*, corresponde a la secuencia ACTTTGGGAGGCCAAGGCGGGCGGA[C/T]CACAAGGTCAGGAGATCGAGAC, con SEQ ID NO: 7.

El término "alelo", tal y como aquí se utiliza, se refiere a uno de dos o más formas de un gen, locus o polimorfismo genético. A veces los diferentes alelos pueden

resultar en diferentes características; sin embargo, otras veces alelos diferentes tendrán el mismo resultado en la expresión de un gen. Los organismos diploides tienen una copia de cada gen (y un alelo) en cada cromosoma del par de cromosomas homólogos.

5 El término “muestra que comprende material genético”, tal y como aquí se utiliza, se refiere a materia biológica aislada de un sujeto que contiene cualquier material biológico adecuado para detectar el SNP deseado. Dicha muestra puede comprender células y/o material no celular del sujeto. En la presente invención, la muestra comprende material genético, por ejemplo, ADN, ADN genómico (ADNg),  
 10 ADN complementario (ADNc), ARN, ARN nuclear heterogéneo (ARNhn), mRNA, etc., del sujeto. La muestra se puede aislar de cualquier tejido o fluido biológico adecuado tal como, por ejemplo sangre, saliva, plasma, suero, orina, líquido cefalorraquídeo (LCR), heces, un hisopo bucal o bucal-faríngeo, una muestra quirúrgica, y un espécimen obtenido a partir de una biopsia. Los métodos para aislar células y  
 15 muestras de tejidos son bien conocidos para los expertos en la técnica. En una realización particular, la muestra se selecciona del grupo que consiste en sangre, suero, plasma, células de la piel, saliva, folículos pilosos o biopsias de tejidos.

La primera etapa del método de la invención opcionalmente comprende la determinación de cualquier polimorfismo que se encuentre en desequilibrio de  
 20 ligamiento con cualquiera de los SNP rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622. Dos o más SNP se encuentran en “desequilibrio de ligamiento” si los genotipos de dichos SNP están estadísticamente asociados. El desequilibrio de ligamiento entre dos o más SNP se puede valorar mediante varias medidas estadísticas, tales como  $D'$ ,  $r^2$ , etc.) (Véase, por ejemplo, Devlin y Risch 1995  
 25 Sep 20; 29(2):311-22). El experto en la materia puede obtener información sobre los SNP que se encuentran en desequilibrio de ligamiento con un SNP determinado utilizando bases de datos apropiadas, por ejemplo, la base de datos International HapMap Project ([www.hapmap.org](http://www.hapmap.org) y <http://hapmap.ncbi.nlm.nih.gov/>) y/o programas apropiados tales como el programa Haploview utilizando el método de intervalos de  
 30 confianza o algoritmo de Gabriel (Gabriel *et al.*, Science, 2002, 296(5576):225-9) o cualquier otra versión más actualizada (Barrett *et al.*, 2005, Bioinformatics 21(2):263-265).

En una realización particular, los SNP que se encuentran en desequilibrio de ligamiento con el SNP rs1257016 se localizan en la región comprendida entre el  
 35 nucleótido 97528365 y el nucleótido 97890168 del cromosoma 2 según las coordenadas de la versión del genoma humano hg19.

En una realización particular, los SNP que se encuentran en desequilibrio de ligamiento con el SNP rs971779 se localizan en la región comprendida entre el nucleótido 66164517 y el nucleótido 66277698 del cromosoma 12 según las coordenadas de la versión del genoma humano hg19.

5 En una realización particular, los SNP que se encuentran en desequilibrio de ligamiento con el SNP rs238547 se localizan en la región comprendida entre el nucleótido 23356312 y el nucleótido 23369090 del cromosoma 16 según las coordenadas de la versión del genoma humano hg19.

10 En una realización particular, los SNP que se encuentran en desequilibrio de ligamiento con el SNP rs1482326 se localizan en la región comprendida entre el nucleótido 69794071 y el nucleótido 69794071 del cromosoma 6 según las coordenadas de la versión del genoma humano hg19.

15 En una realización particular, los SNP que se encuentran en desequilibrio de ligamiento con el SNP rs4760332 se localizan en la región comprendida entre el nucleótido 58219173 y el nucleótido 58337671 del cromosoma 12 según las coordenadas de la versión del genoma humano hg19.

20 En una realización particular, los SNP que se encuentran en desequilibrio de ligamiento con el SNP rs2104388 se localizan en la región comprendida entre el nucleótido 74674871 y el nucleótido 74710374 del cromosoma 13 según las coordenadas de la versión del genoma humano hg19.

25 En una realización particular, los SNP que se encuentran en desequilibrio de ligamiento con el SNP rs2427622 se localizan en la región comprendida entre el nucleótido 62829408 y el nucleótido 62911117 del cromosoma 20 según las coordenadas de la versión del genoma humano hg19.

30 En una realización particular, el método de la invención permite determinar, con una elevada probabilidad, la ascendencia geográfica del sujeto en función del genotipo presente en cada uno de los polimorfismos rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622. El término “genotipo de los dos alelos” o simplemente “genotipo” aplicado a un polimorfismo o SNP, tal y como aquí se utiliza, se refiere a los alelos concretos presentes en los dos cromosomas homólogos para la posición polimórfica de dicho SNP.

Así, en una realización particular,

- el genotipo GG en el polimorfismo rs1257016,
- el genotipo CC en el polimorfismo rs971779,
- 35 - el genotipo TC en el polimorfismo rs238547,
- el genotipo AG en el polimorfismo rs1482326,

- el genotipo CC en el polimorfismo rs4760332,
- el genotipo GG en el polimorfismo rs2104388 y
- el genotipo AG en el polimorfismo rs2427622

5 indican que existe una elevada probabilidad de que la ascendencia geográfica del sujeto es europea.

En otra realización particular,

- el genotipo GG en el polimorfismo rs1257016,
- el genotipo TT en el polimorfismo rs971779,
- el genotipo CC en el polimorfismo rs238547,
- 10 - el genotipo GG en el polimorfismo rs1482326,
- el genotipo CC en el polimorfismo rs4760332,
- el genotipo AA en el polimorfismo rs2104388 y
- el genotipo GG en el polimorfismo rs2427622

15 indican que existe una elevada probabilidad de que la ascendencia geográfica es África sub-sahariana,

En otra realización particular,

- el genotipo GG en el polimorfismo rs1257016,
- el genotipo CC en el polimorfismo rs971779,
- el genotipo CC en el polimorfismo rs238547,
- 20 - el genotipo GG en el polimorfismo rs1482326,
- el genotipo CC en el polimorfismo rs4760332,
- el genotipo AG en el polimorfismo rs2104388 y
- el genotipo GG en el polimorfismo rs2427622

25 indican que existe una elevada probabilidad de que la ascendencia geográfica es Oriente Próximo y Norte de África.

En otra realización particular,

- el genotipo AG en el polimorfismo rs1257016,
- el genotipo CC en el polimorfismo rs971779,
- el genotipo CC en el polimorfismo rs238547,
- 30 - el genotipo AG en el polimorfismo rs1482326,
- el genotipo AC en el polimorfismo rs4760332,
- el genotipo AG en el polimorfismo rs2104388 y
- el genotipo AG en el polimorfismo rs2427622

35 indican que existe una elevada probabilidad de que la ascendencia geográfica es Centro y Sur de Asia.

En otra realización particular,

- el genotipo AG en el polimorfismo rs1257016,
- el genotipo CC en el polimorfismo rs971779,
- el genotipo CC en el polimorfismo rs238547,
- el genotipo AA en el polimorfismo rs1482326,
- 5 - el genotipo AC en el polimorfismo rs4760332, la el genotipo GG en el polimorfismo rs2104388 y
- el genotipo AA en el polimorfismo rs2427622

indican que existe una elevada probabilidad de que la ascendencia geográfica es el este de Asia.

- 10 En otra realización particular,
- el genotipo AA en el polimorfismo rs1257016,
  - el genotipo CC en el polimorfismo rs971779,
  - el genotipo TT en el polimorfismo rs238547,
  - el genotipo AG en el polimorfismo rs1482326,
  - 15 - el genotipo AA en el polimorfismo rs4760332,
  - el genotipo GG en el polimorfismo rs2104388 y
  - el genotipo AA en el polimorfismo rs2427622

indican que existe una elevada probabilidad de que la ascendencia geográfica es América.

- 20 En otra realización particular,
- el genotipo AA en el polimorfismo rs1257016,
  - el genotipo TT en el polimorfismo rs971779,
  - el genotipo CC en el polimorfismo rs238547,
  - el genotipo AA en el polimorfismo rs1482326,
  - 25 - el genotipo CC en el polimorfismo rs4760332,
  - el genotipo AA en el polimorfismo rs2104388 y
  - el genotipo AA en el polimorfismo rs2427622

indican que existe una elevada probabilidad de que la ascendencia geográfica es Oceanía.

- 30 El término “elevada probabilidad”, tal y como aquí se utiliza, significa una probabilidad de al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90%, al menos un 95%, al menos un 97%, al menos un 99% o un 100. En una realización particular, elevada probabilidad es de al menos un 70%, preferiblemente al menos un 75%.

- 35 La proporción de aciertos de la clasificación del método de la invención se puede incrementar añadiendo al panel de los 7 SNP descritos combinaciones de 5 SNP

adicionales, tal y como se muestra en el las Figuras 2 y 3. Así, en una realización particular, el método de la invención comprende adicionalmente la determinación de la secuencia de uno de los siguientes grupos de 5 SNP:

(i) El grupo formado por los polimorfismos:

- 5                   - rs186471, que corresponde a la secuencia  
TATTGATTGGGTGGAATATAAATGT[A/G]GGTATGTGTGTGTGTATA  
AGAC, de SEQ ID NO: 8
- rs7919248, que corresponde a la secuencia  
GCCAAAGAAGAGGCGATTACAGCA[G/T]AAATGAATTTGATAAAA  
10               ATAAA, de SEQ ID NO: 9
- rs2231997, localizado en el gen *DDX49* y que corresponde a la  
secuencia  
ACCACCTGCCAGCCCTGCCTCTCA[C/T]GCTCTGTCCCCCAGAC  
ATGGTG, de SEQ ID NO: 10,
- 15               - rs581468, localizado en el gen *ARHGAP42*, y que corresponde a la  
secuencia  
CAGAAGCTAAACCACCATTCTTTCC[A/G]ACACTGTCTCAGAGACT  
TGGAA, de SEQ ID NO: 11; y
- rs1609997, que corresponde a la secuencia  
20               GCCCTCTTTTAAAGTAACTTTCCTG[A/G]AGACACAACCAAATTCC  
ACTT, de SEQ ID NO: 12.

En una realización particular, el genotipo GG en el polimorfismo rs186471, el genotipo TT en el polimorfismo rs7919248, el genotipo CC en el polimorfismo rs2231997, el genotipo AG en el polimorfismo rs581468 y el genotipo GG en el polimorfismo rs1609997, indican que existe una elevada probabilidad de que la ascendencia geográfica sea europea.

En una realización particular, el genotipo GG en el polimorfismo rs186471, el genotipo TT en el polimorfismo rs7919248, el genotipo CC en el polimorfismo rs2231997, el genotipo AA en el polimorfismo rs581468 y el genotipo AA en el polimorfismo rs1609997, indican que existe una elevada probabilidad de que la ascendencia geográfica sea África sub-sahariana.

En una realización particular, el genotipo GG en el polimorfismo rs186471, el genotipo TT en el polimorfismo rs7919248, el genotipo CC en el polimorfismo rs2231997, el genotipo AA en el polimorfismo rs581468 y el genotipo AG en el polimorfismo rs1609997, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oriente Próximo y Norte de África.

En una realización particular, el genotipo GG en el polimorfismo rs186471, el genotipo TT en el polimorfismo rs7919248, el genotipo CC en el polimorfismo rs2231997, el genotipo AG en el polimorfismo rs581468 y el genotipo AG en el polimorfismo rs1609997, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Centro y Sur de Asia.

En una realización particular, el genotipo AA en el polimorfismo rs186471, el genotipo TG en el polimorfismo rs7919248, el genotipo TC en el polimorfismo rs2231997, el genotipo GG en el polimorfismo rs581468 y el genotipo AA en el polimorfismo rs1609997, indican que existe una elevada probabilidad de que la ascendencia geográfica sea este de Asia.

En una realización particular, el genotipo AG en el polimorfismo rs186471, el genotipo GG en el polimorfismo rs7919248, el genotipo CC en el polimorfismo rs2231997, el genotipo GG en el polimorfismo rs581468 y el genotipo AA en el polimorfismo rs1609997, indican que existe una elevada probabilidad de que la ascendencia geográfica sea América.

En una realización particular, el genotipo AA en el polimorfismo rs186471, el genotipo TG en el polimorfismo rs7919248, el genotipo TT en el polimorfismo rs2231997, el genotipo GG en el polimorfismo rs581468 y el genotipo AA en el polimorfismo rs1609997, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oceanía.

(ii) El grupo formado por los polimorfismos:

- rs1765857, localizado en el gen *LOC105370158* y que corresponde a la secuencia TCTTATGAAAGATGTGCCAATTAAT[A/G]CGATGCTGCATTTTCTTA CTATTCC, de SEQ ID NO: 13,
- rs10497520, localizado en el gen *TTN* y que corresponde a la secuencia AATCCAGGTGCTGTTTCTCCAATT[C/T]AGGTTCTTGAACAAATGC AGTC, de SEQ ID NO: 14,
- rs9552445, localizado en el gen *MICU2* y que corresponde a la secuencia TACAAATAACAAGGTAGTTCTACA[A/G]CAGCATTCTAGGCATTTG CCCT, de SEQ ID NO: 15,
- rs12880675, que corresponde a la secuencia GAACTAATTATTTTATTGATTCAAG[C/G/T]AGCAATGGTTCTAAATC AGA, de SEQ ID NO: 16 y

- rs350886, localizado en el gen *MAK2K2*, y que corresponde a la secuencia

GAGCCCCAGGCGATGGCGGCTCTC[A/G]CCAAAAGGAAGGAGA  
GTGAGGC, de SEQ ID NO: 17.

- 5 En una realización particular, el genotipo AA en el polimorfismo rs1765857, el genotipo CC en el polimorfismo rs10497520, el genotipo AA en el polimorfismo rs9552445, el genotipo CC en el polimorfismo rs12880675 y el genotipo GG en el polimorfismo rs350886, indican que existe una elevada probabilidad de que la ascendencia geográfica sea europea.
- 10 En una realización particular, el genotipo AA en el polimorfismo rs1765857, el genotipo TT en el polimorfismo rs10497520, el genotipo AA en el polimorfismo rs9552445, el genotipo CC en el polimorfismo rs12880675 y el genotipo AA en el polimorfismo rs350886, indican que existe una elevada probabilidad de que la ascendencia geográfica sea África sub-sahariana.
- 15 En una realización particular, el genotipo AA en el polimorfismo rs1765857, el genotipo CC en el polimorfismo rs10497520, el genotipo AA en el polimorfismo rs9552445, el genotipo CC en el polimorfismo rs12880675 y el genotipo AG en el polimorfismo rs350886, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oriente Próximo y Norte de África.
- 20 En una realización particular, el genotipo AA en el polimorfismo rs1765857, el genotipo CC en el polimorfismo rs10497520, el genotipo AA o AG en el polimorfismo rs9552445, el genotipo CC en el polimorfismo rs12880675 y el genotipo GG en el polimorfismo rs350886, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Centro y Sur de Asia.
- 25 En una realización particular, el genotipo GG en el polimorfismo rs1765857, el genotipo TT en el polimorfismo rs10497520, el genotipo GG en el polimorfismo rs9552445, el genotipo TT en el polimorfismo rs12880675 y el genotipo GG en el polimorfismo rs350886, indican que existe una elevada probabilidad de que la ascendencia geográfica sea este de Asia.
- 30 En una realización particular, el genotipo GG en el polimorfismo rs1765857, el genotipo TT en el polimorfismo rs10497520, el genotipo AG en el polimorfismo rs9552445, el genotipo CC en el polimorfismo rs12880675 y el genotipo GG en el polimorfismo rs350886, indican que existe una elevada probabilidad de que la ascendencia geográfica sea América.
- 35 En una realización particular, el genotipo AG en el polimorfismo rs1765857, el genotipo TT en el polimorfismo rs10497520, el genotipo GG en el polimorfismo



En una realización particular, el genotipo TT en el polimorfismo rs2384319, el genotipo TT en el polimorfismo rs7919248, el genotipo GG en el polimorfismo rs2389736, el genotipo AG en el polimorfismo rs2166624 y el genotipo TC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Centro y Sur de Asia.

En una realización particular, el genotipo GG en el polimorfismo rs2384319, el genotipo TG en el polimorfismo rs7919248, el genotipo AG en el polimorfismo rs2389736, el genotipo AG en el polimorfismo rs2166624 y el genotipo CC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la ascendencia geográfica sea este de Asia.

En una realización particular, el genotipo GG en el polimorfismo rs2384319, el genotipo GG en el polimorfismo rs7919248, el genotipo AA en el polimorfismo rs2389736, el genotipo AA en el polimorfismo rs2166624 y el genotipo TC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la ascendencia geográfica sea América.

En una realización particular, el genotipo TT en el polimorfismo rs2384319, el genotipo TG en el polimorfismo rs7919248, el genotipo AA en el polimorfismo rs2389736, el genotipo GG en el polimorfismo rs2166624 y el genotipo CC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oceanía.

(iv) El grupo formado por los polimorfismos:

- rs6542787, localizado en el gen *EDAR* y que corresponde a la secuencia

GAGAACAGAAAGTTGGCACCTTAAA[A/G]TATTTGTGGTGTGATGG  
AGACG de SEQ ID NO: 22,

- rs7919248, que se ha definido previamente,
- rs261532, localizado en el gen *UBE2D2* y que corresponde a la secuencia

GTCTGAATTGAGAACCTGGATTTCT[G/T]GACTGTCCAGTGTATTTT  
CTGG, de SEQ ID NO: 23,

- rs5996039, localizado en el gen *PMM1* y que corresponde a la secuencia

GTGCTCAAGTGTGGTCAATCAGT[A/G]AGGGAACTATCTGGCTG  
CTCTC de SEQ ID NO: 24, y

- rs2268969, que se ha definido previamente.

En una realización particular, el genotipo AA en el polimorfismo rs6542787, el genotipo TT en el polimorfismo rs7919248, el genotipo GG en el polimorfismo rs261532, el genotipo AA en el polimorfismo rs5996039 y el genotipo TT en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la  
5 ascendencia geográfica sea europea.

En una realización particular el genotipo AG en el polimorfismo rs6542787, el genotipo TT en el polimorfismo rs7919248, el genotipo TT en el polimorfismo rs261532, el genotipo GG en el polimorfismo rs5996039 y el genotipo TT en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la  
10 ascendencia geográfica sea África sub-sahariana.

En una realización particular, el genotipo AA en el polimorfismo rs6542787, el genotipo TT en el polimorfismo rs7919248, el genotipo TG en el polimorfismo rs261532, el genotipo AA en el polimorfismo rs5996039 y el genotipo TT en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la  
15 ascendencia geográfica sea Oriente Próximo y Norte de África.

En una realización particular, el genotipo AA en el polimorfismo rs6542787, el genotipo TT en el polimorfismo rs7919248, el genotipo GG en el polimorfismo rs261532, el genotipo AG en el polimorfismo rs5996039 y el genotipo TC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la  
20 ascendencia geográfica sea Centro y Sur de Asia.

En una realización particular, el genotipo GG en el polimorfismo rs6542787, el genotipo TG en el polimorfismo rs7919248, el genotipo GG en el polimorfismo rs261532, el genotipo AA en el polimorfismo rs5996039 y el genotipo CC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la  
25 ascendencia geográfica sea este de Asia.

En una realización particular, el genotipo GG en el polimorfismo rs6542787, el genotipo GG en el polimorfismo rs7919248, el genotipo GG en el polimorfismo rs261532, el genotipo GG en el polimorfismo rs5996039 y el genotipo TC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la  
30 ascendencia geográfica sea América.

En una realización particular, el genotipo GG en el polimorfismo rs6542787, el genotipo TG en el polimorfismo rs7919248, el genotipo TT en el polimorfismo rs261532, el genotipo AA en el polimorfismo rs5996039 y el genotipo CC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la  
35 ascendencia geográfica sea Oceanía.

(v) El grupo formado por los polimorfismos:



polimorfismo rs2268969, indican que existe una elevada probabilidad de que la ascendencia geográfica sea América.

En una realización particular, el genotipo GG en el polimorfismo rs6542787, el genotipo TG en el polimorfismo rs7919248, el genotipo TT en el polimorfismo rs261532, el genotipo TT en el polimorfismo rs8139993 y el genotipo CC en el polimorfismo rs2268969, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oceanía.

(vi) El grupo formado por los polimorfismos:

- rs3784651, localizado en el gen *MCTP2* y que corresponde con la secuencia  
 CCCAGGCCTGAAAGGCCAGAACTCA[C/T]AAAGGGAATACTGACTT  
 TGTAG de SEQ ID NO: 26,
- rs10497520, que se ha definido previamente,
- rs261532, que se ha definido previamente,
- rs12880675, que se ha definido previamente y
- rs4742634, localizado en el gen *PTPRD* y que corresponde a la secuencia  
 GAATTTGAGTTCACCATATCTACAT[C/T]GCTGGCAGTAAGACATA  
 GAAAA, de SEQ ID NO: 27.

En una realización particular, el genotipo TT en el polimorfismo rs3784651, el genotipo CC en el polimorfismo rs10497520, el genotipo GG en el polimorfismo rs261532, el genotipo CC en el polimorfismo rs12880675 y el genotipo CC en el polimorfismo rs4742634, indican que existe una elevada probabilidad de que la ascendencia geográfica sea europea.

En una realización particular, el genotipo CC en el polimorfismo rs3784651, el genotipo TT en el polimorfismo rs10497520, el genotipo TT en el polimorfismo rs261532, el genotipo CC en el polimorfismo rs12880675 y el genotipo TT en el polimorfismo rs4742634, indican que existe una elevada probabilidad de que la ascendencia geográfica sea África sub-sahariana.

En una realización particular, el genotipo TT en el polimorfismo rs3784651, el genotipo CC en el polimorfismo rs10497520, el genotipo TG en el polimorfismo rs261532, el genotipo CC en el polimorfismo rs12880675 y el genotipo TC en el polimorfismo rs4742634, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oriente Próximo y Norte de África.

En una realización particular, el genotipo TC en el polimorfismo rs3784651, el genotipo CC en el polimorfismo rs10497520, el genotipo GG en el polimorfismo

rs261532, el genotipo CC en el polimorfismo rs12880675 y el genotipo TC en el polimorfismo rs4742634, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Centro y Sur de Asia.

En una realización particular, el genotipo CC en el polimorfismo rs3784651, el  
5 genotipo TT en el polimorfismo rs10497520, el genotipo GG en el polimorfismo rs261532, el genotipo TT en el polimorfismo rs12880675 y el genotipo TT en el polimorfismo rs4742634, indican que existe una elevada probabilidad de que la ascendencia geográfica sea este de Asia.

En una realización particular, el genotipo TT en el polimorfismo rs3784651, el  
10 genotipo TT en el polimorfismo rs10497520, el genotipo GG en el polimorfismo rs261532, el genotipo CC en el polimorfismo rs12880675 y el genotipo TT en el polimorfismo rs4742634, indican que existe una elevada probabilidad de que la ascendencia geográfica sea América.

En una realización particular, el genotipo CC en el polimorfismo rs3784651, el  
15 genotipo TT en el polimorfismo rs10497520, el genotipo TT en el polimorfismo rs261532, el genotipo TT en el polimorfismo rs12880675 y el genotipo TT en el polimorfismo rs4742634, indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oceanía.

La determinación del genotipo de los polimorfismos del método de la invención  
20 se lleva a cabo en ambos alelos por medio de diversos procesos conocidos por una persona experta en la materia. En algunas realizaciones, por ejemplo, cuando la determinación de la secuencia de los SNP se lleva a cabo en una muestra de sangre completa, dicha muestra puede utilizar directamente para la detección de dicho SNP. En otras realizaciones, se extrae el material genético de las células que están  
25 presentes en un fluido biológico (por ejemplo, sangre periférica, saliva, fluido sinovial, etc.) como una etapa inicial, y, en tales casos, el material genético total que se extrae a partir de dichas muestras representa el material de trabajo adecuado para la posterior amplificación. El aislamiento del material genético de la muestra se puede realizar por métodos conocidos por la persona experta en la técnica. Dichos métodos  
30 se pueden encontrar, por ejemplo, en Sambrook et al., 2001. "Molecular cloning: a Laboratory Manual", 3rd ed., Cold Spring Harbor Laboratory Press, N.Y., Vol. 1-3. Además, en algunas realizaciones, la generación de ácidos nucleicos para el análisis de muestras requiere amplificación de ácido nucleico. Muchos métodos de amplificación se basan en una reacción en cadena enzimática tal como, por ejemplo,  
35 una reacción en cadena de la polimerasa (PCR), una reacción en cadena de la ligasa (LCR), o una replicación de secuencia autosostenida, amplificación circular, etc.; esta

lista es meramente ilustrativa y en modo alguno limitativa. Métodos de amplificación de ácidos nucleicos se describen en Sambrook et al., 2001 (citado en supra).

Después de aislar y amplificar (si es necesario) el ácido nucleico, se detectan las secuencias de los diferentes SNP del método de la invención. La determinación de las secuencias de los SNP según el método de la invención se puede llevar a cabo mediante cualquier método adecuado, los cuales son conocidos por el experto en la materia, como por ejemplo mediante la realización de secuenciación, mini-secuenciación, hibridación, análisis de fragmentos de restricción, ensayo de ligación de oligonucleótidos, PCR específica de alelo, o una combinación de los mismos. Como tal, los sistemas y métodos para la detección de SNP, en general incluyen, pero no se limitan a, secuenciación de ácidos nucleicos, métodos de hibridación y la tecnología de matriz (por ejemplo, la tecnología disponible de Aclara BioSciences, Affymetrix, Agilent Technologies, Illumina Inc., etc); también se pueden utilizar las técnicas basadas en el cambio de movilidad de los fragmentos de ácido nucleico amplificados, como por ejemplo Single Stranded Conformational Polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE), Chemical Mismatch Cleavage (CMC), Restriction Fragment Polymorphisms (RFLPs), WAVE analysis y similares (Methods Mol. Med. 2004; 108: 173-88). Por supuesto, esta lista es meramente ilustrativa y en modo alguno limitativa. Como es evidente en la técnica, la secuencia de dichos SNP puede ser determinada a partir de cualquiera de las cadenas de ácido nucleico o de ambas cadenas.

En una realización particular, la determinación de la secuencia de los SNP según el método de la invención se lleva a cabo por PCR en tiempo real.

En una realización particular, el método de la invención comprende, para cada uno de los SNP, la puesta en contacto de la muestra que contiene material genético con un ácido nucleico que hibrida de forma selectiva con una región genómica que comprende dicho SNP. En una realización más particular, la muestra que contiene el material genético se somete a un proceso de amplificación de ácidos nucleicos previamente a la puesta en contacto con el ácido nucleico que hibrida de forma selectiva con una región genómica que comprende el polimorfismo, de forma que dicha amplificación se lleve a cabo sobre las regiones genómicas que contienen los polimorfismos a detectar. En una realización aún más particular, el ácido nucleico es un cebador y el método comprende la extensión del cebador usando como la región genómica que comprende el polimorfismo. El término "cebador", tal y como aquí se utiliza, se refiere a un oligonucleótido que actúa para iniciar la síntesis de una hebra de ácido nucleico complementaria. La extensión se puede llevar a cabo mediante una

reacción uniplex, es decir, llevando a cabo la extensión independiente de cada cebador para cada uno de los SNP que se van a detectar, o mediante una reacción multiplex, es decir, llevando a cabo la extensión simultánea de varios cebadores, correspondientes a varios SNP. En una realización preferida, la extensión del cebador es una reacción de extensión multiplex.

En una realización particular, la etapa de determinar la ascendencia geográfica del sujeto comprende calcular para cada división geográfica la suma de los logaritmos de las probabilidades de asignación a cada división geográfica en base al genotipo de los dos alelos determinado en cada SNP, en donde existe una elevada probabilidad de que las ascendencia geográfica de dicho sujeto sea aquella para la cual dicha suma de logaritmos es mayor, en donde las divisiones geográficas son Europa, África subsahariana, Oriente Próximo y Norte de África, Centro-Sur de Asia, Este de Asia, América y Oceanía, y en donde las probabilidades de asignación de una muestra a cada división geográfica para cada genotipo de cada SNP son las frecuencias de dichos genotipos en una población control.

El término "población control", tal y como aquí se utiliza, se refiere a un grupo de sujetos que comparten una ascendencia geográfica conocida. La población control preferiblemente está formada por al menos 25 sujetos. En una realización particular, la población control se obtiene de los individuos que conforman el Human Genome Diversity Panel (HGDP) (Li *et al.* Science 319, 2008: 1100-4). En una realización más particular, las probabilidades de asignación de una muestra a cada división geográfica para cada genotipo de cada SNP son las que figuran en la Tabla 1.

En una realización particular, la confianza de dicha asignación se calcula de la forma siguiente:

$$\lambda = BestFact * v1^{st} + (1.0 - BestFact) * (v1^{st} - v2^{nd})$$

en donde

BestFact tiene un valor de 0,1

$v1^{st}$  es la mayor suma de los logaritmos de probabilidades de asignación de la muestra del sujeto a cada una de las divisiones geográficas en base al genotipo de cada SNP

$v2^{nd}$  es la segunda mayor suma de los logaritmos de probabilidades de asignación de la muestra del sujeto a cada una de las divisiones geográficas en base al genotipo de cada SNP.

		rs1257016				rs971779				rs238547				rs1482326			
Población	AA	AG	GG	TT	TC	CC	TT	TC	CC	TT	TC	CC	AA	AG	GG		
EUR	0,013	0,144	0,844	0,013	0,231	0,756	0,013	0,231	0,756	0,194	0,494	0,313	0,088	0,463	0,450		
AFR	0,174	0,355	0,471	0,959	0,041	0,000	0,959	0,041	0,000	0,000	0,017	0,983	0,000	0,058	0,942		
MEA	0,045	0,307	0,648	0,136	0,409	0,455	0,136	0,409	0,455	0,114	0,381	0,506	0,045	0,256	0,699		
CSA	0,247	0,430	0,323	0,158	0,386	0,456	0,158	0,386	0,456	0,045	0,389	0,567	0,297	0,380	0,323		
EAS	0,326	0,497	0,177	0,116	0,431	0,453	0,116	0,431	0,453	0,000	0,033	0,967	0,718	0,254	0,028		
AME	0,833	0,167	0,000	0,009	0,065	0,926	0,009	0,065	0,926	0,787	0,194	0,019	0,259	0,528	0,213		
OCE	0,972	0,028	0,000	0,722	0,250	0,028	0,722	0,250	0,028	0,000	0,000	1,000	0,889	0,111	0,000		
		rs4760332				rs2104388				rs2427622							
Población	AA	AC	CC	AA	AG	GG	AA	AG	GG	AA	AG	GG	AA	AG	GG		
EUR	0,075	0,375	0,550	0,113	0,363	0,525	0,113	0,363	0,525	0,181	0,438	0,381	0,088	0,463	0,450		
AFR	0,000	0,017	0,983	0,983	0,017	0,000	0,983	0,017	0,000	0,000	0,000	1,000	0,000	0,058	0,942		
MEA	0,011	0,153	0,835	0,273	0,466	0,261	0,273	0,466	0,261	0,040	0,205	0,756	0,000	0,256	0,699		
CSA	0,222	0,494	0,285	0,222	0,430	0,348	0,222	0,430	0,348	0,095	0,513	0,392	0,045	0,256	0,699		
EAS	0,436	0,497	0,066	0,011	0,221	0,768	0,011	0,221	0,768	0,652	0,309	0,039	0,297	0,380	0,323		
AME	0,880	0,120	0,000	0,019	0,259	0,722	0,019	0,259	0,722	0,750	0,185	0,065	0,889	0,111	0,000		
OCE	0,222	0,306	0,472	0,722	0,278	0,000	0,722	0,278	0,000	0,556	0,361	0,083	0,889	0,111	0,000		

Tabla 1. Probabilidad de asignación de cada genotipo de cada SNP. EUR: Europa; AFR: África sub-sahariana; MEA: Oriente Próximo y Norte de África; CSA: Centro-Sur de Asia, EAS: Este de Asia; AME: América; OCE: Oceanía.

En una realización particular, la etapa (ii) del método de la invención se lleva a cabo mediante un ordenador programable. Dicho ordenador programable puede comprender, entre otros, un procesador, un sistema de almacenamiento de datos (incluyendo una memoria volátil y no volátil y/o elementos de almacenamiento), al menos un dispositivo de entrada y al menos un dispositivo de salida.

#### Kit de la invención y uso del kit de la invención

En otro aspecto, la presente invención se refiere a un kit, en adelante “kit de la invención”, que comprende reactivos adecuados para la detección de las secuencia de los polimorfismos de un solo nucleótido (SNP) rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622, en donde dichos reactivos se seleccionan del grupo formado por

- (i) un conjunto de parejas de cebadores adecuados para amplificar las regiones genómicas que comprenden dichos SNP o fragmentos de las mismas que contienen dichos SNP o las secuencias complementarias de las mismas y
- (ii) un conjunto de sondas que hibridan con las regiones genómicas que comprenden las secuencias de dichos SNP o con las secuencias complementarias de las mismas, en donde dichas sondas hibridan de forma específica con una de las variantes de dichos SNP.

Los términos “polimorfismo de un solo nucleótido”, y los SNP rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388 y rs2427622 se han definido previamente en relación con el método de la invención.

En una realización particular, el kit de la invención comprende un conjunto de cebadores adecuados para amplificar

- (i) las regiones genómicas que comprenden los SNP que se van a detectar o
- (ii) fragmentos de las regiones genómicas que comprenden dichos SNP o
- (iii) secuencias complementarias de las mismas.

El término “cebador” se ha definido previamente en relación con el método de la invención.

En otra realización particular, el kit de la invención comprende un conjunto de sondas que hibridan con

- (i) las regiones genómicas que comprenden las secuencias de los SNP que se van a detectar o
- (ii) secuencias complementarias de las mismas,

en donde dichas sondas hibridan de forma específica con una de las variantes de dichos SNP.

El término “sonda”, tal y como aquí se utiliza, se refiere a un oligonucleótido marcado o sin marcar capaz de hibridar selectivamente con un ácido nucleico diana o molde en condiciones adecuadas. Las sondas que pueden formar parte del kit de la invención son sondas capaces de discriminar entre los diferentes alelos o variantes de cada SNP, es decir, sondas que hibridan de forma específica con una de las variantes de los SNP que van a ser detectados.

En una realización particular, los cebadores o sondas adecuados para la detección de la secuencia de los SNP rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622 representan al menos el 10%, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80% o al menos el 90%, al menos el 95%, o al menos el 99% de los cebadores y/o sondas incluidos en el kit.

El kit de la invención puede incluir los reactivos en contenedores adecuados y materiales de embalaje, incluyendo tubos, viales, envases de plástico o similares. Además, el kit de la invención puede contener instrucciones para el uso simultáneo, secuencial o separado de los distintos componentes que se encuentran en el kit. Dichas instrucciones pueden estar en forma de material impreso o en forma de un soporte electrónico capaz de almacenar instrucciones de manera que puedan ser leídas por un sujeto, tales como medios de almacenamiento electrónico (discos magnéticos, cintas y similares), medios ópticos (CD-ROM, DVD) y similares. Adicionalmente o alternativamente, los medios pueden contener las direcciones de Internet que proporcionen dichas instrucciones.

En una realización particular, el kit de la invención comprende adicionalmente reactivos adecuados para la detección de las secuencias de un grupo de 5 polimorfismos adicionales, en donde dicho grupo se selecciona del grupo formado por:

- (i) el grupo formado por los polimorfismos rs186471, rs7919248, rs2231997, rs581468 y rs1609997,
- (ii) el grupo formado por los polimorfismos rs1765857, rs10497520, rs9552445, rs12880675 y rs350886,
- (iii) el grupo formado por los polimorfismos los rs2384319, rs7919248, rs2389736, rs2166624 y rs2268969,
- (iv) el grupo formado por los polimorfismos rs6542787, rs7919248, rs261532, rs5996039 y rs2268969,
- (v) el grupo formado por los polimorfismos rs6542787, rs7919248, rs261532, rs8139993 y rs2268969 y

(vi) el grupo formado por los polimorfismos rs3784651, rs10497520, rs261532, rs12880675 y rs4742634

en donde dichos reactivos se seleccionan del grupo formado por

- 5 (i) un conjunto de parejas de cebadores adecuados para amplificar las regiones genómicas que comprenden dichos SNP o fragmentos de las mismas que contienen dichos SNP o las secuencias complementarias de las mismas y
- (ii) un conjunto de sondas que hibridan con las regiones genómicas que comprenden las secuencias de dichos SNP o con las secuencias complementarias de las mismas, en donde dichas sondas hibridan de forma
- 10 específica con una de las variantes de dichos SNP.

Los polimorfismos rs186471, rs7919248, rs2231997, rs581468, rs1609997, rs1765857, rs10497520, rs9552445, rs12880675, rs350886, rs2384319, rs2389736, rs2166624, rs2268969, rs6542787, rs261532, rs5996039, rs8139993, rs3784651 y rs4742634 se han definido previamente en relación con el método de la invención.

15 En una realización particular, los cebadores o sondas adecuados para la detección de la secuencia de los SNP rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388 y rs2427622 y las secuencias de uno de los grupos de 5 polimorfismos (i) a (vi) representan al menos el 10%, al menos el 20%, al menos el 30%, al menos el 40%, al menos el 50%, al menos el 60%, al menos el 70%, al menos

20 el 80% o al menos el 90%, al menos el 95%, o al menos el 99% de los cebadores y/o sondas incluidos en el kit.

En otro aspecto, la invención se refiere al uso del kit de la invención para determinar la ascendencia geográfica de un sujeto. Los términos “determinar la ascendencia geográfica” y “sujeto” han sido previamente definidos en relación con el

25 método de la invención.

En una realización particular del uso del kit de la invención permite determinar, con una elevada probabilidad, la ascendencia geográfica del sujeto en función de la el genotipo presente en cada uno de los polimorfismos rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622. El término “el genotipo” y “elevada

30 probabilidad” han sido previamente definidos en relación con el método de la invención.

Así, en una realización particular,

- el genotipo GG en el polimorfismo rs1257016,
- el genotipo CC en el polimorfismo rs971779,
- 35 - el genotipo TC en el polimorfismo rs238547,
- el genotipo AG en el polimorfismo rs1482326,

- el genotipo CC en el polimorfismo rs4760332,
- el genotipo GG en el polimorfismo rs2104388 y
- el genotipo AG en el polimorfismo rs2427622

5 indican que existe una elevada probabilidad de que la ascendencia geográfica del sujeto es europea.

En otra realización particular,

- el genotipo GG en el polimorfismo rs1257016,
- el genotipo TT en el polimorfismo rs971779,
- el genotipo CC en el polimorfismo rs238547,
- 10 - el genotipo GG en el polimorfismo rs1482326,
- el genotipo CC en el polimorfismo rs4760332,
- el genotipo AA en el polimorfismo rs2104388 y
- el genotipo GG en el polimorfismo rs2427622

15 indican que existe una elevada probabilidad de que la ascendencia geográfica es África sub-sahariana,

En otra realización particular,

- el genotipo GG en el polimorfismo rs1257016,
- el genotipo CC en el polimorfismo rs971779,
- el genotipo CC en el polimorfismo rs238547,
- 20 - el genotipo GG en el polimorfismo rs1482326,
- el genotipo CC en el polimorfismo rs4760332,
- el genotipo AG en el polimorfismo rs2104388 y
- el genotipo GG en el polimorfismo rs2427622

25 indican que existe una elevada probabilidad de que la ascendencia geográfica es Oriente Próximo y Norte de África.

En otra realización particular,

- el genotipo AG en el polimorfismo rs1257016,
- el genotipo CC en el polimorfismo rs971779,
- el genotipo CC en el polimorfismo rs238547,
- 30 - el genotipo AG en el polimorfismo rs1482326,
- el genotipo AC en el polimorfismo rs4760332,
- el genotipo AG en el polimorfismo rs2104388 y
- el genotipo AG en el polimorfismo rs2427622

35 indican que existe una elevada probabilidad de que la ascendencia geográfica es Centro y Sur de Asia.

En otra realización particular,

- el genotipo AG en el polimorfismo rs1257016,
- el genotipo CC en el polimorfismo rs971779,
- el genotipo CC en el polimorfismo rs238547,
- el genotipo AA en el polimorfismo rs1482326,
- 5 - el genotipo AC en el polimorfismo rs4760332,
- el genotipo GG en el polimorfismo rs2104388 y
- el genotipo AA en el polimorfismo rs2427622

indican que existe una elevada probabilidad de que la ascendencia geográfica es el este de Asia.

- 10 En otra realización particular,
- el genotipo AA en el polimorfismo rs1257016,
  - el genotipo CC en el polimorfismo rs971779,
  - el genotipo TT en el polimorfismo rs238547,
  - el genotipo AG en el polimorfismo rs1482326,
  - 15 - el genotipo AA en el polimorfismo rs4760332,
  - el genotipo GG en el polimorfismo rs2104388 y
  - el genotipo AA en el polimorfismo rs2427622

indican que existe una elevada probabilidad de que la ascendencia geográfica es de América.

- 20 En otra realización particular,
- el genotipo AA en el polimorfismo rs1257016,
  - el genotipo TT en el polimorfismo rs971779,
  - el genotipo CC en el polimorfismo rs238547,
  - el genotipo AA en el polimorfismo rs1482326,
  - 25 - el genotipo CC en el polimorfismo rs4760332,
  - genotipo AA en el polimorfismo rs2104388 y
  - la el genotipo AA en el polimorfismo rs2427622

indican que existe una elevada probabilidad de que la ascendencia geográfica es Oceanía.

- 30 En una realización particular, elevada probabilidad es de al menos un 70%, preferiblemente al menos un 75%.

#### Medio leíble por ordenador de la invención

- 35 En otro aspecto la invención se relaciona con un medio leíble por ordenador, en adelante “medio leíble por ordenador de la invención” que contiene instrucciones ejecutables que, al ejecutarse, resultan en

(i) la recepción por al menos un procesador del genotipo de un sujeto en los dos alelos de los polimorfismos rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622 y

5 (ii) proporcionar una indicación sobre la probabilidad de que un sujeto tenga una determinada ascendencia geográfica.

El término “medio leíble por ordenador”, tal y como aquí se utiliza, se refiere a medios de almacenamiento físicos, incluyendo realización de la invención pueden incluir cualquier medio de almacenamiento leíble por ordenador físico, por ejemplo memoria en estado sólido (tal como memoria flash), medios y dispositivos de  
10 almacenamiento leíbles por ordenador magnéticos y ópticos, y memoria que usa otras tecnologías de almacenamiento persistente. En algunas formas de realización, un medio leíble por ordenador puede ser cualquier medio tangible que permita que el ordenador acceda a datos y programas informáticos. Los medios leíbles por ordenador pueden incluir medios tangibles volátiles y no volátiles, eliminables o no eliminables  
15 implementados en cualquier método o tecnología capaz de almacenar información tal como instrucciones leíbles por ordenador, módulos de programas, programas, datos, estructuras de datos, e información de bases de datos. En algunas formas de realización de la invención, medio leíble por ordenador incluye, pero no está limitado a, RAM (memoria de acceso aleatorio), ROM (memoria de lectura solo), EPROM  
20 (memoria de lectura solo programable borrable), EEPROM (memoria de lectura solo programable borrable eléctricamente), memoria flash u otra tecnología de memoria, CD-ROM (memoria de lectura solo en disco compacto), DVD (discos versátiles digitales) u otro medio de almacenamiento óptico, casetes magnéticos, cinta magnética, almacenamiento de disco magnético u otro medio de almacenamiento  
25 magnético, otros tipos de memoria volátil y no volátil, y cualquier otro medio tangible que se pueda usar para almacenar información y que pueda leer un ordenador incluyendo cualquier combinación adecuada de los anteriores.

El término “procesador”, tal y como aquí se utiliza, se refiere al dispositivo que interpreta las instrucciones de un programa informático mediante la realización de  
30 operaciones aritméticas, lógicas y de entrada/salida del sistema.

Los términos “polimorfismo de un solo nucleótido”, “sujeto” y “ascendencia geográfica” y los SNP rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388 y rs2427622 se han definido previamente en relación con el método de la  
35 invención.

### **Ejemplo**

Los autores de la presente invención han desarrollado una solución eficiente para la asignación no supervisada y precisa de individuos a su origen poblacional reciente, de entre una de las siete divisiones principales de la población mundial, basado en una prueba de ADN consistente en la determinación de un mínimo de 7

5 SNP con características de AIM. Para la deducción de este panel y su identificación como conjunto óptimos de AIM se aplicó un método heurístico combinatorio propio a datos genéticos de gran calidad obtenidos de 642.690 SNP autosómicos determinados en ADN de 938 sujetos de diversas poblaciones (Li *et al*, Science 319, 2008: 1100-4). Las muestras de ADN se coleccionaron de sujetos sin relación de parentesco de 52

10 poblaciones del mundo como parte de estudios del Proyecto de Diversidad del Genoma Humano del Centre d'Etude du Polymorphisme Humain (HGDP-CEPH). El algoritmo empleado sorteaba automáticamente la posible redundancia existente entre SNP escogidos de regiones relativamente cercanas del genoma, manteniendo aquellos no redundantes para encontrar un panel eficiente sin necesitar análisis o

15 filtros adicionales.

La estimación de la capacidad de asignación de individuos se realizó utilizando los datos de HGDP-CEPH agrupados en siete divisiones biogeográficas, tal y como han revelado diversos estudios (Li *et al*. Science 319, 2008: 1100-4): África subsahariana, América, Centro-Sur de Asia, Este de Asia, Europa, Cercano Oriente y

20 Norte de África, y Oceanía. Dicho análisis no consideró las poblaciones de Japón, Kalash y Hazaras, dado su conocido comportamiento distintivo con respecto a su localización geográfica.

Conocida la frecuencia de cada uno de los alelos correspondientes a los AIM del panel en cada grupo de población de HGDP-CEPH, es fácil estimar la probabilidad

25 de que una combinación individual cualquiera haya sido producida por azar en cada una de las siete divisiones biogeográficas (Tabla 1). Se utilizó para ello un método basado en máxima verosimilitud de los datos observados (la el genotipo de la muestra a clasificar). Nótese que existe independencia estadística entre las distribuciones de los AIM, debida al algoritmo por el que fueron seleccionados. Bajo independencia en

30 las observaciones, dicha probabilidad resulta del producto de las probabilidades individuales y, por tanto, es posible averiguar qué subdivisión biogeográfica tiene la probabilidad más alta en la asignación sencillamente sumando los logaritmos de las probabilidades individuales para cada AIM del panel. La confianza de cada asignación se obtuvo mediante la función  $\lambda$  con:

35 
$$\lambda = BestFact * v1^{1st} + (1.0 - BestFact) * (v1^{1st} - v2^{nd})$$

Donde  $v1^{st}$  y  $v2^{nd}$  son, respectivamente, la mayor y la segunda mayor suma de los logaritmos de probabilidades de asignación de la muestra concreta a las siete divisiones biogeográficas. *BestFact* es una constante, con valor de 0,1 tal sugirieron nuestros ensayos piloto, que permite balancear entre el valor  $v1^{st}$  la diferencia entre  $v1^{st}$  y  $v2^{nd}$ . Cuando  $v1^{st}$  y  $v2^{nd}$  son casi iguales (lo que llevaría a plantear dudas en la asignación de una muestra sujeto), el peso de la confianza asociada a la probabilidad de asignación recae en  $v1^{st}$ , indicando el nivel de confianza de la asignación más probable. Para el cálculo del sumando los logaritmos de las probabilidades individuales para cada AIM del panel, en caso de datos perdidos (si alguno de los genotipos ha fallado en la determinación), se asignó  $1/N$  como probabilidad a todas las subdivisiones, siendo  $N$  el tamaño muestral de cada subdivisión. Además, para evitar divisiones por cero, en caso de no haber observaciones, se añadió una corrección de continuidad de una diezmilésima de observación a cada grupo.

Utilizando esta metodología, la validación de la capacidad de asignación del panel fue realizada subdividiendo los datos de HGDP-CEPH en una muestra de aprendizaje (que incluyó el 60% de los individuos) y una muestra de prueba (que incluyó el 40% restante). El acierto de clasificación a uno de las siete divisiones biogeográficas, es decir, la proporción de individuos de la muestra cuya asignación correspondió con la ascendencia real, fue cuantificada en un 76,3% (Figura 1). La capacidad de asignación es superior al 75% en individuos procedentes de África subsahariana, América, Este de Asia, y Oceanía, mientras que la confusión es mayor (máxima confusión de 28,7%) en la asignación de individuos de Centro-Sur de Asia, Europa, Cercano Oriente y Norte de África. Pruebas realizadas con paneles de menor tamaño tuvieron una disminuida capacidad de asignación siendo, en el mejor de los casos del 52,3% de acierto.

Se propone, por tanto, un panel mínimo de 7 AIM minuciosamente seleccionados que mantienen el máximo de información de la ascendencia biogeográfica de un sujeto, permitiendo la clasificación del 76,3% de los sujetos a la población correcta de origen mediante análisis no supervisados.

Los identificadores de referencia (números rs) de los siete SNP adscritos al panel óptimo para la asignación poblacional son: rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622. Los análisis realizados sugieren que, además de ser determinable en reacciones uniplex, el panel consta de secuencias que facilitan la detección simultánea en reacciones multiplex.

Utilizando la misma metodología, se han seleccionado 6 combinaciones de 5 AIM (Tabla 2) que suplementan la capacidad informativa del panel mínimo de 7 AIM

para la asignación de sujetos a una de las siete subdivisiones biogeográficas. Las probabilidades de asignación de cada el genotipo de los suplementos nº 1 a nº 6 se muestran en las tablas 3 a 8.

**Tabla 2.** Combinaciones adicionales de AIM que suplementan la capacidad de asignación del panel óptimo para mejorar la asignación poblacional.

Suplemento nº1	Suplemento nº2	Suplemento nº3	Suplemento nº4	Suplemento nº5	Suplemento nº6
rs186471	rs1765857	rs2384319	rs6542787	rs6542787	rs3784651
rs7919248	rs10497520	rs7919248	rs7919248	rs7919248	rs10497520
rs2231997	rs9552445	rs2389736	rs261532	rs261532	rs261532
rs581468	rs12880675	rs2166624	rs5996039	rs8139993	rs12880675
rs1609997	rs350886	rs2268969	rs2268969	rs2268969	rs4742634

5

Población	rs186471			rs7919248			rs2231997			rs581468		
	AA	AG	GG	GG	TG	TT	CC	TC	TT	AA	AG	GG
EUR	0,000	0,138	0,863	0,000	0,119	0,881	0,488	0,444	0,069	0,344	0,438	0,219
AFR	0,158	0,417	0,425	0,008	0,124	0,868	0,983	0,017	0,000	0,562	0,364	0,074
MEA	0,006	0,188	0,807	0,000	0,182	0,818	0,807	0,182	0,011	0,494	0,420	0,085
CSA	0,114	0,386	0,500	0,108	0,348	0,544	0,443	0,418	0,139	0,203	0,481	0,316
EAS	0,578	0,378	0,044	0,293	0,497	0,210	0,111	0,450	0,439	0,000	0,105	0,895
AME	0,269	0,509	0,222	0,944	0,046	0,009	0,630	0,315	0,056	0,019	0,065	0,917
OCE	0,944	0,028	0,028	0,250	0,389	0,361	0,000	0,114	0,886	0,000	0,000	1,000
	<b>rs1609997</b>											
Población	AA	AG	GG									
EUR	0,063	0,356	0,581									
AFR	0,901	0,099	0,000									
MEA	0,114	0,523	0,364									
CSA	0,158	0,411	0,430									
EAS	0,856	0,144	0,000									
AME	0,778	0,194	0,028									
OCE	0,972	0,028	0,000									

Tabla 3. Probabilidad de asignación de una muestra a una división poblacional en base al genotipo de cada SNP del suplemento nº 1. EUR: Europa; AFR: África sub-sahariana; MEA: Oriente Próximo y Norte de África; CSA: Centro-Sur de Asia, EAS: Este de Asia; AME: América; OCE: Oceanía.

Población	rs1765857				rs10497520				rs9552445				rs12880675					
	AA	AG	GG	CC	TC	TT	AA	AG	GG	CC	TC	TT	AA	AG	GG	CC	TC	TT
EUR	0,888	0,106	0,006	0,781	0,175	0,044	0,725	0,244	0,031	0,700	0,281	0,019						
AFR	0,504	0,355	0,140	0,058	0,372	0,570	1,000	0,000	0,000	0,744	0,231	0,025						
MEA	0,807	0,170	0,023	0,790	0,199	0,011	0,648	0,290	0,063	0,813	0,182	0,006						
CSA	0,538	0,367	0,095	0,468	0,424	0,108	0,405	0,405	0,190	0,582	0,361	0,057						
EAS	0,044	0,304	0,652	0,055	0,354	0,591	0,127	0,409	0,464	0,033	0,337	0,630						
AME	0,009	0,120	0,870	0,019	0,056	0,926	0,204	0,417	0,380	0,657	0,287	0,056						
OCE	0,333	0,361	0,306	0,194	0,306	0,500	0,000	0,167	0,833	0,000	0,056	0,944						
<b>rs350886</b>																		
Población	AA	AG	GG															
EUR	0,131	0,356	0,513															
AFR	0,917	0,083	0,000															
MEA	0,263	0,486	0,251															
CSA	0,127	0,420	0,452															
EAS	0,000	0,144	0,856															
AME	0,000	0,009	0,991															
OCE	0,371	0,429	0,200															

Tabla 4. Probabilidad de asignación de una muestra a una di visión poblacional en base al genotipo de cada SNP del suplemento nº 2.  
 EUR: Europa; AFR: África sub-sahariana; MEA: Oriente Próximo y Norte de África; CSA: Centro-Sur de Asia, EAS: Este de Asia; AME: América; OCE: Oceanía.

Población	rs2384319				rs7919248				rs2389736				rs2166624			
	GG	TG	TT	GG	TG	TT	AA	AG	GG	AA	AG	GG	AA	AG	GG	
EUR	0,013	0,169	0,819	0,000	0,119	0,881	0,000	0,094	0,906	0,175	0,419	0,406				
AFR	0,008	0,298	0,694	0,008	0,124	0,868	0,686	0,264	0,050	0,008	0,000	0,992				
MEA	0,017	0,256	0,727	0,000	0,182	0,818	0,040	0,256	0,705	0,063	0,324	0,614				
CSA	0,051	0,373	0,576	0,108	0,348	0,544	0,000	0,228	0,772	0,089	0,449	0,462				
EAS	0,867	0,110	0,022	0,293	0,497	0,210	0,160	0,436	0,403	0,204	0,431	0,365				
AME	0,833	0,139	0,028	0,944	0,046	0,009	0,630	0,306	0,065	0,963	0,037	0,000				
OCE	0,111	0,194	0,694	0,250	0,389	0,361	0,556	0,306	0,139	0,000	0,000	1,000				
<b>rs2268969</b>																
Población	CC	TC	TT													
EUR	0,044	0,263	0,694													
AFR	0,000	0,000	1,000													
MEA	0,028	0,159	0,813													
CSA	0,165	0,430	0,405													
EAS	0,663	0,304	0,033													
AME	0,306	0,537	0,157													
OCE	0,528	0,444	0,028													

Tabla 5. Probabilidad de asignación de una muestra a una división poblacional en base al genotipo de cada SNP del suplemento nº 3.  
 EUR: Europa; AFR: África sub-sahariana; MEA: Oriente Próximo y Norte de África; CSA: Centro-Sur de Asia, EAS: Este de Asia; AME: América; OCE: Oceanía.

Población	rs6542787			rs7919248			rs261532			rs5996039		
	AA	AG	GG	GG	TG	TT	GG	TG	TT	AA	AG	GG
EUR	0,731	0,244	0,025	0,000	0,119	0,881	0,506	0,381	0,113	0,681	0,288	0,031
AFR	0,248	0,537	0,215	0,008	0,124	0,868	0,008	0,066	0,926	0,165	0,364	0,471
MEA	0,619	0,318	0,063	0,000	0,182	0,818	0,278	0,460	0,261	0,528	0,358	0,114
CSA	0,690	0,259	0,051	0,108	0,348	0,544	0,418	0,411	0,171	0,449	0,487	0,063
EAS	0,017	0,160	0,823	0,293	0,497	0,210	0,735	0,243	0,022	0,818	0,166	0,017
AME	0,046	0,167	0,787	0,944	0,046	0,009	0,981	0,019	0,000	0,009	0,037	0,954
OCE	0,028	0,306	0,667	0,250	0,389	0,361	0,111	0,361	0,528	0,944	0,028	0,028
<b>rs2268969</b>												
Población	CC	TC	TT									
EUR	0,044	0,263	0,694									
AFR	0,000	0,000	1,000									
MEA	0,028	0,159	0,813									
CSA	0,165	0,430	0,405									
EAS	0,663	0,304	0,033									
AME	0,306	0,537	0,157									
OCE	0,528	0,444	0,028									

Tabla 6. Probabilidad de asignación de una muestra a una división poblacional en base al genotipo de cada SNP del suplemento nº 4.  
 EUR: Europa; AFR: África sub-sahariana; MEA: Oriente Próximo y Norte de África; CSA: Centro-Sur de Asia, EAS: Este de Asia; AME: América; OCE: Oceanía.

Población	rs6542787			rs7919248			rs261532			rs8139993		
	AA	AG	GG	GG	TG	TT	GG	TG	TT	CC	TC	TT
EUR	0,731	0,244	0,025	0,000	0,119	0,881	0,506	0,381	0,113	0,031	0,288	0,681
AFR	0,248	0,537	0,215	0,008	0,124	0,868	0,008	0,066	0,926	0,471	0,364	0,165
MEA	0,619	0,318	0,063	0,000	0,182	0,818	0,278	0,460	0,261	0,114	0,358	0,528
CSA	0,690	0,259	0,051	0,108	0,348	0,544	0,418	0,411	0,171	0,063	0,487	0,449
EAS	0,017	0,160	0,823	0,293	0,497	0,210	0,735	0,243	0,022	0,017	0,166	0,818
AME	0,046	0,167	0,787	0,944	0,046	0,009	0,981	0,019	0,000	0,954	0,037	0,009
OCE	0,028	0,306	0,667	0,250	0,389	0,361	0,111	0,361	0,528	0,028	0,028	0,944
<b>rs2268969</b>												
Población	CC	TC	TT									
EUR	0,044	0,263	0,694									
AFR	0,000	0,000	1,000									
MEA	0,028	0,159	0,813									
CSA	0,165	0,430	0,405									
EAS	0,663	0,304	0,033									
AME	0,306	0,537	0,157									
OCE	0,528	0,444	0,028									

Tabla 7. Probabilidad de asignación de una muestra a una división poblacional en base al genotipo de cada SNP del suplemento nº 5.  
 EUR: Europa; AFR: África sub-sahariana; MEA: Oriente Próximo y Norte de África; CSA: Centro-Sur de Asia, EAS: Este de Asia; AME: América; OCE: Oceanía.

Población	rs3784651			rs10497520			rs261532			rs12880675		
	CC	TC	TT	CC	TC	TT	GG	TG	TT	CC	TC	TT
EUR	0,019	0,213	0,769	0,781	0,175	0,044	0,506	0,381	0,113	0,700	0,281	0,019
AFR	0,471	0,347	0,182	0,058	0,372	0,570	0,008	0,066	0,926	0,744	0,231	0,025
MEA	0,045	0,415	0,540	0,790	0,199	0,011	0,278	0,460	0,261	0,813	0,182	0,006
CSA	0,120	0,449	0,430	0,468	0,424	0,108	0,418	0,411	0,171	0,582	0,361	0,057
EAS	0,547	0,409	0,044	0,055	0,354	0,591	0,735	0,243	0,022	0,033	0,337	0,630
AME	0,009	0,278	0,713	0,019	0,056	0,926	0,981	0,019	0,000	0,657	0,287	0,056
OCE	1,000	0,000	0,000	0,194	0,306	0,500	0,111	0,361	0,528	0,000	0,056	0,944
<b>rs4742634</b>												
Población	CC	TC	TT									
EUR	0,600	0,363	0,038									
AFR	0,017	0,174	0,810									
MEA	0,449	0,466	0,085									
CSA	0,190	0,519	0,291									
EAS	0,022	0,271	0,707									
AME	0,009	0,074	0,917									
OCE	0,000	0,028	0,972									

Tabla 8: Probabilidad de asignación de una muestra a una división poblacional en base al genotipo cada SNP del suplemento nº 6. EUR: Europa; AFR: África sub-sahariana; MEA: Oriente Próximo y Nort e de África; CSA: Centro-Sur de Asia, EAS: Este de Asia; AME: América; OCE: Oceanía.

Al suplementar el panel mínimo, los paneles de 12 AIM resultantes mejoraron la proporción de aciertos de clasificación, estando cuantificada en todos los casos en un 87% en los datos de HGDP-CEPH (938 sujetos). Con cualquiera de estos seis suplementos, la capacidad de asignación es superior al 75% en individuos procedentes de cualquiera de las siete regiones geográficas (África sub-sahariana, América, Centro-Sur de Asia, Este de Asia, Europa, Cercano Oriente y Norte de África, y Oceanía), mostrándose en la Figura 2 la asignación resultante de utilizar el panel mínimo con el suplemento nº 1. Por tanto, deben tratarse dichas combinaciones de 5 AIM como suplementos equivalentes en cuanto a su capacidad de clasificación.

Con el fin de confirmar si estas propiedades en el panel suplementado se mantenían en muestras de sujetos independientes, se realizó una validación en datos de 2000 individuos independientes procedentes de la versión 3 del proyecto 1000 Genomas. A pesar de que este proyecto solo dispone de datos genéticos de sujetos procedentes de 4 de las 7 subregiones biogeográficas (África sub-sahariana, Centro-Sur de Asia, Este de Asia, Europa), el panel mostró valores equivalentes de acierto en la asignación (88% en el conjunto de individuos) al obtenido en sujetos de HGDP-CEPH. En la Figura 3 se muestran los resultados de asignación de los 2000 individuos de este proyecto al utilizar el panel mínimo con el suplemento nº 1.

**REIVINDICACIONES**

1. Método para determinar la ascendencia geográfica de un sujeto que comprende
  - 5 (i) determinar en una muestra que comprende material genético de dicho sujeto el genotipo de los dos alelos de los polimorfismos de un solo nucleótido (SNP) rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622 y, opcionalmente, cualquier otro polimorfismo que se encuentre en desequilibrio de ligamiento con cualquiera de los anteriores, en ambos alelos y
  - 10 (ii) determinar la ascendencia geográfica de dicho sujeto en base a los genotipos de dichos SNP.
  
2. Método según la reivindicación 1, en donde
  - 15 (i) el genotipo GG en el polimorfismo rs1257016, el genotipo CC en el polimorfismo rs971779, el genotipo TC en el polimorfismo rs238547, el genotipo AG en el polimorfismo rs1482326, el genotipo CC en el polimorfismo rs4760332, el genotipo GG en el polimorfismo rs2104388 y el genotipo AG en el polimorfismo rs2427622 indican que existe una elevada probabilidad de que la ascendencia geográfica del sujeto sea europea,
  - 20 (ii) el genotipo GG en el polimorfismo rs1257016, el genotipo TT en el polimorfismo rs971779, el genotipo CC en el polimorfismo rs238547, el genotipo GG en el polimorfismo rs1482326, el genotipo CC en el polimorfismo rs4760332, el genotipo AA en el polimorfismo rs2104388 y el genotipo GG en el polimorfismo rs2427622 indican que existe una elevada probabilidad de que la ascendencia geográfica sea África sub-sahariana,
  - 25 (iii) el genotipo GG en el polimorfismo rs1257016, el genotipo CC en el polimorfismo rs971779, el genotipo CC en el polimorfismo rs238547, el genotipo GG en el polimorfismo rs1482326, el genotipo CC en el polimorfismo rs4760332, el genotipo AG en el polimorfismo rs2104388 y el genotipo GG en el polimorfismo rs2427622 indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oriente Próximo y Norte de África,
  - 30 (iv) el genotipo AG en el polimorfismo rs1257016, el genotipo CC en el polimorfismo rs971779, el genotipo CC en el polimorfismo rs238547, el genotipo AG en el polimorfismo rs1482326, el genotipo AC en el polimorfismo rs4760332, el genotipo AG en el polimorfismo rs2104388 y el
  - 35

- genotipo AG en el polimorfismo rs2427622 indican que existe una elevada probabilidad de que la ascendencia geográfica sea Centro y Sur de Asia,
- (v) el genotipo AG en el polimorfismo rs1257016, el genotipo CC en el polimorfismo rs971779, el genotipo CC en el polimorfismo rs238547, el genotipo AA en el polimorfismo rs1482326, el genotipo AC en el polimorfismo rs4760332, el genotipo GG en el polimorfismo rs2104388 y el genotipo AA en el polimorfismo rs2427622 indican que existe una elevada probabilidad de que la ascendencia geográfica sea el este de Asia,
- (vi) el genotipo AA en el polimorfismo rs1257016, el genotipo CC en el polimorfismo rs971779, el genotipo TT en el polimorfismo rs238547, el genotipo AG en el polimorfismo rs1482326, el genotipo AA en el polimorfismo rs4760332, el genotipo GG en el polimorfismo rs2104388 y el genotipo AA en el polimorfismo rs2427622 indican que existe una elevada probabilidad de que la ascendencia geográfica sea de América y
- (vii) el genotipo AA en el polimorfismo rs1257016, el genotipo TT en el polimorfismo rs971779, el genotipo CC en el polimorfismo rs238547, el genotipo AA en el polimorfismo rs1482326, el genotipo CC en el polimorfismo rs4760332, el genotipo AA en el polimorfismo rs2104388 y el genotipo AA en el polimorfismo rs2427622 indican que existe una elevada probabilidad de que la ascendencia geográfica sea Oceanía.
3. Método según las reivindicaciones 1 ó 2 que comprende adicionalmente la determinación de los genotipos de los dos alelos de un grupo de 5 SNP, en donde dicho grupo se selecciona del grupo formado por:
- (i) el grupo formado por los polimorfismos rs186471, rs7919248, rs2231997, rs581468 y rs1609997,
- (ii) el grupo formado por los polimorfismos rs1765857, rs10497520, rs9552445, rs12880675 y rs350886,
- (iii) el grupo formado por los polimorfismos los rs2384319, rs7919248, rs2389736, rs2166624 y rs2268969,
- (iv) el grupo formado por los polimorfismos rs6542787, rs7919248, rs261532, rs5996039 y rs2268969,
- (v) el grupo formado por los polimorfismos rs6542787, rs7919248, rs261532, rs8139993 y rs2268969 y
- (vi) el grupo formado por los polimorfismos rs3784651, rs10497520, rs261532, rs12880675 y rs4742634.

4. El método según cualquiera de las reivindicaciones 1 a 3 en donde para cada uno de los polimorfismos cuyo genotipo de los dos alelos se determina en la etapa (i), el método comprende la puesta en contacto de la muestra que contiene material genético con un ácido nucleico que hibrida de forma selectiva con una región genómica que comprende el polimorfismo.
- 5
5. El método según la reivindicación 4 en donde la muestra que contiene material genético se somete a un proceso de amplificación de ácidos nucleicos previamente a la puesta en contacto con el ácido nucleico que hibrida de forma selectiva con una región cromosómica que comprende el polimorfismo, en donde dicha amplificación se lleva a cabo sobre las regiones genómicas que comprenden los polimorfismos a detectar.
- 10
6. El método según las reivindicaciones 4 o 5 en donde el ácido nucleico que hibrida de forma selectiva con una región genómica que comprende el polimorfismo es un cebador y el método comprende la extensión del cebador usando como molde la región genómica que comprende el polimorfismo y la detección del producto de extensión.
- 15
7. El método según la reivindicación 6 en donde la extensión del cebador es una reacción de extensión multiplex.
- 20
8. El método según cualquiera de las reivindicaciones 1 a 7 en donde la muestra biológica se selecciona del grupo formado por sangre, suero, plasma, células de la piel, saliva, folículos pilosos o biopsias de tejidos.
- 25
9. Método según cualquiera de las reivindicaciones 1 a 8 en donde la etapa de determinar la ascendencia geográfica del sujeto comprende calcular para cada división geográfica la suma de los logaritmos de las probabilidades de asignación a cada división geográfica en base al genotipo de los dos alelos determinado en cada SNP,
- 30
- en donde existe una elevada probabilidad de que la ascendencia geográfica de dicho sujeto sea aquella para la cual dicha suma de logaritmos es mayor,

en donde las divisiones geográficas son Europa, África sub-sahariana, Oriente Próximo y Norte de África, Centro-Sur de Asia, Este de Asia, América y Oceanía, y

5 en donde las probabilidades de asignación de una muestra a cada división geográfica para cada genotipo de cada SNP son las frecuencias de dichos genotipos en una población control.

10. El método según cualquiera de las reivindicaciones 1 a 9 en donde un ordenador programable se emplea para llevar a cabo la etapa (ii).

10

11. Un medio leíble por ordenador que contiene instrucciones ejecutables que, al ejecutarse, resultan en

(i) la recepción por al menos un procesador del genotipo de un sujeto en los dos alelos de los polimorfismos rs1257016, rs971779, rs238547, rs1482326, rs4760332, rs2104388, y rs2427622 y

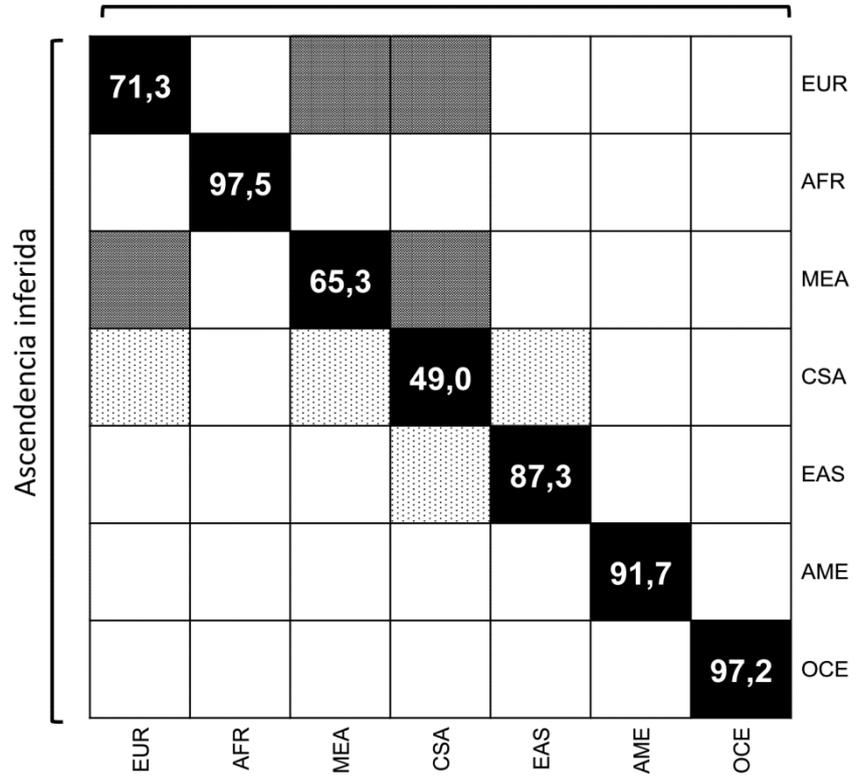
15

(ii) proporcionar una indicación sobre la probabilidad de que un sujeto tenga una determinada ascendencia geográfica.

.....

**FIG. 1**

Ascendencia declarada



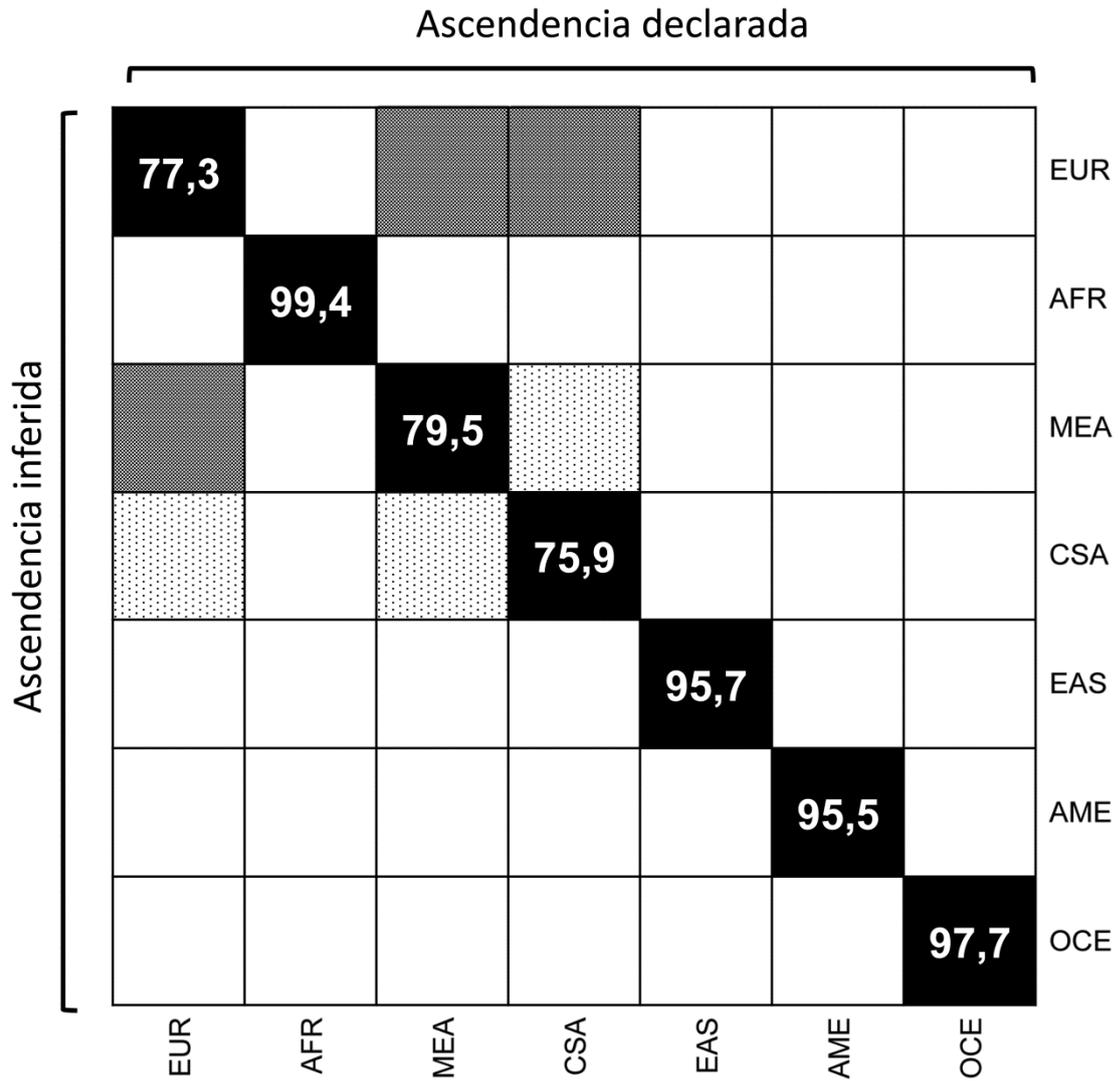


FIG. 2

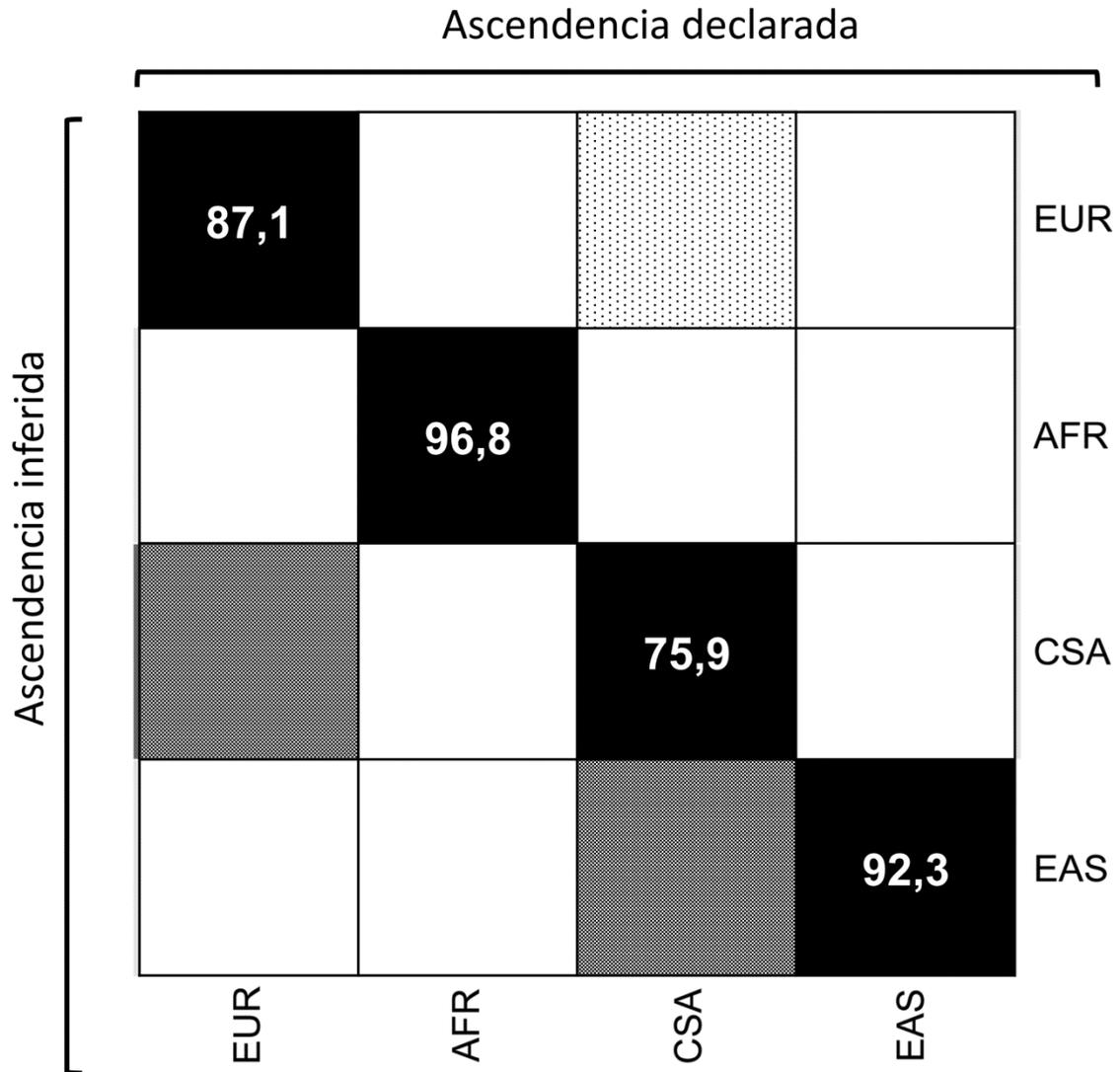


FIG. 3

# ES 2 609 130 B1

## LISTA DE SECUENCIAS

<110> UNIVERSIDAD DE LA LAGUNA  
SERVICIO CANARIO DE SALUD

<120> Método para determinar la ascendencia geográfica de un sujeto

<130> P12378ES00

<140> P201531475

<141> 2015-10-14

<160> 27

<170> PatentIn version 3.5

<210> 1  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 1  
tcaatatata atctttgtct cttgtrcttt ttgacttaa gtttattt 48

<210> 2  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 2  
cactgccatc acaaagaaaa acataygtaa atttcttcaa ggctacaa 48

<210> 3  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 3  
taggcttcaa gaccatggac ttcccygccg tcaccatctg caatgcta 48

<210> 4  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 4  
gctatgtttc aaaaggaatt gaagarcaac atctccatag aatacagt 48

<210> 5  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 5  
taccacaaa atcccaaatg ccttgmata gtatgggaga gaactgct 48

# ES 2 609 130 B1

<210> 6  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 6  
acagctggga aatatgtgta tgtcrtgtg gggaagtta aactttcc 48

<210> 7  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 7  
actttgggag gccaaaggcg gcggaycaca aggtcaggag atcgagac 48

<210> 8  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 8  
tattgattgg gtggaatata aatgtrggta tgtgtgtgtg tataagac 48

<210> 9  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 9  
gccaaagaag aggcgattca cagcakaat gaatttgata aaaataaa 48

<210> 10  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 10  
accacctgcc cagccctgcc tctcaygctc tgtccccag acatgggtg 48

<210> 11  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 11  
cagaagctaa accaccattc tttccracac tgtctcagag acttgaa 48

<210> 12  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 12  
gccctctttt taagtaactt tcctgragac acaaccaaac ttccactt 48

# ES 2 609 130 B1

<210> 13  
<211> 51  
<212> DNA  
<213> Homo sapiens

<400> 13  
tcttatgaaa gatgtgccaa ttaatrcgat gctgcatttt cttactattc c 51

<210> 14  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 14  
aatccagggtg ctgtttctcc aacttyaggt tcttgaacaa atgcagtc 48

<210> 15  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 15  
tacaaataaa caaggtagtt ctacarcagc attctaggca tttgcctt 48

<210> 16  
<211> 46  
<212> DNA  
<213> Homo sapiens

<400> 16  
gaactaatta ttttattgat tcaagbagca atggttctaa atcaga 46

<210> 17  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 17  
gagccccag gcgatggcgg ctctcrcaa aaggaaggag agtgaggc 48

<210> 18  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 18  
accgtgtttt gtttctttaa ctggckttgc tgtgcagcat ctttgtgc 48

<210> 19  
<211> 48  
<212> DNA  
<213> Homo sapiens

# ES 2 609 130 B1

<400> 19  
agtggttggg ttcac tctttc actgargttt gtagatactg tcaaccct 48

<210> 20  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 20  
tgtttcagac atctaattat aagacrttac cacctgtaca ttctatag 48

<210> 21  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 21  
tgaggccagc cagggatcatg gctccygcct cctccttctg tgcccctc 48

<210> 22  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 22  
gagaacagaa agttggcacc ttaaartatt tgtggtgtga tggagacg 48

<210> 23  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 23  
gtctgaattg agaacctgga tttctkgact gtccagtgtga ttttctgg 48

<210> 24  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 24  
gtgctcaagt gtttgtcaaa tcagtragg aactatctgg ctgctctc 48

<210> 25  
<211> 48  
<212> DNA  
<213> Homo sapiens

<400> 25  
ccaaaatcta actaatggag gttcaygtga agagagggaa tctgaaag 48

<210> 26  
<211> 48  
<212> DNA

# ES 2 609 130 B1

<213> Homo sapiens

<400> 26

cccaggcctg aaagggcaga actcayaaag ggaatactga ctttgtag

48

<210> 27

<211> 48

<212> DNA

<213> Homo sapiens

<400> 27

gaatttgagt tcaccatata tacatygctg gcagtaagac atagaaaa

48



- ②<sup>1</sup> N.º solicitud: 201531475  
②<sup>2</sup> Fecha de presentación de la solicitud: 14.10.2015  
③<sup>2</sup> Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤<sup>1</sup> Int. Cl.: **C12Q1/68** (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑤ <sup>6</sup> Documentos citados	Reivindicaciones afectadas
A	US 20070037182 A1 (GASKIN <i>et al.</i> ) 15.02.2007, Página 1, párrafo [0004]; página 3, párrafos [0014], [0015]; página 28, Tabla-I.	1-11
A	US 20070003944 A1 (SINHA <i>et al.</i> ) 04.01.2007, Página 2, párrafos [0021] - [0026]; reivindicaciones 1-9.	1-11

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

**El presente informe ha sido realizado**

para todas las reivindicaciones

para las reivindicaciones n.º:

Fecha de realización del informe  
16.02.2017

Examinador  
M. D. García Grávalos

Página  
1/4

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

C12Q

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, NPL, BIOSIS, MEDLINE, EMBASE, USPTO PATENT DATABASE, PUBMED, GOOGLE PATENTS, EBI.

Fecha de Realización de la Opinión Escrita: 16.02.2017

**Declaración**

<b>Novedad (Art. 6.1 LP 11/1986)</b>	Reivindicaciones 1-11	<b>SI</b>
	Reivindicaciones	<b>NO</b>
<b>Actividad inventiva (Art. 8.1 LP11/1986)</b>	Reivindicaciones 1-11	<b>SI</b>
	Reivindicaciones	<b>NO</b>

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

**Base de la Opinión.-**

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

**1. Documentos considerados.-**

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	US 20070037182 A1	15.02.2007
D02	US 20070003944 A1	04.01.2007

**2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración**

La presente invención divulga un método para determinar la ascendencia geográfica de un sujeto mediante la determinación en una muestra biológica del genotipo de los dos alelos de los polimorfismos de un solo nucleótido *rs1257016*, *rs971779*, *rs238547*, *rs1482326*, *rs4760332*, *rs2104388* y *rs2427622* (reivindicaciones 1-11).

El documento D01 divulga unos cebadores y sondas con combinaciones de polimorfismos de un solo nucleótido (SNP) que se pueden ser útiles como biomarcadores predictivos de la ascendencia u origen geográfico de un sujeto (ver página 1, párrafo [0004]; página 3, párrafos [0014], [0015]; página 28, Tabla-I).

El documento D02 divulga un método para identificar el origen geográfico de un sujeto en muestras de ADN, mediante la inserción de polimorfismos basados en la intercalación de elementos dispersos largos (LINE) y elementos dispersos cortos (SINE), utilizando un análisis de tipo inserciones *Alu* (ver página 2, párrafos [0021] - [0026]; reivindicaciones 1-9).

**1. NOVEDAD Y ACTIVIDAD INVENTIVA (Art. 6.1 y Art. 8.1 LP 11/1986)****1.1. REIVINDICACIONES 1-11**

El documento D01 se considera el más cercano al estado de la técnica ya que anticipa la relación entre la presencia de determinados SNPs que se pueden ser útiles como biomarcadores predictivos del origen geográfico de un sujeto.

La diferencia entre el documento D01 y el objeto técnico de la invención radica en los propios polimorfismos anticipados como marcadores genéticos del origen geográfico de un sujeto que son diferentes y no están relacionados con los reivindicados en la presente solicitud. Además, el número de polimorfismos necesario para llevar a cabo el método de la invención es menor, resolviendo el problema con un panel de siete polimorfismos. De este modo, se considera que el método de la invención proporciona una alternativa diferente a lo divulgado en el estado de la técnica.

En consecuencia, según lo divulgado en D01 las reivindicaciones 1-11 cumplen con los requisitos de novedad y actividad inventiva (**Art. 6.1 y Art. 8.1 LP11/1986**).

Aunque el documento D02 también anticipa polimorfismos como marcadores genéticos del origen geográfico de un sujeto, estos no guardan relación con los reivindicados y se considera que este documento se refiere al estado de la técnica y no es relevante en relación al objeto técnico de la invención.