

UNIVERSIDAD DE LA LAGUNA

**Composición genética y posible origen paterno
de las poblaciones humanas canarias,
deducidos de su polimorfismo en el cromosoma Y**

Autor: Flores Infante, Carlos A.

Director: Vicente Martínez Cabrera

Departamento de Parasitología, Ecología y Genética

Dr. VICENTE MARTÍNEZ CABRERA, PROFESOR TITULAR DE GENÉTICA
DE LA FACULTAD DE BIOLOGÍA DE LA UNIVERSIDAD DE LA LAGUNA,
CON DNI N° 35208729-F

CERTIFICA:

Que Carlos Flores Infante, Licenciado en Biología, ha realizado bajo su dirección el trabajo titulado “Composición genética y posible origen paterno de las poblaciones humanas canarias, deducidos de su polimorfismo en el cromosoma Y”, el cual reúne las condiciones necesarias para ser presentado como Tesis Doctoral.

Fdo. Dr. Vicente Martínez Cabrera

La Laguna, Diciembre de 2001

Este trabajo ha sido realizado gracias al Ministerio de Educación y Cultura, por la financiación del proyecto de investigación “*Composición genética y posible origen paterno de las poblaciones humanas canarias, deducidos de su polimorfismo en el cromosoma Y*” (PB96-1034) y por la concesión de una beca de Formación de Personal Investigador de la cual me he beneficiado entre los años 1997-2001. Igualmente deseo expresar mi gratitud a la Consejería de Educación, Cultura y Deportes del Gobierno de Canarias, por apoyar en el trabajo con el proyecto de cofinanciación COF1999/019.

En primer lugar, deseo agradecer a todos los miembros del Área de Genética de la Facultad de Biología de La Universidad de La Laguna por apostar por mí en la realización de un trabajo tan interesante en el momento más oportuno.

A Marta Wallenius por estar siempre a mi lado, y a Fernando Domínguez por su constante preocupación y apoyo.

Al Dr. L. Espino por la obtención de la mayoría de las muestras de sangre del África Negra, a A. Babou y C. Pérez por la de saharauis y mauritanos, y al Dr. A. Noya por la de árabes y bereberes de Marruecos.

A los Drs. A. Oterino, G. Calderón, A. Romany y A. Rojas por la recolección de las muestras de sangre peninsulares.

A las unidades de hematología de los Centros Hospitalarios y de Salud del Archipiélago por colaborar en la obtención de las muestras de sangre de las islas, y de una manera especial a los Drs. A. Argani, F. Feraz, A. Sosa, A. Benítez, A. Milena, C. González y J. Morilla por todas las facilidades en los muestreos.

A J. Alonso por su ayuda en la creación de los mapas de frecuencia, y al Dr. P. Underhill por sus inestimables consejos y por compartir información antes de su publicación.

A mis padres

En memoria de D. Juan Infante Herrera

ÍNDICE

I. INTRODUCCIÓN.....	1
1. POBLAMIENTO PREHISTORICO Y EVIDENCIAS.....	2
2. CONQUISTA Y COLONIZACIÓN.....	4
3. EVIDENCIAS DE MEZCLA ABORIGEN EN LA POBLACIÓN CANARIA ACTUAL: ESTUDIOS PREVIOS.....	5
3.1. Antropología física.....	5
3.2. Evidencias genéticas.....	6
3.2.1. <i>Marcadores nucleares.....</i>	<i>6</i>
3.2.2. <i>ADN mitocondrial.....</i>	<i>8</i>
4. EL CROMOSOMA Y.....	9
5. OBJETIVOS.....	10
II. MATERIAL Y MÉTODOS.....	13
1. BÚSQUEDA DE VARIACIÓN GENÉTICA.....	14
2. MUESTRAS.....	17
3. EXTRACCIÓN DE ADN.....	19
3.1. Extracción con el método del fenol/cloroformo.....	19
3.2. Extracción con el método de sosa/Tris-HCl.....	20
4. MARCADORES MOLECULARES.....	20
4.1. Marcadores de evolución lenta (UEP).....	21
4.1.1. <i>Amplificaciones de UEP.....</i>	<i>21</i>
4.1.2. <i>Detección de las variantes en los UEP del tipo inserciones/deleciones.....</i>	<i>22</i>
4.1.3. <i>Detección de las variantes en los UEP del tipo SNP.....</i>	<i>23</i>
4.1.4. <i>Marcadores UEP usados.....</i>	<i>24</i>
• Locus <i>DYS287.....</i>	<i>24</i>
• Marcador <i>p12f2.....</i>	<i>24</i>
• Locus <i>DYS271.....</i>	<i>25</i>
• Locus <i>92R7.....</i>	<i>27</i>
• Locus <i>SRY1532.....</i>	<i>28</i>
• Locus <i>SRY2627.....</i>	<i>29</i>
• Locus <i>PN2.....</i>	<i>30</i>

• Locus Tat.....	32
• Locus M9.....	33
• Locus M13.....	34
• Locus M26.....	35
• Locus M34.....	36
• Locus M52.....	38
• Locus M67.....	39
• Locus M70.....	40
• Locus M78.....	41
• Locus M81.....	42
• Locus M89.....	44
• Locus M123.....	45
• Locus M153.....	46
• Locus M170.....	48
• Locus M172.....	49
• Locus M173.....	50
• Locus M201.....	51
4.2. Marcadores de evolución rápida	
(microsatélites o STR).....	53
4.2.1. <i>Amplificación de los STR</i>	53
4.2.2. <i>Detección de variantes en los STR</i>	54
4.2.3. <i>Marcadores STR descritos por primera vez</i>	54
• <i>DYS466</i>	54
• <i>DYS467</i>	55
4.2.4. <i>Marcadores STR tomados de la bibliografía</i>	56
• <i>DYS391</i>	56
• <i>DYS393</i>	57
• <i>DYS434</i>	57
5. PURIFICACIÓN DE PRODUCTOS AMPLIFICADOS.....	58
5.1. Precipitación con polietilenglicol (PEG).....	58
5.2. Precipitación con acetato amónico.....	59
6. SECUENCIACIÓN DE PRODUCTOS AMPLIFICADOS.....	59

6.1. Reacciones de secuenciación.....	59
6.2. Separación de los productos de la reacción de secuenciación.....	61
7. PROGRAMAS Y ANÁLISIS ESTADÍSTICO.....	62
7.1. Construcción de los cebadores.....	62
7.2. Estructura y relaciones poblacionales.....	62
7.2.1. <i>Análisis de varianza y distancias genéticas.....</i>	62
7.2.2. <i>Representaciones gráficas de las distancias genéticas.....</i>	63
7.2.3. <i>Correlaciones entre distancias genéticas y geográficas.....</i>	65
7.2.4. <i>Correlaciones entre variables.....</i>	65
7.2.5. <i>Mapas de frecuencia.....</i>	66
7.3. Estimaciones de mezcla.....	66
7.4. Contribuciones sexuales por área geográfica.....	67
7.5. Redes abreviadas promedio y diversidad en STR asociada a los haplotipos.....	67
7.6. Edad de los haplotipos.....	68
7.7. Porcentaje de divergencia por haplotipo y edad de colonización.....	68
III. RESULTADOS.....	70
1. NUEVOS POLIMORFISMOS.....	71
1.1. <i>Secuenciación de elementos <i>Alu</i>: detección del polimorfismo <i>DYS466</i>.....</i>	71
1.2. <i>Sondeo en las bases de datos: detección del polimorfismo <i>DYS467</i>.....</i>	73
2. CLASIFICACIÓN Y RELACIÓN DE LOS HAPLOTIPOS.....	74
3. VARIANTES EN LOS STR.....	78
4. ÁFRICA.....	78
5. PENÍNSULA IBÉRICA Y NORMANDÍA.....	86
6. ISLAS CANARIAS.....	93
7. RELACIONES GLOBALES DE LAS POBLACIONES	

CONSIDERADAS.....	98
7.1. Canarias y Europa: las repercusiones del proceso de conquista y recolonización en el pool génico paterno.....	98
7.2. La relación de Canarias y la Península Ibérica con las poblaciones africanas.....	102
8. CONTRIBUCIONES PARENTALES A LA POBLACIÓN ACTUAL DE CANARIAS.....	107
IV. DISCUSIÓN.....	112
1. ÁFRICA.....	113
1.1. Los haplotipos del África subsahariana.....	113
1.2. Los haplotipos de asignación norteafricana.....	116
1.3. Haplotipos de origen no africano.....	119
2. PENÍNSULA IBÉRICA Y NORMANDÍA.....	122
2.1. Los haplotipos de asignación europea.....	122
2.2. Los haplotipos de asignación caucásica y de Oriente Próximo.....	125
2.3. Los haplotipos de asignación africana.....	129
2.4. Haplotipos indeterminados.....	132
3. ISLAS CANARIAS.....	134
3.1. Los linajes paternos y la colonización aborigen de Canarias.....	139
3.2. Hipótesis sobre el proceso de colonización aborigen a la luz del análisis del cromosoma Y en las poblaciones canarias actuales.....	144
V. CONCLUSIONES.....	147
VI. BIBLIOGRAFÍA.....	151
APÉNDICE.....	174

I. INTRODUCCIÓN

Los estudios antropológicos y arqueológicos han permitido la construcción de hipótesis sobre el origen, evolución y dispersión de nuestra especie. La reciente incorporación de técnicas de genética molecular a la caja de herramientas del antropólogo ha hecho que la Antropología se transforme en una ciencia multidisciplinar, proporcionando nuevas perspectivas en el estudio de la evolución humana. El gran salto en esta disciplina se ha producido gracias al estudio de la variación directamente a nivel del ADN. Las secuencias de ADN son mucho más informativas que las secuencias de proteínas o las variaciones electroforéticas de éstas, ya que el código genético es degenerado y una gran parte de las secuencias de ADN no codifica proteínas. Tanto las regiones no codificantes como las sustituciones silenciosas en regiones codificantes sólo pueden estudiarse en el análisis de secuencias de ADN. Aún más, las secuencias revelan información detallada sobre el polimorfismo en sí, es decir, si se debe a una sustitución nucleotídica, si es una inserción/delección, etc... Más recientemente, la incorporación de marcadores genéticos transmitidos unisexualmente, como son el ADN mitocondrial (por vía materna) y el cromosoma Y (por vía paterna), junto con el análisis de marcadores nucleares y la información arqueológica, han aportado una visión global de los procesos demográficos humanos que han tenido lugar a lo largo de su historia evolutiva.

El origen y pervivencia de los aborígenes canarios en la población actual ha sido objeto de múltiples estudios de Antropología física. El desarrollo de las técnicas moleculares ha permitido utilizar la variabilidad genética como una herramienta más para el mismo fin. Hasta este momento, los datos de los que se disponía ofrecían una visión basada en loci nucleares y ADN mitocondrial. El objetivo de este trabajo es el uso de marcadores del cromosoma Y para proporcionar una nueva perspectiva del poblamiento y la colonización del Archipiélago Canario.

1. POBLAMIENTO PREHISTÓRICO Y EVIDENCIAS

El Archipiélago Canario pertenece geográficamente al continente africano, encontrándose situado entre las latitudes 27° 37' y 29° 25' N y las longitudes 13° 20' y 18° 10' O. Aunque

parece que se sabía de su existencia desde la literatura clásica (Martínez, 1992), no fue incorporado al mundo europeo hasta el siglo XV (Suárez y col. 1988). Cuando la Corona de Castilla inicia el proceso de conquista de las islas en 1402, se encuentran con unas islas habitadas por aborígenes de tez blanca anclados en una cultura neolítica.

Se han lanzado muchas hipótesis acerca del origen de los aborígenes canarios: pudieron ser de procedencia egipcia, cartaginesa o incluso vikinga; se les ha asociado con un sustrato norteafricano común que incluiría a los vascos; y hasta una colonización forzosa al ser transportados hasta allí por el rey Juba II de Mauritania (Hernández, 1999). Si bien algunas de estas hipótesis pueden estar basadas en la mitología, otras cobran fuerza por la sencilla razón de no haberse logrado demostrar el conocimiento de la navegación en la cultura aborígen. En cualquier caso, las mayores semejanzas antropológicas y culturales de los aborígenes canarios se encuentran en el noroeste de África y el Sahara (González y Tejera, 1990; Martínez, 1996).

El momento en el cual los aborígenes llegaron a las islas es también contradictorio. El registro arqueológico parece indicar que el noroeste africano sufrió una marginación cultural, desarrollando tecnologías autóctonas durante largo tiempo y recibiendo pocas influencias externas (Newman, 1995). Las dataciones apoyadas en características culturales estiman la llegada a Canarias aproximadamente hace 2.500-1.000 años a.C. (Cuscoy, 1963). Sin embargo, la mayoría de las dataciones por C^{14} oscilan entre 500 y 1.000 años d.C., aunque unas pocas parecen datar de ≈ 500 años a.C. (Martínez, 1996). Parece bastante aceptado que la llegada de los aborígenes a Canarias no pudo ser mucho antes del primer milenio a.C. (Martínez, 1996; Hernández, 1999).

En la mayoría de las islas se diferencian varias fases culturales superpuestas (Cuscoy, 1963; Swidetzky, 1963; Onrubia-Pintado, 1987), que habitualmente se atribuyen a distintas entradas de población. Por tanto, parece lógico pensar en los aborígenes como un mosaico étnico y no como una población homogénea. La mayoría de los autores defienden, al menos, dos

llegadas: una primera oleada, procedente probablemente de la zona del Magreb, que crearía la cultura de sustrato en todas las islas; en algunas se detecta una segunda oleada que incorpora elementos culturales saharianos a la cultura de sustrato. Fuere como fuere, lo que sí es cierto es que en cada isla florecen elementos arqueológicos diferenciados. En el conjunto de las Islas Canarias se encuentran herramientas, construcciones y grabados que, en la mayoría de las ocasiones, tienen sus paralelismos en el Magreb y el Sahara, atestiguando la presencia de poblaciones relacionadas con el noroeste africano (Arco y Navarro, 1987).

2. CONQUISTA Y COLONIZACIÓN

El proceso de conquista (1402-1496) se desarrolla en dos etapas. En la primera, contingentes normandos bajo la Corona de Castilla ocupan, en un proceso rápido y casi sin encontrar resistencia, Lanzarote, Fuerteventura y El Hierro, las islas menos pobladas de aborígenes del Archipiélago (Arco y Navarro, 1987). Éstas van a permanecer unos 30 años bajo el dominio normando. Aunque el sometimiento inicial de La Gomera (1420) también se realizó de una manera pacífica durante este siglo, debido a sus repetidas rebeliones se ejecuta y deporta a muchos de los aborígenes gomeros. La segunda fase de la conquista se lleva a cabo directamente por la Corona de Castilla, ocupando, tras violentos enfrentamientos, las islas de Gran Canaria, La Palma y Tenerife (Suárez y col. 1988).

A pesar del efecto devastador de la conquista, muchos de los aborígenes permanecen en las islas como esclavos o incluso libres. Dado que las islas conquistadas por los normandos no albergaban mucha población aborígen ésta se intentó preservar, siendo El Hierro la más afectada. La política de Isabel la Católica orientada a la protección de los indígenas favoreció la supervivencia del pueblo aborígen. Aún así, el proceso de conquista en Gran Canaria, La Palma y Tenerife diezma sus poblaciones (Hernández, 1999).

La forma en que cada isla es conquistada también repercute en la estructura poblacional de las mismas durante la colonización. En general, el pueblo canario, compuesto de aborígenes

que han quedado en Canarias y europeos asentados durante la conquista, comienza a sufrir nuevas influencias en un proceso demográfico que aún hoy perdura. Las islas de Lanzarote, Fuerteventura y El Hierro reciben, en su mayoría, normandos, moriscos y andaluces, aumentando en número de habitantes gracias principalmente a estos últimos. En La Gomera parece que también se asienta, en gran abundancia, población de origen andaluz. Las poblaciones de Gran Canaria, La Palma y Tenerife crecen atrayendo principalmente a portugueses, gallegos, andaluces y en menor medida a italianos, flamencos, judíos, moriscos y subsaharianos (Suárez y col. 1988; Pérez, 1991). Estos últimos eran traídos a las islas principalmente como mano de obra para las plantaciones e ingenios de caña de azúcar, importados desde Cabo Verde o directamente desde el África Negra. Para Gran Canaria se estima la entrada de más de 10.000 esclavos africanos a lo largo del siglo XVI (Lobo-Cabrera, 1993). A finales de este siglo, la población global de Canarias es de unos 35.000 habitantes (desde unos 800 en El Hierro hasta unos 18.000 en Tenerife), de los que dos tercios eran de origen europeo y el tercio restante aborígenes y africanos. El crecimiento demográfico ha continuado desde entonces, llegándose, a principios del siglo XX, a una población de unos 360.000 habitantes y superando el millón y medio en la actualidad.

3. EVIDENCIAS DE MEZCLA ABORIGEN EN LA POBLACIÓN CANARIA ACTUAL: ESTUDIOS PREVIOS

3.1. Antropología física

Desde un principio los estudios antropológicos revelaron, al menos, dos tipos humanos en los restos prehistóricos (Verneau, 1887), el Cro-Magnon y el Mediterraneoide. Algunos caracteres negroides también fueron detectados (Falkenburger, 1939-40), aunque disminuidos en estudios sucesivos (Schwidetsky, 1956; Fusté, 1959). Las mayores afinidades se encontraron con el tipo Mechta-Afalou y con el proto-Mediterranoide, asociados en el Magreb a culturas paleolíticas y neolíticas, respectivamente (Camps-Fabrer, 1989). Por la relativa abundancia de ambos tipos, se

establecieron dos grupos de islas: un grupo central que incluía a La Gomera y Tenerife, donde el tipo Cro-Magnon era más abundante, y un grupo periférico formado por el resto de las islas, donde el tipo Mediterraneo era el importante. Estudios óseos, comparando colecciones prehistóricas y rurales modernas de Gran Canaria y Tenerife, favorecen la persistencia de caracteres prehistóricos en la población actual (Fusté, 1959; Schwidetsky, 1975).

3.2. Evidencias genéticas

3.2.1. Marcadores nucleares

El primer análisis genético de la población canaria moderna se realizó con el grupo sanguíneo Rh (Guasch y col. 1952) y el resultado fue el de una típica población europea con cierta mezcla africana. Esta conclusión quedó corroborada en todos los estudios siguientes usando grupos sanguíneos (Bravo y de las Casas, 1958; Roberts y col. 1966; Parejo, 1966; Rössing, 1967; Pinto y col. 1996c). En el caso concreto del sistema AB0, La Gomera es la isla que presenta una mayor frecuencia del alelo 0 y en un valor muy similar al de las poblaciones del Magreb (Pinto y col. 1996c). A este respecto merece destacarse que, en los años sesenta se obtuvieron las frecuencias para los grupos sanguíneos AB0 de tejidos procedentes de 370 momias aborígenes de Gran Canaria y Tenerife (Schwarzfischer y Liebrich, 1963), encontrándose una incidencia del grupo 0 muy similar a la encontrada en la población actual de La Gomera (Pinto y col. 1996c). Este hecho relaciona claramente a la población canaria, tanto la aborigen como la actual, con las poblaciones del noroeste africano.

Los polimorfismos enzimáticos también confirmaron la existencia de un componente africano en la población canaria. Para estos polimorfismos la certeza es mayor, dado que se encuentran alelos asociados a poblaciones africanas, aunque subsaharianas. Por ejemplo, la variante alélica R del locus ACP-1, presente en Gran Canaria (0'8%) y Lanzarote (0'6%), no fue detectada en la muestra de la Península Ibérica mientras que alcanza valores de 1'1 y 6'8% en poblaciones del noroeste y la región subsahariana del continente africano, respectivamente

(Martell y col. 1986; Morilla y col. 1988; Pérez y col. 1991; Pinto y col. 1994). El alelo negroide G6PD A+, presente en aproximadamente un 2% de los magrebíes y en un 27% de los subsaharianos, también se encontró en todas las islas, con frecuencias que oscilaron entre un 0'5% en La Gomera y un 3% en Lanzarote (Morilla y col. 1988; Afonso y col. 1989). Por su semejanza molecular, estos alelos en Canarias han sido relacionados con influencias subsaharianas más que con el Magreb (Pinto y col. 1996b). La variante G6PD Seattle en Canarias (Cabrera y col. 1996) también se ha relacionado con aportaciones africanas, ya que está presente en el noroeste de África y, aunque se ha detectado en poblaciones europeas, no se encontró en una muestra de canarios descendientes de peninsulares pero sí en la autóctona (Pinto y col. 1994).

También se ha explicado como una contribución aborígen la alta incidencia en Canarias, frente a su distribución en Europa, de la hemoglobinopatía *Hb D-Los Ángeles* (Malcorra y col. 1990). Por otro lado, al igual que en el caso de los polimorfismos enzimáticos, Esteban y col. (1998) sólo pueden determinar una cierta influencia subsahariana en las islas usando los loci KM y GM junto a polimorfismos de restricción en genes que codifican inmunoglobulinas.

En términos generales, la información obtenida de varios marcadores ligados ha sido superior al uso de marcadores independientes. La combinación de un microsatélite del gen CD4 y un elemento polimórfico *Alu* ha proporcionado interesantes resultados para inferir el origen africano y la posterior dispersión del hombre moderno, ya que se detectó una disminución de la diversidad en este sistema con la distancia al continente africano (Tishkoff y col. 1996). Recientemente la aplicación de esta combinación a Canarias ha sido también muy fructífera (Flores y col. 2001b). Los resultados obtenidos detectaron que sólo las dos islas centrales, Gran Canaria y Tenerife, y la más occidental, El Hierro, muestran diferencias significativas con respecto al noroeste africano. Aún más, gracias a una de estas combinaciones, que se había relacionado exclusivamente con poblaciones del noroeste africano (Flores y col. 2000b) y que se

encontró en la mayoría de las islas, se pudo demostrar inequívocamente y por primera vez con marcadores nucleares el parentesco de Canarias con el noroeste africano sin asociarla a variantes negroides.

3.2.2. ADN mitocondrial

El ADN mitocondrial se hereda exclusivamente por vía materna, no presentando fenómenos de recombinación. Por otro lado, exhibe una tasa de mutación \approx 5-10 veces mayor que la nuclear (Ingman y col. 2000). Estas características hacen que haya sido muy usado para estudiar la contribución materna en el origen y dispersión de las poblaciones humanas (Cann y col. 1987; Vigilant y col. 1989; Piercy y col. 1993; Horai y col. 1996; Watson y col. 1997; Richards y col. 1998). En el primer trabajo sobre ADN mitocondrial en Canarias (Pinto y col. 1996a) se manifestó la estrecha relación entre los linajes maternos del noroeste africano y el de las islas y, además, se infirió que la contribución aborigen materna a la población actual canaria era superior a la paterna, inferida indirectamente de marcadores nucleares. Un estudio más exhaustivo de polimorfismos mitocondriales en las siete islas (Rando, 1998) reveló la existencia en todas ellas del grupo de linajes U6, considerado específico del noroeste africano (Côrte-Real y col. 1996; Rando y col. 1998), alcanzando la mayor frecuencia (36%) en La Gomera. Además, dado que el número de secuencias diferentes y la diversidad molecular dentro de este grupo disminuía con la distancia al continente africano, se utilizó esa información para inferir cómo se pudieron colonizar las islas. Los autores detectaron sólo un evento colonizador, proponiendo un proceso de pasos sucesivos entre islas vecinas, empezando por las más próximas a la costa africana y terminando en las más alejadas (Rando y col. 1999). Estudios con ADN mitocondrial antiguo de los restos aborígenes (Tenerife, La Gomera y La Palma) han desvelado que muchos de los linajes que hoy se detectan en Canarias (Rando y col. 1999) también existían en la población aborigen. Más aún, un gran número de ellos están relacionados con poblaciones

saharianas o subsaharianas, apoyando una influencia negroide desde la propia población aborigen (Rando, 1998).

4. EL CROMOSOMA Y

El cromosoma que determina la masculinidad contiene una región no recombinante, que constituye el 95% de su longitud, flanqueada en los dos extremos por regiones pseudoautosómicas (Jobling y Tyler-Smith, 1995). Por tanto, la región no recombinante del cromosoma Y es transmitida estrictamente, salvo por la acumulación gradual de mutaciones, de padres a hijos, siendo la contrapartida del ADN mitocondrial. A diferencia de éste, el cromosoma Y es bastante grande ($\approx 60 \times 10^6$ pares de bases en comparación con las $\approx 16 \times 10^3$ pares de bases del ADN mitocondrial) y posee secuencias muy complejas. Hasta hace unos 2 años había muy poca información acerca de su secuencia. Además, debido a que su tasa de mutación puntual es unas 5-10 veces menor que la del ADN mitocondrial (Thomson y col. 2000) se habían detectado pocos polimorfismos. Sin embargo, gracias al uso de técnicas para la detección selectiva de variantes (Underhill y col. 1997) se han descrito en la actualidad más de doscientos polimorfismos (Underhill y col. 2001b; Hammer y col. 2001) que, asociados a marcadores de evolución rápida, como son los microsátelites (Jobling y Tyler-Smith, 1995), han permitido estudiar la evolución a distintas escalas de tiempo. Otra ventaja de su estudio se debe precisamente a su baja tasa de mutación, ya que minimiza los fenómenos de recurrencia o reversión de variantes y, por tanto, la presencia compartida de una mutación definirá, con alta probabilidad, un grupo de cromosomas relacionados por descendencia. Es más, sus estados ancestrales pueden estimarse por comparación con otras especies de primates. De esta manera, la construcción de árboles filogenéticos resulta más sencilla y exacta (Underhill y col. 2000).

Estudios recientes con el cromosoma Y han demostrado su eficacia para elucidar el origen, evolución y dispersión de las poblaciones humanas de África (Scozzari y col. 1999), Europa (Rosser y col. 2000; Semino y col. 2000), Asia (Su y col. 1999, 2000; Wells y col. 2001),

América (Karafet y col. 1999) y Oceanía (Capelli y col. 2001; Underhill y col. 2001a). Las variantes de este cromosoma tienden, además, a estar más localizadas geográficamente que las de cualquier otro marcador. La fracción de varianza que se distribuye entre las poblaciones mundiales para los autosomas es de 12-18% (Cavalli-Sforza y col. 1994; Barbujani y col. 1997; Stoneking y col. 1997; Jin y col. 1999; Jorde y col. 2000; Watkins y col. 2001), para el ADN mitocondrial es de 18-30% (Stoneking y col. 1990; Seielstad y col. 1998; Jorde y col. 2000) y para el cromosoma Y es de más de un 40% (Seielstad y col. 1998; Hammer y col. 2001). En este contexto, el valor de sólo el 23% de variación encontrado para el cromosoma Y con el polimorfismo p49a,f/*TaqI*, debe considerarse una estima mínima por la recurrencia de las variantes empleadas (Poloni y col. 1997). Este alto nivel de diferenciación era teóricamente esperable, dado que el número de cromosomas Y en la población es un cuarto del número de copias de cualquier autosoma. Por lo mismo, los datos del ADN mitocondrial se esperarían equiparables a los del cromosoma Y. Se ha llegado a la conclusión de que la patrilocalidad (practicada en más del 70% de las sociedades humanas) y, en menor medida, la poliginia son las causas más probables del agrupamiento geográfico del cromosoma Y y de la mayor dispersión del ADN mitocondrial humano (Seielstad y col. 1998). Aunque no se ha demostrado, la selección diferencial también podría causar este efecto (Whitfield y col. 1995; Jobling y Tyler-Smith, 2000; Thomson y col. 2000). A pesar de que los marcadores habitualmente usados son considerados neutrales, la ausencia de recombinación hace que tanto el cromosoma Y como el genoma mitocondrial actúen como unidades frente a la selección. En contradicción con esta posibilidad se ha detectado recientemente una alta similitud en la distribución geográfica, a nivel mundial, de marcadores del cromosoma 21 y linajes del cromosoma Y (Jin y col. 1999).

5. OBJETIVOS

Por la síntesis expuesta anteriormente, es evidente que un estudio de las poblaciones humanas del Archipiélago debe incluir comparaciones con las que dieron lugar al pueblo canario

actual, y por tanto, un análisis obligado de las poblaciones continentales tanto africanas como europeas. Concretamente, las poblaciones africanas de uso necesario son las del noroeste, lugar señalado como el más probable origen de sus colonizadores prehistóricos, así como del área subsahariana, con muestras de Senegal, por ser la zona más cercana, y de aquellas de las que procedieron la gran mayoría de los esclavos negros del Archipiélago, como son los diferentes pueblos del golfo de Guinea. En Europa, las muestras de referencia deben incluir, sobre todo, poblaciones de la Península Ibérica, haciendo énfasis en regiones como Andalucía y el noroeste peninsular, históricamente documentadas como zonas de alta influencia en la colonización de las Islas Canarias. De este modo, en este trabajo se proponen los siguientes objetivos:

1. Búsqueda de variación genética en el cromosoma Y, mediante el uso de dos metodologías:
 - a) La secuenciación de regiones potencialmente variables, como son los elementos *Alu*.
 - b) La localización de regiones con alta probabilidad de variación en bases de datos, por el alto número de repeticiones que alberguen, y su confirmación experimental.
2. Análisis de la variación del cromosoma Y, usando marcadores escogidos selectivamente de la bibliografía junto a los descubiertos durante el desarrollo del trabajo, en muestras recolectadas de África, Península Ibérica e Islas Canarias.
3. Estudio de las relaciones en las poblaciones continentales y en las Islas Canarias, estimando el grado de heterogeneidad dentro y entre ellas, así como la identificación de linajes paternos específicos de cada área.
4. Cuantificación de las contribuciones paternas de las poblaciones continentales a la población canaria en su totalidad y pormenorizada para cada isla, al objeto de diferenciar

el grado de influencia europea y aborígen en cada una de ellas y contrastarla con la evidencia histórica.

5. Estimación de la antigüedad y modo en que se produjo la colonización prehistórica del Archipiélago, en base a la diversidad y distribución de linajes de posible asignación aborígen.

6. Comparación de la información resultante del análisis del cromosoma Y con la obtenida previamente para marcadores genéticos nucleares y de ADN mitocondrial.

II. MATERIAL Y MÉTODOS

1. BÚSQUEDA DE VARIACIÓN GENÉTICA

El escaso número de polimorfismos descubiertos en el cromosoma Y, en el momento en que comenzó este trabajo, obligó a la búsqueda de nuevas variantes útiles para el estudio de poblaciones atlánticas y mediterráneas. Para ello, se seleccionó el clon 264, M, 20 (número de acceso AC004617) de las bases de datos electrónicas, que contiene unas 175.000 pares de bases (pb) específicas de la región no recombinante del cromosoma Y, situadas muy cerca de la región centromérica en Yq11.1, y se adoptaron dos metodologías para el sondeo de posibles variantes en el mismo:

- a) Escrutinio mediante secuenciación directa de 10 elementos *Alu* (Tabla 1), con un tamaño medio de 350 pb, en 28 muestras de sangre de individuos no emparentados de diversas regiones geográficas (Tabla 2). Estos elementos fueron amplificados en reacciones de 50 μ l que constaron de 1 U de Ecotaq ADN polimerasa (Ecogen), 10 pmol de cada cebador, 0'2 mM de deoxinucleótidos trifosfato (dNTP), 2'5 mM de $MgCl_2$, 16'6 mM $(NH_4)SO_4$, 67 mM Tris-HCl pH = 8'8 y 50-100 ng de ADN molde. Un ciclador térmico PTC-100 (MJ Research) mantuvo el programa de amplificación 2 min. a 94° C seguido por 35 ciclos a 94° C durante 30 seg., una temperatura de anillamiento adecuada (ver Tabla 1) durante 30 seg. y 72° C durante 30 seg., para terminar con un ciclo de extensión a 72° C durante 5 min. Dado que las secuencias repetitivas del genoma habitualmente se eliminan en las exploraciones de variabilidad (Shen y col. 2000; Underhill y col. 2000), los posibles polimorfismos que pudieran aparecer serían inequívocamente nuevos. Muchos polimorfismos descritos para autosomas se deben precisamente a la inserción o deleción de estos elementos. Además, existe una cierta asociación entre la aparición de nuevos microsatélites y la cercanía de estas repeticiones *Alu*. Debido a su forma de evolución, las repeticiones

Tabla 1. Fragmentos seleccionados con elementos *Alu* del clon 264, M, 20.

Tipo <i>Alu</i>	Longitud ^a	Inicio ^b	Fin ^b	Cebador f ^c	Cebador r ^c	Anillamiento (°C)
Sx	384	2631	3014	cac aga cga ggc act tgg ta	tta ggc ggc ata gtg aga cc	57
Sg	408	7274	7681	acg cag gcc ata agt agg aa	tcc act tca aaa tca gca caa	57
Jo	173	9319	9491	ttc gta ggc tgt agg ggt gt	tct ggc ctc ttt cac tgt ca	57
Sg	354	9686	10139	ggg acc agc tgg act ttg ta	ggc tgc atg tcc tct aaa cc	57
Sg	361	26185	26545	aca aag cct gtt tgg tgg tc	tgt ctg tgc ttg tcc ctt tg	54
Jb	330	28715	29044	tgg tgt ttt ggg gtt ttg tt	agg gag atg aca ccc act tg	53
Jo	343	49999	50341	tgc gac aat gag gga tat ga	tct gaa gcc tgt ggt tta gga	50
Y	357	59829	60185	tca gca tgg agg aac tgt ca	tta aca gct ggg tgg gtt tc	53
Y	393	86446	86838	gtg acc aaa tct gtg cat gg	aaa gta tct gtt gct gct ttt gc	58
Y	399	174046	174444	cat tta gca acc gga caa gc	tca ccc ttt gta tcc atg ctc	58

^a El tamaño (en pares de bases) no incluye los cebadores.

^b Posiciones relativas al clon.

^c Secuencia 5'-3'.

Tabla 2. Muestras utilizadas en el sondeo de variación para el cromosoma Y.

Área de procedencia	Población	Nº de muestras
Europa	España	6
África	Marruecos (árabe)	4
	Marruecos (bereber)	4
	Sahara Occidental	2
	Mauritania	2
	Senegal	3
Asia	Jordania	7
Total		28

Alu dejan, en los sitios de inserción en el genoma, zonas ricas en adenina (colas de poli-A). Se supone que, por sucesivos cambios en estas regiones poli-A, se originan pequeños grupos de unidades repetitivas que posteriormente derivan en un nuevo microsatélite (Crouau-Roy y Clisson, 2000; Hancock, 1999).

- b) Búsqueda electrónica directa de secuencias con alto número de repeticiones cortas en serie como elementos potencialmente polimórficos. La selección electrónica se efectuó restringiendo el rastreo a secuencias con unidades repetitivas tetranucleotídicas y un número de repeticiones superior a 10. Se sabe que la aparición de polimorfismo en un locus con repeticiones en serie depende del aumento en número de éstas por encima de un umbral (Hancock, 1999). Además, es bien conocido que cuanto mayor sea el número de repeticiones almacenadas, mayor será su tasa de mutación y, por tanto, más polimórfico será el locus (Jin y col 1996; Brinkmann y col. 1998; Hancock, 1999).

2. MUESTRAS

Este trabajo se realizó en varones que, después de haber sido informados de los objetivos del proyecto, cedieron una muestra de su sangre anónimamente. Sólo fueron analizadas las muestras de personas no emparentadas con, al menos, dos generaciones de ancestros paternos conocidos. Los resultados del presente estudio están basados en el análisis de 1471 muestras de sangre recolectadas para la ocasión junto a datos de 362 individuos tomados de la bibliografía (Figura 1).

La muestra de África se compone de 135 individuos del noroeste (29 marroquíes, 46 bereberes de diferentes zonas de Marruecos, 24 saharauis del campo de Tindouf y 36 mauritanos de diferentes etnias) y 90 del área subsahariana (43 wolof de Senegal, 21 senufo de Costa de Marfil y 26 fang de Guinea Ecuatorial). A efectos comparativos, se usaron los resultados publicados por Underhill y col. (2000) para otras muestras africanas (88 de Etiopía, 40 de Sudan, 44 de Malí, 36 de África central, 54 de África del sur y 39 khoisan).

Como muestra europea se consideraron tan sólo individuos de la zona del suroeste del continente. Aunque se dispone de una muestra de normandos (26), como representantes franceses, en su gran mayoría son de la Península Ibérica (568), con representantes del norte (109 portugueses, 60 leoneses, 19 gallegos y 70 cántabros), centro (21 castellanos), este (31 valencianos) y sur (155 sevillanos, 22 onubenses, 28 gaditanos, 27 cordobeses y 26 malagueños). Como parte de la muestra peninsular también se incluyeron vascos (45) y catalanes (16) descritos en la bibliografía (Underhill y col. [2000] y Bosch y col. [2001], respectivamente).

La muestra de Canarias se compone de 652 individuos de las siete islas principales del Archipiélago: 47 de El Hierro, 85 de La Palma, 92 de La Gomera, 178 de Tenerife, 78 de Gran Canaria, 97 de Lanzarote y 75 de Fuerteventura.

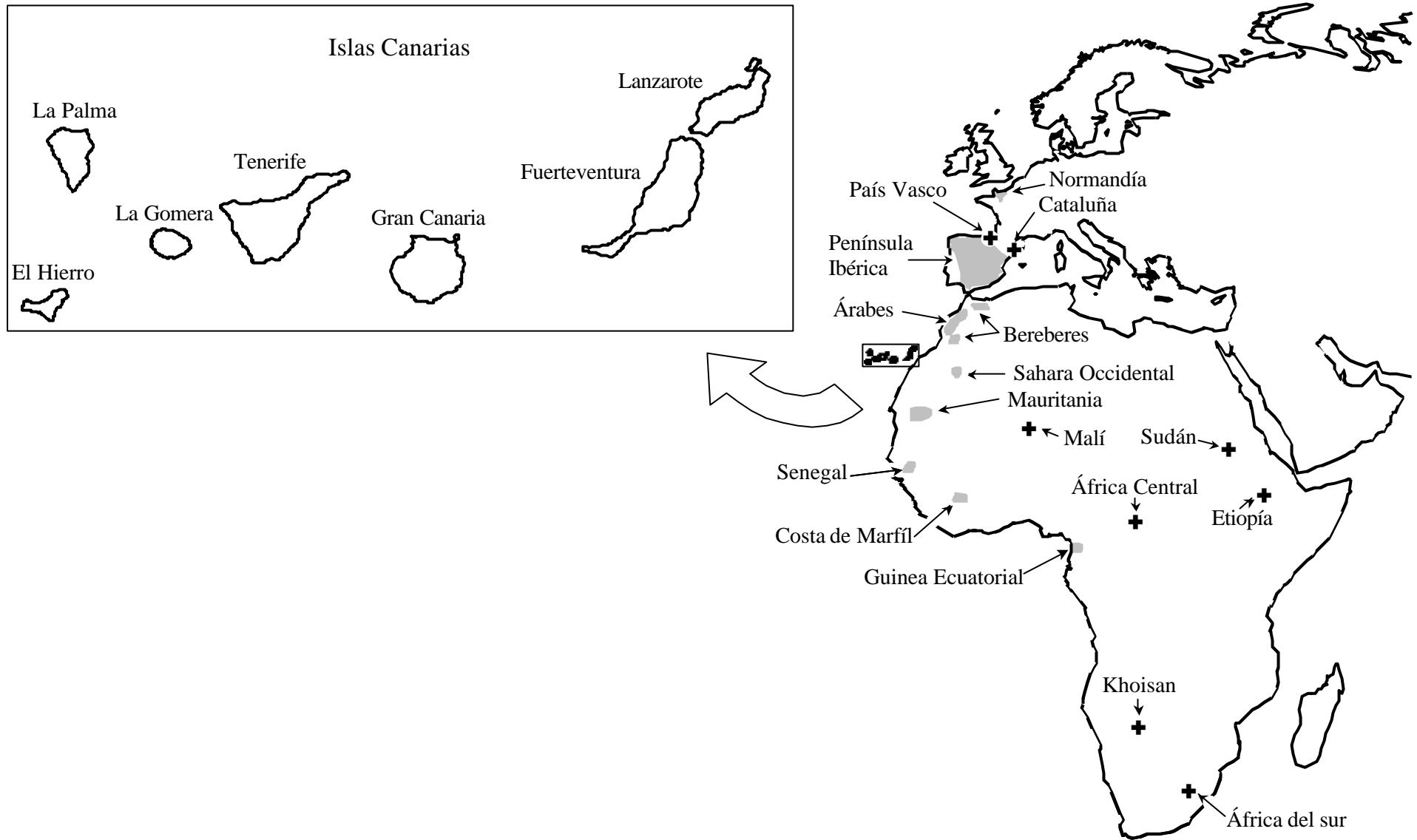


Figura 1. Situación geográfica aproximada de Canarias, de otras poblaciones muestreadas para este trabajo (áreas sombreadas) y de las muestras tomadas de la bibliografía (cruces).

3. EXTRACCIÓN DE ADN

Todas las extracciones de ADN se hicieron a partir de sangre total usando dos protocolos diferentes en función de la disponibilidad de sangre.

3.1. Extracción con el método del fenol/cloroformo

Este método de extracción se usó sólo cuando la cantidad de sangre de la que se disponía era suficiente. Para extraer 40 muestras el protocolo a seguir es:

1. Se preparan 10 ml de tampón de amplificación con 50 mM ClK, 2'5 mM MgCl₂, 10 mM Tris-HCl pH = 8'3, 0'45% Nonidet P-40 y 0'45% Tween-20.
2. Se divide el tampón de amplificación en volúmenes de 250 µl usando tubos estériles de 1'5 ml. A cada alícuota se le añaden 10 µl de dodecil sulfato de sodio (SDS) al 20% y 10 µl de Proteinasa K (20 mg/ml).
3. A cada tubo de mezcla se le añaden 250 µl de sangre agitada.
4. Se agita vigorosamente cada tubo de extracción y se lleva a un baño a 65° C.
5. Tras 3 h. de incubación las muestras se depositan en hielo prensado. Allí se les añaden 500 µl de fenol y se mantienen 3 min.
6. Se centrifugan las muestras a 7.000 rpm durante 5 min.
7. Se recupera la fase acuosa (superior) con una pipeta y se pasa a un tubo estéril de 1'5 ml. A continuación se añaden, a la fase acuosa, 250 µl de fenol y 250 µl de cloroformo (24:1, cloroformo:isoamílico). Nuevamente se agitan vigorosamente las muestras y se mantienen en hielo prensado 3 min.
8. Se realiza una nueva centrifugación a 7.000 rpm durante 5 min.
9. Se recupera la fase acuosa (superior) y se transfiere a un tubo estéril de 1'5 ml. Se les añaden 500 µl de cloroformo, se agitan vigorosamente y se introducen en hielo prensado 3 min.

10. Se centrifugan las muestras a 7.000 rpm durante 5 min.
11. Se recupera la fase acuosa (superior) y se lleva a un nuevo tubo estéril de 1'5 ml.
12. Las muestras se almacenan a 4° C hasta su utilización.

3.2. Extracción con el método de sosa/Tris-HCl

Este método se llevó a cabo tal y como fue publicado por Rudbeck y Dissing (1998) y se usó cuando se disponía de poca sangre. Para extraer 50 muestras de sangre los pasos a seguir son:

1. Se preparan 9 ml de una solución de neutralización 0'04 M Tris-HCl pH = 7'5 y 1 ml de una solución de extracción 0'2 M NaOH.
2. A tubos de 0'6 ml se añaden 5 µl de sangre agitada y 20 µl de la solución de extracción. Se agita vigorosamente cada tubo y se mantienen a temperatura ambiente durante 15 min.
3. A cada muestra se le añaden 180 µl de la solución de neutralización y se agitan vigorosamente.
4. Almacenar las extracciones a 4° C hasta su utilización.

4. MARCADORES MOLECULARES

En la realización del presente trabajo se usaron un total de 29 marcadores moleculares, 2 descubiertos durante la búsqueda de variación de este estudio y 27 de la bibliografía, que se pueden subdividir en dos tipos:

- Marcadores de evolución lenta o UEP (del Inglés Unique Event Polymorphism). Se utilizaron 24 marcadores cuidadosamente escogidos de la bibliografía para ofrecer el máximo de información en las relaciones de las poblaciones de interés para el trabajo. Estos marcadores son sucesos únicos en la historia evolutiva y, como tales, marcan relaciones entre los individuos que los comparten. Las relaciones poblacionales se

han obtenido en base a ellos. Se distinguen dos tipos en función del cambio que originen:

- a) Inserciones o deleciones de fragmentos de ADN: *DYS287* y *p12f2*.
 - b) Cambios de una sola base o SNP (del Inglés Single Nucleotide Polymorphism), que constituyen la gran mayoría de los marcadores estudiados en este trabajo: *92R7*, *DYS271*, *PN2*, *SRY1532*, *SRY2627*, *Tat*, *M9*, *M13*, *M26*, *M34*, *M52*, *M67*, *M70*, *M78*, *M81*, *M89*, *M123*, *M153*, *M170*, *M172*, *M173* y *M201*.
- Marcadores de evolución rápida, como son los microsatélites o STR (del inglés Short Tandem Repeat). Los dos polimorfismos descubiertos en este estudio pertenecen a este tipo de marcadores. Junto a éstos, se usaron otros 3 loci STR de la bibliografía: *DYS391*, *DYS393* y *DYS434*. Estos cinco loci fueron estudiados en algunos individuos previamente caracterizados para los UEP, y sólo para estimar relaciones recientes entre ellos.

4.1. Marcadores de evolución lenta (UEP)

La mayoría de estos polimorfismos de suceso único sólo tienen dos formas alélicas, una ancestral, compartida con los linajes más antiguos de la especie o incluso con otras especies cercanas, y una derivada. A pesar de ello, estos marcadores bialélicos son muy informativos para las relaciones poblacionales, pero lo son aún más cuando se combinan las variantes alélicas de cada uno de ellos para cada individuo. Cada combinación resultante de esta asociación se denomina haplotipo. Para determinar el haplotipo de cada individuo es necesario conocer cuál es la variante alélica que posee en cada uno de los loci.

4.1.1. Amplificaciones de UEP

Las amplificaciones de ADN se llevaron a cabo en reacciones de 10-20 µl que constaron de 0'4-0'8 U de Ecotaq ADN polimerasa (Ecogen), 3-5 pmol de cada cebador, 0'2 mM de

deoxinucleótidos trifosfato (dNTP), 2'5 mM de $MgCl_2$, 16'6 mM $(NH_4)SO_4$, 67 mM Tris-HCl pH = 8'8 y 10-50 ng de ADN molde. Todas las reacciones fueron incubadas en un ciclador térmico PTC-100 (MJ Research) que, mediante un programa, mantuvo la temperatura a 94° C durante 2 min. seguido por 35-40 ciclos a tres condiciones de temperatura y tiempo concretas para cada marcador:

- Un primer paso, de desnaturalización, a 94° C mantenido de 5 a 30 seg.
- Un segundo paso, de anillamiento, en el que la temperatura disminuye para que los cebadores de la reacción se unan a las cadenas de ADN separadas por el calor. La temperatura adecuada para cada pareja de cebadores en cada marcador osciló entre 45 y 68° C, y el tiempo entre 5 y 30 seg.
- Un tercer paso, de elongación de las cadenas de ADN, a 72° C en el que permanecieron entre 5 y 60 seg.

Tras la amplificación se dio un periodo de 5 min. a 72° C para la terminación de la elongación.

4.1.2. Detección de las variantes en los UEP del tipo inserciones/delecciones

Los alelos de los dos loci cuyo polimorfismo se debe a inserciones o delecciones (*DYS287* y *p12f2*) fueron detectados mediante electroforesis directa de productos de amplificación, a los que se añadió una solución de carga A (0'07 g azul de bromofenol, 0'07 g de xileno-cianol, 12 ml de glicerol y llevar con agua estéril hasta 40 ml), en geles de agarosa. Las concentraciones de agarosa usadas fueron del 1'5% para el locus *DYS287* y para el fragmento amplificado con la pareja de cebadores SY1/SY2 del locus *p12f2*, y del 1% para el fragmento amplificado con la pareja de cebadores SY1/SY3 del locus *p12f2* (ver pag. 24). Para la fabricación de un gel de agarosa se mezclan 0'4 g o 0'6 g de agarosa y 40 ml de agua estéril, para geles del 1 y 1'5%, respectivamente. Para disolver la agarosa en el agua se aplicó calor en un microondas 61BL (Basic Line) a 250 W durante 90 seg. Cuando la temperatura descendió hasta

aproximadamente los 50° C se añadieron 800 µl de tampón TAE 50x (para un litro: 242 g de Tris, 57'1 ml de ácido acético glacial y 100 ml de ácido etilendinitrilotetraacético [EDTA] 0'5 M a pH = 8). A continuación, se homogeneizó suavemente la solución y se vertió en un molde de 0'4 x 8'5 x 12'5 cm al que previamente se le había colocado un peine. Tras 30 min. de polimerización, las electroforesis se hicieron con los geles sumergidos en TAE 1x a un voltaje constante de 100 V (65 mA) durante unos 30 min. A continuación, los geles fueron teñidos en una solución de Bromuro de Etidio (1 µg/ml) durante 15 min. para ser visualizados al UV con un transiluminador Gelprinter Plus (TDI).

4.1.3. Detección de las variantes en los UEP del tipo SNP

Dado que las variantes en los loci SNP no originan ningún cambio en el tamaño de los alelos, todos ellos se testaron, una vez amplificados, mediante polimorfismos de longitud de los fragmentos originados con endonucleasas de restricción o RFLP (del inglés Restriction Fragment Length Polymorphism). Las reacciones de RFLP se llevaron a cabo, dependiendo de cada marcador, en volúmenes que oscilaron entre 10 y 15 µl, mezclando de 5 a 10 µl de producto de amplificación con el tampón adecuado para la enzima y 0'5-1 U de endonucleasa de restricción. Una vez preparada la mezcla, la digestión se incubó a la temperatura adecuada para la enzima durante 2 h. Transcurrido este tiempo, se añadió solución de carga A y se corrió la totalidad del volumen en electroforesis verticales usando geles de poliacrilamida (PAGE, del inglés Poly-Acrylamide Gel Electrophoresis) nativa de 3'5, 6 u 8% (29:1, acrilamida:N,N'-metilen-bis-acrilamida) en tampón TBE 1x (10'8 g/l de Tris, 5'5 g/l de ácido bórico, 372 mg/l de EDTA-Na₂ todo a pH = 8'3) usando una cubeta XCell™ MiniCell (NOVEX). Para preparar estos geles la cubeta incluye casetes de 10 x 10 cm con espaciadores incorporados y peines de 10-15 pocillos. Para la mezcla se hizo una solución al 30% (29:1), agregando 29 g de acrilamida y 1 g de N,N'-metilen-bis-acrilamida y llevando con agua estéril hasta 100 ml. Una vez homogeneizada, la

mezcla se hizo pasar por un filtro tipo HA (Millipore) mediante una bomba de vacío. Para construir un gel de acrilamida al 6% se montó el casete y se mezclaron 2 ml de solución de acrilamida/N,N'-metilen-bis-acrilamida al 30%, 1 ml de TBE 10x (108 g/l de Tris, 55 g/l de ácido bórico y 3'72 g/l de EDTA-Na₂ todo a pH = 8'3), 7 ml de agua estéril, 4 µl de N,N,N',N'-tetrametil-etilendiamina (TEMED) y 100 µl de peroxidisulfato de amonio (100 mg/ml). Una vez añadidos los componentes, se homogeneizó suavemente la solución, se vertió en el casete y se selló el conjunto con el peine. Tras 15 min. de polimerización, las electroforesis se corrieron a un voltaje constante de 300 V (60 mA) durante unos 25 min. y luego los geles se tiñeron en una solución de Bromuro de Etidio (1 µg/ml) durante 15 min. para ser visualizados al UV con un transiluminador.

4.1.4. Marcadores UEP usados

- Locus *DYS287*

Este locus, conocido comúnmente como YAP (del inglés, Y Alu Polymorphism), debe su polimorfismo a la presencia (YAP+, variante derivada) o ausencia (YAP-, variante ancestral) de un elemento *Alu*. Las variantes de este marcador se determinaron mediante una amplificación sencilla de ADN y separación de los productos resultantes en electroforesis de agarosa (Hammer y Horai, 1995). Las reacciones de amplificación se llevaron a cabo en volúmenes de 20 µl siguiendo un programa de 35 ciclos a 94° C durante 30 seg., 54° C durante 30 seg. y 72° C durante 30 seg. Los cebadores usados para esta reacción fueron:

YAPa 5'-cag ggg aag ata aag aaa ta -3'

YAPb 5'-act gct aaa agg gga tgg at -3'

Las muestras que tienen el alelo ancestral originan fragmentos de amplificación de 150 pb, mientras que en las que poseen el alelo derivado este fragmento aumenta hasta las 450 pb.

- Marcador p12f2

Este polimorfismo se estudió durante mucho tiempo mediante el uso de hibridaciones con sondas del plásmido p12f sobre digestos de ADN genómico humano (Casanova y col. 1985). Aunque el origen del polimorfismo era incierto, se sabía que por ganancia/pérdida de unas 2 kilobases (kb) se detectaban fragmentos de dos tamaños en estas digestiones: unos individuos con una banda de 10 kb y otros con una de 8 kb. Muy recientemente se ha atribuido este polimorfismo a la ausencia/presencia de un elemento repetitivo LINE en el locus *DYS11* (Sun y col. 2000), con un tamaño aproximado a la diferencia de las bandas detectadas por hibridación. Aún así, hay autores que consideran que la base molecular del polimorfismo es más complicada que la simple presencia/ausencia de este elemento (Rosser y col. 2000). En cualquier caso, no parece haber duda de que la pérdida del elemento LINE sólo ha ocurrido una vez durante la historia evolutiva del hombre. En este trabajo el polimorfismo de este locus fue detectado mediante el uso del protocolo descrito por Sun y col. (2000). Para cada muestra se hicieron dos reacciones de amplificación independientes en volúmenes de 10 µl. En una de ellas, que origina el fragmento SY1187, se incluyeron los cebadores SY1 (5'-caa att gta aga caa gga cat gc-3') y SY2 (5'-gca cgt tgt gca cat gta ctc-3') y se amplificó con un programa de 35 ciclos a 94° C durante 5 seg., 56° C durante 5 seg. y 72° C durante 5 seg. En la otra, que origina el fragmento SY1193, se incluyeron los cebadores SY1 y SY3 (5'-tgt cca tgt tgg atg agg aa-3') y se amplificó con un programa de 35 ciclos a 94° C durante 20 seg., 56° C durante 30 seg. y 72° C durante 60 seg. Sólo las muestras que tienen la variante ancestral (con el inserto LINE, correspondiente a la banda de 10 kb en las hibridaciones) originan producto (fragmento de 272 pb) en la amplificación de SY1187, mientras que, sólo las muestras que poseen el alelo derivado (sin el inserto LINE, correspondiente a la banda de 8 kb en las hibridaciones) originan producto (fragmento de 2078 pb) en la amplificación de SY1193.

- Locus *DYS271*

En este locus, también conocido como sY81, se origina el polimorfismo debido a una transición nucleotídica A→G. El protocolo seguido en este trabajo para su detección fue el descrito por Seielstad y col. (1994). Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores:

1380f 5'-agg cac tgg tca gaa tga ag -3'

1380r 5'-aat gga aaa tac agc tcc cc -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 20 seg., 57° C durante 20 seg. y 72° C durante 20 seg. Los productos de esta amplificación, con una longitud de 209 pb, fueron digeridos directamente con la restrictasa *NlaIII* (CATG/) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, tampón 4 de New England Biolabs (NEB) a concentración 1x y 1 U de enzima *NlaIII* (NEB R125S) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE al 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una A (variante ancestral) en la posición 168 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos en la digestión es de 102, 65 y 42 pb. Cuando es una G (variante derivada) la que aparece en esa posición, entonces el patrón de bandas es de 144 y 65 pb. En el producto de amplificación hay una diana *NlaIII* invariable, que origina el fragmento de 65 pb, y una diana variable, que se destruye al aparecer la variante derivada. La secuencia del producto de amplificación de este locus es:

aggcactggtcagaatgaatggaatggcacacaggacaagtccagaccaggaaggtccagtaacatgggagaagaacggaaggagtt
 ctaaaattcagggtcccttgggctcccctgttataaaatgtaggtttattattatatttcattgtaacaaaagtcc**Mt**gagatctgtggaggat
 aaaggggagctgtattttccatt

Las bases subrayadas corresponden a las zonas de unión de los cebadores usados y las bases en negrita a las dianas reconocidas por *NlaIII*, siendo M la base polimórfica.

- Locus 92R7

Se trata de una secuencia moderadamente repetitiva que, usada como sonda en digestiones genómicas con *HindIII*, detecta un fragmento de longitud variable (Mathias y col. 1994). Actualmente se sabe que este polimorfismo se debe a una transición C→T y se determina por amplificación y digestión (Forster y col. 2000). Las amplificaciones de ADN se realizaron en reacciones de 10 µl, donde se incorporaron los cebadores:

92R7a 5'-tgc atg aac aca aaa gac gta -3'

92R7b 5'-gca ttg tta aat atg acc agc -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 20 seg., 55°C durante 20 seg. y 72° C durante 30 seg. Los productos de esta amplificación, con un tamaño de 55 pb, fueron digeridos directamente con la enzima *HindIII* (A/AGCTT) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 1x de tampón One-Phor-All Buffer PLUS (Pharmacia Biotech R27-0901) y 1 U de enzima *HindIII* (Stratagene R500600) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y se corrió el total del volumen en PAGE al 8% (29:1) en las condiciones descritas. Cuando la muestra tiene una C (variante ancestral) en la posición 26 (desde el extremo 5' del cebador a) del producto de amplificación, el patrón de fragmentos en la digestión es de 30 y 25 pb. Cuando es una T (variante derivada) la que ocupa esa posición, entonces el fragmento aparece sin digerir (55 pb). En el producto de amplificación hay una diana *HindIII*, que se destruye al aparecer la variante derivada. Después de las digestiones siempre permanece producto sin digerir, dado que hay más de una copia de esta secuencia en el cromosoma Y. La secuencia del producto amplificado en estas condiciones es:

tccatgaacacaaaaagacgtaga**agMtt**gtctttctggcatatttaacaatgc

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *HindIII*, siendo M la base polimórfica.

- Locus *SRY1532*

El polimorfismo en este locus, también conocido como *SRY10831* (Whitfield y col. 1995), fue determinado siguiendo el protocolo descrito por Kwok y col. (1996). Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 μ l, donde se incorporaron los cebadores:

SRY1 5'-tcc tta gca acc att aat ctg g-3'

SRY2 5'-aaa tag caa aaa atg aca caa ggc-3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 20 seg., 58° C durante 20 seg. y 72° C durante 30 seg. Los productos de amplificación, con un tamaño de 167 pb, fueron digeridos directamente con la enzima *DraIII* (CACNNN/GTG) en reacciones de 15 μ l que constaron de: 10 μ l de productos de amplificación, 1x de tampón 3 (NEB) y 0'5 U de *DraIII* (NEB R510S) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 μ l de solución de carga A y el total del volumen se corrió en PAGE al 6% (29:1) en las condiciones descritas. Los posibles patrones de fragmentación son o bien una banda sin digerir, que corresponde a la variante con A en la posición 115 (desde el extremo 5' del cebador 1) del producto de amplificación, o bien bandas de 112 y 55 pb, que corresponde a la variante G en la misma posición. Se sabe que en esta posición han ocurrido dos cambios en sentido contrario en momentos evolutivos diferentes. Es decir, primero ocurrió una mutación A→G en un linaje muy antiguo de la especie humana, y más recientemente una retromutación G→A en un linaje más joven. De esta manera, conociendo el entorno de variantes que acompaña a la mutación, este polimorfismo puede distinguir dos

grupos de linajes. En cualquier caso, en el producto de amplificación hay una diana *DraIII* variable, que se destruye cuando A ocupa la posición polimórfica. La secuencia del producto de amplificación en estas condiciones es:

tccttagcaaccattaatctggttttagtctctattcattgcctttcctggatattcatatacatgggatcattcagtatctggcctctgtatctgac
 ttttcacacagtMtaacattttcaaggtcacctatgtggtgccttgtctcatttttgcattt

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *DraIII*, siendo M la base polimórfica.

- Locus *SRY2627*

El polimorfismo en este locus, debido a una transición C→T, fue determinado siguiendo un protocolo de amplificación y digestión descrito por Hurles y col. (1999). Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores:

2627f 5'-cgc ggc ttt gaa ttt caa gct ctg -3'

2627r 5'-cca ggg ccc cga ggg act ctt -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 30 seg., 68° C durante 30 seg. y 72° C durante 30 seg. Los productos de amplificación, con una longitud de 391 pb, fueron digeridos directamente con la restrictasa *BanI* (G/GPyPuCC) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, tampón 4 (NEB) a concentración 1x y 0'5 U de *BanI* (NEB R118S) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE al 3'5% (29:1) en las condiciones descritas. Cuando la muestra tiene una C (variante ancestral) en la posición 130 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos en la digestión es de 264, 86 y 41 pb. Cuando es una T (variante derivada) la que ocupa esa posición, entonces el patrón de

bandas es de 350 y 41 pb. En el producto de amplificación hay una diana *BanI* invariable, que produce el fragmento de 41 pb, y otra variable, que se destruye al aparecer la variante derivada.

La secuencia del producto de amplificación en estas condiciones es:

cgcgcttgaattcaagctctggttctgtccttgggcacctgcgcgtgaatcgttgcgaggtggccaagtaaggccccacgc
 agtttgcttccgggccaaggaagccccacag**gggtgcM**ccacagggtgaagccccatgcctacagggtgaagcggctgaagctggt
 agtggttccgaggaagcgggtcaaagtcccgtccagaggtcctcttcttgggtgtcactcccgaacccccgccaggggtctggcttccca
 tcgacacctctctctgttcagtcacatcacgtatcccacgaggccggacggcactgccacctcgtcgccccgtagactcgggccccgaa
 actacacttaagagtcctcgggcccctgg

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *BanI*, siendo M la base polimórfica.

- Locus PN2

Este locus se encuentra muy cerca del *DYS287* y su variación se debe a una transición C→T (Hammer y col. 1997). Habitualmente este polimorfismo se detecta por hibridación con una sonda, dado que el cambio no crea ni destruye diana alguna de las restrictasas conocidas. Su caracterización en este trabajo se realizó mediante un protocolo de amplificación y digestión. Para ello, las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, en un programa de 35 ciclos a 94° C durante 15 seg., 54° C durante 15 seg. y 72° C durante 15 seg., que originaron un fragmento de 104 pb. Para que la mutación creara una diana, se diseñó uno de los cebadores con una base cerca de 3' diferente a la de la secuencia real (base desapareada o "mismatch"). Así, todos los productos amplificados tendrán esta nueva base diferente a la secuencia real que, acompañada de una de las variantes alélicas, originará una diana reconocible por la restrictasa *AluI*. Los cebadores usados en estas amplificaciones fueron:

*PN2f 5'-aag gag cat taa taa aac taa **agc**-3'

PN2r 5'-gaa ctc ccg att ccc ctc ta-3'

El cebador marcado con asterisco tiene una A cerca del extremo 3' (señalada en negrita), sustituyendo a una T en la secuencia original de este segmento de ADN, que origina una diana *AluI* (AG/CC) variable en los productos de esta amplificación. Estos productos modificados fueron digeridos en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 1x de tampón One-Phor-All Buffer PLUS y 0'5 U de *AluI* (Boehringer Mannheim R239275) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE al 8% (29:1) en las condiciones descritas. Cuando la muestra tiene una C (variante ancestral) en la posición 25 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos en la digestión es de 69 y 35 pb. Cuando es una T

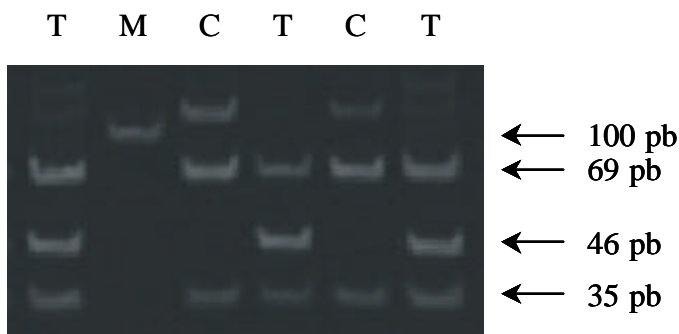


Figura 2. Detección de los alelos C y T del marcador PN2. M es un marcador de peso molecular. El fragmento de 69 pb todavía aparece en los digestos de las muestras con el alelo T, mientras que el fragmento de 23 pb no es visible en la fotografía.

(variante derivada) la que ocupa esa posición, entonces el patrón de bandas es de 46, 35 y 23 pb (Figura 2). En el producto de amplificación hay una diana *AluI* invariable, que origina el fragmento de 35 pb, y una diana variable, que se crea cuando aparece la variante derivada. La secuencia del producto de amplificación en estas condiciones es:

aaggagcattaataaaaactaaagcMttctcctcctaggggcacctctcttaaggaaaagatggctggagctgattctcctgaggtagaggg
gaatcgggagttc

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *AluI*, siendo M la base polimórfica.

- Locus Tat

El polimorfismo de este locus, debido a una transición T→C, fue detectado siguiendo el protocolo descrito por Zerjal y col. (1997). Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores:

Tat 1 5'-gac tct gag tgt aga ctt gtg a-3'

Tat 3 5'-gaa ggt gcc gta aaa gtg tga a-3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 15 seg., 54° C durante 15 seg. y 72° C durante 15 seg. Los productos de esta amplificación, con una longitud de 112 pb, fueron digeridos directamente con *NlaIII* en reacciones de 15 µl como las descritas para el marcador *DYS271*. Los productos de la digestión también fueron tratados de la misma manera. Cuando la muestra tiene una T (estado ancestral) en la posición 28 (desde el extremo 5' del cebador 1) del producto de amplificación, el patrón de fragmentos de digestión es de 86 y 26 pb. Cuando es una C (estado derivado) la que aparece en esta posición, entonces el fragmento no es digerido (112 pb). En el producto de amplificación hay una diana *NlaIII* variable, que se destruye al aparecer la variante derivada. Esta posición fue determinada también por secuenciación directa en los individuos en que apareció la variante derivada. La secuencia del producto amplificado es:

gactctgagtgtagacttgtgaattca**M**gtgttttaatttaatttcagagcacagcaagagagaagggtaaatattcaggtgattattcac
acttttacggcaccttc

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *NlaIII*, siendo M la base polimórfica.

- Locus M9

Éste y todos los SNPs que a continuación se mencionan (denominados con una M seguido de un número) han sido descubiertos recientemente con una técnica denominada “Cromatografía líquida desnaturizante a alta presión” o DHPLC (del inglés Denaturing High Performance Liquid Cromatography; Underhill y col. 1997). Este polimorfismo, producido por una transversión C→G, se determinó mediante amplificación y digestión. Las amplificaciones de ADN se llevaron a cabo en volúmenes de 10 µl, donde se incorporaron los cebadores:

M9f 5'-gca gca tat aaa act ttc agg -3'

M9r 5'-aaa acc taa ctt tgc tca agc -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 30 seg., 54° C durante 30 seg. y 72° C durante 60 seg. Los productos de amplificación, con 341 pb de longitud, fueron digeridos directamente con la enzima *HinfI* (G/ANTC) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 1x de tampón 2 (NEB) y 0.5 U de enzima *HinfI* (NEB R155S) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y todo el volumen se corrió en PAGE al 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una C (estado ancestral) en la posición 68 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos en la digestión es de 182, 95 y 64 pb. Cuando es una G (variante derivada) la que ocupa esa posición, entonces el patrón de bandas es de 246 y 95 pb. En el producto amplificado hay una diana *HinfI* invariable, que origina el fragmento de 95 pb, y una diana variable, que se destruye al aparecer la variante derivada. La secuencia del producto de amplificación es:

gcagcatataaaaacttfcaggaccctgaaatacagaactgcaaagaacggcctaagatggttgaatMctctttattttcttttaatttagaca
tgttcaaacgttcaatgtcttacatacttagttatgtaagtaaggtagcgttacttcattatgcatttcaataactcaaaaaaaattcctttgtgaaat

gttgaaatattttctaactctgtttcacgagcttcaaaaatgaggaaaaa**gattc**agtttacatttcagcaaatgcctcttttaacggattatg
 ttacttaacatttacagtagcatttacgcttgagcaaaagtaggttt

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *HinfI*, siendo M la base polimórfica.

- Locus M13

Este polimorfismo, que también se produce por una transversión G→C (Underhill y col. 1997), fue detectado por amplificación y digestión. Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores:

M13f 5'-tcc taa cct ggt ggt ctt tc -3'

M13r 5'-agc cat gat ttt atc caa cc -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 20 seg., 54° C durante 20 seg. y 72° C durante 20 seg. Los productos de amplificación, con 231 pb, fueron digeridos directamente con *NdeII* (/GATC) en reacciones de 10 µl que constaron de: 5 µl de productos de amplificación, 1x de tampón específico para la enzima (Boehringer Mannheim) y 0'5 U de enzima *NdeII* (Boehringer Mannheim R1040235) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE al 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una G (estado ancestral) en la posición 157 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos en la digestión es de 160 y 71 pb. Cuando es una C (variante derivada) la que ocupa dicha posición, entonces el producto aparece sin digerir (231 pb). En el producto de amplificación hay una diana *NdeII* variable, que se destruye al aparecer la variante derivada. La secuencia del fragmento amplificado es (invertida y complementaria con respecto a los cebadores):

agccatgattttatccaaccacatttgcaaaaaatataacaatgagaaaattataacagtaacagagaactgatMtaaccgtcttgccctaac
ctacctactgccctgttcattcctgggcataaactaacctaacttgagaaaaatttagtctatagtttaaaggagaatagtccttcccaaaacta
aatcacctttgtaaaacaatgaaagaccaccaggttagga

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *NdeII*, siendo M la base polimórfica.

- Locus M26

El polimorfismo de este locus, producido por una transición G→A (Underhill y col. 2000), no puede detectarse directamente por RFLP, dado que ninguna restrictasa reconoce la secuencia donde reside la variación. Aún así, el polimorfismo se detectó mediante amplificación y digestión siguiendo la metodología descrita para el locus PN2. Las amplificaciones de ADN se realizaron en reacciones de 10 µl, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una G en negrita cerca del extremo 3' que sustituye a una T en la secuencia real):

*M26f 5'- ttt ttc tga att aga atg atc -3'

M26r 5'- gta cac ctt tct tag gtt gc -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 10 seg., 48° C durante 10

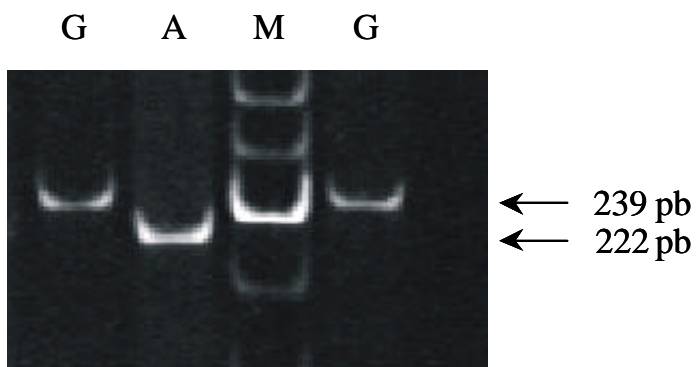


Figura 3. Detección de los alelos G y A del marcador M26. M es un marcador de peso molecular. El fragmento de 17 pb no se aprecia en la fotografía.

seg. y 72° C durante 10 seg. Los productos de esta amplificación, con 239 pb de longitud, fueron digeridos directamente con la enzima *BclI* (T/GATCA) en reacciones de 15 µl que constaron de:

10 µl de productos de amplificación, 1x de tampón 3 (NEB) y 0'5 U de enzima *BcII* (NEB R160S) por tubo. Tras 2 h. de incubación a 50° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE al 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una G (estado ancestral) en la posición 22 (desde el extremo 5' del cebador f) del producto de amplificación, la digestión resulta en un fragmento de 239 pb sin digerir. Cuando es una A (estado derivado) la que ocupa dicha posición, entonces el patrón de bandas es de 222 y 17 pb (Figura 3). En el producto de amplificación hay una diana *BcII* variable, que se crea en presencia de la variante derivada. La secuencia del fragmento amplificado es:

ttttctgaattagaatgatcMcagagaacactgaatggcctatgaaattcaattttgctgcagatttcgcatgtttcttaatgaacataact
aacttctaatacacaagataaattcttgccctatgtgcaaaaacttagtgctgcatccttggtatggttttaaaaagtgtcaaaaactggcccctcatg
tcaatacagccccaattagggaggcaacctaagaaaggtgtac

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *BcII*, siendo M la base polimórfica.

- Locus M34

La variación en este locus, debida a una transversión G→T (Underhill y col. 2000), se determinó mediante amplificación y digestión. Aunque en este caso el cambio de base produce una diferencia en el patrón de corte de la enzima *HinfI*, los cebadores usados en este trabajo no son los mismos que en el trabajo de referencia (Underhill y col. 2000). Los cebadores fueron tomados de otras zonas, más cercanas al polimorfismo en cuestión, para subir la temperatura de anillamiento en los ciclos de amplificación y evitar posibles productos secundarios. Las amplificaciones se hicieron en volúmenes de 10 µl, donde se incorporaron los cebadores:

M34f 5'- atg tta atg cct ggc ttc ca -3'

M34r 5'- agt cat tcc agg gac atc ca -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 5 seg., 55° C durante 5 seg. y 72° C durante 5 seg. Los productos de amplificación, con una longitud de 259 pb, fueron digeridos directamente con *HinfI* en las condiciones descritas anteriormente para el marcador M9. Los productos de esta digestión también fueron tratados de la misma manera. Cuando la muestra tiene una G (variante ancestral) en la posición 29 (desde el extremo 5' del cebador f)

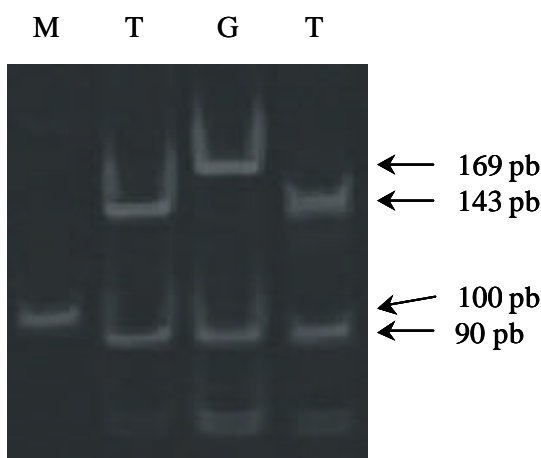


Figura 4. Detección de los alelos G y T del marcador M34. M es un marcador de peso molecular. El fragmento de 26 pb no se distingue de la nube de productos inespecíficos de la reacción.

del producto de amplificación, el patrón de fragmentos en la digestión es de 169 y 90 pb. Cuando es una T (variante derivada) la que ocupa dicha posición, entonces el patrón de bandas es de 143, 90 y 26 pb (Figura 4). En el producto de amplificación hay una diana *HinfI* invariable, que origina el fragmento de 90 pb, y una diana variable, que se crea al aparecer la variante derivada.

La secuencia del producto de amplificación en estas condiciones es:

atettaatgcctggctccaccaggag**M**cacatgtggtgtgctgcaaataaagtgttatgattattggggtccccaagctggacctgta
tccatgtcaagtgccacagggttacttgctttagcatggctccttggtgctgtaagtgaataattaact**gagtc**tttttgcaggagcta
actgagaccaatcaatcagtcattttcccttctgtgtgtaacacaagctgatgctcctggaatgact

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a las dianas reconocidas por la restrictasa *HinfI*, siendo M la base polimórfica.

- Locus M52

La variación en este locus, debida a una transversión A→C (Underhill y col. 2000), fue detectada siguiendo la metodología usada para el marcador PN2 en este trabajo. Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una C en negrita cerca del extremo 3' que sustituye a una G en la secuencia original):

*M52f 5'- ata cct ata aga ata ttg cct cca -3'

M52r 5'- gac gaa gca aac att tca ag -3'

El programa de amplificación usado fue de 40 ciclos a 94° C durante 15 seg., 50° C durante 15 seg. y 72° C durante 15 seg. Los productos de esta amplificación, con una longitud de 82 pb, fueron digeridos directamente con la enzima *StyI* (C/C[A,T][A,T]GG) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 2x de tampón One-Phor-All Buffer PLUS y 1 U de la enzima *StyI* (Boehringer Mannheim R1047744) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE al 8% (29:1) en las condiciones descritas. Cuando la muestra tiene una A (estado ancestral) en la posición 25 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de bandas en la digestión es de 60 y 22 pb. Cuando es una C (variante derivada) la que ocupa dicha posición, entonces el resultado es una banda de 82 pb sin digerir. En el producto de amplificación hay una diana *StyI* variable que se destruye en presencia de la variante derivada. La secuencia del producto de amplificación en estas condiciones es:

atacctataagaatattgcctcca**M**ggatggttgataggtttcttgatatttcattctctcttgaatggttgcttcgtc

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *StyI*, siendo M la base polimórfica.

- Locus M67

La variación en este locus, debida a un cambio del tipo A→C (Underhill y col. 2000), fue determinada, mediante amplificación y digestión, siguiendo la metodología empleada para el marcador PN2. Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una C en negrita cerca del extremo 3' que sustituye a una T en la secuencia original):

M67f 5'- gac aaa ctc ccc tgc aca ct -3'

*M67r 5'- gtt cgt gga ccc ctc tac at -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 10 seg., 58° C durante 10 seg. y 72° C durante 10 seg. Los productos de esta amplificación, con una longitud de 264 pb, fueron digeridos directamente con la enzima *NdeI* (CA/TATG) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 1x de tampón 4 (NEB) y 1 U de restrictasa *NdeI* (NEB R111S) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una A (estado ancestral) en la posición 244 (desde el extremo 5' del cebador f) del producto de amplificación, el resultado de la digestión es un fragmento de 264 pb sin digerir. Cuando es una T (estado derivado) la que ocupa dicha posición, entonces el patrón de bandas es de 243 y 21 pb. En el producto de amplificación hay una diana *NdeI* variable que se crea cuando aparece la variante derivada. La secuencia del producto de amplificación en estas condiciones es:

gacaaactccctgcacactggtttgtgcgacacctaataaaaggagaaaaagcagaaaagggtgtgggtcagaactaatgggcca
 tgtgaactcaaagatgtctctagatgctgtaacagatgtaggaagagtggaaaggctctatctcaagtacgtgtcctaaaagaaaaatgaga
 ttgtgaatttaaagtgtattcatagaaaagtactcaaaatatgtgtaattcaaaaaaca**M**atgtagaggggtccacgaac

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *NdeI*, siendo M la base polimórfica.

- Locus M70

El polimorfismo en este locus, debido a una transversión A→C (Underhill y col. 2000), fue detectado mediante amplificación y digestión. Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores usados por Underhill y col. (2000):

M70f 5'- ggt tat cat agc cca cta tac ttt g -3'

M70r 5'- atc ttt att ccc ttt gtc ttg ct -3'

El programa de amplificación usado fue de 40 ciclos a 94° C durante 10 seg., 48° C durante 10 seg. y 72° C durante 10 seg. Los productos de esta amplificación, con una longitud de 257 pb, fueron digeridos directamente con la enzima *BsaI* (GGTCTCN_{1/5}) en reacciones de 15 µl que constaron de: 10 µl de productos de reacción, 1x de tampón 3 (NEB) y 1 U de enzima *BsaI* (NEB R535S) por tubo. Tras 2 h. de incubación a 50° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una A (estado ancestral) en la posición 45 (desde el extremo 5' del cebador f) del producto de amplificación, el resultado de la digestión es una banda de 257 pb sin digerir. Cuando es una C (estado derivado) la que aparece en dicha posición, entonces el patrón de bandas es de 218 y 39 pb. En el producto de amplificación hay una diana *BsaI* variable, que se

crea al aparecer el estado derivado. La secuencia del producto que se amplifica en estas condiciones es:

ggttatcatagcccactatactttggactcatgtctccatgagaMctaaagactaccacaacagaatccctatagtccagccctcagatcacat
 acatgtacagggcatgttgaagtagtcggacttgaaggaatcagccatttcacaaaactctgcaaactgtactcctgggtagcctgttcaaate
 caaaagcttcaggaggctgtttactcctgaaataaaatatttcagcaagacaaaggaataaagat

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *BsaI*, donde M es la base polimórfica.

- Locus M78

El polimorfismo de este locus, producido por una transición C→T (Underhill y col. 2000), fue detectado mediante amplificación y digestión. Las amplificaciones de ADN se realizaron en reacciones de 10 µl, donde se incorporaron los cebadores diseñados por Underhill y col. (2000):

M78f 5'- ctt cag gca tta ttt ttt ttg gt -3'

M78r 5'- ata gtg ttc ctt cac ctt tcc tt -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 20 seg., 50° C durante 20 seg. y 72° C durante 20 seg. Los productos de amplificación, de 301 pb de longitud, fueron digeridos directamente con la enzima *AciI* (C/CGC) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 1x de tampón 3 (NEB) y 1 U de enzima *AciI* (NEB R551S) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una C (estado ancestral) en la posición 197 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos de la digestión es de 196 y 105 pb. Cuando es una T

(estado derivado) la que ocupaba dicha posición, entonces el resultado es un fragmento de 301 pb sin digerir (Figura 5). En el producto de amplificación hay una diana *AciI* variable, que se destruye al aparecer la variante derivada. La secuencia del producto amplificado en estas condiciones es:

cttcaggcattatTTTTTggttctccactacaggagaaatgtaaagtgatgagtcagaatttaggatggctgtatgggttctttgactaataca
 agaaatcactttgtaatgaatgaaatcagtggttctgcattactccgtatgttcgacatgaacacaaattgatacacttaacaaagatacttcttc
Mgcccttccaaatatttcaaaataagctggtcatagtacttgcttttcataaaaagatggtaagcttccaatatttagatttaaggaaaggtgaa
ggaacactat

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *AciI*, donde M es la base polimórfica.

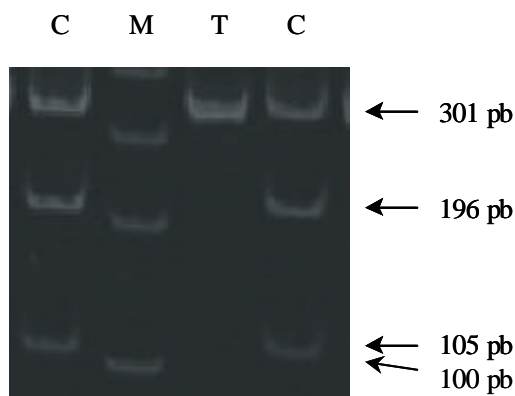


Figura 5. Detección de los alelos C y T del marcador M78. M es un marcador de peso molecular.

- Locus M81

El polimorfismo de este locus, producido por una transición C→T (Underhill y col. 2000), fue determinado mediante amplificación y digestión. Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores descritos por Underhill y col. (2000):

M81f 5'-act taa ttt ata gtt tca atc cct ca -3'

M81r 5'- ttc atg gag atg tct gta tct gg -3'

El programa de amplificación usado fue de 40 ciclos a 94° C durante 20 seg., 54° C durante 20 seg. y 72° C durante 20 seg. Los productos de esta amplificación, con una longitud de 422 pb, fueron digeridos directamente con la enzima *Hpy*CH4IV (A/CGT) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 1x de tampón 1 (NEB) y 0.5 U de la enzima *Hpy*CH4IV (NEB R619S) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una C (estado ancestral) en la posición 147 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos producidos por la digestión es de 276 y 146 pb. Cuando es una T (estado derivado) la que ocupa esa posición, entonces el resultado es el fragmento completo de 422 pb (Figura 6). En el producto de

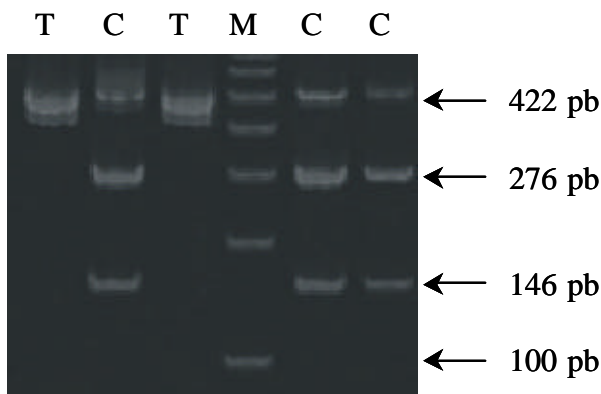


Figura 6. Detección de los alelos C y T del marcador M81. M es un marcador de peso molecular.

amplificación hay una diana *Hpy*CH4IV variable que se destruye al aparecer el estado derivado.

La secuencia del producto amplificado en estas condiciones es:

acttaatttatagttcaatccctcagtaattttaacttactctattttaagaactataaccaaactatctgtaagactttaagcactatcatactcag
ctacacatctcttaacaaaagaggtaaattttgtcctttttgaa**Mgt**catagagtatactcacacaaaccaagaagaacaatctactacata
cctacgctatatggatataactattgctcctagctacaaattagtgcgacactattgtactgaatattataggccatgtaacacaatggtttaag

tatctgtgcctctaaacacagaaaagatatagtgaaagtacagtattgctcctttattaaactcaaaatgttatgcagcatatgaccgactataaa
 atagcgcttatccagatacagacatctccatgaa

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por *Hpy*CH4IV, con M como posición polimórfica.

- Locus M89

El polimorfismo de este locus, producido por una transversión C→T (Underhill y col. 2000), se determinó, por amplificación y digestión, siguiendo la metodología empleada para el marcador PN2. Las amplificaciones de ADN se realizaron en reacciones de 10 µl, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una C en negrita cerca del extremo 3' que sustituye a una G en la secuencia original):

M89f 5'- aag aac tct gcc cca ttc aa-3'

*M89r 5'- caa ctc agg caa agt gag aca t -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 30 seg., 58° C durante 30 seg. y 72° C durante 30 seg. Los productos de la amplificación, con un tamaño de 191 pb, fueron digeridos directamente con la enzima *Nla*III en las condiciones descritas anteriormente para el marcador *DYS271*. Los productos de esta digestión también fueron tratados de la misma manera. Cuando la muestra tiene una C (estado ancestral) en la posición 169 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos de digestión es de 93, 79 y 19 pb. Cuando es una T (estado derivado) la que ocupa dicha posición, entonces el patrón de bandas es de 98 y 93 pb. En el producto de amplificación hay una diana *Nla*III invariable, que origina el fragmento de 93 pb, y una diana variable, que se destruye al aparecer la variante derivada. La secuencia del producto amplificado en estas condiciones es:

aagaactctgccccattcaagcttggcccaacactgttccaggggctggctggaattccaagctggtgagtcttatcctatgaggtgcatg
 aaagtggggccacagaaggatgctgctcagcttctggattcagctctcttctaagggtatgtacaaaaatct**Matgtctcactttgcctga**
gttg

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a las dianas reconocidas por *NlaIII*, donde M denota la base polimórfica.

- Locus M123

El polimorfismo en este locus se debe a una transición G→A (Underhill y col. 2000). Este cambio modifica el patrón de corte de la enzima *SfaNI* pero, dado que es muy cara, se decidió crear una diana artificial siguiendo la metodología usada para el marcador PN2. Las

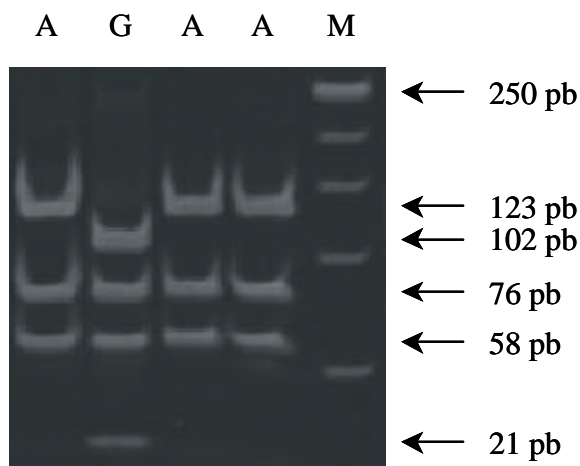


Figura 7. Detección de los alelos G y A del marcador M123. M es un marcador de peso molecular.

amplificaciones de ADN se realizaron en reacciones de 10 μ l, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una C en negrita cerca del extremo 3' que sustituye a una A en la secuencia original):

*M123f 5'- cac agt atc tga act agc atc tca -3'

M123r 5'- cag cga att aga ttt tct tgc -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 20 seg., 50° C durante 20 seg. y 72° C durante 20 seg. Los productos amplificados, con un tamaño de 257 pb, fueron

digeridos directamente con la enzima *DdeI* (C/TNAG), una enzima mucho más barata que *SfaNI*, en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 2x de tampón One-Phor-All Buffer PLUS y 1 U de *DdeI* (Boehringer Mannheim R835293) por tubo. Tras 2 h. de incubación a 37° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una G (estado ancestral) en la posición 25 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de bandas en la digestión es de 102, 76, 58 y 21 pb. Cuando es una A (estado derivado) la que ocupa dicha posición, entonces el patrón de fragmentos es de 123, 76 y 58 pb (Figura 7). El fragmento amplificado incluye dos dianas *DdeI* invariables, que producen los fragmentos de 76 y 58 pb, y una diana variable, que se destruye cuando aparece el estado derivado. La secuencia del producto amplificado es:

cacagtatctgaactagcatctca**M**catcgctgaatacctagaaatgcaaattcctgggcaacaccagaatctaacaagcaaaaaacta
 tggggggaacaggaagtcggttaataact**gag**tttgcaacctcaacttgccttataggaaagcaaaatctcaatatgataaagtttc
 ttaacaaaact**ctgag**ataactatgtgagggaaagaagttgatcacatgcaagaaaatctaattc**gctg**

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a las dianas reconocidas por *DdeI*, con M como base polimórfica.

- Locus M153

El polimorfismo de este locus, debido a una transversión T→A (Underhill y col. 2000), se determinó por amplificación y digestión. Las amplificaciones de ADN se llevaron a cabo en reacciones de 10 µl, donde se incorporaron los cebadores descritos por Underhill y col. (2000):

M153f 5'- tta ctg ata atg cca tat tgt ttt g -3'

M153r 5'- ttc tca gac acc aat ggt cct -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 10 seg., 45° C durante 10 seg. y 72° C durante 10 seg. Los productos de amplificación resultantes, con un tamaño de 459 pb, fueron digeridos directamente con la enzima *Tsp509I* (/AATT) en reacciones de 10 µl que costaron de: 5 µl de productos de amplificación, 1x de tampón 1 (NEB) y 0'5 U de enzima *Tsp509I* (NEB R576S) por tubo. Tras 2 h. de incubación a 65° C se añadieron 2 µl de solución de carga A y el total del volumen se corrió en PAGE 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una T (variante ancestral) en la posición 427 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos en la digestión es de 168, 130, 82 y 79 pb. Cuando es una A (estado derivado) la que ocupa dicha posición, entonces el patrón de fragmentos es de 168, 130, 79, 48 y 34 pb (Figura 8). En el producto de amplificación hay tres

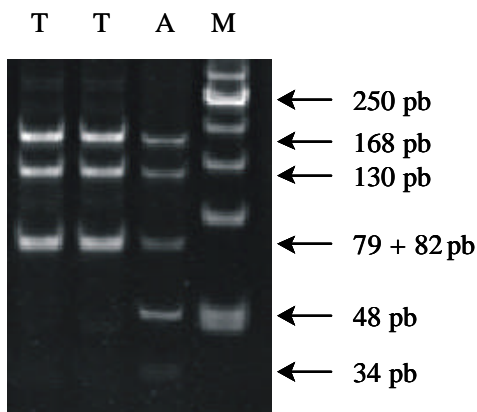


Figura 8. Detección de los alelos T y A del marcador M153. M es un marcador de peso molecular. El fragmento de 34 pb se aprecia ligeramente. Los fragmentos de 79 y 82 pb resultan indistinguibles en las condiciones de separación usadas.

dianas *Tsp509I* invariables, que producen los fragmentos de 168, 130 y 79 pb, y una diana variable, que se crea al aparecer la variante derivada. La secuencia del producto amplificado en estas condiciones es:

ttactgataatgccatattgtttggcttaatatcaggctaagtaaccacagtattctgatttaaaaaaacatactagagagcaagttattgac
 aaatcttaggaacttcaggtacagcatatgatttctgaactatgtgtgtaaataaggttttgtttattca**aat**taacacagggtagctctgtgatg
 cctccgattgatagctctaataaaacactttaatagtagcatatcaata**aat**ttatcatcatcgattttcttctaataatgaataacacatattg

tgattttctaagagtcaaaatctcaaaaatcattttaggtataaaatatacccccgaagttttatttccatttata**aatt**aatctgacttgaaa
 ggggaaaaaagctcaaagggtatgtgaaca**Mtt**cattaagataggaccattggtctcagaa

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a las dianas reconocidas por la enzima *Tsp509I*, siendo M la base polimórfica.

• Locus M170

El polimorfismo de este locus, debido a una transversión A→C (Underhill y col. 2000), se determinó, por amplificación y digestión, siguiendo la metodología usada para el marcador PN2. Las amplificaciones de ADN se realizaron en reacciones de 10 µl, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una A en negrita cerca del extremo 3' que sustituye a una T en la secuencia original):

*M170f 5'- cta ttt tat tta ctt aaa aat cat tga tc -3'

M170 r 5'- aga cca cac aaa aac agg tc -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 10 seg., 50° C durante 10 seg. y 72° C durante 10 seg. Los productos resultantes de esta amplificación, con un tamaño de 88 pb, fueron digeridos directamente con la enzima *BclI* en las mismas condiciones descritas para el marcador M26. El tratamiento posterior de los productos de la digestión también idéntico al del marcador M26 excepto que fueron resueltos en PAGE 8% (29:1). Cuando la muestra tiene

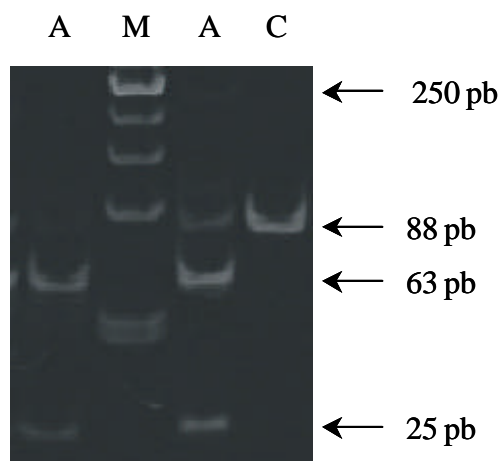


Figura 9. Detección de los alelos A y C del marcador M170. M es un marcador de peso molecular.

una A (estado ancestral) en la posición 30 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos de la digestión es de 63 y 25 pb. Cuando es una C (estado derivado) la que ocupa dicha posición, entonces el resultado es un fragmento sin digerir de 88 pb (Figura 9). En el producto de amplificación hay una diana *BclI* variable, que se destruye cuando aparece la variante derivada. La secuencia de producto amplificado en estas condiciones es:

ctattttatttacttaaaaaatcattgatcMttttttcagtggtgggtgtgtctcactgtaaaatgaggacctgtttttgtgtgtct

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por la enzima *BclI*, siendo M la base polimórfica.

- Locus M172

El polimorfismo en este locus, debido a una transversión T→G (Underhill y col. 2000), se determinó, mediante amplificación y digestión, siguiendo la metodología usada para el marcador PN2. Las amplificaciones de ADN se realizaron en reacciones de 10 µl, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una G en negrita cerca del extremo 3' que sustituye a una A en la secuencia original):

M172f 5'- aaa tta gga gcc aga tga cc -3'

*M172r 5'- aat aat aat tga aga cct ttt **gag** t -3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 5 seg., 52° C durante 5 seg. y 72° C durante 5 seg. Los productos de esta amplificación, con una longitud de 176 pb, fueron digeridos directamente con la enzima *HinfI* en reacciones de 15 µl como las descritas para el marcador M9. Los productos de esta digestión también fueron tratados de la misma manera. Cuando la muestra tiene una T (estado ancestral) en la posición 151 (desde el extremo 5' del cebador f) del producto de amplificación, el resultado de la digestión es un fragmento sin digerir

de 176 pb. Cuando es una G (estado derivado) la que ocupa dicha posición, entonces el patrón de fragmentos es de 151 y 25 pb. En el producto de amplificación hay una diana *HinfI* variable, que se crea al aparecer la variante derivada. La secuencia del producto amplificado en estas condiciones es:

aaattaggagccagatgaccaggatgccccagatgagcatgagccctctccatcagaagatgccccattatcctcattcacctgcctctca
 gtatcaacaggtaaaaaggattttcattttatccccaaacccatttgatgctt**M**actcaaaagggtcttcaattattatt

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por la restrictasa *HinfI*, siendo M la base polimórfica.

- Locus M173

El polimorfismo de este locus, debido a una transversión A→C (Underhill y col. 2000), se determinó, por amplificación y digestión, siguiendo la metodología empleada para el marcador PN2. Las amplificaciones de ADN se realizaron en reacciones de 10 µl, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una C en negrita cerca del extremo 3' que sustituye a una G en la secuencia original):

M173f 5'- aag ttg atg cca ctt ttc ag -3'

*M173r 5'- ttc tga ata tta aca gat **cac** aaa g -3'

El programa de amplificación usado fue de 40 ciclos a 94° C durante 30 seg., 48° C durante 30 seg. y 72° C durante 30 seg. Los productos de esta amplificación, con un tamaño de 199 pb, fueron digeridos directamente con la enzima *DraIII* en reacciones de 15 µl como las descritas para el marcador *SRY1532*. Los productos de esta digestión también fueron tratados de la misma manera. Cuando la muestra tiene una A (estado ancestral) en la posición 174 (desde el extremo 5' del cebador f) del producto de amplificación, los fragmentos resultantes son de 178 y 21 pb.

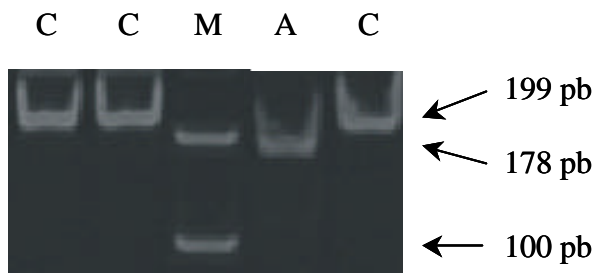


Figura 10. Detección de los alelos A y C del marcador M173. M es un marcador de peso molecular. El fragmento de 21 pb no se aprecia en la fotografía.

Cuando es una C (estado derivado) la que ocupa dicha posición, entonces se produce un fragmento de 199 pb sin digerir (Figura 10). En el producto de amplificación hay una diana *DraIII* variable, que se destruye al aparecer la variante derivada. La secuencia del producto amplificado en estas condiciones es:

aagttgatgccacttttcagaaaaatggttggtttgtacaaattgaaatacattgtttaaataaagcacagtactcacttaggtttgccatat
aaattfactgtaacttctagaaaattggaataaagtaagaaaaattttctacaattcaagggcatttagaac**Mctttgtgatctg**taaatattc
agaa

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por la restictasa *DraIII*, siendo M la posición variable.

- Locus M201

El polimorfismo de este locus, debido a una transversión G→T (Semino y col. 2000), se detectó, mediante amplificación y digestión, siguiendo la metodología empleada para el marcador PN2. Las amplificaciones de ADN se realizaron en reacciones de 10 µl, donde se incorporaron los cebadores (el cebador marcado con asterisco tiene una C en negrita cerca del extremo 3' que sustituye a una T en la secuencia original):

*M201f 5'- cta ata atc cag tac caa ctg agg -3'

M201r 5'- tga aag ttc aaa cgt caa aca g-3'

El programa de amplificación usado fue de 35 ciclos a 94° C durante 10 seg., 55° C durante 10 seg. y 72° C durante 10 seg. Los productos de amplificación resultantes, con un tamaño de 207 pb, fueron digeridos directamente con la enzima *Bs**I* (CCNNNNN/NNGG) en reacciones de 15 µl que constaron de: 10 µl de productos de amplificación, 1x de tampón 3 (NEB) y 0.5 U de enzima *Bs**I* (NEB R555S) por tubo. Tras 2 h. de incubación a 55° C se añadieron 3 µl de solución de carga A y el total del volumen se corrió en PAGE 6% (29:1) en las condiciones descritas. Cuando la muestra tiene una G (estado ancestral) en la posición 25 (desde el extremo 5' del cebador f) del producto de amplificación, el patrón de fragmentos es de 108, 78 y 21 pb. Cuando es una T (estado derivado) la que ocupa dicha posición, entonces el patrón de fragmentos es de 129 y 78 pb (Figura 11). En el producto de amplificación hay una diana *Bs**I* invariable, que produce el fragmento de 78 pb, y una diana variable, que se destruye cuando

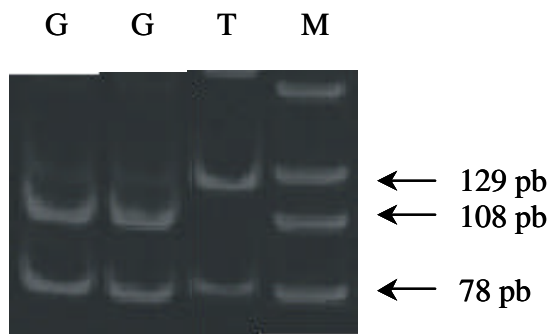


Figura 11. Detección de los alelos G y T del marcador M201. M es un marcador de peso molecular. El fragmento de 21 pb no aparece en la fotografía.

aparece la variante derivada. La secuencia del producto amplificado en estas condiciones es:

ctaataatccag**taccaactgagg**Mtttcgtaataggtacttagtggttgatgaagctgataggatgctggatatgggattgaacctcaga
 tacgtcgtatagttgaacaagatactatg**ccaccaaagggc**gttcgtcacacatgatgtttagtctactttcctaaggaaatacaggtact
gttgacgttgaacttca

Las bases subrayadas corresponden a las zonas de unión de los cebadores y las bases en negrita a la diana reconocida por la restrictasa *Bs**I*, siendo M la base polimórfica.

4.2. Marcadores de evolución rápida (microsatélites o STR)

Los microsatélites o STR son marcadores multialélicos constituidos por una secuencia corta (2-6pb) repetida varias veces seguidas que se encuentran distribuidos por el genoma de muchas especies, incluidos los humanos. Su abundancia, alto índice de polimorfismo y su facilidad de detección los han hecho una herramienta muy usada tanto para mapeo genético como para investigaciones forensicas o estudios evolutivos. Su peculiar estructura parece ser la responsable de su elevada tasa de mutación, así como de la manera en que lo hacen. Está ampliamente aceptado que evolucionan de una forma alélica a otra, preferentemente mediante pasos cortos, involucrando en la mayoría de las ocasiones ganancias o pérdidas en una sola repetición. Además, las ganancias o pérdidas de repeticiones no son infinitas, sino que oscilan en un rango determinado de tamaños.

Tanto los STR escogidos de la bibliografía como los descubiertos en este trabajo son tetranucleotídicos (las unidades repetitivas tienen una secuencia de cuatro nucleótidos), no se encuentran interrumpidos por ningún elemento (base o conjunto de bases que no pertenezca a ninguna repetición del microsatélite) y los productos de su amplificación nunca son superiores a 140 pb. Estas tres características hacen que su separación en electroforesis sea rápida y sencilla, pero además, que la determinación de una forma alélica esté regida únicamente por su tamaño. Habitualmente se usan algunos STR que no son homogéneos en toda su longitud (pueden estar interrumpidos por una repetición degenerada en alguna base), funcionando como conjuntos de repeticiones que evolucionan por separado. En estos casos, el encontrar un alelo de igual tamaño en dos individuos no implicaría directamente que compartiesen la misma forma alélica, con lo que las relaciones que se puedan deducir de ellos pueden no ser reales.

4.2.1. Amplificación de los STR.

Las amplificaciones se llevaron a cabo en reacciones de 10 μ l compuestas de 0'4 U de Ecotaq ADN polimerasa (Ecogen), 3 pmol de cada cebador, 0'2 mM de deoxinucleótidos

trifosfato (dNTP), 2'5 mM de $MgCl_2$, 16'6 mM $(NH_4)SO_4$, 67 mM Tris-HCl pH = 8'8 y 10-50 ng de ADN molde. Todas las reacciones fueron incubadas en un ciclador térmico PTC-100 (MJ Research). El programa de amplificación fue de 94° C durante 2 min. seguido por 35 ciclos de 20 seg. a 94° C, 20 seg. a la temperatura adecuada a la pareja de cebadores en concreto y 20 seg. a 72° C. Tras la amplificación se dio un periodo de 5 min. a 72° C para la terminación de la elongación.

4.2.2. Detección de variantes en los STR.

A los 10 μ l de productos de amplificación se añadieron 2 μ l de solución de carga A para ser separados a continuación mediante PAGE al 6% (19:1, acrilamida: N,N'-metilen-bis-acrilamida) usando una cubeta X-Cell™ MiniCell (NOVEX) y tampón TBE 1x. La preparación de estos geles fue exactamente igual a los anteriormente descritos para la detección de variantes de UEP, aunque partiendo de una mezcla de acrilamida al 20%. Esta solución al 20% (19:1) se preparó agregando 19 g de acrilamida y 1 g de N,N'-metilen-bis-acrilamida y llevando con agua estéril hasta 100 ml. Las electroforesis se corrieron a un voltaje constante de 300 V (60 mA) durante unos 25 min. y luego los geles fueron teñidos en una solución de Bromuro de Etidio (1 μ g/ml) durante 15 min. para ser visualizados al UV con un transiluminador Gelprinter Plus (TDI). La combinación de alelos secuenciados de un STR dado, de los que por tanto se conocía su tamaño y número de repeticiones, y su inclusión en las electroforesis, a modo de marcador, sirvió para crear un patrón de referencia contra el cual contrastar los productos de amplificación.

4.2.3. Marcadores STR descritos por primera vez

- *DYS466* (N° de acceso 11500744)

Este STR está constituido por unidades del tipo $(GTTT)_n$ que se repiten entre 5 y 8 veces (Figura 12). Fue descubierto durante la secuenciación directa de fragmentos con repeticiones del tipo *Alu* (ver pag. 71). La forma alélica que se encuentra en el clon 264, M, 20 en las condiciones

de detección de este trabajo, se corresponde con un fragmento de 116 pb y contiene 7 repeticiones. La temperatura de anillamiento en las amplificaciones fue de 58° C y los cebadores usados fueron:

M20tr1f 5'- gtg acc aaa tct gtg cat gg -3'

M20tr1d 5'- cag agc aag act ctg tct caa aac -3'

Los productos de amplificación oscilaron entre 108 y 120 pb, correspondiendo a cuatro formas alélicas.

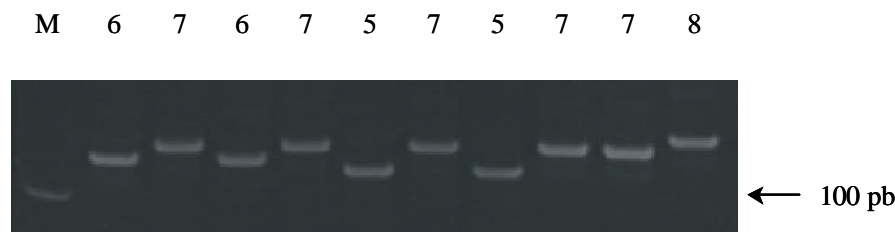


Figura 12. Detección rutinaria de los cuatro alelos detectados para el marcador *DYS466*. Los alelos están definidos por el número de repeticiones. M es un marcador de peso molecular.

- *DYS467* (N° de acceso 11500746)

Este STR está constituido por unidades del tipo (GATA)_n que se repiten entre 7 y 16 veces (Figura 13). Fue descubierto en la búsqueda electrónica y amplificado con cebadores obtenidos de las regiones flanqueantes (ver pag. 73). Dado que, hay una zona homóloga a la zona exterior donde se une uno de los cebadores diseñados, las amplificaciones originan dos fragmentos de distinto tamaño, el más pesado de los cuales (≈ 235 pb) engloba las repeticiones del ligero (≈ 115 pb). En cualquier caso, el fragmento pesado no fue considerado en este estudio y el fragmento de 115 pb del producto ligero incluye 14 repeticiones (forma alélica incluida en el clon de referencia). La temperatura de anillamiento en los ciclos de amplificación fue de 56° C y los cebadores usados fueron:

M20tr2f 5'- gat gag agt tgg ata cag aag tag g -3'

M20tr2r 5'- tgt caa tga ttt tct gtc ttc tat ca -3'

Los productos de amplificación considerados oscilaron entre 87 y 123 pb, correspondiendo a ocho formas alélicas.

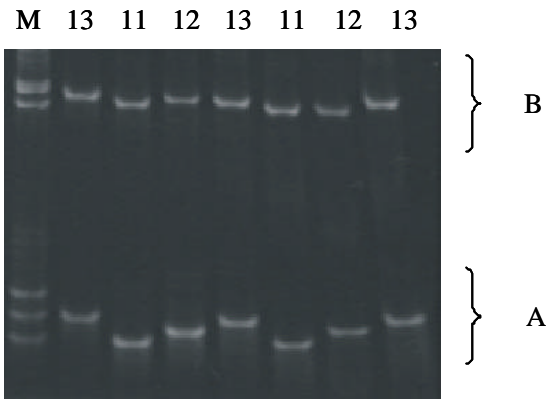


Figura 13. Detección de algunos alelos del marcador *DYS467* (A) y de los fragmentos polimórficos más pesados que se originan en la misma amplificación (B). Los alelos están definidos por el número de repeticiones. M es un marcador construido con los alelos de 11, 13 y 15 repeticiones del propio locus. Se desconoce el tamaño exacto de los fragmentos de la parte B.

4.2.4. Marcadores STR tomados de la bibliografía

- *DYS391* (N° de acceso AF140637)

Este STR está constituido por unidades del tipo (GATA)_n que se repiten entre 8 y 12 veces (Figura 14). El producto amplificado con los cebadores publicados (Jobling y Tyler-Smith, 1995) es de unos 280 pb. Para este estudio, este fragmento fue reducido hasta un tamaño aproximado de 100 pb mediante la construcción de un nuevo cebador en 3'. El alelo al que corresponde el número de acceso tiene una longitud de 283 pb y contiene 10 repeticiones. Este



Figura 14. Detección de algunos alelos del marcador *DYS391*. Los alelos están definidos por el número de repeticiones. M es un marcador de peso molecular.

alelo se corresponde con el fragmento de 100 pb de este trabajo. La temperatura de anillamiento en los ciclos de amplificación fue de 55° C y los cebadores usados fueron:

391f 5'- cta ttc att caa tca tac acc ca -3'

391r 5'- agg tag gca ggc aga tag gc -3'

Los productos de amplificación resultantes oscilaron entre 92 y 108 pb, correspondiendo a cinco formas alélicas.

- *DYS393* (N° de acceso AF140639)

Este STR está constituido por unidades repetitivas del tipo (GATA)_n que se repiten entre 11 y 15 veces (Figura 15). El alelo al que corresponde el número de acceso incluye 12 repeticiones y tiene un tamaño de 119 pb. La temperatura de anillamiento para su amplificación fue de 54° C y los cebadores usados fueron:

393f 5'- gtg gtc ttc tac ttg tgt caa tac -3'

393r 5'- aac tca agt cca aaa aat gag g -3'

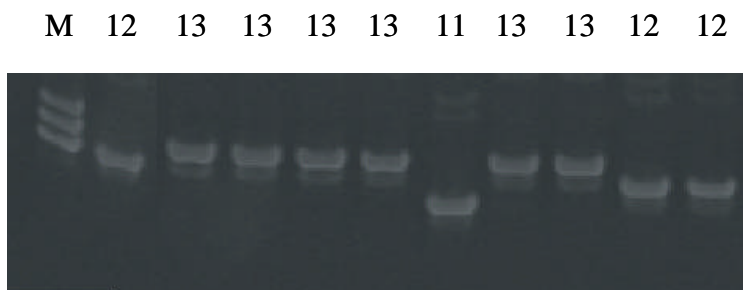


Figura 15. Detección de algunos alelos del marcador *DYS393*. Los alelos están definidos por el número de repeticiones. M es un marcador construido con los alelos de 12, 13 y 14 repeticiones del propio locus.

Los productos de esta amplificación oscilaron entre 115 y 131 pb, correspondiendo a cinco formas alélicas.

- *DYS434* (N° de acceso 9899860)

Este STR está constituido por unidades repetitivas del tipo (CTAT)_n que se repiten entre 8 y 11 veces (Ayub y col., 2000). El alelo al que corresponde el número de acceso incluye 9 repeticiones y tiene una longitud de 114 pb. La temperatura de anillamiento para su amplificación fue de 58° C y los cebadores usados fueron:

434L 5'- cac tcc ctg agt gct gga tt -3'

434R 5'- gga gat gaa tga atg gat gga -3'

Los fragmentos originados en las amplificaciones oscilaron entre 110 y 118 pb, correspondiendo a tres formas alélicas (Figura 16).



Figura 16. Detección rutinaria de los alelos del marcador *DYS434*. Los alelos están definidos por el número de repeticiones.

5. PURIFICACIÓN DE PRODUCTOS AMPLIFICADOS

Las purificaciones de ADN para su secuenciación se hicieron siguiendo dos metodologías diferentes en función del tamaño. Cuando el tamaño de los fragmentos a purificar superaba las 250 pb se usó una precipitación con polietilenglicol (PEG), mientras que, cuando el tamaño de los fragmentos no superaba las 250 pb se usó una precipitación con acetato amónico.

5.1. Precipitación con polietilenglicol (PEG)

1. Se añaden 30 μ l de una solución de polietilenglicol (20% PEG 8000 y 2'5 M de NaCl) a 45 μ l de producto amplificado.
2. Se homogeniza la mezcla agitando vigorosamente y se incuba durante 30 min. en hielo prensado.
3. Se centrifuga la mezcla a 13.000 rpm durante 15 min.
4. Se retira toda la fase líquida con una pipeta Pasteur.
5. Se añaden 100 μ l de etanol al 70% muy frío y se vuelve a centrifugar la mezcla a 13.000 rpm durante 5 min.
6. Se retira el alcohol con una pipeta Pasteur. El precipitado se lleva a una estufa de secado donde permanece a 37° C durante 10 min.

7. Se añaden 25 μ l de agua estéril y se deja resuspendiendo a 4° C hasta el día siguiente.

Posteriormente se conserva el producto a -70° C.

5.2. Precipitación con acetato amónico

1. A 45 μ l de productos de amplificación se le añaden 25 μ l de una solución 7.5 M de acetato amónico y se agita vigorosamente.
2. Se añaden 75 μ l de isopropanol a la mezcla anterior y se homogeniza.
3. Se incuba la mezcla durante 1 h. en hielo prensado.
4. Se centrifuga a 13.000 rpm durante 15 min.
5. Se retira la fase líquida con una pipeta Pasteur.
6. Se añaden 100 μ l de etanol al 70% muy frío y se vuelve a centrifugar la mezcla a 13.000 rpm durante 10 min.
7. Se retira todo el líquido con una pipeta Pasteur. El precipitado se lleva a una estufa de secado donde permanece a 37° C durante 10 min.
8. Se añaden 25 μ l de agua estéril y se deja resuspendiendo a 4° C hasta el día siguiente.

Posteriormente se conserva el producto a -70° C.

6. SECUENCIACIÓN DE PRODUCTOS AMPLIFICADOS

6.1. Reacciones de secuenciación

Los productos de amplificación purificados fueron usados para su secuenciación directa por ambas cadenas mediante una reacción de elongación y terminación. En todos los casos se usaron los mismos cebadores y las mismas condiciones de reacción que en las respectivas amplificaciones, con la excepción de que los cebadores fueron marcados con el isótopo radioactivo $\gamma^{32}\text{P}$ -ATP por el extremo 5' y los reactivos usados provenían del paquete comercial fmol[®] DNA Sequencing System (Promega).

Para el marcaje de los cebadores se hicieron reacciones de 10 μ l que contenían 10 pmol del cebador en cuestión, 10 pmol de $\gamma^{32}\text{P}$ -ATP, 5-10 U de T4 Polinucleótido Kinasa, 50 mM Tris-HCl pH = 7.5, 10 mM MgCl_2 , 5 mM de di-tio-treitol (DTT) y 0.5 mM de espermidina. Una vez homogeneizadas las mezclas se llevaron a un baño de 37° C, donde se incubaron 30 min. para la incorporación del ^{32}P en los cebadores, y luego a otro a 90° C, donde permanecieron 2 min. para inactivar las Kinasa. Los cebadores marcados fueron posteriormente almacenados a -20° C hasta el momento de su utilización.

El sistema de secuenciación utilizado requiere en la parada de la reacción de elongación la presencia, en la mezcla de reacción, de di-deoxinucleótidos (ddNTP). Así, para la lectura por separado de las cuatro bases de ADN fue necesario montar cuatro reacciones por cada muestra a secuenciar, cada una de ellas con un ddNTP distinto. Para distinguir entre las cuatro reacciones se usaron tubos de 0.6 ml coloreados (rojos para ddGTP, azules para ddATP, amarillos para ddTTP y verdes para ddCTP). De esta manera cada muestra a secuenciar constó de cuatro reacciones (cada una en un tubo de color diferente), cada una de ellas con \approx 50-100 ng de ADN molde purificado, 1 pmol de cebador marcado, 2 μ l de la correspondiente mezcla de d/ddNTP, 0.25 U de Taq ADN polimerasa de secuenciación, 50 mM Tris-HCl pH = 9 y 2 mM MgCl_2 .

Una vez finalizado el periodo de elongación y terminación, las reacciones se pararon usando una solución de N,N-dimetilformamida con 1 mM EDTA pH = 8, 1 mg/ml de azul de bromofenol y 1 mg/ml de xileno-cianol. A continuación, las reacciones se incubaron a 98° C durante 2 min., para desnaturalizar las cadenas de ADN, tras lo cual se introdujeron rápidamente en hielo prensado con etanol, para impedir su renaturalización, y se almacenaron a -20° C hasta su utilización.

6.2. Separación de los productos de la reacción de secuenciación

La separación se efectuó mediante PAGE al 6% (19:1) en condiciones desnaturalizantes (7 M de urea). Para la construcción de los geles de poliacrilamida se dispuso de casetes constituidos por dos cristales de 41 x 33 cm (uno de ellos con una muesca en la parte superior para permitir la carga de las muestras), espaciadores de 44 x 1 x 0'4 cm y peines de “dientes de tiburón”. Los cristales que contenían las muescas fueron tratados con dicloro-dimetil-silano al 5% en cloroformo, en la cara que entraba en contacto con el gel, antes de su utilización en la construcción de los geles. Este producto actúa como un repelente del gel de poliacrilamida para que, al finalizar la electroforesis, sea más fácil su separación del casete.

La mezcla para la elaboración de los geles de poliacrilamida al 6% (19:1) se preparó con 28'5 g de acrilamida, 1'5 g de N,N'-metilen-bis-acrilamida, 210 g de urea, 5'4 g de Tris, 2'75 g de ácido bórico, 186 mg de EDTA-Na₂, 175 µl de TEMED y agua estéril hasta 500 ml. Una vez homogeneizada, se hizo pasar la mezcla por un filtro tipo HA (Millipore) mediante una bomba de vacío. Para comenzar la polimerización de cada gel, una vez montado los casetes que sirvieron de molde para la poliacrilamida, a 60 ml de la mezcla anterior se le añadieron 400 µl de peroxidisulfato de amonio (100 mg/ml), se agitó suavemente y se vertió en el molde. A continuación, los peines de “dientes de tiburón”, puestos con los dientes hacia fuera, sirvieron para hacer una cámara cerrada en cada molde, evitando que el agente reductor (peroxidisulfato de amonio) reaccionara con el oxígeno atmosférico. Transcurrida 1 h. los geles, ya polimerizados, se limpiaron de residuos y se montaron en cubetas de secuenciación verticales S4S (Owl Scientific). Las cubetas de secuenciación se llenaron con tampón TBE 1x y se hizo un pre-calentamiento, suministrándoles corriente eléctrica a potencia constante de 120W (1800 V y 75 mA), para homogeneizar las concentraciones de urea y la temperatura en cada gel. Una vez terminado se dio la vuelta a los peines de “dientes de tiburón” y se presionó en las muescas hasta que éstos se clavaron en los geles. En este momento se cargaron los productos de la

secuenciación siguiendo el orden: rojo, azul, amarillo y verde (correspondientes a las reacciones que dejarían al final de cada cadena las bases G, A, T y C, respectivamente). Cada cuatro reacciones se dejó un hueco vacío para separar las reacciones correspondientes a cada muestra. Una vez cargadas todas las reacciones, se comenzó la electroforesis que, en las mismas condiciones que el pre-calentamiento, duró entre 1'5 y 3 h aproximadamente. Una vez terminada la electroforesis, los geles se separaron del casete usando papel de filtro. A continuación se cubrieron con láminas de plástico transparente para evitar su deterioro y se desecaron en secadores Drygel Sr. (Hofer SE1160) aplicando calor (75° C) y vacío. Tras 2 h. de secado se les retiró el plástico y se incubaron en contacto directo con una película radiográfica Super RX (Fuji) durante 1-4 días. Transcurrido este tiempo las películas fueron reveladas para visualizar las secuencias.

7. PROGRAMAS Y ANÁLISIS ESTADÍSTICO

7.1. Construcción de cebadores

Tanto en la búsqueda de nuevos polimorfismos como para la modificación de los protocolos de amplificación de marcadores descritos, fue necesario el diseño de nuevos cebadores. En todos los casos se usó el programa Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) restringiendo el tamaño del fragmento a amplificar, según los casos, entre 80 y 500 pb. A continuación, mediante el uso del programa Generunner ver. 3.0 (Hastings Software, Inc.), se localizaron las dianas de restricción que se originarían en los productos de las amplificaciones y se estimaron los tamaños de los fragmentos generados por las reacciones de digestión.

7.2. Estructura y relaciones poblacionales

7.2.1. Análisis de varianza y distancias genéticas

En ausencia de selección, las diferencias en frecuencias haplotípicas de dos poblaciones se deben sencillamente a que han sufrido diferentes eventos demográficos (mezclas con otras

poblaciones, efectos fundadores, endogamia...) o mutacionales. Una manera habitual de cuantificar la cantidad de divergencia entre dos poblaciones o grupos es mediante el uso de los estadísticos F (Wright, 1965). Estos estadísticos permiten un análisis jerárquico de modo que, si éste se realiza con las poblaciones agrupadas, se puede cuantificar qué porcentaje de la varianza total se debe a diferencias entre grupos y qué porcentaje a diferencias dentro de grupos. Estas estimas se obtuvieron con el subprograma AMOVA (Excoffier y col. 1992) disponible en el paquete ARLEQUIN 2.000 (Schneider y col. 2000; <http://anthro.unige.ch/arlequin>). En estos cálculos se pueden tener en cuenta no sólo las diferencias de frecuencia, sino también las diferencias a nivel molecular entre los caracteres que se comparan, en este caso haplotipos. Esta información molecular no ha sido usada, dado que las relaciones entre las poblaciones que han divergido recientemente se explican mejor por sus diferencias demográficas. El programa usado también permite calcular la probabilidad de diferenciación entre poblaciones dos a dos mediante la comparación de frecuencias en tablas de contingencia. El método que ejecuta el programa es análogo al exacto de Fisher (Raymond y Rousset, 1995). Las distancias genéticas fueron calculadas como análogos de los estadísticos F_{ST} (Reynolds y col. 1983) mediante el subprograma GENEDIST del paquete PHYLIP ver.3.5 (Felsenstein, 1993) accesible en la dirección <http://evolution.genetics.washington.edu/phylip.html>.

7.2.2. Representaciones gráficas de las distancias genéticas

Se usaron tres tipos de representaciones gráficas para estimar las relaciones poblacionales. Dos de estos métodos representan a las poblaciones en redes, directamente relacionadas dos a dos, el otro es un método exploratorio en el que no se unen directamente las poblaciones comparadas.

- Neighbor-Joining (NJ) – Este es un método de agrupamiento que utiliza un algoritmo de agrupación por vecinos más próximos (Saitou y Nei, 1987), sin asumir tasas de evolución constantes. Para su construcción se necesitan distancias, por lo que se

produce pérdida de información. Las distancias genéticas fueron calculadas a partir de 1000 réplicas o bootstraps (Efron, 1982), para testar la significación, mediante el uso de los subprogramas SEQBOOT y GENEDIST. A continuación, las redes fueron construidas con el subprograma NEIGHBOR y una red consenso mediante el subprograma CONSENSE. La topología resultante fue representada con el programa TreeView ver.1.6.6 (Page, 1996), accesible en la dirección <http://taxonomy.zoology.gla.ac.uk/rod/rod.html>.

- Maximum Likelihood (ML) – Es un método de agrupamiento que maximiza las relaciones entre la matriz de datos y la red escogida, sin asumir tasas de evolución constantes. En este proceso se producen muchas redes, pero se escoge sólo la que tiene más probabilidad de haber producido los datos observados (Page y Holmes, 1998). Dado que no necesita del cálculo previo de distancias, no se produce pérdida de información. El modelo en el que está basado es el de movimiento browniano, asumiendo que las variables evolucionan por pura deriva génica (Cavalli-Sforza y Edwards, 1967). Las redes fueron calculadas a partir de los 1000 bootstraps obtenidos mediante los subprogramas SEQBOOT y CONTML. A continuación, la topología consenso se construyó mediante el uso del subprograma CONSENSE y fue representada con el programa TreeView ver.1.6.6.
- Análisis mutidimensional (Kruskal y Wish, 1978) - Dado que los métodos NJ y ML imponen un modelo de bifurcación en las relaciones poblacionales, que puede no ser adecuado para poblaciones muy estrechamente relacionadas, también se usó un método de representación exploratorio en el que no hay uniones entre las poblaciones comparadas. Este método trata de encontrar la estructura de un conjunto de medidas de distancia entre las poblaciones. Asigna las poblaciones a posiciones específicas en un espacio conceptual, normalmente de dos dimensiones, de modo que las distancias

entre los puntos en el espacio concuerdan al máximo con las distancias genéticas. Este método permite determinar una medida de bondad de ajuste mediante el cálculo de: valor de stress y correlación entre las distancias. El valor de stress da una idea de la calidad de la representación. Cuando el valor de stress está entre 0'15 y 0'1 la representación se considera buena. Cuando este valor disminuye por debajo de 0'1, entonces los datos se encuentran muy bien representados en las dimensiones escogidas. En caso contrario habría que aumentar la dimensionalidad (<http://www.analytictech.com/borgatti/mds.htm>). Por otro lado, cuanto más cercano a 1 sea el valor de correlación entre la distancia genética y la distancia espacial, mejor representadas estarán las relaciones entre las poblaciones.

Los subprogramas SEQBOOT, GENEDIST, NEIGHBOR, CONTML y CONSENSE son parte del paquete PHYLIP ver.3.5. La representación multidimensional se llevó a cabo mediante el uso de ALSCAL incluido en el paquete SPSS 9.0 (SPSS, Inc.).

7.2.3. Correlaciones entre distancias genéticas y geográficas

Las distancias genéticas obtenidas como análogos de F_{ST} fueron correlacionadas con las distancias geográficas mediante el test de Mantel (1967), que ejecuta también el ARLEQUIN 2.000. Estas distancias geográficas se calcularon a su vez con el programa Great Circle Distance (<http://www.mercury.demon.co.uk/dist/dodist.html>) introduciendo las coordenadas geográficas que figuran en la dirección <http://www.calle.com> de las ciudades principales de las regiones muestreadas.

7.2.4. Correlaciones entre variables

Los métodos de correlación bi-variados más conservadores, por tener menos asunciones, son los no paramétricos. Por esta razón se calculó, mediante el programa SPSS 9.0, el coeficiente de correlación por rangos de Spearman (r_s).

7.2.5. Mapas de frecuencia

Los mapas de frecuencia fueron contruidos con el procedimiento Kriging (Delfiner, 1976) usando el programa Surfer v.7 (Golden Software, Inc.). La finalidad de esta metodología es usar el valor de frecuencia de un haplotipo, o conjunto de haplotipos, en zonas geográficas muestreadas, para obtener estimas en regiones adyacentes no estudiadas por carencia de muestras. El resultado es una malla de puntos geo-referenciados con estimas en puntos no muestreados, resultantes de la combinación lineal de los valores observados en el resto del área, y de los valores observados en las zonas analizadas. La superficie de frecuencias se construye incorporando los valores de la malla al área geográfica en cuestión, asignando un tono de color a regiones que contengan las mismas frecuencias. Gracias a los gradientes, que se obtienen de estas representaciones, se pueden deducir eventos demográficos pasados que puedan haberse producido en las poblaciones objeto de estudio.

7.3. Estimas de mezcla

El problema de estimar las contribuciones genéticas de las poblaciones parentales a poblaciones híbridas o mezcladas ha sido estudiado desde hace más de 60 años (Bernstein, 1931). Se han propuesto muchos estimadores de estas contribuciones parentales relativas. La gran mayoría de ellos están basados sólo en la comparación de frecuencias de marcadores entre las poblaciones parentales y la híbrida. Suponen que la frecuencia que observamos en la población híbrida debe ser una combinación lineal de las frecuencias en las poblaciones parentales, asumiendo que la deriva génica que ha ocurrido después de la mezcla es despreciable. Dado que los distintos haplotipos del cromosoma Y estudiados en este trabajo se ha producido en lugares concretos, a partir de los cuales se han se han expandido por procesos migratorios, se ha utilizado el estadístico m_L (Long 1991) como estimador de los porcentajes de mezcla, que está basado sólo en las frecuencias haplotípicas. Este estimador fue determinado mediante el programa ADMIX.PAS, amablemente cedido por el Dr. Long. Además de la estima de la

aportación, el cálculo de este estimador produce un error, que puede subdividirse en un error de muestreo y un error debido a la varianza no explicada por el evento de mezcla y que, en ausencia de selección, da un indicio de la deriva génica, en forma de distancia genética (F_{ST}), que ha sufrido la población híbrida.

7.4. Contribuciones sexuales por área geográfica

Con las estimas de mezcla paternas, a partir de los datos del cromosoma Y de este trabajo, maternas, con los datos de ADN mitocondrial (Pinto y col. 1996a; Rando y col. 1999), y totales, con los marcadores nucleares (Roberts y col. 1966; Pinto y col. 1994), se han estimado las contribuciones teóricas por sexo y área geográfica. Usando las ecuaciones de Hsieh y Sutton (1992) se puede averiguar si las poblaciones parentales implicadas en la mezcla han contribuido por igual con ambos sexos.

7.5. Redes abreviadas promedio y diversidad en STR asociada a los haplotipos

En algunos haplotipos, contruidos a partir de las combinaciones alélicas de los marcadores UEP, se estudió la diversidad de STR construyendo sublinajes mediante la combinación de los estados alélicos de los diferentes microsátélites. Debido a su alta tasa de mutación, y al propio modo en que mutan, estos marcadores producen muchos paralelismos que habitualmente resultan mejor representados mediante su relación en redes. Dado que existe incertidumbre en los pasos mutacionales que se han producido para originar los linajes que se observan, este método mantiene “camino” paralelos e igual de probables a modo de reticulaciones. Normalmente estas redes suelen ser muy complicadas pero, teniendo en cuenta criterios de frecuencia, pueden resolverse algunos de estos eventos paralelos (Bandelt y col. 1995). Estas redes fueron construidas usando el algoritmo RM con el programa NetworkWin (<http://www.fluxus-engineering.com/sharenet.htm>). El número medio de mutaciones en los STR por haplotipo, usado como un indicador de la diversidad, se estimó por medio del programa ARLEQUIN 2.000 (Schneider y col. 2000).

7.6. Edad de los haplotipos

Para determinar cuando ocurrió una mutación en la filogenia construida para el cromosoma Y, en base a los marcadores bialélicos, se siguió el método descrito por Bertranpetit y Calafell (1996). Dentro de cada uno de los haplotipos bialélicos se puede estimar cuanto tiempo ha transcurrido para que se acumule la variación en los microsatélites que se observa en la actualidad. Comparando todas las variantes STR con un linaje raíz conocido se puede estimar el número medio de pasos (π) que han ocurrido. El linaje raíz escogido para este cálculo fue siempre aquel con el menor número de pasos mutacionales al resto de cromosomas. Nótese que de esta manera las estimas resultantes serán las más conservadoras. Dado que se conoce la tasa de mutación en los microsatélites del cromosoma Y ($\mu = 2.1 \times 10^{-3}$, $0.6-4.9 \times 10^{-3}$ 95% CI; Heyer y col. 1997), se puede estimar el número de generaciones (τ) que han transcurrido para generar tal diversidad mediante la ecuación:

$$\pi = \mu \cdot \tau$$

Suponiendo que transcurren unos 30 años por generación, se puede obtener una estima de tiempo en años.

7.7. Porcentaje de divergencia por haplotipo y edad de colonización

Las estimas usadas para el cálculo del porcentaje de divergencia, y por lo tanto del tiempo en que ésta se inició, están basadas en distancias genéticas ASD (del Inglés Average Square Distance), donde $ASD = \mu\tau$ (Thomas y col. 1998). Por un lado, se ha calculado ASD entre el linaje STR raíz y el resto de linajes observados, que por definición es una estima de la edad de los haplotipos muy similar a la descrita por Bertranpetit y Calafell (1996). Por otro, se ha calculado la distancia ASD entre el pool de linajes canarios y el del conjunto Península-noroeste de África. En este caso $ASD = 2\mu\tau$, puesto que los dos conjuntos divergen desde una raíz. Dividiendo este valor por dos veces la divergencia desde la raíz se obtiene un porcentaje de

tiempo de separación (Hurles y col. 1999). Los valores ASD fueron calculados con el programa Microsat 1.5d (Minch, 1997) accesible en la dirección <http://hpgl.stanford.edu/projects/microsat/>.

III. RESULTADOS

1. NUEVOS POLIMORFISMOS

Los dos métodos empleados para la búsqueda de polimorfismos dieron como resultado la determinación de dos loci polimórficos no descritos hasta el momento.

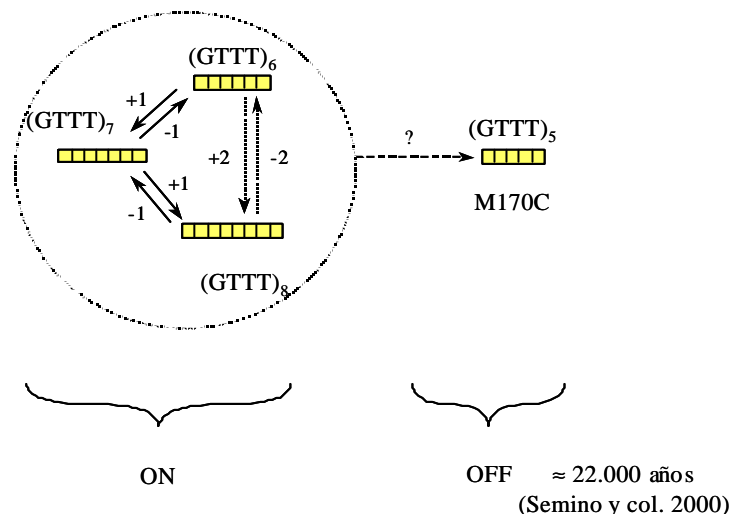
1.1. Secuenciación de elementos *Alu*: detección del polimorfismo *DYS466*

Aunque se usaron muestras de sangre de individuos de diversa localización geográfica, aprovechando las características de patrilocalidad del cromosoma Y, tan sólo se detectó una secuencia variante entre las 280 secuencias analizadas. En una de las seis muestras españolas se reveló la existencia de un acortamiento en la secuencia repetitiva GTTT de la región flanqueante 5' del elemento *Alu*_Y, insertado en el clon 264, M, 20 entre las posiciones 86.446 y 86.838. Mientras que en la secuencia del clon el número de repeticiones GTTT es de 7, en esta muestra el número disminuyó a 5. La confirmación de la existencia de polimorfismo en este elemento *Alu* en las poblaciones se realizó mediante PCR-RFLP digiriendo con *Hinf*I (mismas condiciones que las descritas para el marcador M9). La digestión con esta enzima origina siempre dos fragmentos: uno de 332 pb y otro de \approx 100 pb, que varía en tamaño según el número de repeticiones GTTT que contenga. Para que la detección rutinaria fuera directamente por amplificación se diseñó un nuevo cebador r, anclado en 3' a una repetición GTTT para evitar la aparición de productos inespecíficos por su homología con regiones de elementos *Alu*, que fue usado junto al cebador f previamente diseñado. La secuencia del elemento *Alu* y zonas flanqueantes amplificadas es:

gtgaccaaactctgcatggtgagaccttctaaactacagctatctactgttggtttgtttattg**tttgtttgtttgtttgtttgtttgtttgaga**
caGAGTCttgctctgtcaccaggctgaagtgcagtggtgcgatcttggtcactgcaagctccgctctcaggttcacgccattcct
gcctcagcctcccgagtagctgggaatacaggaaccaccaccatgcctggctaattttgtattttcatagagatggggttttaccatgta
gccaggatggtcttgatctctgacctcatgatccaccaccacagcctccaaagtactgggattacaggtgtggaccaccatgccagcc
aactacagctaccttttagtggttatccttattg**caaaagcagcaacagatacttt**

Las bases subrayadas son las zonas de unión de los cebadores, siendo las doble subrayadas las del cebador r usado en las amplificaciones rutinarias, las bases en negrita son las repeticiones GTTT polimórficas y las bases en mayúscula corresponden a la diana reconocida por *HinfI*. Por su situación y alto contenido en T, el origen de este polimorfismo pudo ser causado por la aparición de mutaciones en la cola de poli-A asociada al elemento *Alu*, como ya se ha apuntado en otros casos (Nadir y col. 1996; Ayub y col. 2000; Crouau-Roy y Clisson, 2000). En Canarias, las formas alélicas detectadas contienen 5, 6, 7 u 8 repeticiones GTTT, siendo los alelos de 6 y 8 repeticiones los más raros (entre los dos sumaron un 1%), y los de 5 y 7 repeticiones los mayoritarios, 9'7 y 89'3%, respectivamente. La distribución de las frecuencias de los alelos de este locus no es la habitual (a modo de campana de Gauss), según predicen los modelos de evolución en los microsatélites. Al analizar la distribución de sus alelos en un trasfondo de marcadores bialélicos resulta que, mientras los alelos de 6, 7 y 8 repeticiones se asocian con diferentes haplotipos, el alelo de 5 repeticiones siempre se encuentra en el mismo trasfondo, teniendo un ligamiento total con la variante M170C. Dado que M170C parece haber ocurrido una sola vez, hace unos 20.000 años (Semino y col. 2000), en la historia evolutiva del hombre, este alelo de 5 repeticiones tuvo que originarse también una sola vez, pero además, permaneciendo estable en esta forma alélica, puesto que no se han detectado haplotipos con M170C con otros alelos del locus.

Figura 17. Representación esquemática de la evolución de los alelos detectados para el marcador *DYS466*.



En definitiva, este locus tiene tres alelos (de 6, 7 y 8 repeticiones) que evolucionan al modo del resto de los microsatélites descritos, y un alelo (de 5 repeticiones) que se comporta como un evento único que mimetiza la evolución de M170C (Figura 17). Si bien el polimorfismo de regiones repetidas en serie aparece cuando se supera un umbral, fijado en >7 repeticiones (Meissier y col. 1996), parece que, al menos en este locus, las 5 repeticiones marcan el límite mínimo para la posibilidad de mutación. Precisamente por su forma de evolución, este locus no fue tenido en cuenta en los cálculos de coalescencia.

1.2. Sondeo en las bases de datos: detección del polimorfismo *DYS467*

Aprovechando la existencia de la gran cantidad de secuencias disponibles en las bases de datos, se han efectuado búsquedas de nuevos polimorfismos STR que han dado buenos resultados (Ayub y col. 2000). En una primera búsqueda Ayub y col. (2000) detectaron una región muy heterogénea del clon 264, M, 20, que incluía al STR *DYS389*, quedando descartada de posteriores exploraciones. Esta razón llevó a seleccionar dicha región del clon, junto a regiones adyacentes (≈ 1100 pb), para el escrutinio de nuevos polimorfismos. Tras la caracterización de las zonas con polimorfismo conocido dentro de la secuencia, se detectó una región en el extremo 5' de *DYS389* con un número elevado de repeticiones tetranucleotídicas (14) del tipo (GATA)_n no interrumpidas (Figura 18), por lo que se diseñaron cebadores para su

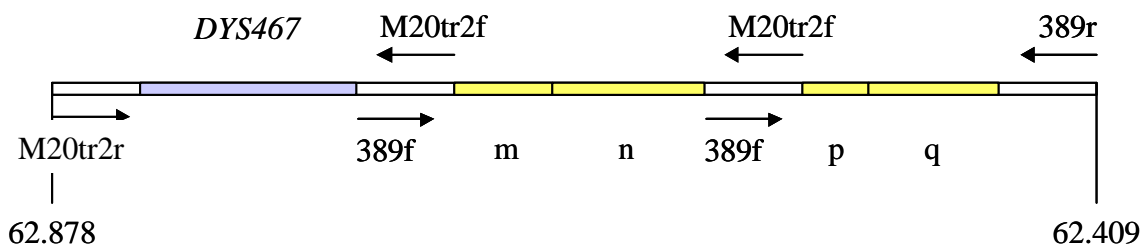


Figura 18. Representación esquemática de la región situada entre las posiciones 62409 y 62878 del clon 264, M, 20 (AC004617) donde se localiza el marcador *DYS467* (azul) descubierto en este trabajo. Las regiones repetitivas correspondientes al marcador *DYS389* (amarillo) se han subdividido en las regiones m, n, p y q. Las flechas indican las zonas de homología de los cebadores que se usan (la dirección marca el extremo 3') en las amplificaciones respectivas de ambos marcadores.

amplificación. Al igual que ocurre en el *DYS389*, uno de los cebadores empleados (M20tr2f) tiene homología en una zona duplicada, amplificándose dos productos: uno más pesado, que engloba a parte del *DYS389* (fragmentos m y n; Forster y col. 2000) junto al *DYS467*, y uno más ligero, que sólo contiene el *DYS467* (Figura 18).

El estudio se centró en el fragmento ligero por integrar sólo el STR de interés, ya que el fragmento pesado ofrece problemas en la interpretación de su variación al incorporar, a la vez, el polimorfismo de tres STR. Se han descrito 4 STR distintos, por interrupciones dentro de *DYS389*, denominados m, n, p y q (Forster y col. 2000). Las formas alélicas detectadas en el total de la muestra para *DYS467* contienen 7, 10, 11, 12, 13, 14, 15 y 16 repeticiones GATA. El alelo más frecuente en Canarias contiene 12 repeticiones, encontrándose en un 52'9% de los individuos. En orden decreciente de frecuencia le siguen: 13 repeticiones en 21'1%, 14 repeticiones en 12'9%, 11 repeticiones en 8'0%, 15 repeticiones en 4'0%, 10 repeticiones en 0'6%, 7 repeticiones en 0'3% y 16 repeticiones en 0'2%. El alelo de 7 repeticiones quizás se haya producido en un evento único puesto que no se detectaron alelos intermedios entre él y el de 10 repeticiones, y por la disminución que conllevaría en la tasa de mutación. En cualquier caso, su estructura sin interrupciones, el número de alelos detectados (7 en este trabajo) y la heterocigosidad que alcanza en las poblaciones (0'651 en Canarias) hacen de este nuevo marcador un locus muy útil para trabajos evolutivos y forensicos. Nótese que habitualmente las estimas de diversidad obtenidas de STR del cromosoma Y son menores que las de autosomas. White y col. (1999) describieron seis nuevos STR del cromosoma Y con una diversidad media de 0'663, lejos de la diversidades comúnmente detectadas en los STR nucleares usados en los casos forensicos (>0'750).

2. CLASIFICACIÓN Y RELACIÓN DE LOS HAPLOTIPOS

Los estados alélicos de cada marcador UEP se combinaron para construir haplotipos que, a continuación, se relacionaron mediante una red de máxima parsimonia (Figura 19). Dado que

se conocen perfectamente los estados ancestrales de los polimorfismos usados (Underhill y col. 2000), también conocemos cuales son los linajes más antiguos de la filogenia. Usando determinados polimorfismos, Underhill y col. (2000) dividieron la variabilidad del cromosoma Y mundial en diez grupos (I-X): los grupos I y II, comprenden los linajes más antiguos, que se encuentran casi exclusivamente en África; el grupo III, que es el más abundante en África y cuya frecuencia disminuye con el aumento de la distancia a este continente; el grupo IV, que está confinado casi exclusivamente a Japón; el grupo V, que se encuentra distribuido por Oceanía y centro-este de Asia; el grupo VI, que teniendo su mayor frecuencia en el oeste de Asia, también

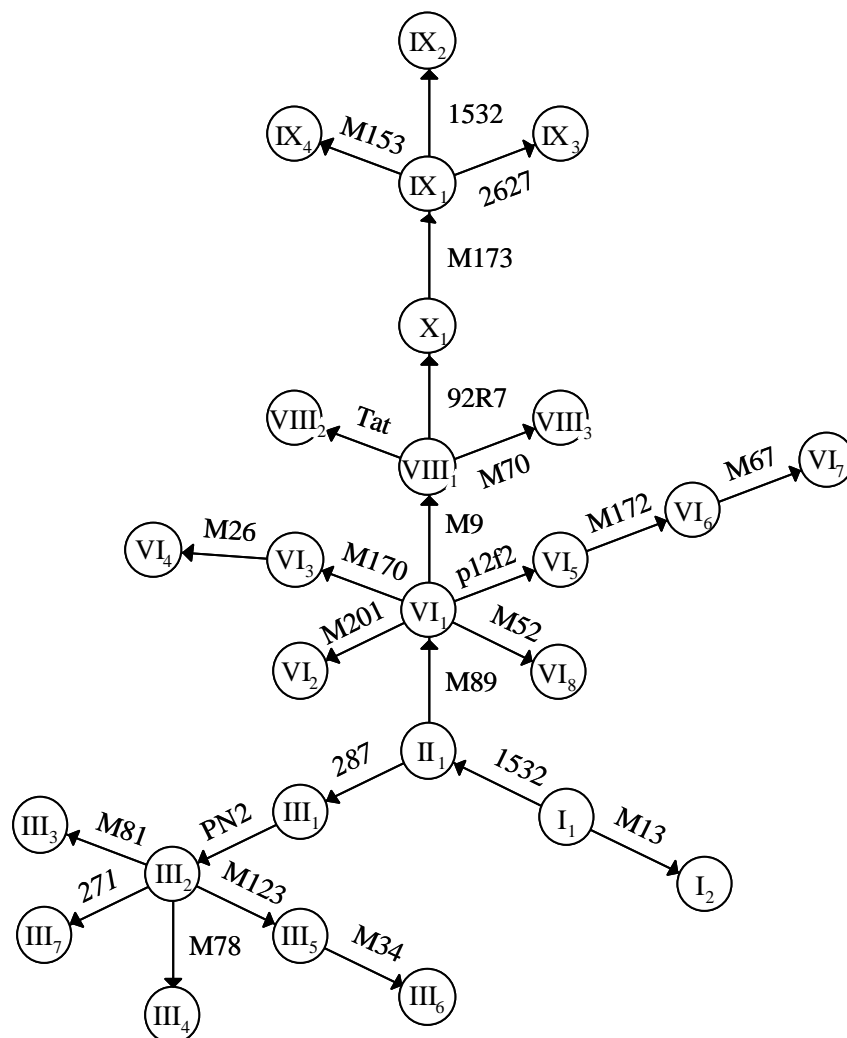


Figura 19. Red de máxima parsimonia relacionando los haplotipos del cromosoma Y (círculos). Las flechas denotan eventos únicos, indicado por los marcadores bialélicos, apuntando hacia el estado derivado. Las letras de los marcadores *DYS271*, *DYS287*, *SRY1532* y *SRY2627* se han omitido.

se encuentra distribuido por África y Europa; el grupo VII, que se localiza en el este asiático; el grupo VIII, que se distribuye desde Europa hasta Oceanía; el grupo IX, que alcanza su frecuencia más alta en Europa, aunque se distribuye hasta el oeste de Asia; y el grupo X, que se distribuye por Asia central y el continente americano. Por lo anteriormente expuesto se puede deducir que la distribución geográfica de estos grupos no es al azar.

Usando este método de clasificación, los grupos que se podrían detectar con los marcadores escogidos para el presente estudio son:

- Grupo I, por la presencia de *SRY1532A* (M42 en Underhill y col. 2000) en un trasfondo M89C.
- Grupo II, por la presencia de *SRY1532G* en un trasfondo M89C y YAP-. Aunque estos haplotipos podrían ser asignables al grupo V, su distribución geográfica lo descartaría.
- Grupo III, por la presencia de YAP+. El grupo IV también se detectaría en este trabajo por el mismo marcador, pero su asignación al este de Asia descarta su aparición en el rango geográfico de este trabajo.
- Grupo VI, por M89T en ausencia de la mutación M9.
- Grupo VIII, por M9G en ausencia de la mutación en 92R7 (similar a M45 en Underhill y col. [2000]).
- Grupo IX, por la presencia de M173C.
- Grupo X, por la presencia de 92R7T en un trasfondo M173A.

La nomenclatura usada para designar los diferentes haplotipos se basa: primero en su clasificación en uno de los grandes grupos, designándolo con el número romano correspondiente, y a continuación, se les asigna un número arábico, a modo de subíndice, para denotar la presencia de otras mutaciones diferenciadoras. En los 1471 individuos analizados se han detectado los siguientes grupos (Figura 20):

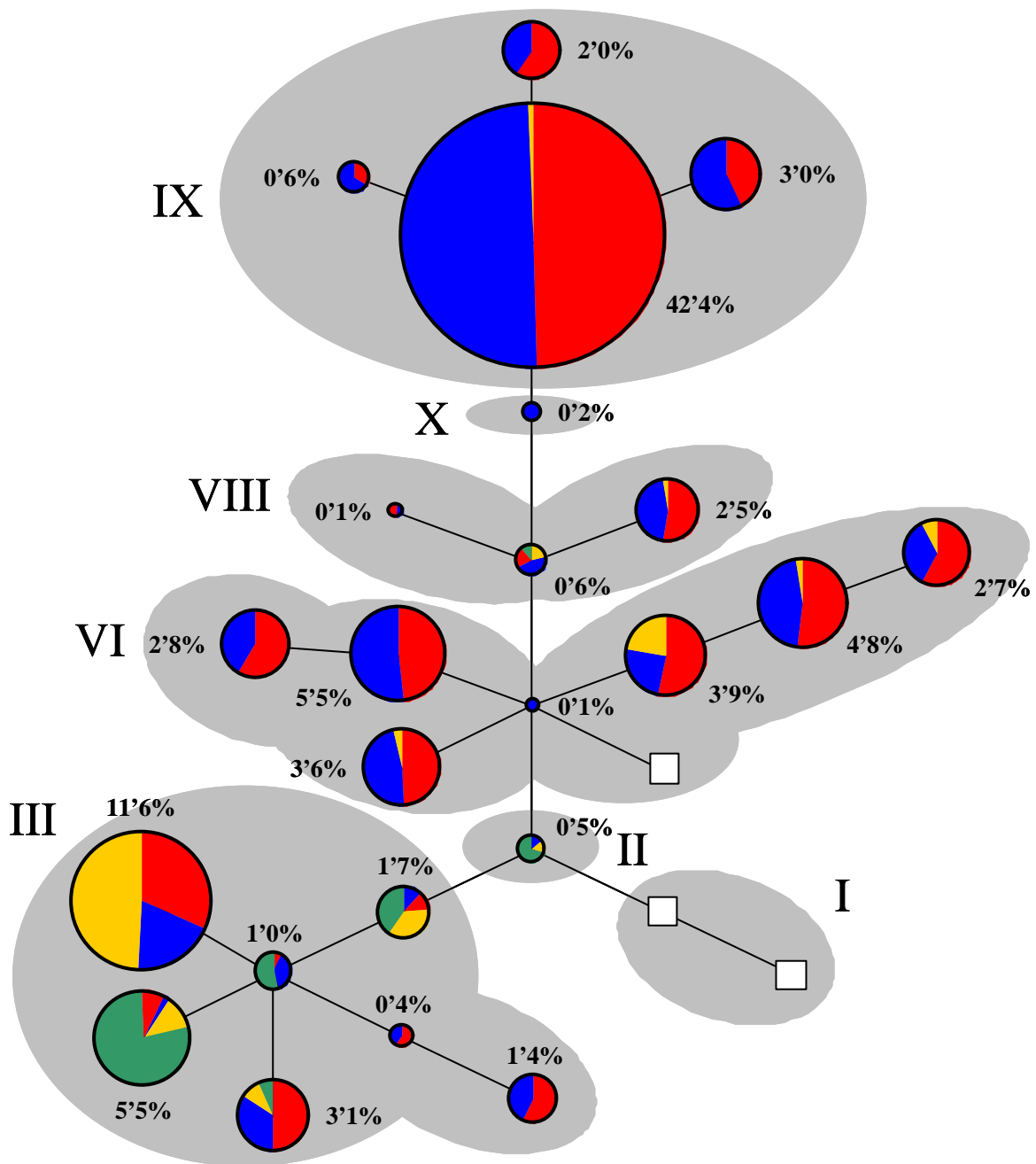


Figura 20. Se presenta la misma red de máxima parsimonia relacionando los haplotipos del cromosoma Y de la Figura 19, indicando las frecuencias para cada haplotipo de las muestras analizadas de Europa (azul), Canarias (rojo), norte de África (amarillo) y área subsahariana (verde). Los círculos son proporcionales a las frecuencias en el conjunto total y los cuadrados representan variantes no detectadas.

Grupo II: 0'5% (7 individuos)

Grupo VI: 23'5% (346 individuos)

Grupo III: 24'6% (362 individuos)

Grupo VIII: 3'2% (47 individuos)

Grupo IX: 48% (706 individuos)

Grupo X: 0'2% (3 individuos)

Dada la alta diferenciación detectada entre las poblaciones continentales que aquí se estudian (Flores y col. 2001c), es conveniente hacer un análisis previo de cada zona por separado.

3. VARIANTES EN LOS STR

Las variantes alélicas de los cinco marcadores STR sólo han sido estudiadas en algunos de los haplotipos definidos en base a los marcadores bialélicos (III₁, III₃, III₄, III₅, III₆, III₇, VI₄, VI₅, VI₆, VI₇ y VIII₃), por ser de especial interés para el poblamiento del Archipiélago Canario (ver Apéndice). De forma secundaria, esta información también se ha usado para el estudio de las otras áreas geográficas analizadas. El objeto de estos análisis es la datación, en ningún caso se han usado para estimar relaciones poblacionales.

4. ÁFRICA

La gran mayoría de los 225 individuos africanos analizados resultó pertenecer al grupo III (84'4%). El resto de los grupos de linajes se encontraron por debajo del 10%: grupo VI en 8'9%, grupo II en 2'7%, grupo IX en 2'2% y grupo VIII en 1'8%. Dentro del grupo mayoritario diferentes haplotipos se asocian a distintas zonas geográficas (Tabla 3). Tal y como se había descrito anteriormente, el haplotipo III₇ (definido por *DYS271*) es el mayoritario en las poblaciones subsaharianas, alcanzando el máximo de frecuencia en el oeste de África (Hammer y col. 1997, 1998; Underhill y col. 2000). Sin embargo, en el norte de África el panorama es totalmente diferente. Si bien el grupo III sigue siendo mayoritario, no se trata de los mismos haplotipos. En este caso el conjunto mayoritario lo constituyen los haplotipos YAP+/PN2T, tal y como se había descrito con anterioridad (Hammer y col. 1997; Passarino y col. 1998; Bosch y col. 1999; Scozzari y col. 1999; Flores y col. 2001c), la gran mayoría de los cuales incluyen la mutación M81 (haplotipo III₃). Resultados similares se obtuvieron con el análisis de sondas del sistema 49a,f/*TaqI*. Los haplotipos p49a,f ht4 y ht5 se han asociado al marcador *DYS271* y al

Tabla 3. Frecuencia (%) de haplotipos del cromosoma Y en África.

Grupo	Haplotipo	MD ^a	Poblaciones												
			Árabes M.	Bereberes M.	Sahara Oc.	Mauritania	Malí	Sudán	Etiopía	Senegal	C. Marfil	G.Ecuat.	Khoisan	A.Central	A.Sur
I	1	- ^b					2'3		1'1				43'6		5'6
	2	M13						45'0	12'5						
II	1	<i>SRY1532</i>	3,4				11'4	15'0	10'2	2,3	14,3	3,8	28'2	33'3	13'0
III	1	<i>DYS287</i>		4,3	12,5	11,1	34'1	7'5	17'1	14,0	9,5	7,7		8'3	14'8
	2	PN2							25'0	7,0	9,5	7,7	10'3		3'7
	3	M81	62,0	76,1	66,7	41,7	29'5	5'0							
	4	M78	3,4	6,5				17'5	22'7	4,6		3,8			
	6	M34							2'3						
	7	<i>DYS271</i>				25,0	20'4		3'4	72,1	66,7	73,2	17'9	58'4	62'9
	VI	2	M201	7,0						1'1 ^c					
5		12f2			20,8	22,2		10'0	3'4 ^c						
6		M172	3,4	2,2											
7		M67	7,0	2,2											
VIII	1	M9	3,4	2,2							3,8				
	3	M70	3,4												
IX	1	M173	7,0	6,5			2'3		1'1						
Tamaño muestral			29	46	24	36	44	40	88	43	21	26	39	36	54

^aÚltimo marcador definitorio; ^bEstado ancestral para todos los marcadores; ^cDeterminado en Semino y col. (2002).

haplotipo YAP+/PN2T, respectivamente, alcanzando frecuencias cercanas a las detectadas en este trabajo (Pasarino y col. 1998; Lucotte y col. 2000). Para el resto, mientras el grupo de linajes II se detecta principalmente en las poblaciones subsaharianas, los demás grupos se encuentran casi en su totalidad asociados a las poblaciones del norte de África (VI, VIII y IX). El grupo I no se ha detectado en este trabajo, pero sí aparece en algunas poblaciones de referencia tanto en el norte como en el sur de África (Tabla 3).

Un análisis de varianza incluyendo las muestras africanas analizadas desveló que un alto porcentaje de la variación que se observa se distribuye entre las poblaciones (32'4%, $P < 0'00001$). Sin embargo, cuando se realizó un análisis de varianza sólo para el noroeste africano se obtuvo que tan sólo un 8'1% ($P < 0'00001$) de la variación es atribuible a diferencias entre poblaciones, reduciéndose hasta el 2'4% ($P = 0'08016$) cuando se elimina Mauritania del cálculo. El valor de estructuración obtenido para las tres poblaciones subsaharianas fue de -2'07% ($P = 0'88400$). Si lo que se comparan son los dos grupos de poblaciones, subsaharianas y norteafricanas, la variación entre grupos llega hasta el 41'7% ($P < 0'00001$), mientras que las diferencias entre las poblaciones dentro de una misma región sólo son del 2'89% ($P < 0'00001$). El valor de estructuración obtenido en primera instancia no se debe a las poblaciones pertenecientes a una misma área geográfica, sino a las diferencias existentes entre el noroeste africano y las poblaciones subsaharianas. Este resultado también se observa al comparar las distancias genéticas de las poblaciones dos a dos (Tabla 4). Las distancias más grandes son siempre entre las poblaciones norteafricanas y las subsaharianas, mostrando Mauritania una posición intermedia entre ambos grupos. La estructuración geográfica también es visible al realizar un análisis de correlación entre distancias genéticas y geográficas (Coef. = 0'580, $P = 0'009$).

Combinando las poblaciones del presente trabajo con las tomadas de la bibliografía, el valor de estructuración de las poblaciones africanas sigue siendo muy alto (26'7%, $P < 0'00001$).

Tabla 4. Distancias genéticas (bajo la diagonal) y probabilidades de diferenciación (sobre la diagonal) entre las poblaciones africanas.

	Árabes M.	Bereberes M.	Sahara Oc.	Mauritania	Malí	Sudán	Etiopía	Senegal	C.Marfil	G.Ecuat.	Khoisan	A.Central	A.Sur
Árabes M.	-	0'3154	0'0047	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000
Bereberes M.	0'0309	-	0'0075	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000
Sahara Oc.	0'0692	0'0706	-	0'0282	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000
Mauritania	0'1254	0'1804	0'0947	-	0'0001	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000
Malí	0'1773	0'2429	0'1869	0'0845	-	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000
Sudán	0'3107	0'4008	0'3473	0'2472	0'2219	-	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000
Etiopía	0'2807	0'3675	0'3175	0'2058	0'1507	0'1088	-	0'0000	0'0000	0'0000	0'0000	0'0000	0'0000
Senegal	0'4751	0'5650	0'5170	0'2819	0'2557	0'3970	0'3052	-	0'3812	0'8708	0'0000	0'0022	0'1100
C.Marfil	0'4418	0'5352	0'4875	0'2594	0'2283	0'3603	0'2720	0'0224	-	0'8347	0'0000	0'1465	0'7335
G.Ecuat.	0'4761	0'5679	0'5226	0'2881	0'2740	0'4013	0'3139	0'0065	0'0189	-	0'0000	0'0101	0'1807
Khoisan	0'3528	0'4516	0'4080	0'2726	0'2247	0'2618	0'1896	0'3347	0'2749	0'3315	-	0'0000	0'0000
A.Central	0'4252	0'5226	0'4753	0'2637	0'2122	0'3280	0'2654	0'1120	0'0471	0'1062	0'2329	-	0'0846
A.Sur	0'4204	0'5124	0'4620	0'2363	0'1916	0'3367	0'2546	0'0252	0'0100	0'0301	0'2415	0'0452	-

Nota.- Valores de $P < 0'05$ en negrita.

Un valor ligeramente superior (38'8%) se obtuvo en otro trabajo (Scozzari y col. 1999) también para el conjunto del continente africano. Las discrepancias entre ambos valores reflejan probablemente la diferencia en las muestras analizadas, ya que Scozzari y col. (1999) incluyeron grupos de África central muy diferenciados (46'3% de variación entre ellos), y a que en sus cálculos se tuvo en cuenta no sólo las diferencias de frecuencias, sino también las diferencias moleculares entre haplotipos. En cualquier caso, dado que la geografía se correlaciona mejor que el lenguaje con la estructura genética de las poblaciones africanas (Scozzari y col. 1999), se distinguen tres grupos de poblaciones: noroeste (árabes y bereberes de Marruecos, Sahara Occidental, Mauritania y Malí), noreste (Sudán y Etiopía) y el área subsahariana (Senegal, Costa de Marfil, Guinea Ecuatorial, África Central, África del Sur y khoisan). Con estos tres grupos casi se explica el total de la variabilidad inicial (22'7%, $P < 0'00001$), aunque aún se detectan diferencias significativas dentro de los grupos (8'8%, $P < 0'00001$). En las poblaciones del norte africano, al incluir Malí con las poblaciones del oeste, la variación resulta casi igual que sin ella (11'8%, $P < 0'00001$) y muy similar a la que se detecta en el este (9'0%, $P < 0'00001$) o al detectado por Scozzari y col. (1999) en todo el norte de África (8'2%). Sin embargo, la incorporación de nuevas poblaciones al grupo subsahariano hace que la variación entre poblaciones pase de no ser significativamente diferente de cero a un 13'0% ($P < 0'00001$). Este cambio se debe exclusivamente a la incorporación de los khoisan. Sin ellos, la variación entre poblaciones subsaharianas se reduce hasta un 1'6% ($P = 0'09775$), es decir, son homogéneas.

Revisando las distancias genéticas (Tabla 4) se observa que Malí guarda las menores distancias con Mauritania, pero es bastante divergente de las poblaciones de Marruecos y el Sahara Occidental. En el caso de los khoisan, resulta llamativo que presenten distancias similares al ser comparados con poblaciones del norte de África y del área subsahariana. Es destacable que precisamente la menor distancia la tengan con la muestra de Etiopía, quizás debido a la cantidad de linajes ancestrales que comparten (Underhill y col. 2000; Semino y col. 2002). Aún así, las

relaciones poblacionales a nivel geográfico y genético siguen estando significativamente correlacionadas (Coef. = 0'274, P = 0'024).

La alta diferenciación detectada se acentúa al comparar determinados grupos de poblaciones. Los más diferenciados son el conjunto de poblaciones subsaharianas, con o sin los khoisan, con respecto a las poblaciones del noroeste de África (Tabla 5), quizás en gran parte debido al nivel de aislamiento que sufren ambos grupos. Este mismo resultado puede deducirse de los análisis de correspondencia de las poblaciones africanas de Scozzari y col. (1999).

Tabla 5. Análisis de varianza entre grupos de poblaciones africanas.

Comparaciones	% de variación	Significación
Noroeste vs. Noreste	17'9	0'04400
Noroeste vs. Subsaharianas ^a	28'0	0'00098
Noroeste vs. Subsaharianas ^b	33'0	0'01466
Noreste vs. Subsaharianas ^a	19'5	0'04008
Noreste vs. Subsaharianas ^b	26'6	0'04790

^a Khoisan incluidos; ^b Khoisan excluidos.

Las relaciones entre las poblaciones africanas se aprecian muy bien en las representaciones gráficas (Figura 21), siendo casi idénticas en ambas redes. En los dos casos se observan los tres grupos de poblaciones que se destacaron por su relación geográfica: por un lado, las poblaciones del noroeste africano con las muestras de Marruecos en el extremo y las de la zona del Sahel más cercanas a las subsaharianas; por otro lado se distingue el grupo de las dos poblaciones del noreste africano; y en último lugar, las poblaciones del área subsahariana, con las del oeste africano en un extremo y los khoisan en una posición intermedia. Nótese que las relaciones que se aprecian en la red concuerdan perfectamente con el nivel de variación detectado entre los grupos, en especial la relación de las poblaciones del oeste donde conviven

los grupos más divergentes entre sí. Los análisis multidimensionales también ofrecen la misma perspectiva, aunque acercan mucho más a los khoisan a la zona del noreste (Figura 22).

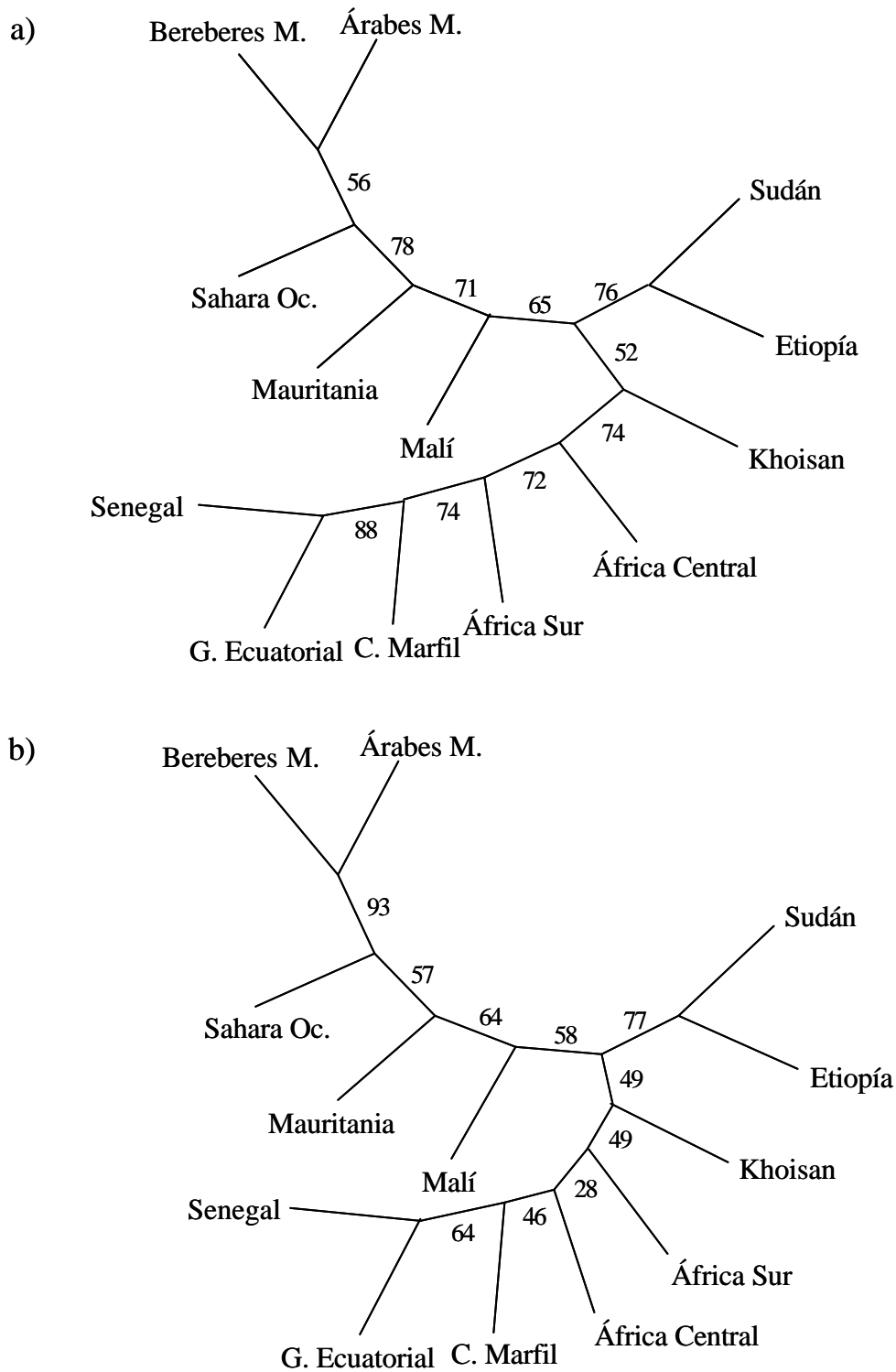


Figura 21. Relaciones genéticas entre las poblaciones africanas en base al método NJ (a) y ML (b). Los valores de las ramas indican el porcentaje de bootstraps a partir de 1000 replicaciones.

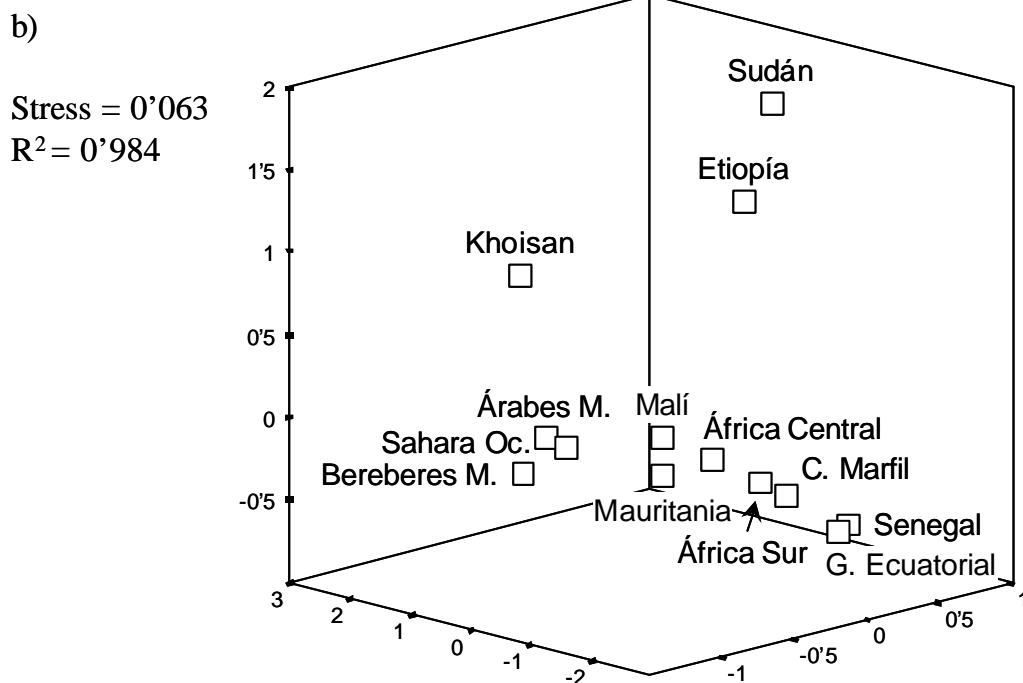
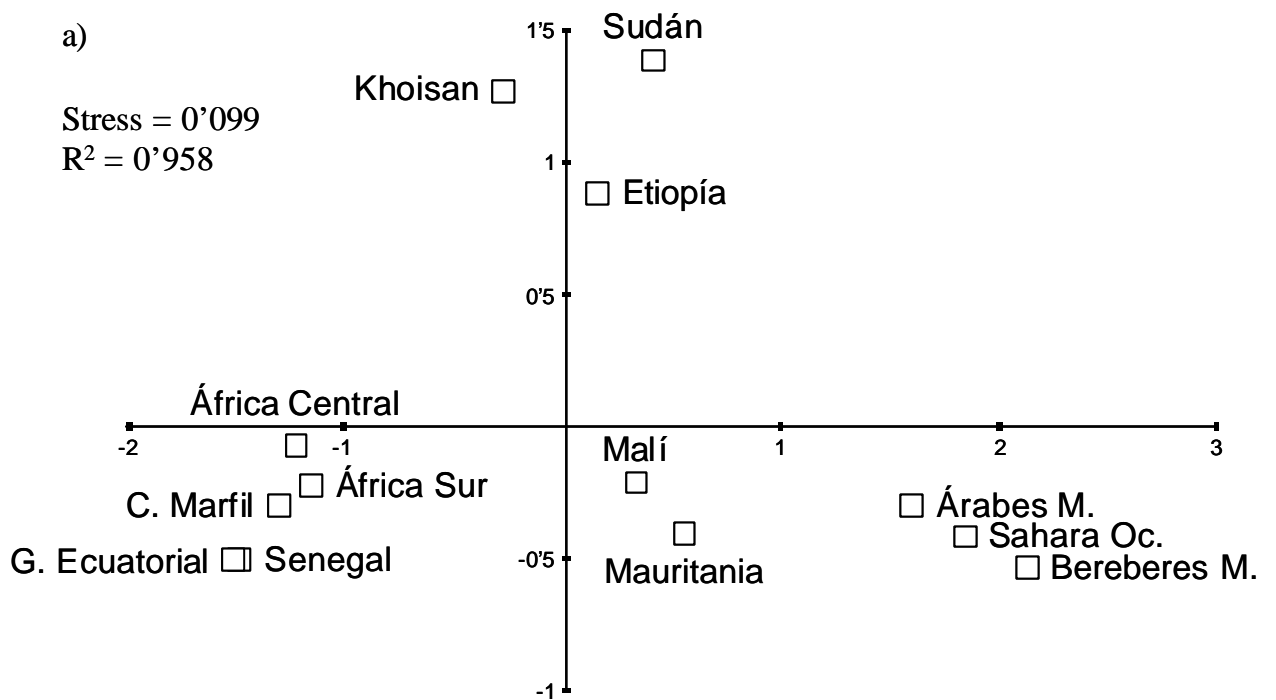


Figura 22. Análisis multidimensional en dos (a) y tres (b) dimensiones de las distancias genéticas entre las poblaciones africanas consideradas.

5. PENÍNSULA IBÉRICA Y NORMANDÍA

Un 59'4% de los 594 individuos analizados de este área geográfica se clasificaron dentro del grupo IX, constituyendo así el grupo mayoritario. A éste le siguen en frecuencia el grupo VI (24'6%) y el grupo III (11'8%). El resto de grupos detectados no superan en conjunto el 5% de frecuencia: 3'5% del grupo VIII, 0'5% del grupo X y 0'2% del grupo II. Tal y como se había descrito en trabajos anteriores, el grupo más abundante en la región está definido por la mutación M173 (Tabla 6). El haplogrupo 1 y el haplotipo p49a,f ht15, linajes estrechamente relacionados con este grupo, se detectan en Europa en un rango de frecuencia de 40-80% (Lucotte y Hazout, 1996; Lucotte y Loirat, 1999; Hill y col. 2000; Rosser y col. 2000; Flores y col. 2001c; Scozzari y col. 2001; Wilson y col. 2001), alcanzando sus máximos en el oeste. Dentro del grupo IX, los haplotipos que se distinguen están muy localizados geográficamente. Los haplotipos IX₃ y IX₄, definidos por las mutaciones *SRY2627* y M153 respectivamente, se localizan principalmente en la Península Ibérica y regiones relacionadas geográfica e históricamente. Por ello se les ha atribuido un origen reciente en la región (Hurles y col. 1999; Bosch y col. 2001). Ambos haplotipos alcanzan su máximo en el noreste peninsular (Tabla 6). Por otra parte, el haplotipo IX₂, definido por la mutación *SRY1532*, se encuentra principalmente asociado a las poblaciones del noreste de Europa, superando el 50% en Rusia y Polonia (Semino y col. 2000; Rosser y col. 2000). Sin embargo, en la Península Ibérica es escaso y, además, no parece asociado a ninguna región en particular (Tabla 6).

Dentro del grupo VI se engloban haplotipos de asignación europea (VI₃ y VI₄), y haplotipos relacionados con Oriente Próximo (VI₅-VI₇) y con el Cáucaso (VI₂)(Semino y col. 2000). Todos ellos se encuentran muy distribuidos por la geografía peninsular, aunque algunos alcanzan frecuencias relativamente altas en el norte (VI₃ y VI₄) o en el sur (VI₆).

Los haplotipos del grupo III se encuentran también bastante dispersos por la geografía peninsular aunque todos ellos alcanzan sus máximas frecuencias en Andalucía, a excepción de

Tabla 6. Frecuencia (%) de haplotipos del cromosoma Y en la Península Ibérica y Normandía.

Grupo	Haplotipo	MD ^a	Poblaciones													
			Sevilla	Huelva	Cádiz	Córdoba	Málaga	Portugal	León	Galicia	Cantabria	Valencia	Castilla	Cataluña	P. Vasco	Normandía
II	1	<i>SRY1532</i> ^b							0'9							
III	1	<i>DYS287</i>					3'8				2'9					
	2	PN2	0'6					0'9		5'3	1'4	3'2				
	3	M81	4'5	4'5		7'4	11'5	5'5	5'0	10'5	8'6	6'5	4'8		2'2	
	4	M78	0'6	4'5	3'6	3'7	7'7	6'4	3'3	5'3						
	5	M123						1'8								
	6	M34	1'3				3'8	0'9	1'7	10'5		6'5				
	7	<i>DYS271</i>		4'5			3'8									
VI	1	M89						0'9	1'7							
	2	M201	4'0				3'8	7'3	6'7		8'6			18'7		
	3	M170	9'7	9'2	3'6	11'0		3'7	1'7		4'3	9'7	14'3	6'2	4'4	26'9
	4	M26	2'6		10'7	3'7		1'8	1'7		1'4	3'2	19'0		4'4	
	5	12f2	3'2		3'6			3'7	1'7	5'3	2'9					
	6	M172	5'2	13'7	14'3	11'0	7'7	4'6	3'3		2'9	9'7				
	7	M67	2'6			3'7	7'7	3'7	1'7				9'5			
VIII	1	M9				3'7		0'9	3'3							
	2	Tat	0'6													
	3	M70	4'5	4'5	10'7			0'9			4'3	3'2				
IX	1	M173	57'4	54'6	42'8	55'8	31'0	50'5	58'2	63'1	48'4	45'1	38'1	43'8	62'3	61'5
	2	<i>SRY1532</i> ^b	0'6				7'7		6'7		4'3	3'2				3'9
	3	<i>SRY2627</i>	2'6	4'5	7'1		7'7	2'8	3'3		8'6	6'5	4'8	31'3	11'1	7'7
	4	M153			3'6		3'8				1'4	3'2	9'5		15'6	
X	1	92R7						2'8								
Tamaño muestral			155	22	28	27	26	109	60	19	70	31	21	16	45	26

^aÚltimo marcador definitorio; ^bMutación recurrente.

los definidos por las mutaciones M123 y M34 (haplotipos III₅ y III₆). Estos dos haplotipos no se detectan en África, salvo en Etiopía (Underhill y col. 2000; Semino y col. 2002); pero en conjunto el grupo III no alcanza en la Península Ibérica frecuencias equiparables a las detectadas en África.

Dentro del grupo VIII, salvo el haplotipo VIII₃, que se encuentra asociado principalmente a Andalucía (Tabla 6) como ya ha sido descrito (Semino y col. 2000), el resto de haplotipos se encuentran en muy baja frecuencia y/o distribuidos en poblaciones geográficamente no relacionadas.

El análisis de varianza dentro de esta región desveló que tan sólo un 0'9% ($P = 0'03226$) de la variación que se detecta se distribuye entre las poblaciones. Este valor de estructuración implica que la migración entre áreas ha sido suficiente para borrar casi todas sus diferencias y, por tanto, homogeneizar casi totalmente a la Península Ibérica. La inclusión de poblaciones peninsulares más diferenciadas, como vascos y catalanes, no hacen variar la baja estructuración que presenta esta región geográfica (1'35% de variación, $P = 0'00391$). Tal y como era de esperar de estos resultados, no existen diferencias significativas entre las poblaciones de la cornisa cantábrica y Andalucía (-0'06% de variación entre grupos, $P = 0'33200$), ni entre la cornisa atlántica y mediterránea (0'29% de variación entre grupos, $P = 0'08000$). Sin embargo, dentro de regiones, los niveles globales de diferenciación peninsular se mantienen tanto en Andalucía, aunque no significativamente (1'12% de variación entre poblaciones, $P = 0'10166$), como en la cornisa cantábrica (1'20% de variación entre poblaciones, $P = 0'01197$). Estos valores de estructuración, si bien ligeramente inferiores, son comparables a los detectados en algunas regiones de África analizadas en este trabajo. Este resultado era de esperar dado que, en las comparaciones de datos globales del cromosoma Y, Europa es el continente en el que menor diferenciación se detecta (Hammer y col. 2001; Kayser y col. 2001).

Al comparar las distancias genéticas por parejas (Tabla 7) se observa que las poblaciones

Tabla 7. Distancias genéticas (bajo la diagonal) y probabilidades de diferenciación (sobre la diagonal) en las poblaciones de la Península Ibérica y Normandía.

	Sevilla	Huelva	Cádiz	Córdoba	Málaga	Portugal	León	Galicia	Cantabria	Valencia	Castilla	Cataluña	P. Vasco	Normandía
Sevilla	-	0'4849	0'0852	0'6380	0'0007	0'0347	0'0374	0'2244	0'0638	0'1754	0'0095	0'0360	0'0002	0'4065
Huelva	0'0120	-	0'7246	0'9862	0'5372	0'6267	0'3424	0'2823	0'3565	0'9724	0'0786	0'0407	0'0052	0'0896
Cádiz	0'0356	0'0267	-	0'2813	0'1002	0'0762	0'0365	0'0344	0'0775	0'6011	0'1492	0'0375	0'0046	0'0059
Córdoba	0'0098	0'0092	0'0373	-	0'1997	0'7569	0'5110	0'2557	0'1197	0'8281	0'1881	0'0049	0'0049	0'0699
Málaga	0'0680	0'0585	0'0455	0'0631	-	0'0190	0'1704	0'3551	0'0727	0'4402	0'1007	0'1335	0'0000	0'0005
Portugal	0'0115	0'0188	0'0312	0'0168	0'0392	-	0'7038	0'6355	0'0712	0'1119	0'0260	0'1793	0'0001	0'0545
León	0'0123	0'0249	0'0489	0'0211	0'0610	0'0109	-	0'5507	0'4360	0'2249	0'0084	0'1261	0'0027	0'0629
Galicia	0'0303	0'0468	0'0811	0'0410	0'0925	0'0321	0'0250	-	0'1718	0'5795	0'0102	0'0014	0'0050	0'0008
Cantabria	0'0158	0'0252	0'0316	0'0291	0'0356	0'0110	0'0153	0'0440	-	0'3751	0'0258	0'6306	0'0116	0'1605
Valencia	0'0191	0'0163	0'0198	0'0194	0'0331	0'0201	0'0299	0'0492	0'0159	-	0'3266	0'1184	0'0282	0'2319
Castilla	0'0582	0'0699	0'0438	0'0564	0'0547	0'0528	0'0750	0'1078	0'0529	0'0391	-	0'0240	0'0449	0'0165
Cataluña	0'0919	0'0985	0'0901	0'1153	0'0850	0'0753	0'0887	0'1420	0'0532	0'0781	0'1061	-	0'0193	0'0204
P. Vasco	0'0361	0'0491	0'0647	0'0501	0'0984	0'0461	0'0377	0'0554	0'0426	0'0485	0'0733	0'0994	-	0'0090
Normandía	0'0370	0'0515	0'0954	0'0470	0'1287	0'0657	0'0625	0'0904	0'0643	0'0568	0'0894	0'1205	0'0670	-

Nota.- Valores de $P < 0'05$ en negrita.

más diferenciadas son las que se encuentran en el noreste peninsular (catalanes y vascos), castellanos y normandos. A pesar del bajo grado de estructuración de las poblaciones peninsulares, las distancias genéticas están correlacionadas con las geográficas (Coef. = 0'418, $P = 0'027$). La correlación se mantiene en el resto del continente europeo en valores similares a los detectados en este trabajo (0'387; Rosser y col. 2000).

Las representaciones gráficas resultantes (Figura 23) reflejan los mismos resultados que los análisis de varianza: la alta mezcla de estas poblaciones hace que se muevan de rama en rama en las diferentes comparaciones, disminuyendo los valores de bootstrap. Las topologías resultantes de ambas representaciones no son exactamente iguales, aunque en ambos casos se conserva la asociación de algunas poblaciones del noreste peninsular y Normandía. La red de ML (Figura 23b) distingue tres grupos de poblaciones con una estrecha relación geográfica: normandos, vascos, castellanos, catalanes y cántabros por un lado; valencianos, sevillanos, cordobeses, onubenses y gaditanos en medio; y portugueses, leoneses y gallegos por otro. Sin embargo, los malagueños quedarían incluidos en el grupo de poblaciones del noroeste peninsular. En esta misma red se observa que el valor de bootstrap más alto obtenido, 50%, separa a los vascos y castellanos del resto de las poblaciones consideradas.

Los análisis multidimensionales (Figura 24) ofrecen una situación similar. Descartando la Figura 24a, que por su valor de stress no refleja una buena representación, se distinguen, a grandes rasgos, cuatro grupos de poblaciones: por un lado, los vascos, normandos y catalanes hacia un extremo; por otro, el sur de Andalucía (representado por gaditanos y malagueños) y los castellanos a una ligera distancia; a los gallegos como un grupo aislado; y por último, el resto de poblaciones.

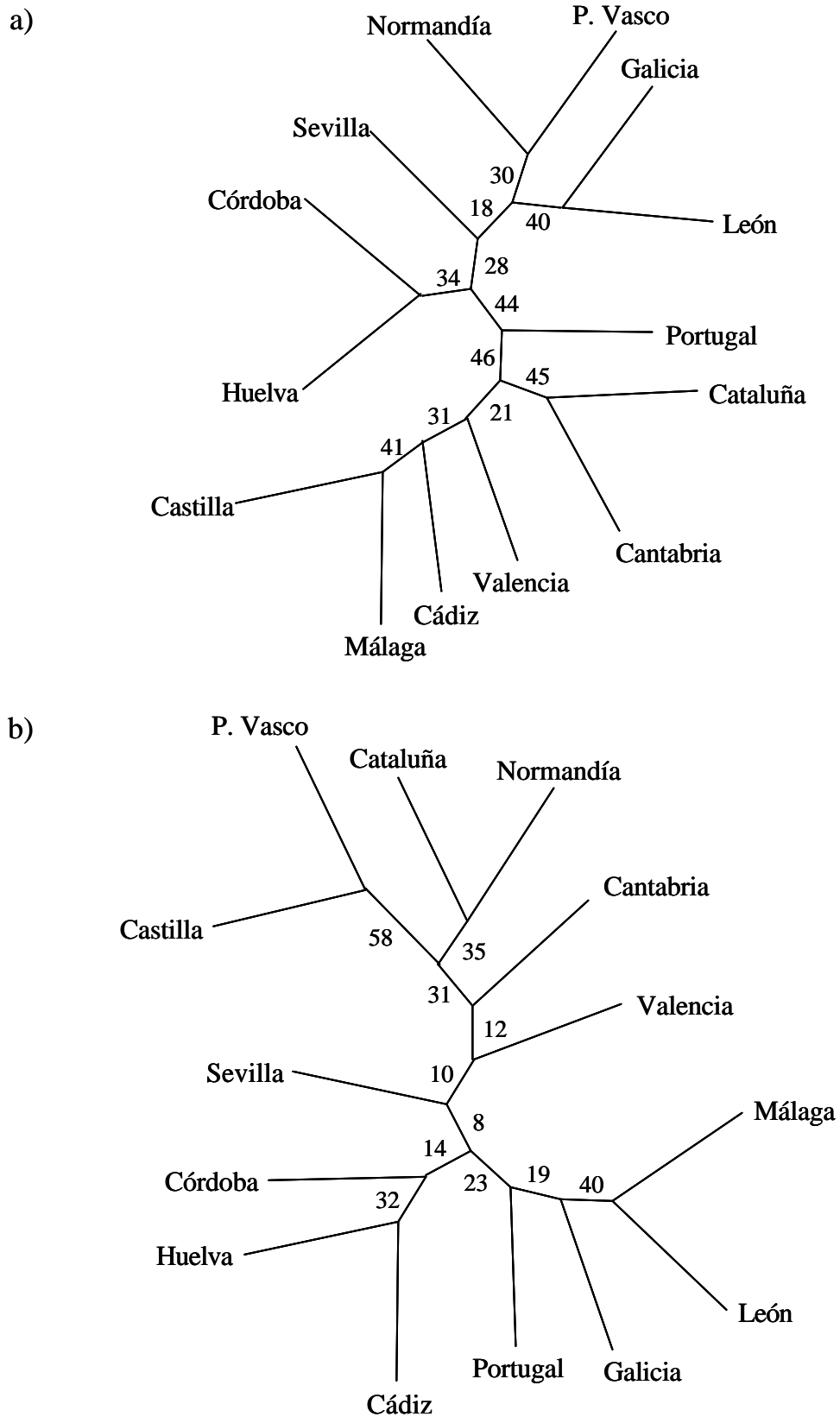


Figura 23. Relaciones genéticas entre las poblaciones peninsulares y normandos en base al método de NJ (a) y ML (b). Los valores de las ramas indican el porcentaje de bootstraps a partir de 1000 replicaciones.

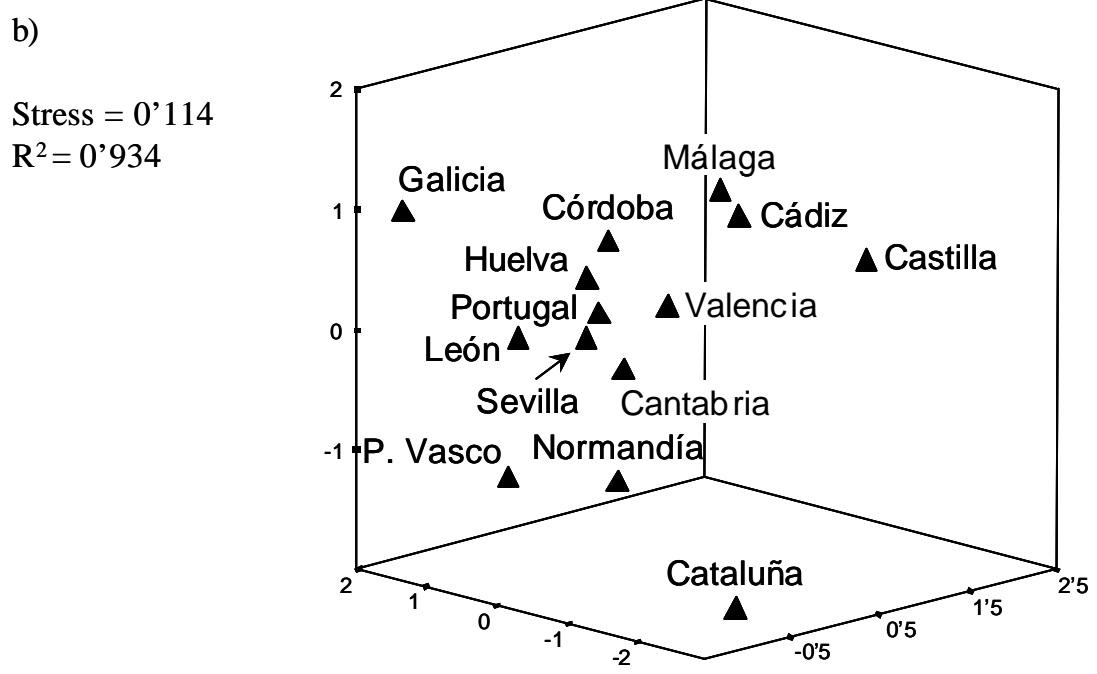
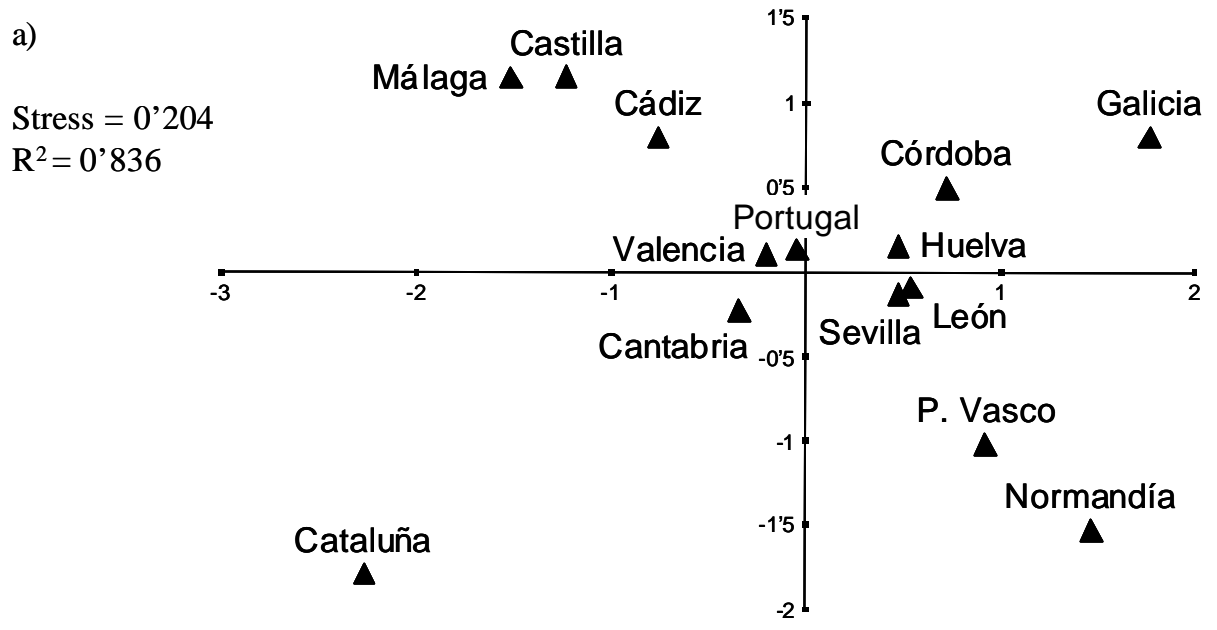


Figura 24. Análisis multidimensional en dos (a) y tres (b) dimensiones de las distancias genéticas entre las poblaciones peninsulares y normandos.

6. ISLAS CANARIAS

El 53'4% de los 652 individuos de procedencia canaria pertenecen al grupo IX. A este conjunto le siguen en frecuencia los grupos VI (27'6%) y III (15'6%). El 3'4% restante está incluido dentro del grupo VIII. El grupo mayoritario en la Península Ibérica también es el mayoritario en Canarias (Tabla 8), aunque comparativamente las frecuencias en el Archipiélago son estadísticamente menores ($P < 0'05$, prueba U de Mann-Whitney). Dentro de este grupo el haplotipo IX₁ se encuentra casi siempre por debajo del 50%, excepto en Tenerife y Gran Canaria. Como era de esperar, los haplotipos de origen peninsular, IX₃ y IX₄, también se detectan en las islas, incluso en algunas en frecuencias llamativas. Es el caso del haplotipo IX₃ en El Hierro y Fuerteventura, que aparece en un 12'7% y 8%, respectivamente. Este haplotipo aparece en frecuencias equiparables en la Península Ibérica sólo en el noreste (vascos y catalanes). El haplotipo IX₂, a diferencia de la Península Ibérica, aparece consistentemente en todas las islas, aunque en frecuencias que no destacan de las continentales.

Con respecto a los haplotipos del grupo VI, sus distribuciones en las islas son similares a las observadas para la Península Ibérica, aunque nuevamente ciertos haplotipos destacan en ciertas islas. Es el caso del haplotipo VI₃ y su derivado VI₄ en La Gomera. Frecuencias equiparables a la de esta isla sólo se detectan en Castilla, Normandía y Cádiz. Otro caso es el haplotipo VI₅, que aparece en las tres islas más occidentales en frecuencias mucho más altas de las detectadas en la Península Ibérica.

Los haplotipos del grupo III, asociados principalmente al continente africano, también se detectan en el Archipiélago, superando en ocasiones a los valores máximos peninsulares, así, por ejemplo, la frecuencia de III₃ en Fuerteventura y Gran Canaria supera los valores detectados en Málaga y Galicia, pero nunca el nivel detectado en África. Los haplotipos asociados principalmente al África subsahariana, III₁ y III₇, se detectan en todas las islas excepto en Tenerife y La Gomera.

Tabla 8. Frecuencia (%) de haplotipos del cromosoma Y en Canarias.

Grupo	Haplotipo	MD ^a	Poblaciones						
			El Hierro	La Palma	La Gomera	Tenerife	G. Canaria	Lanzarote	Fuerteventura
III	1	<i>DYS287</i>						1'0	2'7
	2	PN2				0'6			
	3	M81	2'1	5'9	4'3	10'7	11'6	6'2	13'3
	4	M78	6'4	2'4	4'3	3'4	3'8	3'1	2'7
	5	M123				1'7			
	6	M34	2'1	3'5	2'2	2'2	2'6		
	7	<i>DYS271</i>	4'3	2'4			1'3	1'0	
VI	2	M201	4'3	2'4	5'4	3'9	3'8	5'2	2'7
	3	M170	2'1	4'7	12'0	5'6	5'1	7'2	2'7
	4	M26		4'7	8'7	1'7	1'3	6'2	2'7
	5	12f2	6'4	7'0	7'6	2'2	3'8	5'2	4'0
	6	M172	8'5	7'0	3'3	5'1	3'8	7'2	6'7
	7	M67	4'3	2'4	7'6	1'1	1'3	3'1	8'0
	VIII	1	M9				0'6		1'0
2		Tat		1'2					
3		M70		3'5	1'1	5'6	5'1		1'3
IX	1	M173	44'7	47'0	41'3	50'0	53'9	49'5	39'9
	2	<i>SRY1532</i>	2'1	2'4	2'2	2'2	1'3	4'1	5'3
	3	<i>SRY2627</i>	12'7	3'5		1'7	1'3		8'0
	4	M153				1'7			
Tamaño muestral			47	85	92	178	78	97	75

^aÚltimo marcador definitorio.

En referencia al grupo VIII, la mayoría de los linajes están incluidos en el haplotipo VIII₃, concentrados principalmente en las dos islas centrales, Tenerife y Gran Canaria, y en valores equiparables a los de Andalucía.

Al hacer un análisis de varianza se detecta una homogeneidad total en Canarias, ya que tan sólo un 0'3% ($P = 0'13300$) de la variación es atribuible a diferencias entre las islas. Sin embargo, hay que destacar que ciertas poblaciones son significativamente diferentes de otras, como es el caso de La Gomera y El Hierro (Tabla 9). El patrón de distancias genéticas resultante no guarda un patrón correlacionado con la geografía (Coef. = $-0'081$, $P = 0'612$), aunque si es destacable que ciertas parejas cercanas geográficamente se asemejen genéticamente. Esto queda

Tabla 9. Distancias genéticas (bajo la diagonal) y probabilidades de diferenciación (sobre la diagonal) entre las poblaciones de las Islas Canarias.

	El Hierro	La Palma	La Gomera	Tenerife	G. Canaria	Lanzarote	Fuerteventura
El Hierro	-	0'6375	0'0046	0'0033	0'0894	0'0307	0'3528
La Palma	0'0118	-	0'2727	0'2954	0'8862	0'5256	0'3076
La Gomera	0'0263	0'0120	-	0'0027	0'0638	0'5224	0'0082
Tenerife	0'0227	0'0065	0'0204	-	0'9971	0'0756	0'0243
G. Canaria	0'0266	0'0087	0'0257	0'0021	-	0'2497	0'1646
Lanzarote	0'0201	0'0048	0'0102	0'0078	0'0097	-	0'1204
Fuerteventura	0'0164	0'0136	0'0201	0'0167	0'0230	0'0179	-

Nota.- Valores de $P < 0'05$ en negrita.

reflejado perfectamente en las representaciones gráficas NJ y ML (Figura 25). En ambos casos la topología es idéntica, pero además, los valores de bootstrap son también bastante altos en algunas ramas, como la unión de las dos islas centrales, la unión de Lanzarote y La Gomera o la unión de El Hierro y Fuerteventura, esta última sólo en el NJ. Los análisis multidimensionales (Figura 26) dejan ver un patrón similar, aunque se aprecia que algunas de las islas se encuentran

bastante diferenciadas. Es el caso de La Gomera por un lado, que se sitúa en uno de los extremos de la representación, y de Fuerteventura por el otro.

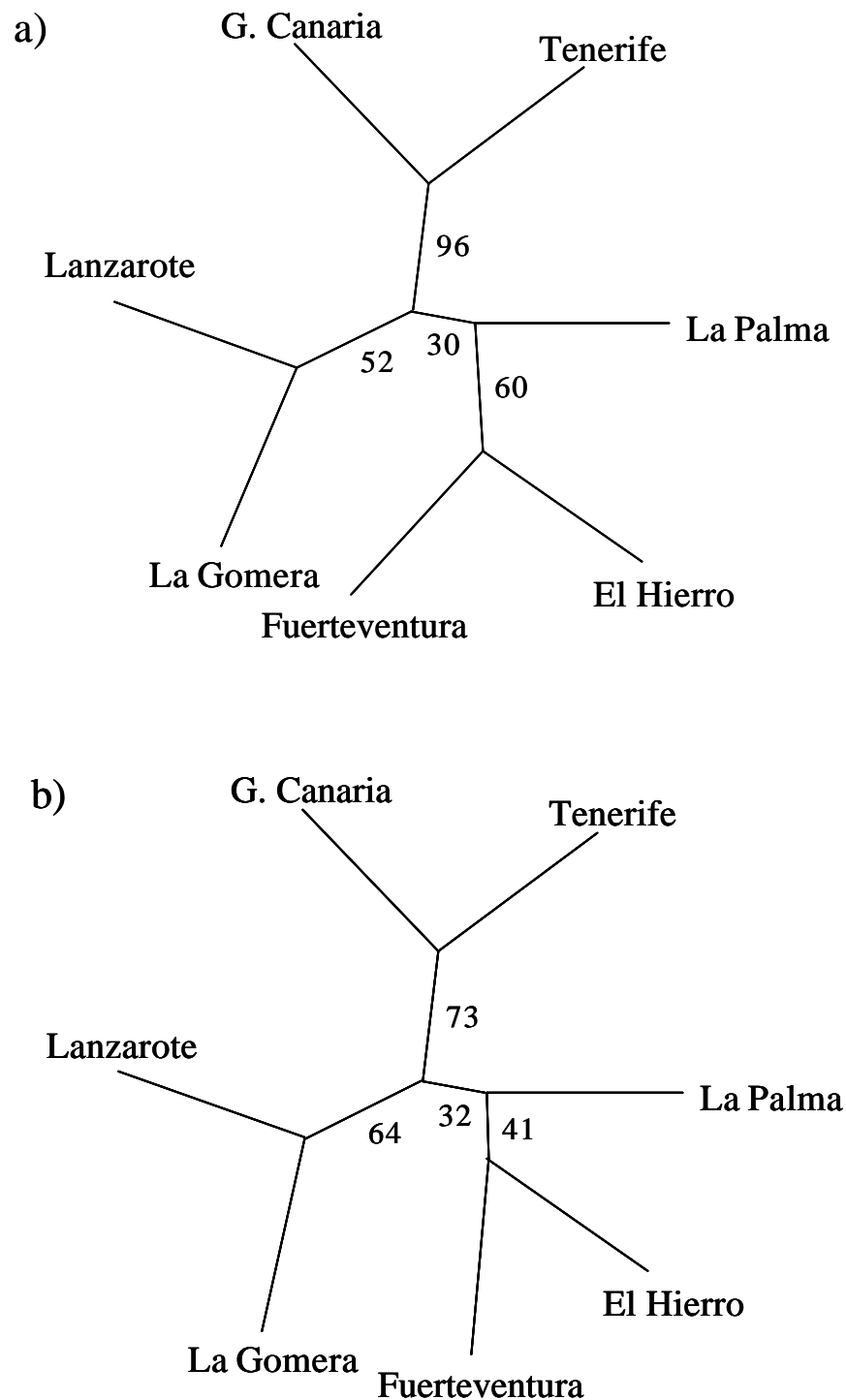


Figura 25. Relaciones entre las poblaciones de las Islas Canarias en base al método de NJ (a) y ML (b). Los valores de las ramas indican el porcentaje de bootstraps a partir de 1000 replicaciones.

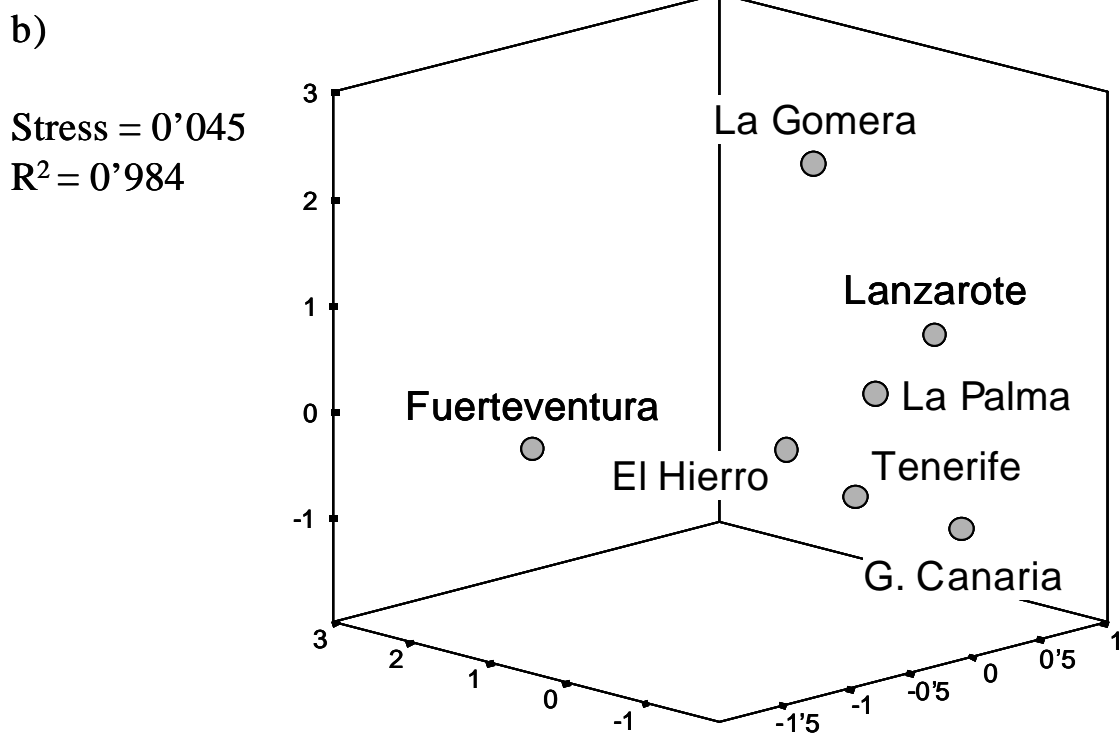
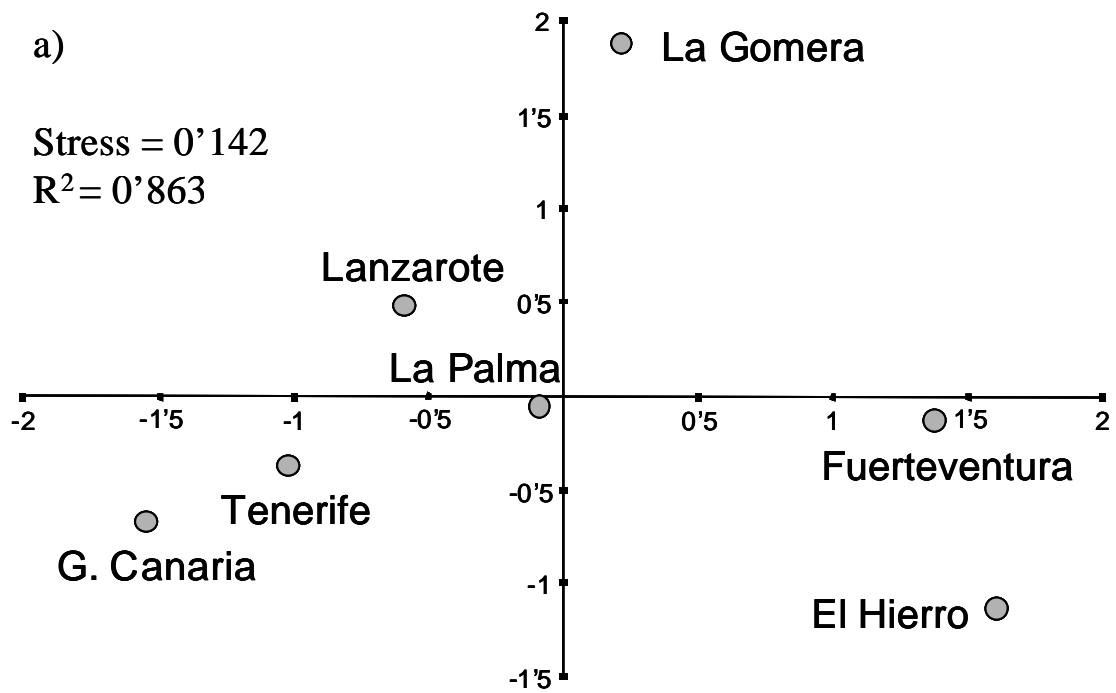


Figura 26. Análisis multidimensional en dos (a) y tres (b) dimensiones de las distancias genéticas entre las poblaciones de las Islas Canarias.

7. RELACIONES GLOBALES DE LAS POBLACIONES CONSIDERADAS

Al comparar el conjunto de poblaciones tratadas en este trabajo se observa que un 17'4% de la variación ($P < 0'00001$) se distribuye entre ellas. Esta variación se encuentra siempre por debajo del valor de diferenciación detectado entre las áreas del continente africano. Ello se debe a que Canarias y las poblaciones del suroeste europeo consideradas son muy homogéneas, aunque significativamente diferentes ($P = 0'00840$). A este respecto, se ha considerado conveniente analizar primero las relaciones de las Islas Canarias con la Península Ibérica y Normandía, sus vecinos genéticos más cercanos, antes de proceder a una comparación global.

7.1. Canarias y Europa: las repercusiones del proceso de conquista y recolonización en el pool génico paterno

De los resultados anteriores se desprende que las islas han recibido su mayor influencia paterna de las poblaciones europeas. Por islas, las únicas que se diferencian significativamente de la Península Ibérica son La Gomera ($P = 0'02230$) y Fuerteventura ($P = 0'02835$). Pero ¿cuál de las regiones europeas analizadas se acerca más a Canarias? Al calcular las distancias genéticas se observa que las diferentes islas no tienen las mismas afinidades con las poblaciones continentales (Tabla 10). El Hierro, La Palma, La Gomera, Tenerife y Lanzarote tienen sus mayores afinidades con Portugal. Gran Canaria se podría incluir en este grupo puesto que, aunque su distancia menor es con Sevilla, la distancia a Portugal no es mucho mayor. El mayor parecido genético de Fuerteventura se detecta con Málaga, precisamente las dos poblaciones no africanas estudiadas con mayor incidencia de haplotipos norteafricanos del grupo III. La población más alejada genéticamente de las islas es la catalana que, exceptuando El Hierro, presenta, de todas las comparaciones posibles, las mayores distancias con todas ellas. De las relaciones gráficas se deduce un patrón bastante similar. Al representar estas distancias en la red NJ (Figura 27), la topología resume perfectamente lo anterior: Fuerteventura aparece muy relacionada con Málaga, El Hierro con Cataluña, y el resto, a excepción de La Gomera, muy

Tabla 10. Distancias genéticas entre las poblaciones insulares y las europeas consideradas.

	El Hierro	La Palma	La Gomera	Tenerife	G. Canaria	Lanzarote	Fuerteventura
Sevilla	0'0300**	0'0125	0'0263*	0'0096	0'0079	0'0104	0'0359**
Huelva	0'0244	0'0161	0'0378	0'0152	0'0163	0'0175	0'0364
Cádiz	0'0255	0'0175	0'0304*	0'0278	0'0370	0'0263*	0'0321*
Córdoba	0'0354	0'0178	0'0288	0'0142	0'0148	0'0112	0'0350
Málaga	0'0259	0'0325	0'0358**	0'0374*	0'0486*	0'0399**	0'0118
Portugal	0'0148	0'0082	0'0177	0'0071	0'0081	0'0068	0'0199
León	0'0285	0'0185	0'0379*	0'0144	0'0121	0'0138	0'0354
Galicia	0'0576	0'0373	0'0651	0'0300	0'0213	0'0405	0'0634
Cantabria	0'0172	0'0112	0'0256**	0'0081	0'0112	0'0131*	0'0152
Valencia	0'0205	0'0100	0'0219*	0'0116	0'0187	0'0154	0'0193
Castilla	0'0578*	0'0420	0'0229	0'0503*	0'0613*	0'0406*	0'0421
Cataluña	0'0514	0'0814	0'0903**	0'0880*	0'0961*	0'0906**	0'0746
P. Vasco	0'0563**	0'0501**	0'0729**	0'0453**	0'0457**	0'0511**	0'0713**
Normandía	0'0807*	0'0667	0'0678*	0'0624	0'0619*	0'0590*	0'0118**

Nota.- Los valores en negrita son las distancias mínimas entre las islas y las poblaciones continentales; Los asteriscos indican probabilidades de diferenciación significativas al nivel $P < 0'05$ (*) y $P < 0'01$ (**).

cercanas a Portugal. En el gráfico ML (Figura 27b) la incorporación de las islas no modifica las relaciones entre las poblaciones peninsulares, conservándose los tres grupos citados con anterioridad. Mientras que El Hierro, La Palma y Fuerteventura quedan incluidas en el grupo del sur peninsular, el resto son acogidas por el grupo del noroeste. Es significativo que ninguna de las islas, excepto El Hierro, se asocia al grupo de poblaciones del noreste en ambas representaciones. Las gráficas multidimensionales (Figura 28) proporcionan relaciones intermedias entre ambas redes. La Gomera y Fuerteventura se acercan al sur de Andalucía y el resto de islas quedan en un conjunto formado por localidades del sur y noroeste peninsular. La gran similitud que existe entre las poblaciones del noroeste y Andalucía hace que sus relaciones con las diferentes islas no resulten claras. Lo que sí se refleja en todos los gráficos es que los grupos europeos más divergentes son los vascos, catalanes y normandos, poblaciones que deben

haber contribuido muy poco a la distribución de los linajes paternos que se observan hoy en Canarias.

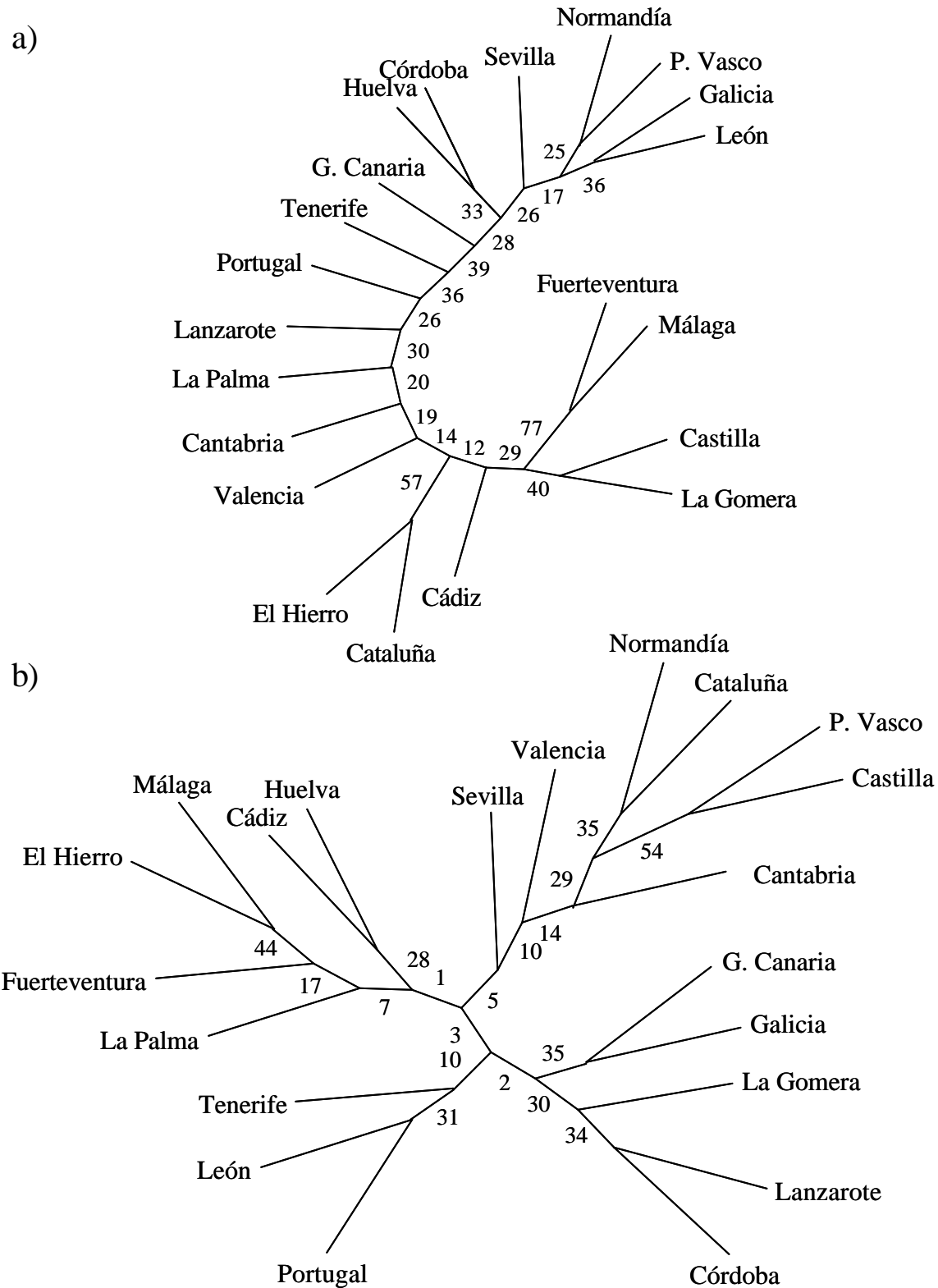


Figura 27. Relaciones genéticas entre las Islas Canarias y las poblaciones Europeas en base al método NJ (a) y ML(b). Los valores de las ramas indican el porcentaje de bootstraps a partir de 1000 replicaciones.

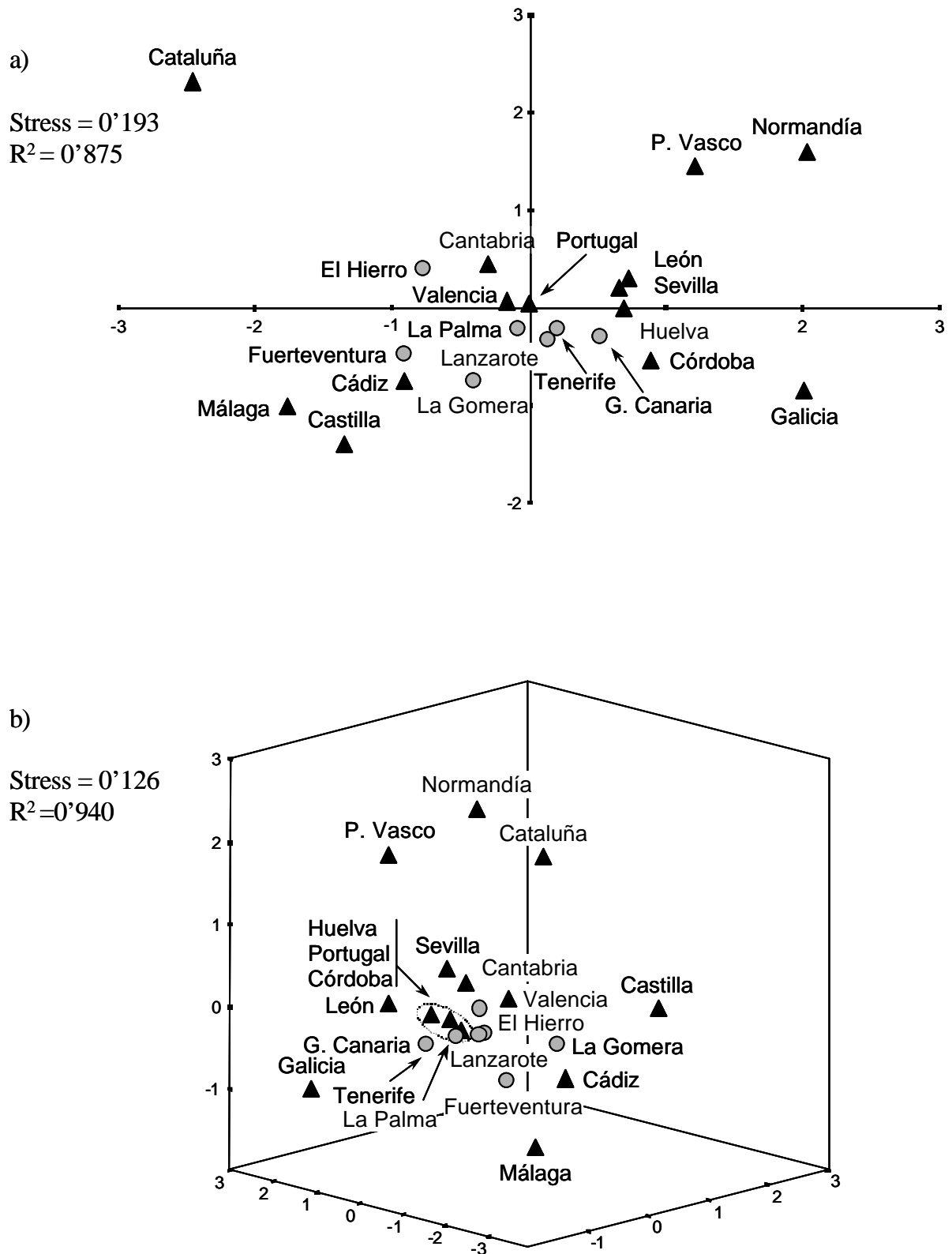


Figura 28. Análisis multidimensional en dos (a) y tres (b) dimensiones relacionando a las poblaciones de las Islas Canarias (círculos) y europeas (triángulos) en base a las distancias genéticas.

7.2. La relación de Canarias y la Península Ibérica con las poblaciones africanas

En estos análisis Canarias ha sido tratada como una población conjunta por su homogeneidad genética. Por las grandes diferencias detectadas en el continente africano es difícil dar una estima global de divergencia, pero en líneas generales los análisis de varianza oscilaron entre un 19'1 y un 36'0%. Tal y como se observa en la Tabla 11, las comparaciones de las áreas africanas con Canarias originan valores de diferenciación menores que los obtenidos al compararlas con la Península Ibérica. Nótese que, los menores valores de diferenciación en ambos casos se detectan al compararlas con el noreste de África y no con el noroeste, más relacionado histórica y geográficamente con Canarias y el sur de Europa. Este resultado se debe a que tanto Europa y, por tanto, indirectamente también Canarias, como el noreste africano han sufrido fuertes influencias desde el Oriente Próximo, haciendo que sus relaciones genéticas aumenten.

Tabla 11. Análisis de varianza de las poblaciones europeas y canarias con respecto a las zonas africanas.

Área africana	Canarias		Península Ibérica	
	Variación (%)	P ^a	Variación (%)	P ^a
Noreste africano ^b	19'1	0'03226	22'1	0'00684
Noroeste africano ^c	22'5	0'00098	26'8	0'00000 ^e
Área subsahariana ^d	29'3	0'00000 ^e	32'2	0'00000 ^e
Área subsahariana (sin khoisan)	33'1	0'00293	36'0	0'00000 ^e

^aProbabilidad; ^bIncluye Etiopía y Sudán; ^cIncluye poblaciones desde Marruecos hasta Mauritania y Malí;

^dIncluye Senegal, Costa de Marfil, Guinea Ecuatorial, África Central, Sudáfrica y khoisan; ^eP < 0'000005.

Las distancias genéticas (Tabla 12) y las representaciones gráficas NJ y ML (Figura 29) ofrecen el mismo panorama, dejando a Canarias perfectamente incluida entre las poblaciones europeas y, aunque sean de las más cercanas al conjunto africano, están muy diferenciadas de las

tres áreas africanas claramente delimitadas. En el caso concreto de la red ML, aunque su topología coincide perfectamente con la distribución geográfica de las poblaciones representadas, la unión de las europeas a África es mediante las poblaciones del noroeste africano, contrariamente a los resultados del análisis de varianza y de las distancias genéticas.

Los análisis multidimensionales (Figura 30) apoyan nuevamente la mayor cercanía de Canarias y la Península Ibérica a las poblaciones del este africano, aunque en este caso se pone en evidencia el problema de las mezclas étnicas en este tipo de comparaciones. Nótese que en ambos casos las poblaciones africanas menos fijadas para un haplotipo concreto (las menos aisladas genéticamente) son las más cercanas al grupo Canarias-Península Ibérica. Sin embargo, las poblaciones del África Negra y el noroeste africano ocupan situaciones casi extremas en ambas representaciones.

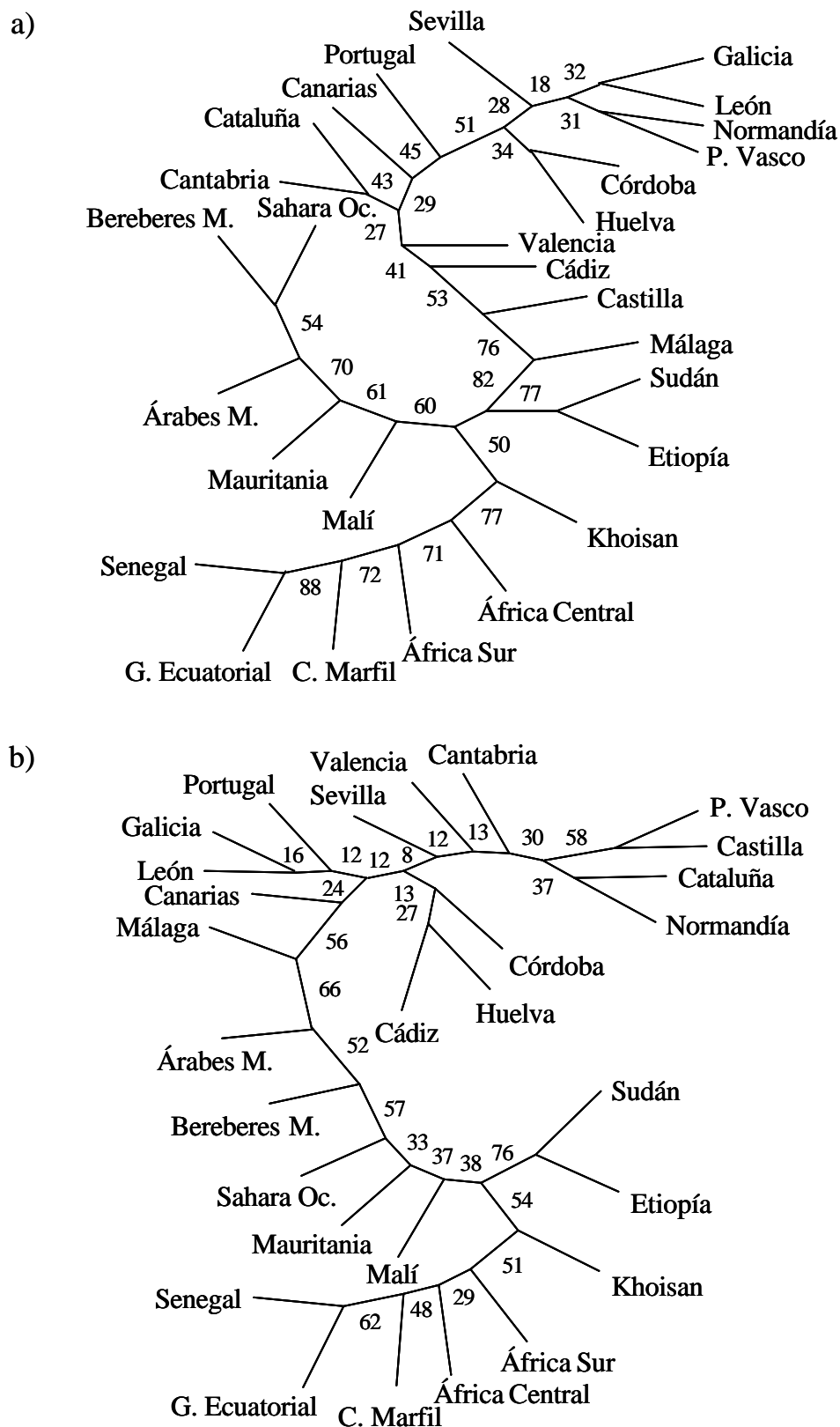


Figura 29. Relaciones genéticas globales considerando a Canarias en conjunto en base al método de NJ (a) y ML (b). Los valores de las ramas indican el porcentaje de bootstraps a partir de 1000 replicaciones.

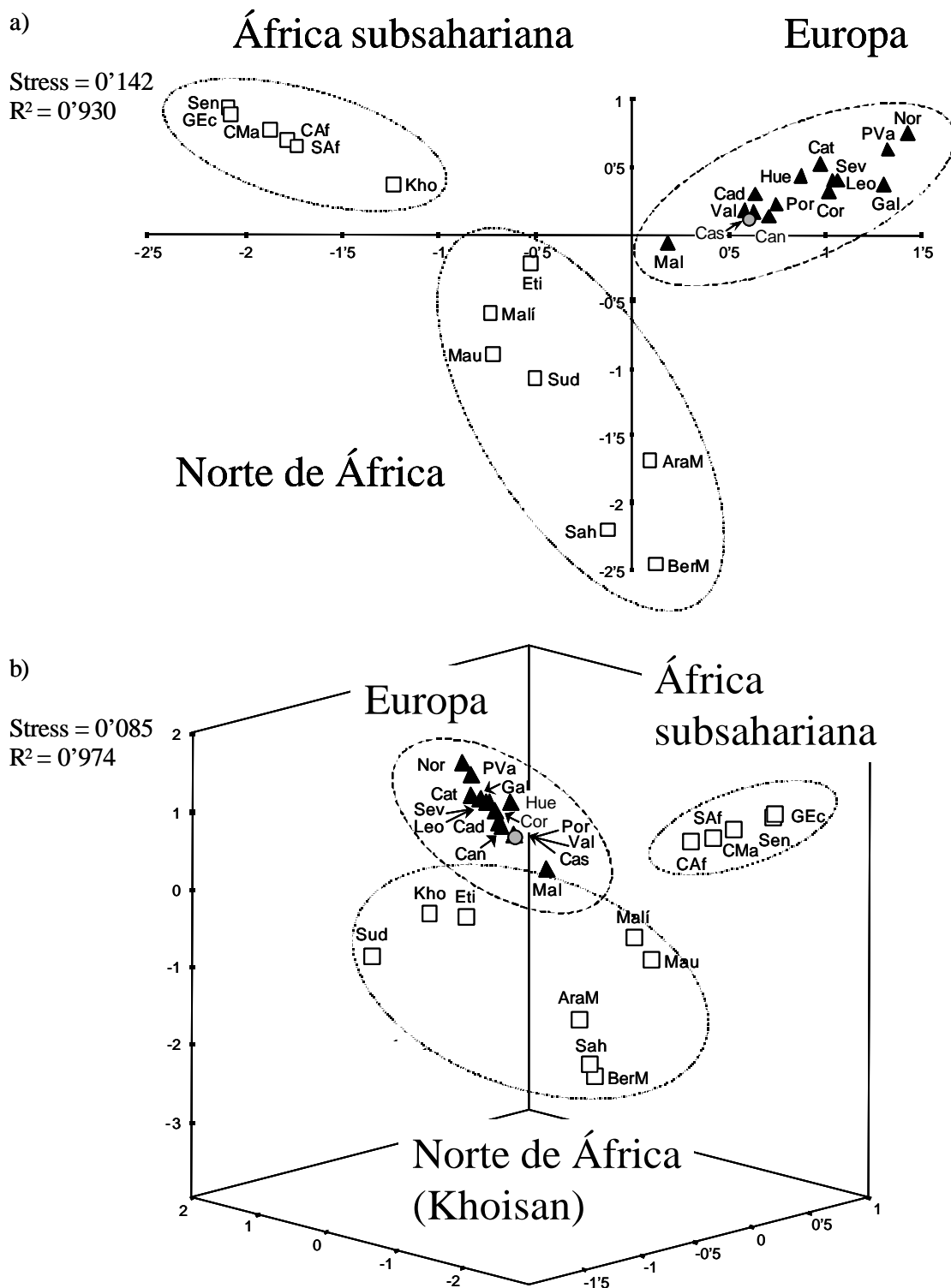


Figura 30. Análisis multidimensional global en dos (a) y tres (b) dimensiones de las poblaciones africanas (cuadrados), europeas (triángulos) y Canarias (círculo), en base a las distancias genéticas calculadas. Las claves usadas son: Sen, Senegal; GEc, Guinea Ecuatorial; CMa, Costa de Marfil; CAf, África Central; SAF, África Sur; Kho, khoisan; Eti, Etiopía; Sud, Sudán; Mau, Mauritania; Sah, Sahara Occidental; BerM, bereberes de Marruecos; AraM, árabes de Marruecos; Nor, Normandía; PVa, País Vasco; Cat, Cataluña; Can, Cantabria; Leo, León; Gal, Galicia; Por, Portugal; Cas, Castilla; Val, Valencia; Sev, Sevilla; Hue, Huelva; Cad, Cádiz; Cor, Córdoba; Mal, Málaga.

8. CONTRIBUCIONES PARENTALES A LA POBLACIÓN ACTUAL DE CANARIAS

La gran mayoría de los haplotipos detectados en el trabajo muestran una elevada especificidad geográfica. Sin embargo, la población actual de Canarias es producto de una mezcla de poblaciones que, a su vez, también se han mezclado entre sí, con lo que la asignación de haplotipos presentes en Canarias a zonas geográficas determinadas puede ser sólo aproximado. Existen varias formas de calcular estimas de influencia, pero la más sencilla consiste en contabilizar los linajes procedentes de una u otra población parental. Dado que las principales poblaciones parentales de Canarias, Península Ibérica y oeste africano, son altamente divergentes en la distribución de los linajes paternos (Tabla 11), si se asume que los primeros colonizadores de Canarias tenían el mismo pool génico que los que hoy habitan el noroeste africano, se puede reconocer un porcentaje elevado de haplotipos que pueden haber llegado a la población canaria actual desde ambas poblaciones. En una cuantificación rudimentaria se puede separar como una contribución africana a los haplotipos del grupo III preferentemente detectados en África (excepto III₅ y III₆, que presentan mayores frecuencias en la Península Ibérica que en África), y como una contribución europea al resto de haplotipos. Entre las europeas se incluyen haplotipos de diverso origen, incluso algunos que también se encuentran en el norte de África.

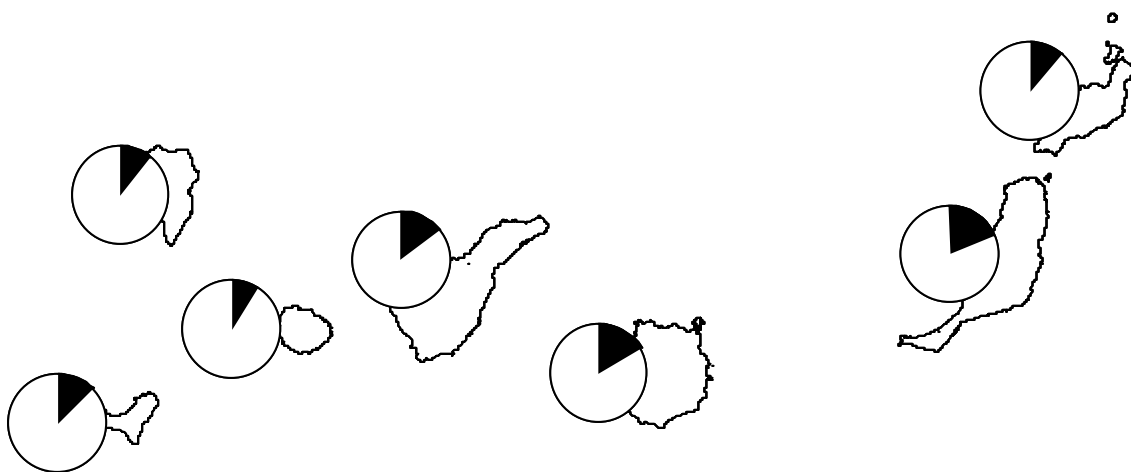


Figura 31. Proporción y distribución de haplotipos del cromosoma Y en Canarias con asignación europea (blanco) y africana (negro).

Del mismo modo entre las africanas existen algunos haplotipos que podrían proceder de Europa. El problema de estas asignaciones es que siempre conllevan un grado de incertidumbre, porque no se puede conocer exactamente cuánto procede de un lugar y cuánto del otro. En cualquier caso, es una primera aproximación para determinar las contribuciones africanas y europeas en el Archipiélago. Aceptando los criterios mencionados se observa que la gran mayoría de los haplotipos en las poblaciones canarias podrían tener su origen en Europa (86'7%), y tan sólo un 13'3% en África. En este punto parece apropiado incidir en que en todas las islas, excepto en La Gomera, se detectan haplotipos de asignación africana por encima de la media detectada en la Península Ibérica. Se observa, además, que la contribución africana es menor en las islas occidentales, con el mínimo (8'7%) en La Gomera, que en las orientales, con el máximo (18'7%) en Fuerteventura (Figura 31).

Otra estima sencilla de influencia consiste en detectar una variante característica de una de las poblaciones parentales y que las diferencias en frecuencia de esta variante entre las poblaciones parentales sea alta. En este trabajo el haplotipo III₃ (definido por M81) se encuentra mayoritariamente confinado al noroeste de África y su diferencia en frecuencia con respecto a la región europea es de 49-64'5%, según se tome todo el noroeste africano o sólo el Magreb respectivamente (árabes y bereberes de Marruecos junto a saharauis). Así, las contribuciones norteafricanas se obtienen mediante la ecuación de Bernstein (1931), según la cual:

$$m_B = (P_C - P_E) / (P_A - P_E)$$

Donde P_C, P_E y P_A son las frecuencias observadas para el haplotipo III₃ en el total de Canarias o por Isla, en el total de las poblaciones europeas consideradas (0'052) y en el total de las poblaciones norteafricanas consideradas (0'542 o 0'697, según se considere todo el noroeste africano o sólo la región del Magreb, respectivamente). Los resultados son muy similares,

independientemente de si se toma una u otra región africana como parental (Tabla 13), detectando una contribución africana en Canarias entre el 4'8 y el 6'3%. Por islas, la contribución africana estimada decrece de este a oeste siendo totalmente nula en El Hierro y La Gomera. Estas influencias no deben considerarse tampoco valores definitivos, ya que se han obtenido a partir de una sola variante y, por tanto, están sujetas a grandes errores.

Tabla 13. Contribuciones africanas basadas en la frecuencia del ht III₃.

Isla	% ht III ₃	Parental africana	
		Noroeste ^a	Magreb ^b
El Hierro	2'1	-6'3	-4'8
La Palma	5'9	1'4	1'1
La Gomera	4'3	-1'8	-1'4
Tenerife	10'7	11'2	8'5
G. Canaria	11'5	12'9	9'8
Lanzarote	6'2	2'0	1'6
Fuerteventura	13'3	16'5	12'6
Canarias Total	8'3	6'3	4'8

^a Incluye poblaciones desde Marruecos hasta Mauritania y Malí.

^b Incluye poblaciones de Marruecos y el Sahara Occidental.

Un estadístico de mezcla más elaborado es la estima de Long (1991). Este estimador permite detectar la fiabilidad de la estima puesto que calcula un error, pero, además, asigna parte de ese error a procesos estocásticos no contemplados en el evento de mezcla y, por tanto, permite determinar el efecto de la deriva en la diferenciación de la población híbrida. Se ha considerado un modelo (Modelo 1) en el que las poblaciones insulares se habrían generado por mezcla de tres poblaciones parentales bien diferenciadas:

1. Total de las poblaciones del suroeste de Europa analizadas.

2. Poblaciones del Magreb, incluyendo sólo a bereberes y árabes de Marruecos y a saharauis. Estas tres poblaciones, totalmente homogéneas (2'4% de variación, $P = 0'08016$), constituirán el pool del noroeste africano libre, casi totalmente, de influencias negroides.
3. Poblaciones subsaharianas, excluyendo los khoisan, puesto que sin ellos forman otro conjunto homogéneo (1'6% de variación, $P = 0'09775$).

Los resultados obtenidos (Tabla 14) indican, dada la existencia de contribuciones negativas, que la mayoría de las poblaciones insulares no se ajustan a este modelo de triple procedencia. Este es el caso de las poblaciones de El Hierro, La Gomera, Tenerife y Fuerteventura. En La Gomera, Tenerife y Fuerteventura la contribución africana procede sólo del Magreb, mientras para El Hierro queda perfectamente patente que las contribuciones africanas proceden sólo del África Negra. Las contribuciones negativas, representadas en la Tabla 14 en negrita, indican que dicha población parental no participa en la población híbrida. Para estos casos el modelo se redujo a dos poblaciones parentales (Modelo 2): para El Hierro considerando Europa y el área subsahariana, y para las otras tres a Europa y el Magreb. Nuevamente, los valores de mezcla indican que la influencia del noroeste africano es mayor en las islas orientales, alcanzando el máximo (17%) en Fuerteventura, que en las occidentales, reduciéndose al mínimo en El Hierro. Casi al contrario sucede con la influencia negroide, que alcanza su máximo (4%) en El Hierro.

Dado que también se detectan influencias de origen negroide en las poblaciones del noroeste de África, no es posible discernir cuáles de los linajes negroides llegaron a Canarias con los aborígenes y cuáles llegaron posteriormente con los esclavos subsaharianos. Así pues, la división que se ha hecho, en influencias del noroeste africano o Magreb y subsaharianas, no implica que estas contribuciones ocurrieran en diferente momento o por diferentes poblaciones

en la historia de Canarias, y deben tomarse sólo como un indicativo de que es posible detectarlas gracias precisamente a los haplotipos distintivos de una y otra comunidad.

A partir de los valores de F_{ST} de la Tabla 14 también se deduce que no se puede explicar toda la variación que se encuentra en las islas por el simple evento de mezcla. Para justificar estas desviaciones no explicadas por el modelo hay que recurrir a procesos como la deriva génica, probablemente con los efectos más fuertes en islas pequeñas como La Gomera, donde se detecta el valor más alto.

Tabla 14. Porcentajes de mezcla (m_L) en Canarias.

Isla	M ^a	Europeo	Magreb	Subsahariano	F_{ST} ^b
El Hierro	(1)	99'57 ± 4'97	-3'65 ± 3'71	4'06 ± 3'35	0'00217
	(2)	95'93 ± 3'40	-	4'07 ± 3'40	
La Palma	(1)	96'14 ± 5'18	1'86 ± 4'80	2'01 ± 2'25	0'00601
La Gomera	(1)	100'18 ± 5'50	0'28 ± 5'48	-0'46 ± 0'10	0'01843
	(2)	99'66 ± 5'70	0'34 ± 5'70	-	
Tenerife	(1)	93'09 ± 3'93	7'33 ± 3'91	-0'42 ± 0'07	0'00320
	(2)	92'62 ± 4'08	7'38 ± 4'08	-	
G. Canaria	(1)	89'76 ± 4'13	9'31 ± 4'04	0'93 ± 1'15	-0'00509
Lanzarote	(1)	95'67 ± 4'44	3'54 ± 4'30	0'79 ± 1'37	0'00211
Fuerteventura	(1)	83'24 ± 10'53	17'13 ± 10'48	-0'38 ± 0'04	0'00584
	(2)	82'94 ± 7'22	17'06 ± 7'22	-	
Canarias Total	(1)	93'48 ± 2'88	5'93 ± 2'82	0'58 ± 0'77	0'00297

^a Modelo. ^b Variación no explicada por el modelo.

IV. DISCUSIÓN

Los haplotipos que se han detectado en este trabajo se generaron en un momento de la historia evolutiva del hombre en un lugar concreto. Para determinar su posible procedencia es necesario el estudio de sus distribuciones en las poblaciones del resto del globo. A muchos de ellos ya se les ha asignado el lugar más probable de origen e incluso cuándo se originaron. Usando esta información se puede rastrear un haplotipo o conjunto de haplotipos en la geografía (filogeografía) y de esta manera obtener relaciones entre las poblaciones estudiadas, clasificando el acervo génico por población y datando las influencias externas. Este análisis permite además detectar procesos demográficos impredecibles a partir sólo de las distancias genéticas.

1. ÁFRICA

1.1. Los haplotipos del África subsahariana

Los grupos I y II se consideran los grupos de linajes paternos más antiguos del hombre moderno (Underhill y col. 2000), que se habrían originado en el África subsahariana hace unos 40.000-140.000 años (Thomson y col. 2000). El grupo I, el más ancestral de todos, no tiene una frecuencia uniforme en África sino que se encuentra casi exclusivamente asociado a los khoisan, y, en menor medida, a las poblaciones del noreste africano (Scozzari y col. 1999; Underhill y col. 2000; Semino y col. 2002; este trabajo). Precisamente algunos de los linajes de los khoisan aparecen cerca de la raíz propuesta para la filogenia del ADN mitocondrial (Vigilant y col. 1991; Penny y col. 1995). Nuevamente, los datos moleculares son una evidencia más de que el hombre anatómicamente moderno apareció en África. Un trabajo reciente con el cromosoma Y en un gran número de muestras del este de Asia, continente donde la hipótesis multiregional tiene mayor apoyo por la aparente continuidad en los fósiles humanos hallados, demostró que todos los cromosomas analizados tenían la mutación M168, que coalesce en África y, por tanto, la ausencia de linajes ancestrales en el pool actual de Asia (Ke y col. 2001). El grupo II alcanza frecuencias más homogéneas entre las poblaciones que lo comparten, que suelen ser de origen subsahariano. Fuera de África ambos grupos se encuentran escasamente representados. Se han

detectado sólo en Portugal (0'9% del grupo II, este trabajo), Cerdeña (1'3% del grupo I, Semino y col. 2000; 0'9%, Passarino y col. 2001), Pakistán (2'3% del grupo II, Underhill y col. 2000) y Palestina (1'4% del grupo I, Nebel y col. 2001), todas ellas poblaciones geográficamente cercanas o influenciadas por el continente africano. Debido a su distribución ambos grupos se consideran subsaharianos. También los haplotipos definidos por la mutación *DYS287* y los que portan la mutación *DYS271*, haplotipos III₁ y III₇ respectivamente, se encuentran casi exclusivamente asociados a poblaciones negroides, pudiendo atribuirse las excepciones a contactos recientes con ellas. Aunque el haplotipo III₁ está definido en este trabajo sólo por la mutación *DYS287*, probablemente debe estar definido además por M33 o M75, mutaciones que sólo se encuentran en África, puesto que no se ha descrito ningún otro haplotipo con la mutación *DYS287* ni en esta región ni en Europa (Underhill et al. 2001b). También éstos se consideran marcadores de origen subsahariano, constituyendo la mayoría de su pool génico (Seielstad y col. 1994; Hammer y col. 1997, 1998; Scozzari y col. 1999; Thomas y col. 2000; Bosch y col. 2001). Así pues, la gran mayoría del acervo génico paterno subsahariano está caracterizado por un conjunto de haplotipos ancestrales (grupos I y II), no compartidos globalmente, en combinación con haplotipos más derivados (haplotipos III₁ y III₇), muy compartidos en África pero no globalmente. Este mismo resultado también se ha obtenido para el ADN mitocondrial (Cann y col. 1987; Ingman y col. 2000; Maca-Meyer y col. 2001) y más recientemente para la región *Dys44* autosómica (Labuda y col. 2000). El conjunto de estos linajes (grupos I y II, y haplotipos III₁ y III₇) representa un 43'6% de las muestras de origen africano analizadas para el presente trabajo (60'3% del total de las africanas) alcanzando sus más altas frecuencias en las regiones subsaharianas del Golfo de Guinea, desde Senegal hasta Sudáfrica (Figura 32). A partir de la franja del Sahel, a medida que se asciende en latitud la abundancia de estos linajes disminuye, encontrándose en muy baja frecuencia en las poblaciones de Marruecos. Esta distribución concuerda perfectamente con la de los linajes L de ADN mitocondrial, en la que las poblaciones

subsaharianas estarían caracterizadas por un 60-100% de estos linajes, disminuyendo paulatinamente hacia el norte hasta un 24'7% en Etiopía (Passarino y col. 1998) y un 12% en bereberes de Marruecos (Rando y col. 1998).

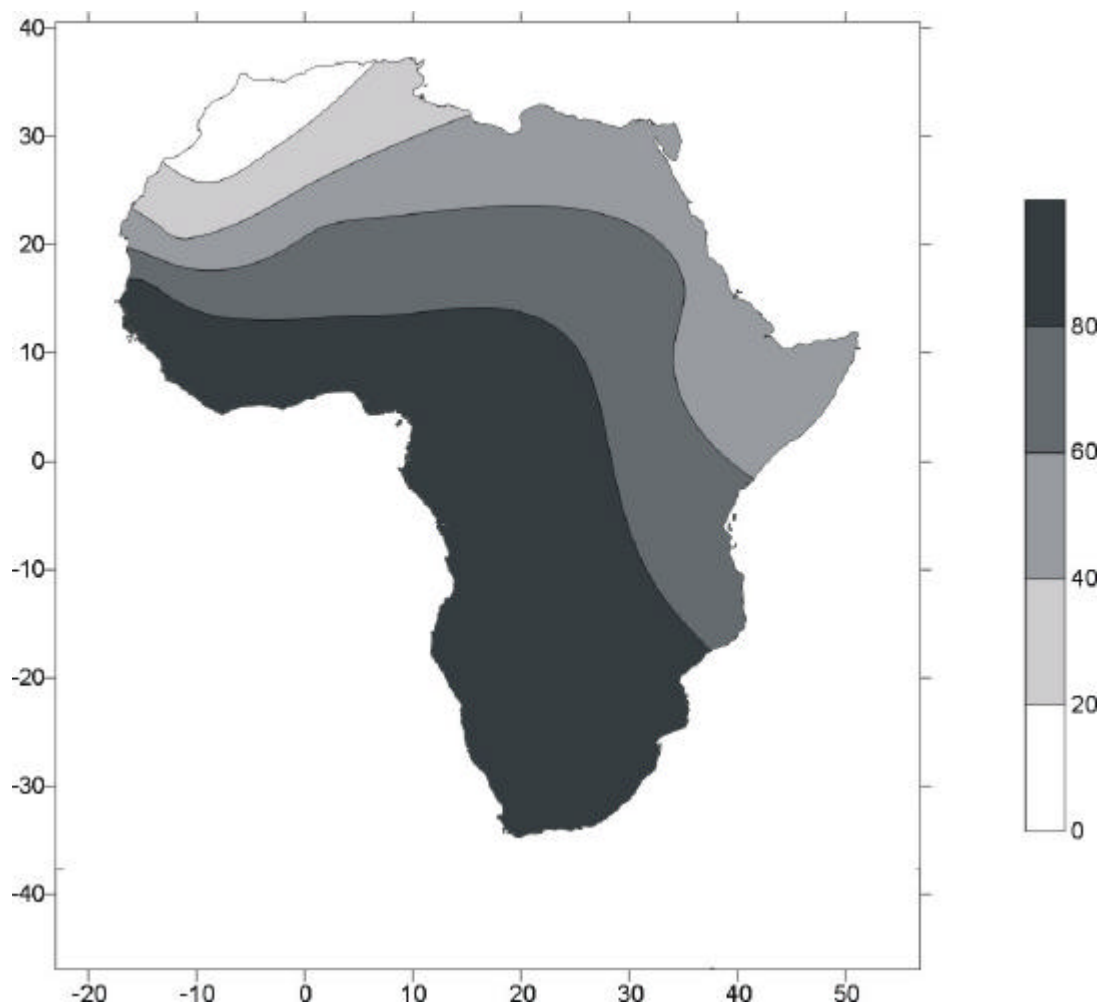


Figura 32. Mapa de África mostrando la distribución de frecuencias de los linajes con asignación subsahariana (grupos I y II, y haplotipos III₁ y III₇). La escala de la derecha relaciona un rango de frecuencias para este conjunto de linajes con un color.

Es probable que estas distribuciones se deban a eventos demográficos mezclados, con expansiones muy antiguas, que dispersarían los grupos I y II, y quizás también el III₁, por África (Underhill y col. 2001b), y otras más recientes, donde los linajes con la mutación *DYS271* reemplazarían casi totalmente la variación anterior (Hammer y col. 1997, 1998). El ADN mitocondrial y los marcadores autosómicos están de acuerdo en asignar la gran mayoría de esos

movimientos en África a una fecha no anterior a 15.000-20.000 años (Rando y col. 1998; Sánchez-Mazas, 2001). Sin embargo, para la expansión de los grupos I y II del cromosoma Y en África se han sugerido fechas mucho más tempranas (Underhill y col. 2001b). Para el haplotipo III₇, aunque Scozzari y col. (1999) concluyen que se originó hace más de 20.000 años, varias estimas apuntan a que su expansión por África fue mucho más reciente. La edad de coalescencia obtenida en el presente trabajo la sitúa en 4.622 años (1.981-16.176), estima similar a la obtenida por otros autores (Bosch y col. 1999; Thomas y col. 2000). Por su distribución y tiempo de coalescencia este haplotipo pudo haberse extendido por África con la expansión Bantú hace unos 3.000-5.000 años (Cavalli-Sforza y col. 1994; Thomas y col. 2000; Underhill y col. 2001b). La edad que se obtiene para el haplotipo III₁, de unos 5.252 años (2.251-18.382), coincide también con la expansión Bantú, aunque Underhill y col. (2001b) lo sitúa en estas poblaciones con anterioridad.

1.2. Los haplotipos de asignación norteafricana

Desde el comienzo del estudio de la variabilidad del cromosoma Y se detectaron linajes que alcanzaban la máxima frecuencia del globo en el norte de África, estando presentes en poblaciones no africanas (Hammer y col. 1997; Bosch y col. 1999; Flores y col. 2000a, 2001c; Lucotte y col. 2000). Más recientemente (Underhill y col. 2000; Bosch y col. 2001) se ha determinado que existe un conjunto de linajes *DYS287/PN2T* del grupo III que es mayoritario en la región norteafricana (Hammer y col. 1997, 1998; Bosch y col. 2001). Estos linajes están definidos por la mutación M35 (Underhill y col. 2000), de la que derivan M78, M81, M123 y M281. Ésta última mutación ha sido descrita recientemente y sólo detectada en dos individuos de Etiopía (Semino y col. 2002). Los linajes definidos por la mutación M123 y sus derivados han sido poco estudiados aún, pero los resultados de este trabajo y los hasta ahora publicados (Underhill y col. 2000; Bosch y col. 2001; Semino y col. 2002) sugieren que son más abundantes fuera que dentro de África. Para ser conservadores han sido tratados como haplotipos de origen

no africano (ver pag. 133). Dado que se han determinado los tres sublinajes principales de M35 (Underhill y col. 2000) pero no un haplotipo definido sólo por esta mutación, éste puede considerarse equiparable al haplotipo III₂ en el presente análisis. El conjunto de estos tres haplotipos (III₂, III₃ y III₄, sin contar a III₅ y III₆ que están caracterizados por la mutación M123) constituye un 43'6% de las muestras analizadas para la región africana (31'9% incluyendo las africanas de la bibliografía). Tal y como se observa en la Figura 33, su distribución es eminentemente norteafricana, con un foco de dispersión al noroeste y otro en el este. Ambos focos se deben a una distribución diferencial de III₃, con altísimas frecuencias en Marruecos, y III₂ junto con III₄, con sus mayores frecuencias en el otro extremo norte. Esta diferenciación

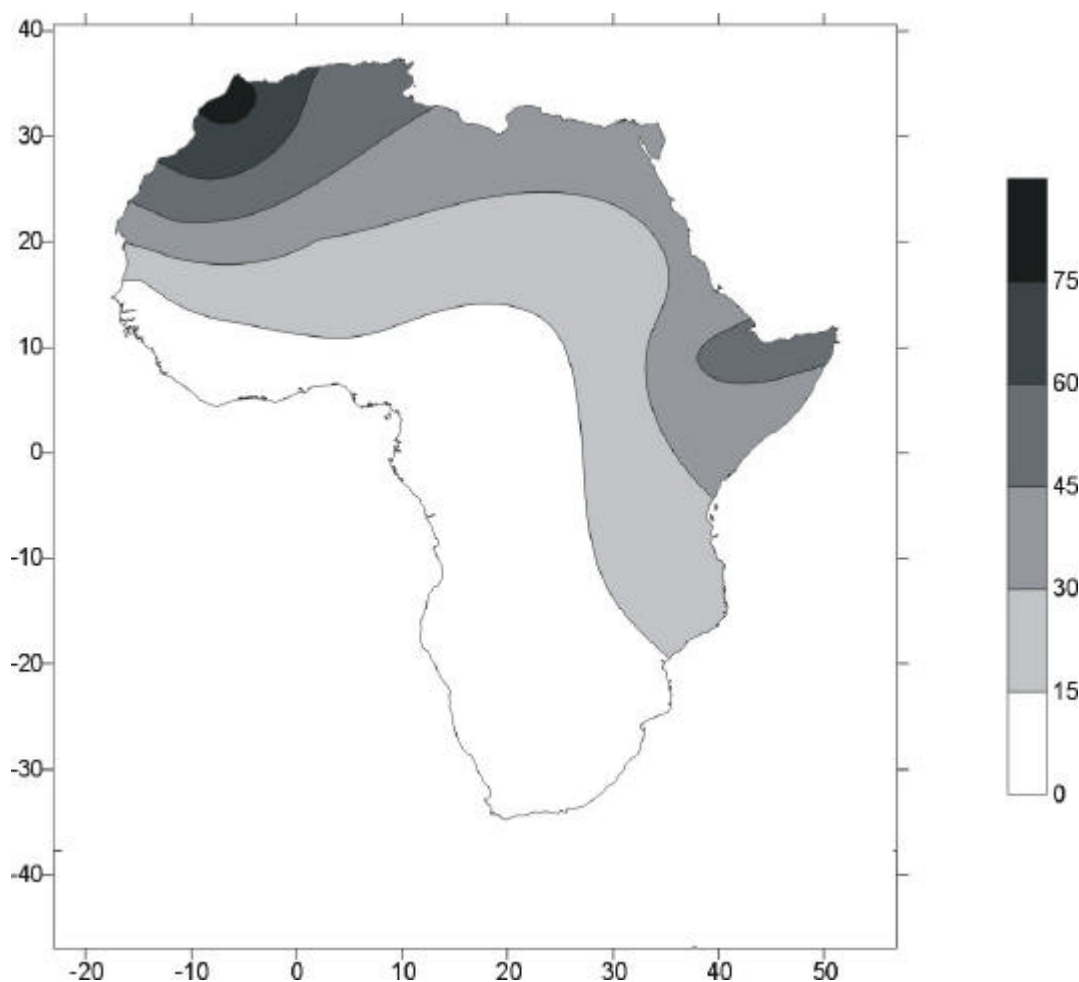


Figura 33. Mapa de África mostrando la distribución de frecuencias de los linajes con asignación norteafricana (haplotipos III₂, III₃ y III₄). La escala de la derecha relaciona un rango de frecuencias para este conjunto de linajes con un color.

este-oeste se observa incluso para marcadores clásicos (Bosch y col. 1997). A pesar de que el haplotipo 5 del p49a,f se encuentra ligado a estos linajes (YAP+/DYS271A en Pasarino y col. 1998), su distribución en el norte de África no consigue distinguir este patrón, aunque sí un decrecimiento gradual en frecuencia desde el oeste (Lucotte y col. 2000). En el caso concreto del linaje definido con la mutación M81 (haplotipo III₃), en la Figura 33 también se observa una gradación norte-sur en su frecuencia en el oeste, disminuyendo a medida que se desciende en latitud, a la inversa de lo que ocurre con los linajes negroides. A pesar del decrecimiento, este linaje sigue siendo el haplotipo mayoritario hasta el Sahara Occidental y Mauritania. Contrariamente algunos estudios con marcadores *Alu* autosómicos (Comas y col. 2000) y con ADN mitocondrial (Rando y col. 1998) concluyen que los saharauis son ya más parecidos a los pueblos subsaharianos que a los norteafricanos. Sin embargo, usando STR autosómicos Bosch y col. (2000) llegan a la misma conclusión que con el cromosoma Y, de forma que los saharauis se agrupan con el resto de norteafricanos.

El origen del conjunto de linajes *DYS287/PN2T* se ha estimado en unos 20.000 años en el norte de África, desde donde más tarde (\approx 10.000 años) se expandiría hacia Europa y Oriente Próximo (Hammer y col. 1998). Las edades estimadas por Bosch y col. (2001) para estos linajes también situaría su origen en el Paleolítico norteafricano ($>$ 10.000 años). Ambas estimas fueron calculadas en base a la posición de los haplotipos en la filogenia y su distribución en las poblaciones estudiadas, pero contrastan con las obtenidas a partir de la coalescencia de la variabilidad asociada a los STR. De este modo, Scozzari y col. (2001) estiman en unos 1.400 años la edad de expansión de un grupo de cromosomas íntimamente relacionado con el haplotipo III₃, el mayoritario en el noroeste africano. Con los datos del presente trabajo la edad de coalescencia resultante para este haplotipo en la región es de 2.041 años (875-7.143), muy similar al valor anterior. Para el haplotipo III₄ la coalescencia es ligeramente superior, 3.571 años (1.531-12.500), pero nunca en el rango de edades paleolíticas. Si bien existen enormes

discrepancias entre ambos resultados, estos valores deben tomarse con precaución, pues en ambos casos se han estimado a partir de premisas que pueden no ser correctas. A parte de los supuestos, hay que tener en cuenta que el área de estudio de la variabilidad está restringida a una región geográfica. Quizás las estimas cambien ampliando el análisis a más poblaciones.

Para otros marcadores independientes también se han detectado linajes que se pudieron originar en el noroeste de África, o que al menos se expandieron desde allí. Es el caso de ciertos haplotipos autosómicos (Flores y col. 2000b) o del linaje U6 para el ADN mitocondrial (Rando y col. 1998), aunque trabajos recientes sitúan sus raíces fuera de África (Maca-Meyer y col. 2001). En cualquier caso, tanto los trabajos con el cromosoma Y (Bosch y col. 2001; Flores y col. 2001c) como con marcadores nucleares (Bosch y col. 1997; Flores y col. 2000b) o con ADN mitocondrial (Rando y col. 1998) están de acuerdo en afirmar que, además de las posibles influencias recientes, debe tenerse en cuenta una diferenciación mesolítica, o incluso más antigua, para explicar el grado de divergencia del conjunto de poblaciones del noroeste de África. De hecho, se sabe que grupos relacionados con las poblaciones caucasoides han habitado en el Magreb con continuidad desde hace al menos unos 20.000 años, cuando floreció la cultura Iberomauritana en la región. Parece bastante aceptado que, incluso después de la incorporación del Neolítico en la región, se retuvieron en las poblaciones caracteres paleolíticos (Cavalli-Sforza y col. 1994; Newman, 1995).

1.3. Haplotipos de origen no africano

Puesto que todos los haplotipos con M89T se distribuyen principalmente fuera de África y que la mutación M168, filogenéticamente anterior a ésta, marca la salida de África del hombre anatómicamente moderno (Underhill y col. 2000, 2001b), el conjunto de grupos VI, VIII y IX pueden atribuirse a influencias externas en África. Como parte de este conjunto también se han incluido los linajes del grupo III con la mutación M123, en África sólo constituido por el haplotipo III₆, dada su ausencia casi total en el continente (excepto un 2'3% de Etiopía) mientras

que se encuentra muy distribuido por Europa y Oriente Próximo (Underhill y col. 2000; resultados de este trabajo). El conjunto de estos linajes (grupos VI, VIII y IX, y el haplotipo III₆) no supone más que un 12'9% de las muestras africanas analizadas en este trabajo (tan sólo un 7'8% si se incluyen las analizadas por otros autores). De su distribución se puede apreciar que las influencias externas han involucrado principalmente a las poblaciones del norte de África y, en mayor medida, al noroeste (Figura 34). A pesar de que Scozzari y col. (1999) encuentran haplotipos con la mutación 92R7 en África central, y por tanto asignables a los grupos IX o X,

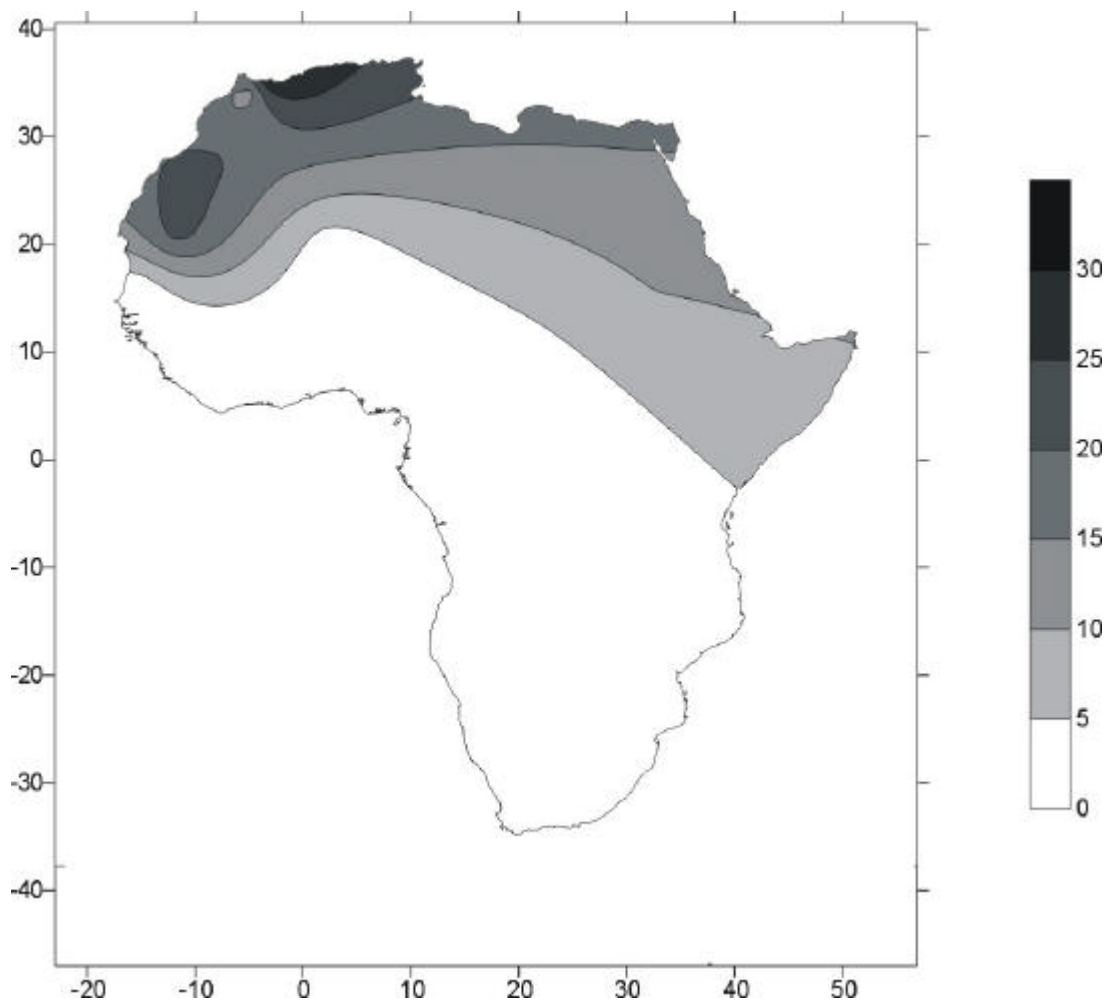


Figura 34. Mapa de África mostrando la distribución de frecuencias de los linajes de asignación no africana (grupos VI, VIII y IX, y haplotipo III₆). La escala de la derecha relaciona un rango de frecuencias para este conjunto de linajes con un color.

en el resto del área subsahariana no se han detectado influencias caucásicas (Casanova y col. 1985; Santachiara-Benerecetti y Semino, 1996; Spurdle y Jenkins, 1996; Passarino y col. 1998; Scozzari y col. 1999). Probablemente, también, en esta ocasión se estén sumando varios eventos demográficos.

Si se tiene en cuenta su origen geográfico, los linajes se pueden separar en influencias de Oriente Próximo, relacionadas con los haplotipos VI₅-VI₇, e influencias procedentes de Europa, en base a los haplotipos VI₂, VIII₃ y IX₁. Las influencias de Oriente Próximo pueden deberse a eventos recientes, como la ocupación islámica (comienzo en el año 643 d.C.), o antiguos, como la difusión de poblaciones neolíticas desde el este. Se sabe que la cultura neolítica se expandió a Egipto y desde allí al oeste 7.500-5.000 años a.C., pero también se conoce una fase más temprana hacia el interior, en el sur de Argelia y el norte de Chad, alrededor de 9.000 años a.C. (Cavalli-Sforza y col. 1994). Dado que los linajes del grupo VI procedentes de Oriente Próximo son diferentes en la zona de Marruecos (todos con la mutación M172) a los presentes en Mauritania, Sahara Occidental, Sudán y Etiopía (todos sin la mutación M172), quizás ambos conjuntos de haplotipos estén reflejando este proceso dual. El haplotipo VI₅ se pudo expandir en el evento más temprano y los haplotipos VI₆ y VI₇ en el tardío. Sin embargo, si la edad de coalescencia estimada de los haplotipos con la mutación M172 está en el rango de edades establecido, 4.286 años (1.837-15.000), la del grupo de cromosomas sin M172 queda muy por debajo de esa fecha propuesta para el desarrollo temprano del neolítico en la región, 3.297 años (1.423-11.538). La coalescencia del conjunto de cromosomas que contienen la mutación p12f2 en la región es de 7.143 años (3.061-25.000), estima muy similar a la obtenida para el conjunto de linajes de ADN mitocondrial no africanos detectados en la región (\approx 8.000 años, Rando y col. 1998).

En cuanto a las influencias procedentes de Europa, éstas pudieran ser debidas a las incursiones romanas en la costa norte de África o a la llegada, en tiempos coloniales, de

españoles, portugueses o franceses. Bosch y col. (2001) han datado esta influencia, estudiando la divergencia en los haplotipos del grupo IX, en unos 2.000 años.

A pesar de las diferencias en la frecuencia del conjunto de haplotipos de origen no africano en el noroeste, denotado en los mapas de frecuencia con islotes de alta y baja influencia (Figura 34), sus valores nunca llegan a ser del mismo nivel que los detectados con el ADN mitocondrial, que atribuyen la gran mayoría del pool génico del Magreb (> 60%) a influencias europeas y de Oriente Próximo (Rando y col. 1998; Brakez y col. 2001). Parece aceptable, al menos para los linajes paternos, que el Neolítico y resto de influencias desde el este, incluida la arabización, no han tenido mucha repercusión en el pool génico del noroeste de África. Esta conclusión también se ve apoyada por los resultados obtenidos para marcadores nucleares (Barbujani y col. 1994; Bosch y col. 1997, 2000; Comas y col. 2000; Flores y col. 2000b; Gómez-Casado y col. 2000). Por otra parte, la mayoría de los linajes del cromosoma Y en el norte de África se distribuyen en un eje norte-sur, al igual que sucede con otros marcadores, como es el caso del ADN mitocondrial (Lalueza, 1997; Rando y col. 1998; Krings y col. 1999), inserciones *Alu* autosómicas (Comas y col. 2000) y haplotipos autosómicos *Alu/CD4* (Flores y col. 2000b). Bosch y col. (1997) detectan este patrón norte-sur con marcadores clásicos sólo en el tercer componente principal (12'8% de la variación), mientras que Cavalli-Sforza y col. (1994) lo detectan en el primero (34' % de la variación) y, además, extrapolado al conjunto del continente. Así, todos los datos parecen señalar para el norte de África la existencia de flujo génico subsahariano hacia el norte y de caucasoide hacia el sur.

2. PENÍNSULA IBÉRICA Y NORMANDÍA

2.1. Los haplotipos de asignación europea

Linajes paternos localizados exclusivamente en las poblaciones de Europa se determinaron ya incluso con el sistema de sondas p49a,f. Con este sistema se descubrió que un haplotipo, denominado p49a,f ht15, se distribuía exclusivamente por Europa, alcanzando las

mayores frecuencias en el oeste, concretamente en vascos (89%; Lucotte y Hazout, 1996) descendiendo en gradiente hacia el sureste (Semino y col. 1996; Lucotte y Loirat, 1999). Del mismo modo, otros haplotipos y alelos ligados a éste se han detectado en patrones clinales en toda Europa: diferentes alelos de STR (Quintana-Murci y col. 1999), Hg 1 (Hill y col. 2000; Rosser y col. 2000), red 3.1G (Malaspina y col. 2000) y el conjunto de linajes M173/M17Gins (Semino y col. 2000). La interpretación que se le ha dado a estos resultados es que este conjunto de haplotipos probablemente constituyen el legado pre-neolítico de Europa (Semino y col. 1996, 2000; Lucotte y Loirat, 1999; Rosser y col. 2000; Wilson y col. 2001). Las edades de coalescencia estimadas a partir de la variación de los STR asociados a estos linajes apoyan esta hipótesis, arrojando una ancestría para el conjunto de Europa de unos 30.000 años (Malaspina y col. 2000; Semino y col. 2000) y de unos 8800-9.500 años para la región atlántica, incluido los vascos (Wilson y col. 2001). Dado que los ancestros de los linajes con la mutación M173 se encuentran en Asia central (Wells y col. 2001), estos linajes se habrían dispersado desde allí a Europa hace unos 30.000-40.000 años. Este conjunto de resultados, apoyados en datos de cambios climáticos y patrones similares detectados en la flora y la fauna de Europa, han llevado a sugerir que, después del Último Pleniglacial, hace menos de 15.000-20.000 años, estos linajes se habrían expandido por Europa desde un refugio localizado en la Península Ibérica (Semino y col. 2000; Underhill y col. 2001b; Wells y col. 2001). En un periodo más reciente se habrían originado en esta población otras mutaciones como *SRY2627*, hace unos 2.300-3.500 años (Hurles y col. 1999), M153 (Bosch y col. 2001), o incluso una reversión en el sitio *SRY1532*, hace unos 7.500-15.000 años (Karafet y col. 1999; Passarino y col. 2001; Wells y col. 2001). El haplotipo definido por esta última mutación también ha sido implicado en un proceso de expansión tras el Último Pleniglacial, pero desde un refugio en la zona de Ucrania, donde alcanza su máxima frecuencia y diversidad asociada a los STR (Rosser y col. 2000; Semino y col. 2000).

Otro conjunto de linajes, relacionados por la mutación M170, se consideran también de posible origen europeo. Su distribución, alcanzando las mayores frecuencias en el centro-este de Europa, su exclusividad en Europa, salvo por su aparición en otras poblaciones debido a contactos recientes, y su edad de coalescencia, calculada a partir de la diversidad asociada a los STR en unos 22.000 años, sugieren que estos linajes también pudieron existir en las poblaciones del paleolítico europeo (Semino y col. 2000). En el suroeste de Europa, además, se ha acumulado sobre este linaje la mutación M26, asociada principalmente a la Península Ibérica (este trabajo) y Cerdeña (Semino y col. 2000). La edad global estimada para esta mutación también coincide con las fechas del Último Pleniglacial (Semino y col. 2000), aunque la diversidad de STR acumulada en la Península Ibérica y Canarias arrojaría unos 3.223 años (1.381-11.280). Más recientemente la edad de un haplotipo íntimamente asociado a esta mutación se estimó en unos 5.700 años, incluyendo en los cálculos poblaciones de toda Europa (Scozzari y col. 2001). Quizás las

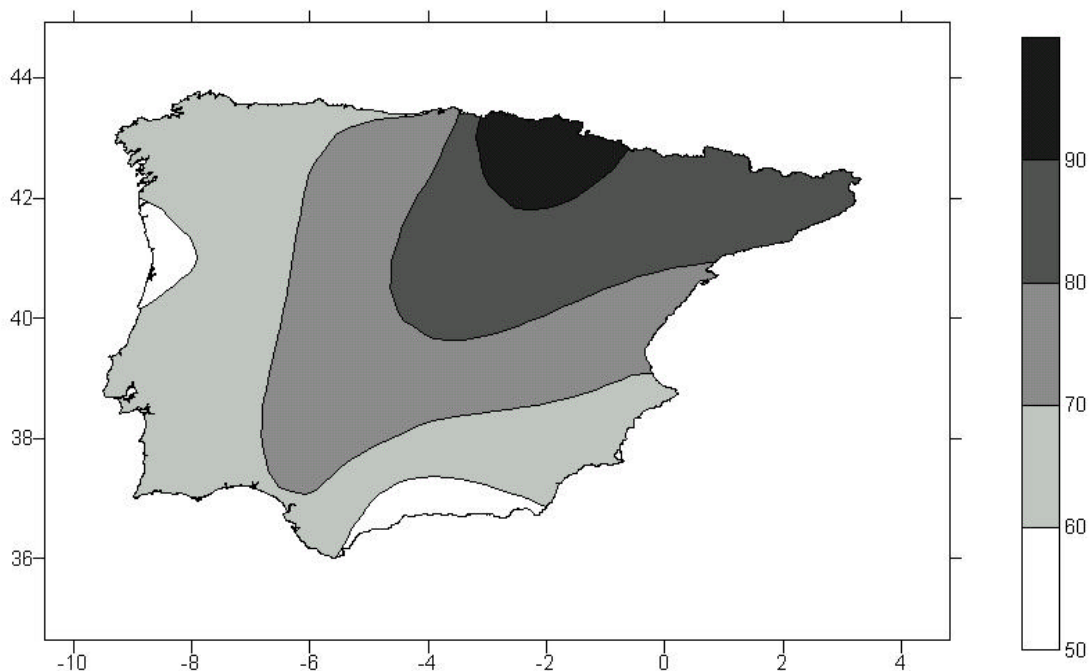


Figura 35. Mapa de la Península Ibérica mostrando la distribución de frecuencias de los linajes de asignación europea (grupo IX y haplotipos VI₃ y VI₄). La escala de la derecha relaciona un rango de frecuencias para este conjunto de linajes con un color.

discrepancias se deban a la utilización de STR diferentes o a las asunciones del propio estimador. Estos dos conjuntos de linajes (grupo IX completo y haplotipos VI₃ y VI₄) conformarían el pool Paleolítico de Europa y representan un 69'4% de las muestras de procedencia europea de este trabajo (71'6% al considerar a los vascos y a los catalanes) distribuyéndose por toda la Península, aunque concentrados mayoritariamente en el noreste alcanzando el máximo en el País Vasco, y Normandía (Figura 35). Esta distribución, con el País Vasco como población extrema conteniendo la mayor proporción de linajes de ancestría paleolítica y de la que parten las bandas de frecuencias concéntricas, concuerda perfectamente con la representación del primer componente principal sobre un mapa geográfico para marcadores nucleares (Bertranpetit y Cavalli-Sforza, 1991) y con los datos disponibles para el ADN mitocondrial, que sitúan a esta región como extremo en los rangos de ancestría paleolítica del resto de Europa (Richards y col. 2000). Para la gran mayoría de los estudios realizados sobre marcadores nucleares, los vascos aparecen como una de las poblaciones extremas de Europa, lo cual se ha atribuido también a una conservación casi total de ancestría pre-neolítica favorecida por aislamiento y deriva génica (Bertranpetit y Cavalli-Sforza, 1991; Cavalli-Sforza y col. 1994; Comas y col. 1998). Este patrón de frecuencias, con las mayores densidades en el noreste peninsular, no debe interpretarse como una expansión, sino como una diferencia entre las poblaciones que han conservado más el acervo existente desde el Mesolítico y las más influenciadas externamente.

2.2. Los haplotipos con asignación caucásica y del Oriente Próximo

Los linajes definidos por la mutación p12f2 se relacionaron desde un principio con Oriente Próximo, donde alcanzan su máxima frecuencia (> 50%). Además, esta mutación se detectó distribuida en toda Europa, decreciendo en frecuencia desde el sureste al oeste, al contrario de lo detectado para los linajes p49a,f ht15 (Semino y col. 1996), Hg 1 (Rosser y col. 2000) y M173/SRY1532G (M173/M17Gins en Semino y col. 2000). La edad de coalescencia calculada para la mutación varía entre 15.000-20.000 años en un estudio global (Hammer y col.

2000), 7.500-10.500 años analizando sólo poblaciones de Oriente Próximo (Quintana-Murci y col. 2001; Nebel y col. 2001), y 4.048 años (1.735-14.167) calculado para las poblaciones de la Península Ibérica (este trabajo). Por su distribución actual y su edad de coalescencia, los linajes con la mutación p12f2 han sido atribuidos en Europa a la contribución, en época Neolítica, de la expansión de los agricultores desde Oriente Próximo (Semino y col. 1996, 2000; Rosser y col. 2000; Nebel y col. 2001; Scozzari y col. 2001). Este periodo comenzó en la zona de Oriente Próximo-Anatolia-Cáucaso (hay controversia en cuanto al lugar exacto desde donde se produjo; Renfrew, 1990) 10.000 años a.C. y llegó a la Península Ibérica 2.800-5.500 años a.C. (Renfrew, 1990; Cavalli-Sforza y col. 1994; Zilhão, 2001), fechas que concuerdan perfectamente con las calculadas a partir de los datos moleculares.

Recientemente el conjunto de linajes paternos definidos por la mutación M201 también ha sido relacionado con la difusión del Neolítico en Europa por su coalescencia, calculada en unos 15.000 años, y distribución, con las máximas frecuencias en el Cáucaso (Semino y col. 2000; Wells y col. 2001). Así pues, al menos estos dos conjuntos de linajes, uno portando la mutación p12f2 y otro con la mutación M201, originados en diferentes zonas de Asia, habrían llegado a Europa con la expansión neolítica.

El hecho de que este conjunto de linajes no muestre una distribución homogénea en la Península Ibérica, con el definido por la mutación M201 más abundante en la zona norte, y el conjunto definido por la mutación p12f2 en Andalucía ($P = 0'023$ en una prueba exacta de Fisher), hace pensar en una ruta de dispersión diferente para cada conjunto, como ya ha sido propuesto para el resto de Europa (Renfrew, 1990). No obstante, podría tratarse de un sólo proceso pero trazado por un conjunto heterogéneo de población, o por el efecto de la deriva génica. En cualquier caso, el conjunto de estos linajes (haplotipos VI₂, VI₅-VI₇) constituye un 14'3% de la muestras europeas de este trabajo (13'4% al considerar a vascos y catalanes) y se distribuye en sus más altas frecuencias en tres zonas costeras (Cataluña, Andalucía y la costa

Atlántica de Portugal) disminuyendo progresivamente hacia el interior (Figura 36). Probablemente esta distribución de máximos entrecortados no sea real, sino que se deba sencillamente a la ausencia de puntos poblacionales en las regiones intermedias. Pereira y col. (2000b) encontraron un alto porcentaje de haplotipos definidos por la mutación p12f2 en el sur de Portugal ($\approx 29\%$), lo que continuaría el patrón detectado por la costa suroeste.

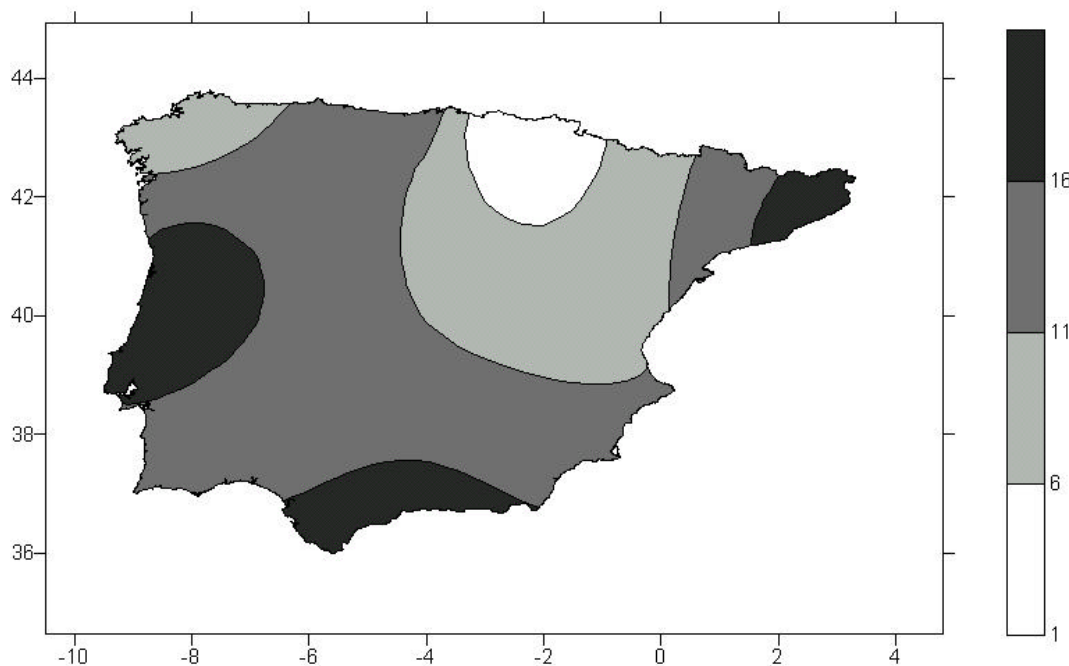


Figura 36. Mapa de la Península Ibérica mostrando la distribución de frecuencias de los linajes originarios del oeste asiático (haplotipos VI₂, VI₅-VI₇). La escala de la derecha relaciona un rango de frecuencias para este conjunto de linajes con un color.

Precisamente esta distribución concuerda a la perfección con la forma sugerida de extensión del Neolítico en la Península Ibérica, entrando por el Mediterráneo y distribuyéndose estrictamente a lo largo de la costa hasta el Atlántico, donde dejaría la mayoría de los asentamientos, y llegando más tarde a las zonas interiores (Cavalli-Sforza y col. 1994).

Algunos autores, basados en datos de marcadores nucleares, están a favor de que la difusión de la agricultura en el Neolítico es el proceso demográfico de mayor importancia para la

estructuración de las poblaciones europeas actuales, y que la gran mayoría del pool génico de estas poblaciones habría sido reemplazado, en el proceso, por el de los agricultores de Oriente Próximo (Sokal y Menozzi, 1982; Cavalli-Sforza y col. 1994; Barbujani y col. 1998; Chikhi y col. 1998; Barbujani y Bertorelle, 2001). Sin embargo, tanto el ADN mitocondrial (Corte-Real y col. 1996; Richards y col. 1996, 2000; Torroni y col. 1998) como el cromosoma Y (Lucotte y Loirat, 1999; Hill y col. 2000; Malaspina y col. 2000; Rosser y col. 2000; Semino y col. 2000; Scozzari y col. 2001; Wilson y col. 2001) favorecen más la ancestría pre-neolítica de Europa frente a la neolítica, atribuyendo para el continente, en conjunto, no más de un 25% de contribución neolítica. Los linajes del cromosoma Y procedentes de la difusión neolítica en la Península Ibérica (sumando el conjunto de haplotipos VI₂, VI₅-VI₇) suponen aproximadamente la mitad (13'4%) de lo calculado para el resto de Europa, resultado que también apoyan los datos sobre ADN mitocondrial (6'7-12%; Richards y col. 2000). Aunque muy coincidentes, estas estimas deben tomarse como valores mínimos puesto que, al menos para los datos del cromosoma Y, muchos linajes no se han considerado por la ignorancia actual que se tiene sobre ellos (ver pag. 132). Del mismo modo, estas estimas podrían reducirse aún más, puesto que algunos haplotipos que se han considerado procedentes de Oriente Próximo podrían tratarse igualmente de influencias desde el norte de África o de una mezcla de ambos eventos. En especial, esto podría ser válido para el haplotipo VI₅ por varias razones: es bastante frecuente en el noroeste africano, su distribución en la Península Ibérica es muy coincidente con la del haplotipo III₃ (Tabla 6), un haplotipo norteafricano, y su edad de coalescencia en la Península Ibérica es de 1.276 años (547-4.464), más coincidente con la observada para el haplotipo III₃ (ver pag. 132) que con la de una dispersión neolítica. En contrapartida, la variación STR asociada a este haplotipo en la Península está bastante diferenciada de su homólogo en el noroeste africano (ver pag. 141). La edad resultante para los haplotipos VI₆ y VI₇ es de 4.911

años [2.104-17.188] y 3.827 años [1.640-13.393] respectivamente, fechas muy ajustadas a la propuesta para la dispersión del neolítico en la región.

Del compendio de resultados obtenidos para el cromosoma Y y el ADN mitocondrial en la distribución de linajes de posible procedencia paleolítica o neolítica, y favorecido por los resultados de marcadores nucleares (Bertranpetit y Cavalli-Sforza, 1991), se puede extraer que la expansión del Neolítico tuvo menor impacto genético en la Península Ibérica que en el resto de Europa, manteniendo en su pool génico una gran proporción de linajes diferenciados durante el Paleolítico. Si bien la expansión del Neolítico no fue la única influencia exterior recibida por las poblaciones paleolíticas de la Península Ibérica, está bastante aceptado que el resto, producidas mayoritariamente durante el último milenio a.C., tuvieron poca influencia genética debido a la densidad poblacional que soportaba ya la zona (Cavalli-Sforza y col. 1994).

2.3. Los haplotipos de asignación africana

Las influencias africanas también han moldeado el acervo genético de la Península Ibérica (Arnaiz-Villena y col. 1995, 1997, 1999; Côté-Real y col. 1996; Rocha y col. 1999; Flores y col. 2000a, 2000b; Gómez-Casado y col. 2000; Bosch y col. 2000, 2001; Larruga y col. 2001), aunque éstas no se detectaron en un análisis de componentes principales con marcadores clásicos (Bertranpetit y Cavalli-Sforza, 1991). Aún existiendo incertidumbre en cuánto afectó y el momento en el cual se produjo, en las muestras de la Península Ibérica del presente trabajo se detectan haplotipos del cromosoma Y indudablemente originarios de África. Se han encontrado haplotipos de origen subsahariano, raramente detectados fuera del continente, en el norte peninsular (un individuo del grupo II en Portugal, dos individuos del haplotipo III₁ en Cantabria) y en Andalucía (un individuo del haplotipo III₁ y dos individuos del haplotipo III₇). Pero la gran mayoría de los haplotipos de origen africano que se observan en la Península pertenecen a los tres haplotipos característicos del norte del continente africano (III₂-III₄), siendo el más extendido y el que alcanza las mayores frecuencias el III₃, que se relaciona casi exclusivamente

con el noroeste africano. Hay que destacar que la mutación M78, que define al haplotipo III₄, también se encuentra asociada a la mayoría de los haplotipos YAP+ europeos (Semino y col. 2002), por lo que parte de su aparición en la Península podría achacarse a ancestría del propio continente o de Oriente Próximo. Por otra parte, no es posible reconocer si los haplotipos de origen norteafricano llegaron al mismo tiempo o en eventos diferentes que los de origen subsahariano, dado que éstos se encuentran mezclados en las poblaciones del norte de África. Debido a esto, se han tratado todos estos haplotipos, tanto los de asignación subsahariana como los de asignación norteafricana, como indicadores de la influencia africana. En conjunto estos linajes (grupo II y haplotipos III₁-III₄ y III₇) constituyen un 10'1% de las muestras de origen europeo (9'3% al considerar a vascos y catalanes) y se distribuyen (Figura 37) casi justo al contrario que el conjunto de linajes atribuidos a ancestría paleolítica (grupo IX y haplotipos VI₃ y VI₄). Dado que estos linajes no constituyen el 100% del pool masculino de las poblaciones de las que proceden, la influencia africana sería superior. Tomando como referencia al conjunto de linajes africanos de los bereberes de Marruecos (86'9%), una estima sencilla de la influencia total africana en la Península Ibérica sería de un 10'7% ($[0'093/0'869] \times 100$), con valores oscilando en un rango de más del 20% en Málaga y Galicia, a menos de un 6% en Cádiz, Castilla, Cataluña y País Vasco. Del mismo modo, los datos actuales de ADN mitocondrial arrojan influencias africanas en la región de 7-10% (Côrte-Real y col. 1996; Pereira y col. 2000a; Larruga y col. 2001). Sin embargo, el 32 y 20% de influencia africana calculada por Bosch y col. (2000) y Flores y col. (2000b), ambas en base a marcadores autosómicos, son ligeramente superiores. Las diferencias entre los valores calculados con marcadores autosómicos y los uniparentales (cromosoma Y y ADN mitocondrial) podrían explicarse sencillamente por un mayor error asociado a los primeros, debido a los pequeños valores de diferenciación que permiten detectar. Sin ir más lejos, en la comparación de las poblaciones peninsulares con las del

noroeste africano, mientras el cromosoma Y arroja $F_{ST} \cong 0.300$, los marcadores autosómicos detectan $F_{ST} \cong 0.019-0.008$ (Bosch y col. 2000; Comas y col. 2000; Flores y col. 2001b).

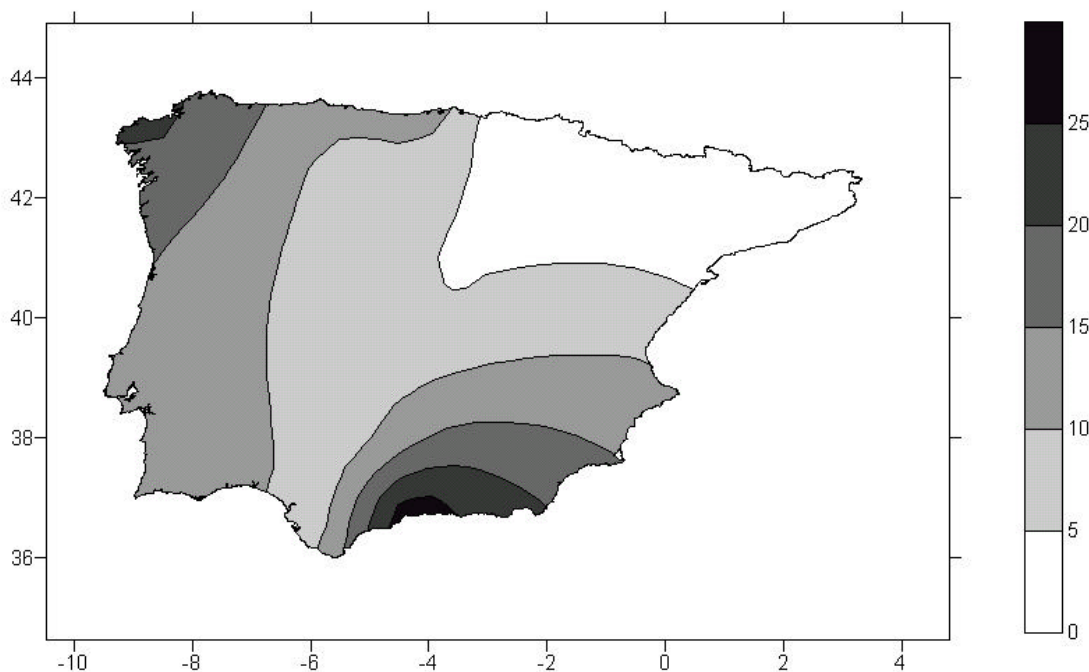


Figura 37. Mapa de la Península Ibérica mostrando la distribución de frecuencias de los linajes originarios del continente africano (grupo II y haplotipos III₁-III₄ y III₇). La escala de la derecha relaciona un rango de frecuencias para este conjunto de linajes con un color.

Apoyando el hecho de que no se observen diferencias en la distribución de estos linajes entre las poblaciones de la cornisa cantábrica y Andalucía ($P = 0.170$ en un a prueba exacta de Fisher), se dibujan claramente dos zonas con máximos de influencia, una centrada en el este de Andalucía y otra en el noroeste peninsular (Figura 37). El gradiente del sur peninsular concuerda con la región de máxima influencia árabe en la ocupación histórica que duró siete siglos (Chejne, 1980). El segundo máximo, situado en el noroeste peninsular, donde la ocupación árabe fue transitoria, no puede explicarse de la misma manera. Las influencias africanas en Andalucía ya se habían puesto de manifiesto en estudios previos del cromosoma Y (Bosch y col. 2001; Lucotte y col. 2001), de marcadores autosómicos (Bosch y col. 2000) y de ADN mitocondrial (Côte-

Real y col. 1996). En el norte peninsular, recientemente se han detectado posibles influencias africanas para el cromosoma Y. Scozzari y col. (2001) detectaron el haplotipo 25-2, muy relacionado con el III₃ de este trabajo, en una frecuencia altísima (42%) en una muestra de pasiegos. Del mismo modo, estudios con marcadores autosómicos ya habían detectado influencias africanas también en las poblaciones del norte peninsular: Arnaiz-Villena y col. (1995) detectaron algunos haplotipos raros del sistema HLA compartidos entre vascos y bereberes argelinos; Esteban y col. (1998) encontraron haplotipos de inmunoglobulinas GM de asignación africana en vascos, cántabros y franceses. Pero los resultados más llamativos de todos son los pertenecientes a trabajos con ADN mitocondrial. Uno de los linajes maternos, denominado U6 y con clara asignación al noroeste de África (Rando y col. 1998), se encuentra en la Península Ibérica. Curiosamente, y aunque se ha encontrado en la población andaluza, su frecuencia es mayor en la zona norte, detectándose incluso en los maragatos, una población aislada de la comunidad leonesa (Larruga y col. 2001). Del mismo modo, Salas y col. (1998) y Pereira y col. (2000a) lo detectan en Galicia y norte de Portugal, pero no en el centro ni en el sur de Portugal. Del conjunto de resultados para todo tipo de marcadores moleculares, parece probable que, además de las influencias históricas documentadas, habría que recurrir a otros procesos demográficos para explicar la aparición de influencias africanas en el norte de la Península Ibérica. Arnaiz-Villena y col. (1995, 1997, 1999) concluyen que las semejanzas genéticas para el sistema HLA entre la Península Ibérica y los bereberes son debidas a que ambas poblaciones proceden de un ancestro común pre-neolítico norteafricano. En cualquier caso, las edades de coalescencia calculadas para los haplotipos III₃ y III₄ en la Península Ibérica, 1.407 años (603-4.924) y 2.902 años (1.244-10.156) respectivamente, apoyan más que los contactos hayan sido históricos, o quizás neolíticos, que no debidos a una relación paleolítica.

2.4. Haplotipos indeterminados

Se han incluido en esta sección haplotipos detectados en este trabajo de los cuales no es posible extraer información basada en los estudios actuales del cromosoma Y. Para ser lo más conservador posible ninguno de ellos ha sido tratado en el análisis filogeográfico. Estos linajes (haplotipos III₅, III₆, VI₁ y X₁, y grupo VIII) no suponen más que un 5'6% del total de la muestra y, por tanto, su incorporación a cualquiera de los grupos de asignación anteriores no cambiará los patrones detectados.

De los haplotipos III₃ y III₆, ambos con la mutación M123, ya se ha hablado anteriormente. No se han incorporado al pool de haplotipos de asignación africana porque se han detectado, hasta el momento, casi exclusivamente (excepto los dos individuos Etíopes), compartidos entre Asia y Europa, pero además en frecuencias muy similares (siempre como haplotipos raros) en ambos continentes. Por lo tanto, tampoco pueden atribuirse con claridad a Oriente Próximo o Europa (Underhill y col. 2000; este trabajo). Quizás sean estos haplotipos los que estén implicados en la clina de frecuencias detectada para un conjunto del cromosoma Y con la mutación *DYS287* en Europa (Hammer y col. 1997, 1998; Rosser y col. 2000; Semino y col. 2000), aunque las edades de coalescencia calculadas para el conjunto de haplotipos con la mutación M123 en la Península Ibérica, 6.494 años (1.855-22.727), no favorecen esta hipótesis.

Los haplotipos VI₁ y X₁ se han detectado en frecuencias muy bajas tanto en la Península Ibérica como en el resto de Europa (Semino y col. 2000). Incluso VI₁ se ha detectado recientemente en una muestra de Senegal (Semino y col. 2002). Por sus situaciones en la red de máxima parsimonia (Figura 19) se puede deducir que ambos son haplotipos muy ancestrales. Dado que casi toda la diversidad detectada en Europa derivaría de ellos, pudiera ser que se tratase de haplotipos dispersados con las expansiones que llevaron al grupo IX y a los haplotipos con la mutación M170 a toda Europa en el Paleolítico. En cualquier caso, en este punto esta idea es sólo una conjetura.

El grupo VIII está constituido por haplotipos muy distribuidos, salvo el caso del VIII₂. Del haplotipo VIII₁ sería difícil concluir algo, puesto que es un grupo de cromosomas no definido por ninguna mutación terminal y además se encuentra distribuido por Europa (Semino y col. 2000; Underhill y col. 2000; Bosch y col. 2001; este trabajo), África (Underhill y col. 2000; Bosch y col. 2001; este trabajo), Asia (Underhill y col. 2000; Wells y col. 2001) y Oceanía (Underhill y col. 2000, 2001a). Parece que se originó fuera de África y que su presencia en éste continente se debe a influencias externas (Semino y col. 2002). Quizás por esta distribución global estos haplotipos pudieron llegar a Europa desde el Paleolítico, aunque nuevamente se estaría especulando. La distribución del haplotipo VIII₂, definido por la mutación Tat, sí se conoce bien. Este haplotipo alcanza sus máximas frecuencias en el norte de Europa y Asia central, y se ha propuesto como una conexión entre ambas áreas (Zerjal y col. 1997). Su aparición en la Península Ibérica contrasta mucho con su distribución conocida, si bien recientemente también se ha detectado en Turquía (Semino y col. 2000). Si su aparición no se debe a un error de muestreo (un solo individuo), quizás llegó al suroeste de Europa con el haplotipo IX₂, aunque la edad de coalescencia calculada (\approx 2.000-4.000 años, Zerjal y col. 1997) lo situaría en un periodo más reciente, no coincidente con la expansión poblacional desde la zona de Ucrania después del Último Pleniglacial (Semino y col. 2000). En referencia al haplotipo VIII₃, dado que se encuentra consistentemente en todo el sur de Europa, África y Oriente Próximo (Semino y col. 2000, 2002; Underhill y col. 2000; este trabajo), aunque siempre en frecuencias muy bajas, podría tratarse de contactos recientes o bien, por su edad de coalescencia, 3.125 años (1.339-10.938), de un representante superviviente de algún linaje temprano originado en la región antes del Último Pleniglacial.

3. ISLAS CANARIAS

El panorama resultante para Canarias en cuanto a la distribución de los linajes paternos es de una gran homogeneidad. Aunque algunas parejas de poblaciones insulares se diferencian

significativamente, ninguna en especial lo hace de una manera reiterada con todas las islas. Este mismo resultado se ha obtenido recientemente para algunos marcadores autosómicos (Flores y col. 2001b). En contrapartida, en los resultados del ADN mitocondrial, La Gomera destaca por su gran divergencia con respecto al resto (Rando, 1998).

De la distribución de linajes paternos en las islas se puede deducir fácilmente, por el gran parecido con respecto a la de la Península Ibérica (Figura 38), que la gran mayoría del pool

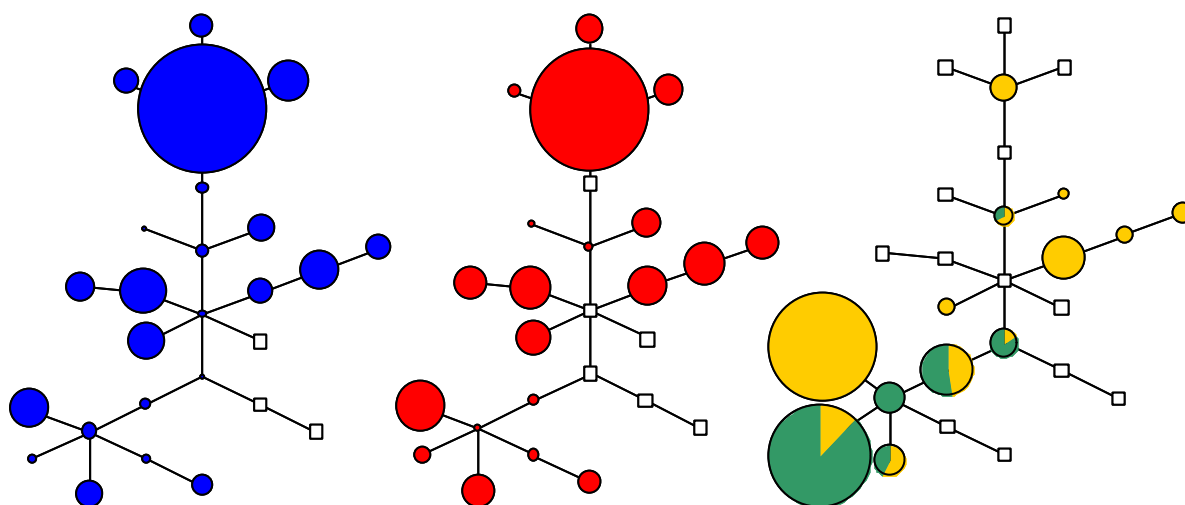


Figura 38. Red de máxima parsimonia establecida para los UEP relacionando los haplotipos del cromosoma Y de las muestras analizadas en este trabajo de Europa (azul), Canarias (rojo), noroeste de África (amarillo) y área subsahariana (verde). Los círculos son proporcionales a las frecuencias y los cuadrados representan variantes no detectadas. Los haplotipos están situados en la misma posición que en la Figura 19.

génico paterno de las islas procede de Europa, tal como se detecta en las contribuciones parentales. Estos resultados están totalmente de acuerdo con el proceso de conquista y recolonización del Archipiélago Canario, donde los contingentes peninsulares han tenido las mayores repercusiones.

En este punto hay que destacar el parecido de las poblaciones insulares con respecto a las andaluzas y portuguesas. Las menores distancias genéticas de las diferentes islas son con Portugal y Sevilla, salvo en el caso de Fuerteventura que la tiene con la muestra malagueña.

Aunque este resultado puede deberse a procesos de deriva génica, también podría explicarse por el gran efecto que han tenido ambas regiones en la historia de Canarias. Andaluces, portugueses y gallegos fueron los pueblos que mayor incidencia tuvieron durante el periodo de conquista, y aún más en el proceso recolonizador del Archipiélago (Suárez y col. 1988). Aún así, todavía hoy se siguen detectando en Canarias linajes paternos de posible origen subsahariano y norteafricano, incluso en frecuencias superiores a las detectadas en la Península Ibérica. Son destacables las frecuencias del haplotipo III₃ y III₄, ambos con probable origen norteafricano, alcanzando en Canarias un 8'3 y 3'5%, respectivamente, en comparación con un 5'2 y 2'4%, respectivamente, en la Península Ibérica. El haplotipo mayoritario en las poblaciones subsaharianas, III₇, se detecta en Canarias, aunque está presente también en las poblaciones peninsulares (0'9 y 0'3%, respectivamente). El resto de haplotipos, salvo excepciones (ver pag. 140), pueden encuadrarse como originarios de Europa aunque, como ya se ha mostrado en resultados, suelen aparecer en menor frecuencia que en la Península Ibérica. También han sido detectadas influencias africanas en Canarias con marcadores nucleares (Martell y col. 1986; Morilla y col. 1988; Pérez y col. 1991; Pinto y col. 1994, 1996b, 1996c; Moral y col. 1997; Flores y col. 2001b) y con ADN mitocondrial (Pinto y col. 1996a; Rando y col. 1999). Mientras que para los marcadores nucleares no se ha podido distinguir entre influencias subsaharianas y norteafricanas, no ocurre lo mismo para los marcadores uniparentales. Por islas, los linajes paternos de posible procedencia subsahariana se detectaron en El Hierro, La Palma, Gran Canaria, Lanzarote y Fuerteventura, en un rango del 1 al 4'3%. La incidencia es bastante baja considerando la cantidad de esclavos negros que entraron en las islas durante los años siguientes a la conquista (Lobo-Cabrera, 1993). Sin embargo, los trabajos de ADN mitocondrial también han detectado linajes subsaharianos en algunas islas del Archipiélago (salvo en Fuerteventura y La Palma), y sólo en frecuencias ligeramente superiores (3-12%; Rando, 1998). En cualquier caso, ni con el

cromosoma Y ni con el ADN mitocondrial puede esclarecerse si estos linajes subsaharianos llegaron a Canarias con los esclavos o si se encontraban ya en la población aborigen.

Por el contrario, los linajes maternos de origen norteafricano sí que se han encontrado en las islas en mucha mayor proporción (8-39%; Rando, 1998) que los paternos. A este respecto, el dato de más increíble contraste entre la distribución de los linajes paternos y maternos para Canarias se encuentra en la población gomera. Mientras que La Gomera es la isla que mejor conserva los linajes maternos de posible origen norteafricano (39%), siendo la contribución bereber del 100% (Rando, 1998), para el cromosoma Y es de las que menos, detectándose una contribución norteafricana de $0.34 \pm 5.70\%$. También para esta isla se detectaron las frecuencias más altas para el alelo 0 del grupo ABO, lo que se relacionó directamente con su parentesco norteafricano (Pinto y col. 1996c). Si bien La Gomera es la isla donde más discrepancia existe entre los marcadores uniparentales, el resto del Archipiélago sigue el mismo patrón.

Recientemente Flores y col. (2001a) estimaron en un 9% (8% norteafricano y 1% subsahariano) la contribución africana total a las Islas Canarias usando marcadores del cromosoma Y. Incluso estando basado tan sólo en nueve marcadores bialélicos y, por tanto, con una menor resolución en la determinación de haplotipos, su estima no varía mucho con respecto a la del presente trabajo ($\approx 7\%$, con aproximadamente un 6% norteafricano y un 1% subsahariano). Ambas están muy por debajo de las detectadas usando tanto marcadores nucleares, donde África contribuye con un promedio del 29%, como secuencias de ADN mitocondrial, donde África contribuye con un 51% como media (Tabla 15). En referencia al ADN mitocondrial, los datos de Pinto y col. (1996a) han sido reanalizados recientemente por medio de nuevos estimadores teniendo en cuenta las diferencias moleculares, pero de esta manera la incidencia africana (73%) es aún más extrema (Dupanloup y Bertorelle, 2001). El aumento de esta estima con respecto a las dos anteriores se puede deber a que las diferencias en posiciones con baja tasa de mutación, que definen linajes conservados por descendencia, se han

considerado con el mismo peso que posiciones con alta tasa de mutación, que pueden aparecer en linajes no relacionados. Aún sin considerar esta última estima por su carácter extremo, las discrepancias entre los resultados del cromosoma Y y los calculados en base al resto de marcadores son notables. La explicación más sencilla para este conjunto de resultados es asumir un proceso de asimetría sexual a raíz de la conquista y la recolonización. A partir de las estimas reales de contribuciones parentales medias, se han inferido las contribuciones teóricas proporcionales de cada región geográfica por sexo en la población canaria (Tabla 16). Tanto si se usan los datos del ADN mitocondrial como los del cromosoma Y, se observa que mientras los pueblos de la Península Ibérica han contribuido principalmente con población masculina, ambas

Tabla 15. Estimaciones de mezcla (%) en la población actual de Canarias según la población parental.

Marcador	Península Ibérica	Noroeste africano	África subsahariana
Grupos sanguíneos (Roberts y col. 1966)	72	22	6
Enzimas (Pinto y col. 1994)	70	20	10
Nucleares ^a	71	21	8
ADN mitocondrial (1) (Pinto y col. 1996a)	36	43	21
ADN mitocondrial (2) (Rando y col. 1999)	62	33	5
ADN mitocondrial (3) (Dupanloup y Bertorelle, 2001)	17	69	14
ADN mitocondrial ^b	49	38	13
Cromosoma Y (Flores y col. 2001a)	91	8	1

^aValores medios de las estimas de grupos sanguíneos y enzimas; ^bValores medios de las estimas 1 y 2 de ADN mitocondrial.

poblaciones africanas han contribuido casi exclusivamente con hembras. La historia de Canarias documenta perfectamente este resultado. Los aborígenes canarios, de donde probablemente provienen estas influencias africanas, fueron sometidos, incluso desde antes del comienzo de la conquista, a ataques e incursiones piratas en busca de esclavos. Las continuas batallas, que se prolongaron durante casi un siglo de conquista, las deportaciones y las revueltas disminuyeron drásticamente la población masculina aborígen en el Archipiélago (Suárez y col. 1988; Hernández, 1999). Tras la conquista, el reino de Castilla intentó preservar la población aborígen que permaneció. El caso más conocido es el de la isla de El Hierro en la que, desde el comienzo de la conquista, su población aborígen se encontró diezmada, principalmente dotada de mujeres y ancianos (Suárez y col. 1988), que posteriormente pudieron contribuir a la recolonización de la isla.

Tabla 16. Contribuciones paternas teóricas: porcentajes por sexo y área geográfica.

Marcadores	Península Ibérica	Noroeste africano	África subsahariana
ADN mitocondrial & nucleares			
Varones	46'5 (65%)	2 (10%)	1'5 (19%)
Hembras	24'5 (35%)	19 (90%)	6'5 (81%)
Cromosoma Y & nucleares			
Varones	46'5 (65%)	3 (14%)	0'5 (6%)
Hembras	24'5 (35%)	18 (86%)	7'5 (94%)

3.1. Los linajes paternos y la colonización aborígen de Canarias

A pesar de que la distribución de los linajes paternos de Canarias sea casi indistinguible de la de las poblaciones peninsulares que intervinieron en su conquista, aún hoy se detectan huellas que parecen atestiguar el modo en el que fue colonizado el Archipiélago desde el continente africano. De entre los linajes del grupo III, los haplotipos III₃ y III₄, definidos por las mutaciones M81 y M78 respectivamente, parecen haber llegado al Archipiélago desde África,

probablemente como un legado aborigen. Las evidencias que apuntan hacia esta suposición son varias. Por un lado, aunque se encuentran en la Península Ibérica, la frecuencia media de estos dos haplotipos en Canarias es superior. Un proceso de deriva génica en las islas podría explicar las diferencias, pero este proceso por sí solo no puede justificar el aumento sistemático de ambos en cada una de las islas. Por otro lado, los dos se distribuyen en un gradiente de frecuencias en el Archipiélago. Mientras el haplotipo III₃ disminuye en frecuencia hacia las islas occidentales, el haplotipo III₄ hace justo lo contrario (Tabla 8). En este punto además merece destacar que el haplotipo VI₅, definido por la mutación p12f2, tiene una distribución en las islas que mimetiza a la del haplotipo III₄, es decir, disminuyendo en frecuencia hacia las islas orientales. Aunque VI₅, de origen asiático, también se encuentra en la Península Ibérica su frecuencia en el noroeste africano es mucho más alta, sobre todo en el Sahara Occidental y Mauritania. Las distribuciones de III₄ y VI₅ también son coincidentes en África. Ambos tienen sus mayores frecuencias en las poblaciones del borde sur del Sahara, en el este, donde III₄ es mayoritario, y en el oeste, donde el mayoritario es VI₅. El haplotipo VI₅ también es más abundante en el este, habiéndose descrito hasta en un 33'4% en un muestra de Etiopía (Semino y col. 2002), más cercana a Oriente Próximo desde donde se pudo dispersar. Un estudio de las poblaciones del sur de Argelia, Libia

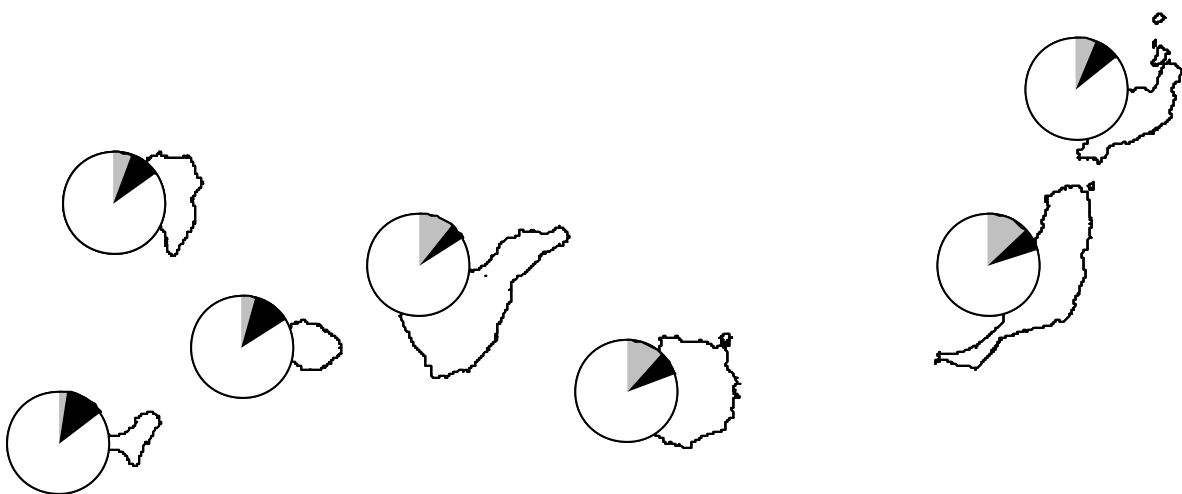


Figura 39. Distribución en Canarias de linajes masculinos indicadores de influencia magrebí (gris) y sahariana (negro).

y Egipto podría esclarecer aún más la relación entre ambos haplotipos, pero lamentablemente hasta el momento no han sido estudiadas con los marcadores adecuados. Dado que el haplotipo III₃ se encuentra estrechamente ligado al Magreb y los haplotipos III₄ y VI₅ a las poblaciones saharianas, se han distinguido estos dos tipos de indicadores en el Archipiélago (Figura 39). Al comparar sus distribuciones en las islas se observa que ambas se encuentran inversamente correlacionadas ($r_s = -0.893$, $P = 0.006$). Ello se debe a que las db s tienen, a su vez, una alta correlación con la distancia a la costa africana (desde el Cabo Juby). Aunque, mientras la contribución sahariana aumenta con la distancia ($r_s = 0.714$, $P = 0.072$), la contribución magrebí disminuye ($r_s = -0.857$, $P = 0.014$). También para el ADN mitocondrial (Rando y col. 1999) y marcadores autosómicos (Flores y col. 2001b) se han detectado patrones correlacionados con la distancia al continente africano, siempre asociados a un decrecimiento de heterocigosidad hacia las islas occidentales.

A pesar de que el haplotipo VI₅ no es un marcador genuino del norte de África, su incidencia en Canarias parece que se debe, al menos en parte, a influencia africana. Igual que éste, otros haplotipos de origen euroasiático pueden haber llegado a Canarias con los aborígenes, aunque su cuantificación es muy difícil. Quizás un análisis futuro con más STR permita detectar la implicación de algunos otros haplotipos. De lo anterior se deduce que el componente africano calculado en las estimas de mezcla (Long, 1991) deben tomarse como valores mínimos y, por tanto, muy conservadores.

El análisis de la variación de alelos de STR favorece la implicación de, al menos, estos tres haplotipos en la colonización aborígen del Archipiélago. En los tres haplotipos, una parte de la red de variación canaria se relaciona estrechamente con la variación norteafricana (Figura 40), incluso con la poca diversidad que muestra el haplotipo III₃. El número medio de diferencias en los tamaños de los alelos STR asociados, como una medida de diversidad, también sitúa a Canarias a medio camino entre el noroeste de África y la Península Ibérica (Tabla 17), excepto

para el haplotipo VI₅, donde el valor de Canarias es superior. Este último resultado puede explicarse por la divergencia existente en la diversidad de STR para las dos poblaciones parentales (Figura 40), mientras que Canarias comparte linajes con ambas. Nótese que en la red del haplotipo III₄ sucede algo similar, dando como resultando una diversidad en Canarias casi del mismo nivel que la norteafricana. En cualquier caso, los valores de diversidad estimados contradicen la posible implicación de la deriva génica en el aumento de sus respectivas frecuencias en las islas.

Tabla 17. Número medio de mutaciones en los STR por área geográfica.

Área	Haplotipo		
	III ₃	III ₄	VI ₅
Península Ibérica	0'648	1'010	0'582
Canarias	0'825	1'960	1'794
Norte de África	0'863	2'000	1'385

Dado que Canarias es una población mezcla de poblaciones que, a su vez, también están mezcladas, los tres linajes están presentes en los dos continentes de donde ha recibido las influencias. Por tanto, un tiempo de coalescencia en los haplotipos implicados no será un estima real del momento en que ocurrió la colonización aborigen. En este caso, las estimas serían un valor poco conservador, puesto que habrían incluidas mezclas antiguas y recientes. Una manera alternativa a la coalescencia es calcular cuánto divergen los patrones STR en los tres haplotipos entre Canarias y el conjunto Norte de África-Península Ibérica y comparar esta divergencia con el tiempo de coalescencia total del haplotipo en concreto. En este caso, la estima tampoco es perfecta puesto que, si bien se tienen en cuenta eventos recientes, esto no es así para la diversidad inicial en la colonización. De cualquier forma, la estima será conservadora. Tomando

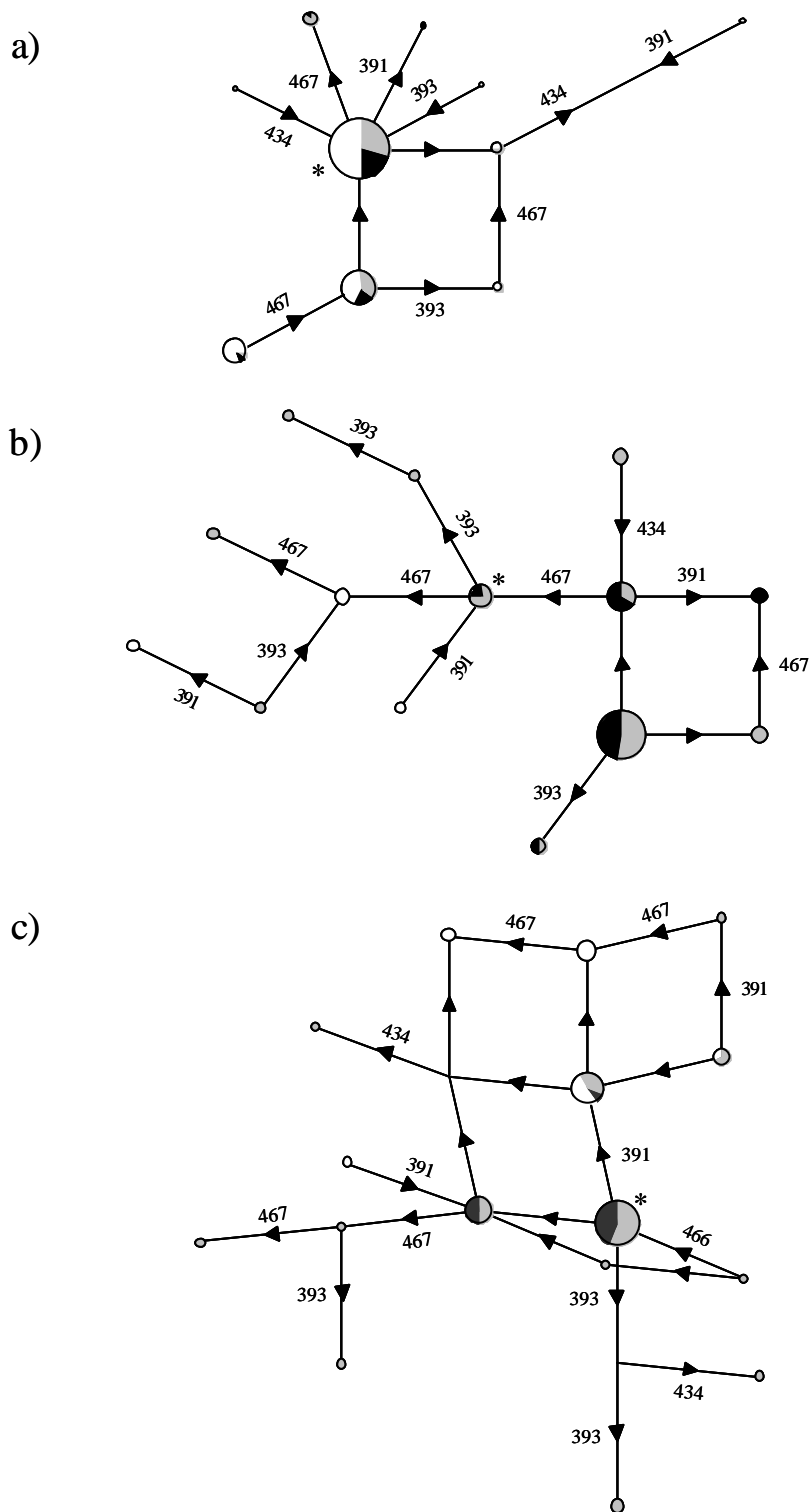


Figura 40. Redes RM de diversidades STR en los haplotipos III₃ (a), III₄ (b) y VI₅ (c). Los esquemas muestran los conjuntos STR-haplotipos, representados por círculos proporcionales a la frecuencia, relacionados por líneas que simbolizan los STR que cambian (indicado en las ramas). Las líneas paralelas expresan el mismo cambio. Las letras *DYS* de los STR han sido omitidas. Las flechas indican el sentido del aumento en una repetición en el STR. Las claves son: Canarias, gris; Península Ibérica, negro; norte de África, blanco. El asterisco simboliza la raíz (ver Apéndice) a partir del cual se ha estimado el tiempo de coalescencia.

todos los haplotipos III₃, independientemente de la población donde se encontró, la edad de coalescencia resultante es de 1.838 años (788-6.433). El porcentaje de divergencia de Canarias para este haplotipo es de 31'6%, lo que da un periodo de divergencia de ? 580 años. Los porcentajes de divergencia para los haplotipos III₄ y VI₅ son 52'6 y 52'7%, respectivamente. Dado que los tiempos de coancestría son diferentes para III₄ y VI₅, 6.060 años (2.598-21.221) y 3.760 años (1.610-13.147) respectivamente, el periodo de divergencia resultante también lo es. Para el haplotipo III₄ este periodo se calcula en unos 3.190 años, mientras que para el haplotipo VI₅ resulta en unos 1.982 años.

3.2. Hipótesis sobre el proceso de colonización aborigen a la luz del análisis del cromosoma Y en las poblaciones canarias actuales

Un conjunto de resultados obtenidos en el presente trabajo implica a tres haplotipos en la colonización aborigen del Archipiélago Canario. Su mayor frecuencia en las islas que en la Península Ibérica, sus distribuciones relacionadas con la distancia al continente africano, y sus diversidades asociadas a STR, favorecen la hipótesis de que los haplotipos III₃, III₄ y VI₅ llegaron a Canarias con la colonización prehispánica, al menos en parte. Sin embargo, las relaciones entre estos tres linajes son heterogéneas, tanto en Canarias como en su continente de procedencia. Los haplotipos III₄ y VI₅ se distribuyen en Canarias asociados mayoritariamente a las islas occidentales, en África ambos se relacionan principalmente a las poblaciones saharianas, y su divergencia global en Canarias es casi idéntica para ambos respecto a la detectada en las dos áreas continentales. Por otro lado, el haplotipo III₃ se encuentra en Canarias más ligado a las islas orientales, en África se distribuye mayoritariamente en el Magreb y su divergencia global en Canarias es menor que para los otros dos haplotipos. A partir de lo expuesto se abre la posibilidad de que dos oleadas africanas intervinieran en la colonización del Archipiélago Canario. Una primera oleada partiría de la región sahariana hace unos 2.000-3.000 años colonizando todo el Archipiélago. Más tarde, hace unos 600 años, desde la zona del Magreb

partiría una segunda oleada que afectó al menos a las islas orientales. Por la distribución actual de ambos indicadores, quizás esta segunda oleada desplazó parcialmente a la población residente. El proceso démico que se ha propuesto para la dispersión de la agricultura en Europa (Ammerman y Cavalli-Sforza, 1984) podría explicar perfectamente lo detectado en Canarias. En este proceso, la población residente es desplazada por nuevos contingentes con tecnologías más avanzadas. A medida que el punto de influencia se situara más lejos del foco de dispersión, la influencia sería menor. De esta manera las islas más occidentales habrían conservado mejor el sustrato arcaico, lo cual está de acuerdo con los estudios de Antropología física (Hernández, 1999).

Si bien la fecha estimada por Rando y col. (1999) de la colonización pionera en Canarias usando ADN mitocondrial, de 2.800 ± 900 años, está de acuerdo con la datación de la primera oleada calculada con el cromosoma Y, no puede decirse lo mismo de la segunda. Rando y col. (1999) sólo detectan un proceso colonizador común a todo el Archipiélago. Es más, concluyen que si hubieron otras influencias, éstas no dejaron una gran huella en la población actual de Canarias. Datos recientes usando haplotipos autosómicos *Alu/CD4* también apoyan que el asentamiento aborigen se produjo en un solo proceso (Flores y col. 2001b). En cualquier caso, la colonización más antigua detectada en este trabajo para el cromosoma Y y la única del ADN mitocondrial, concuerdan con la estimada a partir de las características culturales de los restos arqueológicos, que la sitúan en 2.500-1.000 años a.C. (Martínez, 1996; Hernández, 1999). En base a diferencias culturales entre los restos arqueológicos, una gran cantidad de trabajos defienden, al menos, una oleada más aparte de la llamada “de sustrato” (Schwidetzky, 1963; Onrubia-Pintado, 1987; Navarro y col. 1990; Martín, 1992; Navarro, 1992; Martínez, 1996), aunque, al menos para algunas islas, otros atribuyen estas diferencias sólo a movimientos culturales dentro de las mismas (Armay y González-Reimers, 1985-1987). En definitiva, en lo que atañe al momento y número de colonizaciones aborígenes de Canarias, las heterogeneidades

detectadas con los diferentes marcadores moleculares tienen sus paralelismos en las heterogéneas interpretaciones culturales.

V. CONCLUSIONES

Las conclusiones del presente trabajo pueden resumirse en los siguientes puntos:

1. El continente africano muestra una distribución heterogénea de linajes paternos con un elevado porcentaje de diferenciación (26'7%) entre sus poblaciones, correlacionada ésta con la distancia geográfica.
2. La mayoría de los linajes paternos en África pertenecen al grupo III, aunque en las diferentes áreas geográficas el haplotipo mayoritario es diferente: III₃ lo es en el noroeste, III₄ en el noreste y III₇ en el África Negra, con la excepción de los khoisan, en los que los haplotipos del grupo I son los predominantes.
3. La distribución de variantes del cromosoma Y en este continente permite distinguir un eje principal de diferenciación en sentido norte-sur, asociado a un aumento de linajes negroides y disminución de caucasoides hacia el sur, y otro secundario, localizado sólo en el norte, en sentido este-oeste.
4. La presencia de los grupos VI, VIII, IX, y probablemente del haplotipo III₆, detectados principalmente en el norte del continente, constituyen una evidencia clara de influencias eurasiáticas en África.
5. Aunque de menor nivel, la distribución de los linajes paternos entre las poblaciones de la Península Ibérica y Normandía permite detectar cierta heterogeneidad en el área (1'35%), correlacionada con la distancia geográfica.
6. Los linajes paternos del grupo IX y los del grupo VI definidos por la mutación M170, ambos presuntamente constituyendo la ascendencia pre-neolítica de las poblaciones europeas, muestran frecuencias más altas en Normandía y el noreste peninsular que en el resto de la Península Ibérica.
7. Los linajes VI₂, VI₅, VI₆ y VI₇, originales del Cáucaso y Próximo Oriente, y relacionados con el impacto de la expansión neolítica, se encuentran principalmente en las poblaciones costeras del Mediterráneo y Atlántico peninsular.

8. La aparición en la Península Ibérica de miembros del grupo II y de algunos del III constituye una evidencia clara de la influencia africana cuantificada en aproximadamente un 11%. Además, la distribución de estos linajes, con máximos de frecuencia en Andalucía y el noroeste peninsular, no pueden explicarse sólo por el proceso histórico de ocupación árabe.
9. La distribución conjunta de linajes paternos entre las poblaciones de las Islas Canarias es homogéneo.
10. Los linajes con una clara ancestría europea (haplotipos VI₃ y VI₄ junto al grupo IX) constituyen la mayoría del acervo génico paterno de las poblaciones de las Islas Canarias, alcanzando su valor máximo en Lanzarote (67%) y el mínimo en Fuerteventura (58'6%).
11. La presencia de haplotipos III₁ y III₇ atestigua la influencia subsahariana en las islas, con un máximo en El Hierro (4'3%) y el mínimo en La Gomera y Tenerife (0%), aunque no queda demostrada su procedencia por la importación de esclavos negros al Archipiélago.
12. La aparición del conjunto de haplotipos III₃, III₄ y VI₅ con mayor frecuencia en Canarias que en la Península Ibérica, se explica, al menos en parte, por influencia directa del norte de África, siendo, en conjunto, máxima en Fuerteventura (20%) y mínima en Lanzarote (14'5%).
13. La correlación positiva detectada entre las frecuencias en las islas de los haplotipos III₄ y VI₅ y la distancia de éstas al continente africano, y la divergencia acumulada en Canarias apoyan la existencia de una primera colonización del Archipiélago a partir del área sahariana.
14. La correlación negativa detectada entre las frecuencias en las islas del haplotipo III₃ y la distancia de éstas al continente africano, y la divergencia acumulada en Canarias

apoyan la existencia de una segunda colonización del Archipiélago a partir de la zona del Magreb.

15. Se ha estimado que la población canaria actual se compone de un 93% de linajes paternos de origen europeo y un mínimo de 7% de procedencia africana, de los cuales un 6% procedería del noroeste y un 1% del área subsahariana.
16. La discrepancia entre la composición de la población canaria deducida de los linajes paternos y la obtenida a partir de ADN mitocondrial, con una aportación media de 51% de linajes maternos africanos, se explica por la asimetría sexual producto de la recolonización europea.

VI. BIBLIOGRAFÍA

- Afonso, JM, Hernández, M, Larruga, JM, Cabrera, VM, y González, AM (1989). Human enzyme polymorphism on the Canary Islands. III. Tenerife Island population. *Hum Biol* 61:542-549.
- Ammerman, AJ, y Cavalli-Sforza, LL (1984). *The Neolithic transition and the genetics of population in Europe*. Princeton University Press, Princeton.
- Arco, MC, y Navarro, JF (1987). *Historia popular de Canarias. Vol. 1. Los aborígenes*. Centro de la Cultura Popular Canaria, S/C de Tenerife.
- Arnaiz-Villena, A, Benmamar, D, Alvarez, M, Diaz-Campos, N, Varela, P, Gomez-Casado, E, y Martinez-Laso, J (1995). HLA allele and haplotype frequencies in Algerians. *Hum Immunol* 43:259-268.
- Arnaiz-Villena, A, Martínez-Laso, J, y Alonso-García, J (1999). Iberia: Population Genetics, Anthropology, and Linguistics. *Hum Biol* 71:725-743.
- Arnaiz-Villena, A, Martínez-Laso, J, Gómez-Casado, E, Díaz-Campos, N, Santos, P, Martinho, A, y Breda-Coimbra, H (1997). Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by HLA allelic frequencies and haplotypes. *Immunogenetics* 47:37-43.
- Arnay, M, y González-Reimers, E (1985-1987). *La cerámica decorada prehispanica de Tenerife*. *Tabona* 6:241-277.
- Ayub, Q, Mohyuddin, A, Qamar, R, Mazhar, K, Zerjal, T, Mehdi, SQ, y Tyler-Smith, C (2000). Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res* 28:e8.
- Bandelt, H-J, Forster, P, Sykes, BC, y Richards, MB (1995). Mitochondrial portraits of human populations using median networks. *Genetics* 141:743-753.
- Barbujani, G, y Bertorelle, G (2001). Genetics and the population history of Europe. *Proc Natl Acad Sci USA* 98:22-25.

- Barbujani, G, Bertorelle, G, y Chikhi, L (1998). Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet* 62:488-491.
- Barbujani, G, Magagni, A, Minch, E, y Cavalli-Sforza, LL (1997). An apportionment of human DNA diversity. *Proc Natl Acad Sci USA* 94:4516-4519.
- Barbujani, G, Pilastro, A, de Domenico, S, y Renfrew, C (1994). Genetic variation in North Africa and Eurasia: Neolithic demic difusión vs. Paleolithic colonisation. *Am J Phys Anthropol* 95:137-154.
- Bernstein, F (1931). Die geographische Verteilung der Blutgruppen und ihre anthropologische Bedeutung. En: Comitato Italiano per lo studio dei problemi della popolazione. Istituto Poligrafico dello Stato, Roma, pp. 227-243.
- Bertranpetit, J, y Calafell, F (1996). Genetic and geographic variability in cystic fibrosis: evolutionary considerations. En: *Variation in the human genome*, Chadwick, D y Cardew, G (eds). John Wiley and Sons, Chichester, pp. 97-118.
- Bertranpetit, J, y Cavalli-Sorza, LL (1991). A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann Hum Genet* 55:51-67.
- Bosch, E, Calafell, F, Comas, D, Oefner, PJ, Underhill, PA, y Bertranpetit, J (2001). High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between Northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* 68:1019-1029.
- Bosch, E, Calafell, F, Pérez-Lezaun, A, Clarimón, J, Comas, D, Mateu, E, Martínez-Arias, R, Morera, B, Brakez, Z, Akhayat, O, Sefiani A, Hariti, G, Chambón-Thomsen, A, y Bertranpetit, J (2000). Genetic structure of North-West Africa revealed by STR analysis. *Eur J Hum Genet* 8:360-366.
- Bosch, E, Calafell, F, Pérez-Lezaun, A, Comas, D, Mateu, E, y Bertranpetit, J (1997). Population history of North Africa: evidence from classical genetic markers. *Hum Biol* 69:295-311.

- Bosch, E, Calafell, F, Santos, FR, Pérez-Lezaun, A, Comas, D, Benchemsi, N, Tyler-Smith, C, y Bertranpetit, J (1999). Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *Am J Hum Genet* 65:1623-1638.
- Brakez, Z, Bosch, E, Izaabel, H, Akhayat, O, Comas, D, Bertranpetit, J, y Calafell, F (2001). Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol* 28:295-307.
- Bravo, AM, y de las Casas, MT (1958). Distribución de grupos sanguíneos en los habitantes de la isla de La Palma. *Acta Médica de Tenerife* 10:114-117.
- Brinkmann, B, Klitschar, M, Neuhuber, F, Hühne, J, y Rolf, B (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 62:1408-1415.
- Cabrera, VM, González, P, y Salo, WL (1996). Human enzyme polymorphism on the Canary Islands. VII G6PD Seattle in Canarians and North African Berbers. *Hum Hered* 46:197-200.
- Camps-Fabrer, H (1989). Capsien et Natoufien au Proche-Orient. Colloque "L'homme maghrébin et son environnement depuis 100.000 ans" Maghnia (Algérie) et Trav. du LAPMO:71-104.
- Cann, RL, Stoneking, M, y Wilson, AC (1987). Mitochondrial DNA and human evolution. *Nature* 325:31-36.
- Capelli, C, Wilson, JF, Richards, M, Stumpf, MPH, Gratrix, F, Oppenheimer, S, Underhill, P, Pascali, VL, Ko, T-M, y Goldstein, DB (2001). A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am J Hum Genet*. 68:432-443.

- Casanova, M, Leroy, P, Boucekkine, C, Weissenbach, J, Bishop, C, Fellous, M, Purrello, M, Fiori, G, y Siniscalco, MA (1985). Human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 230:1403-1406.
- Cavalli-Sforza, LL, y Edwards, AWF (1967). Phylogenetic analysis: Models and estimation procedures. *Am J Hum Genet* 19:233-257.
- Cavalli-Sforza, LL, Menozzi, P, y Piazza, A (1994). The history and geography of human genes. Princeton University Press, Princeton.
- Chejne, AG (1980). Historia de España musulmana. Ediciones Cátedra, Madrid.
- Chikhi, L, Destro-Bisol, G, Bertorelle, G, Pascali, V, y Barbujani, G (1998). Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc Natl Acad Sci USA* 95:9053-9058.
- Comas, D, Calafell, F, Benchemsi, N, Helal, A, Lefranc, G, Stoneking, M, Batzer, MA, Bertranpetit, J, y Sajantila, A (2000). *Alu* insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* 107:312-319.
- Comas, D, Calafell, F, Mateu, E, Pérez-Lezaun, A, y Bertranpetit, J (1998). HLA evidence for the lack of genetic heterogeneity in Basques. *Ann Hum Genet* 62:123-132.
- Côrte-Real, HBSM, Macaulay, VA, Richards, MB, Hariti, G, Issad, MS, Cambon-Thomsen, A, Papiha, S, Bertranpetit, J, y Sykes, BC (1996). Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 60:331-350.
- Crouau-Roy, B, y Clisson, I (2000). Evolution of an *Alu* DNA element of type Sx in the lineage of primates and the origin of an associated tetranucleotide microsatellite. *Genome* 43:1-7.
- Cuscoy, LD (1963). Paleontología de las Islas Canarias. Publicaciones del Museo Arqueológico de Tenerife, Cabildo Insular de Tenerife.

- Delfiner, P (1976). Linear estimation of non-stationary spatial phenomena. En: *Advanced Geostatistics in the mining industry*, Guarasio, M, David, M y Haijbechts, C (eds). Reidel, Dordrecht, the Netherlands, pp. 49-68.
- Dupanloup, I, y Bertorelle, G (2001). Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol* 18:672-675.
- Efron, B (1982). *The Jackknife, the Bootstrap and other resampling plans*. Society for Industrial and Applied Mathematics, Philadelphia.
- Esteban, E, Dugoujon, JM, Valveny, N, González-Reimers, E, y Moral, P (1998). Spanish and African contribution to the genetic pool of the Canary islanders: data from GM and KM haplotypes and RFLPs in the immunoglobulin IGHG loci. *Ann Hum Genet* 62:33-45.
- Excoffier, L, Smouse, PE, y Quattro, J (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Falkenburger, F (1939-40). Essai d'une nouvelle classification craniologique des anciens habitants des Iles Canaries. *Anthropologie* 49:332-62, 523-41.
- Felsenstein, J (1993). *Phylogeny inference package (PHYLIP)*, version 3.5. University of Washington, Seattle.
- Flores, C, Hernández, M, González, AM, y Cabrera, VM (2000a). Genetic affinities among human populations inhabiting the sub-Saharan area, Northwest Africa and the Iberian Peninsula. En: *Prehistoric Iberia: Genetics, Anthropology and Linguistics*, Arnaiz-Villena, A. (ed). Kluwer Academic / Plenum Press, New York, pp. 33-50.
- Flores, C, Larruga, JM, González, AM, Hernández, M, Pinto, FM, y Cabrera, VM (2001a). The origin of the Canary Island aborigines and their contribution to the modern population: A molecular genetics perspective. *Curr Anthropol* 42:749-754.

- Flores, C, Maca-Meyer, N, González, AM, y Cabrera, VM (2000b). Northwest African distribution of the CD4/Alu microsatellite haplotypes. *Ann Hum Genet* 64:321-327.
- Flores, C, Maca-Meyer, N, Pérez, JA, y Cabrera, VM (2001b). The peopling of the Canary Islands: a CD4/Alu microsatellite haplotype perspective. *Hum Immunol* 62: 949-953.
- Flores, C, Maca-Meyer, N, Pérez, JA, Hernández, M, y Cabrera, VM (2001c). Y-Chromosome differentiation in Northwest Africa. *Hum Biol* 73:513-524.
- Forster, P, Röhl, A, Lünemann, P, Brinkmann, C, Zerjal, T, Tyler-Smith, C, y Brinkmann, BA (2000). Short tandem repeat-based phylogeny for the human Y chromosome. *Am J Hum Genet* 67:182-196.
- Fusté, M (1959). Contribution à l'Anthropologie de la Grande Canarie. *L'Anthropologie* 63:295-318.
- Gómez-Casado, E, del Moral, P, Martínez-Laso, J, García-Gómez, A, Allende, L, Silvera-Redondo, C, Longas, J, González-Hevilla, M, Kandil, M, Zamora, J, y Arnaiz-Villena, A (2000). HLA genes in Arabic-speaking Moroccans: close relatedness to Berbers and Iberians. *Tissue Antigens* 55:239-249.
- González, R, y Tejera, A (1990). Los aborígenes canarios. Gran Canaria y Tenerife. Ediciones Istmo, Oviedo.
- Guasch, J, Florensa, G, Diaz de Yraola, G, Gavilanes, C, del Rio, R, Tabuena, J, y Raichs, A (1952). Los factores hemáticos en España, excepto en el País Vasco. *Medicina Clínica* 18:268-271.
- Hammer, MF, y Horai, S (1995). Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 56:951-962.

- Hammer, MF, Karafet, T, Rasanayagam, A, Wood, ET, Altheide, TK, Jenkins, T, Griffiths, RC, Templeton, AR, y Zegura, SL (1998). Out of Africa and back again: Nested Cladistic Analysis of human Y chromosome variation. *Mol Biol Evol* 15:427-441.
- Hammer, MF, Karafet, TM, Redd, AJ, Jarjanazi, H, Santachiara-Benerecetti, S, Soodyall, H, y Zegura, SL (2001). Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18:1189-1203.
- Hammer, MF, Redd, AJ, Wood, ET, Bonner, MR, Jarjanazi, H, Karafet, T, Santachiara-Benerecetti, S, Oppenheim, A, Jobling, MA, Jenkins, T, Ostrer, H, y Bonn -Tamir, B (2000). Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc Natl Acad Sci USA* 97:6769-6774.
- Hammer, MF, Spurdle, AB, Karafet, T, Bonner, MR, Wood, ET, Novelletto, A, Malaspina, P, Mitchell, RJ, Horai, S, Jenkins, T, y Zegura, SL (1997). The geographic distribution of human Y chromosome variation. *Genetics* 145:787-805.
- Hancock, JH (1999). Microsatellites and other simple sequences: genomic context and mutational mechanisms. En: *Microsatellites*, Goldstein, DB (ed). Oxford University Press, Oxford.
- Hern ndez, PH (1999). *Natura y cultura de las Islas Canarias*. Tafor publicaciones, La Laguna, Tenerife.
- Heyer, E, Puymirat, J, Dieltjes, P, Bakker, E, y de Knijff, P (1997). Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Mol Genet* 6:799-803.
- Hill, EW, Jobling, MA, y Bradley, DG (2000). Y-chromosome variation and Irish origins. *Nature* 404:351-352.

- Horai, S, Murayama, K, Hayasaka, K, Matsubayashi, S, Hattori, Y, Fucharoen, G, Harihara, S, Park, KS, Omoto, K, y Pan IH (1996). MtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. *Am J Hum Genet* 59:579-590.
- Hsieh, CL, y Sutton, HE (1992). Mitochondrial and nuclear variants in a U.S. Black population: origins of a hybrid population. *Ann Hum Genet* 56:105-112.
- Hurles, ME, Veitia, R, Arroyo, E, Armenteros, M, Bertranpetit, J, Pérez-Lezaun, A, Bosch, E, Shlumukova, M, Chambón-Thomsen, A, McElreavey, K, López de Munain, A, Röhl, A, Wilson, IJ, Singh, L, Pandya, A, Santos, FR, Tyler-Smith, C, y Jobling, MA (1999). Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am J Hum Genet* 65:1437-1448.
- Ingman, M, Kaessmann, H, Pääbo, S, y Gyllensten, U (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-713.
- Jin, L, Macaubas, C, Hallmayer, J, Kimura, A, y Mignot, E (1996). Mutation rate varies among alleles at a microsatellite locus: phylogenetic evidence. *Proc Natl Acad Sci USA* 93:15285-15288.
- Jin, L, Underhill, PA, Doctor, V, Davis, RW, Shen, P, Cavalli-Sforza, LL, y Oefner, PJ (1999). Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. *Proc Natl Acad Sci USA* 96:3796-3800.
- Jobling, MA, y Tyler-Smith, C (1995). Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11:449-456.
- Jobling, MA, y Tyler-Smith, C (2000). New uses for new haplotypes. *Trends Genet* 16:356-362.
- Jorde, LB, Watkins, WS, Bamshad, MJ, Dixon, ME, Ricker, CE, Seielstad, MT, y Batzer, MA (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 66:979-988.

- Karafet, TM, Zegura, SL, Posukh, O, Osipova, L, Bergen, A, Long, J, Goldman, D, Klitz, W, Harihara, S, de Knijff, P, Wiebe, V, Griffiths, RC, Templeton, AR, y Hammer, MF (1999). Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817-831.
- Kayser, M, Krawczak, M, Excoffier, L, Dieltjes, P, Corach, D, Pascali, V, Gehrig, C, Bernini, LF, Jespersen, J, Bakker, E, Roewer, L, y de Knijff, P (2001). An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* 68:990-1018.
- Ke, Y, Su, B, Song, X, Lu, D, Chen, L, Li, H, Qi, C, Marzuki, S, Deka, R, Underhill, P, Xiao, C, Shriver, M, Lell, J, Wallace, D, Wells, RS, Seielstad, M, Oefner, P, Zhu, D, Jin, J, Huang, W, Chakraborty, R, Chen, Z, y Jin, L (2001). African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* 292:1151-1153.
- Krings, M, Salem, AH, Bauer, K, Geisert, H, Malek, AK, Chaix, L, Simon, C, Welsby, D, Di Rienzo, A, Utermann, G, Sajantila, A, Pääbo, S, y Stoneking, M (1999). mtDNA analysis of Nile river valley populations: a genetic corridor or a barrier to migration? *Am J Hum Genet* 1166-1176.
- Kruskal, JB, y Wish, M (1978). *Multidimensional scaling*. Sage Publications, Beverly Hills, CA.
- Kwok, C, Tyler-Smith, C, Mendonca, BB, Hughes, I, Berkovitz, GD, Goodfellow, PN, y Hawkins, JR (1996). Mutation analysis of the 2 kb 5' to SRY in XY females and XY intersex subjects. *J Med Genet* 33:465-468.
- Labuda, D, Ziê tkiewicz, E, y Yotova, V (2000). Archaic lineages in the history of modern humans. *Genetics* 156:799-808.
- Lalueza, C (1997). mtDNA analysis in ancient Nubians supports the existence of gene flow between sub-Saharan and North Africa in the Nile valley. *Ann Hum Biol* 24:217-227.

- Larruga, JM, Díez, F, Pinto, FM, Flores, C, y González, AM (2001). Mitochondrial DNA characterisation of European isolates: The Maragatos from Spain. *Eur J Hum Genet* 9:708-716.
- Lobo-Cabrera, M (1993). La esclavitud en Fuerteventura en los Siglos XVI y XVII. V Jornadas de estudios sobre Fuerteventura y Lanzarote, I:13-40.
- Long, JC (1991). The genetic structure of admixed populations. *Genetics* 127:417-428.
- Lucotte, G, Aouizérate, A, y Berriche, S (2000). Y-chromosome DNA haplotypes in North African populations. *Hum Biol* 72:473-480.
- Lucotte, G, Gérard, N, y Mercier, G (2001). North African genes in Iberia studied by Y-chromosome DNA haplotype V. *Hum Immunol* 62:885-888.
- Lucotte, G, y Hazout, S (1996). Y-chromosome DNA haplotypes in Basques. *J Mol Evol* 42:472-475.
- Lucotte, G, y Loirat, F (1999). Y-chromosome DNA haplotype 15 in Europe. *Hum Biol* 71:433-439.
- Maca-Meyer, N, González, AM, Larruga, JM, Flores, C, y Cabrera, VM (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics* 2:13.
- Malaspina, P, Cruciani, F, Santolamazza, P, Torroni, A, Pangrazio, A, Akar, N, Bakalli, V, Brdicka, R, Jaruzelska, J, Kozlov, A, Malyarchuk, B, Medí, SQ, Michalodimitrakis, E, Varesi, L, Memmi, MM, Vona, G, Villems, R, Parik, J, Romano, V, Stefan, M, Stenico, M, Terrenato, L, Novelletto, A, y Scozzari, R (2000). Patterns of male-specific Inter-population divergence in Europe, West Asia and North Africa. *Ann Hum Genet* 64:395-412.
- Malcorra, JJ, Balda, MI, y Díaz, JM (1990). A survey of structural hemoglobinopathies in the Canary Islands. *Hemoglobin* 14:477-478.

- Mantel, N (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209-220.
- Martell, M, Padrón, G, Hernández, M, Afonso, JM, y Cabrera, VM (1986). Human enzyme polymorphism on the Canary Islands I. Gran Canaria Island population. *Hum Hered* 36:41-44.
- Martín, E (1992). La prehistoria de Canarias. Vol. 3. La Palma y los Auritas. Centro de la Cultura Popular Canaria, S/C de Tenerife.
- Martínez, M (1992). Historia popular de Canarias. Vol. 11. Canarias en la Mitología. Centro de la Cultura Popular Canaria, S/C de Tenerife.
- Martínez, VMF (1996). *Arqueología prehistórica de África*. Editorial Síntesis, Madrid.
- Mathias, N, Bayes, M, y Tyler-Smith, C (1994). Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 3:115-123.
- Meissier, E, Li, S-H, y Stewart, C-B (1996). The birth of microsatellite. *Nature* 381:483.
- Minch, E (1997). *Microsat 1.5d*. Department of Genetics, University of Stanford, Stanford.
- Moral, P, Esteban, E, Vives, S, Valveny, N, Toja, DI, y González, D (1997). Genetic study of the population of Tenerife (Canary Islands, Spain): protein markers and review of classical polymorphisms. *Am J Phys Anthropol* 102:337-349.
- Morilla, JM, Afonso, JM, Hernández, M, Pestano, JM, y Larruga, JM (1988). Human enzyme polymorphism on the Canary Islands II. African influence. *Hum Hered* 38:101-105.
- Nadir, E, Margalit, H, Gality, T, y Ben-Sasson, SA (1996). Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. *Proc Natl Acad Sci USA* 93:6470-6475.
- Navarro, JF (1992). La prehistoria de Canarias. Vol. 5. La Gomera y Los Gomereros. Centro de la Cultura Popular Canaria, S/C de Tenerife.

- Navarro, JF, Martín, E, y Rodríguez, A (1990). La primera etapa del programa de excavaciones en las cuevas de San Juan y su aportación a la diacronía de la prehistoria de La Palma. *Investigaciones arqueológicas en Canarias II*: 189-220.
- Nebel, A, Filon, D, Brinkmann, B, Majumder, PP, Faerman, M, y Oppenheim, A (2001). The Y chromosome pool of Jews as a part of the Genetic Landscape of the Middle East. *Am J Hum Genet* 69:1095-1112.
- Newman, JL (1995). *The peopling of Africa: A geographic interpretation*. Yale University Press, London.
- Onrubia-Pintado, J (1987). Les cultures préhistoriques des Îles Canaries, état de la question. *L'Anthropologie* 91:653-678.
- Page, RD (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12:357-358.
- Page, RD, y Holmes, EC (1998). *Molecular Evolution: A phylogenetic approach*. Blackwell Science, Oxford.
- Parejo, M (1966). El sistema ABO en la población actual de las Islas Canarias. *Actas del V Congreso Panafricano de prehistoria y estudios del Cuaternario II*:213-217.
- Passarino, G, Semino, O, Quintana-Murci, L, Excoffier, L, Hammer, M, y Santachiara-Benerecetti, AS (1998). Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 62:420-434.
- Passarino, G, Semino, O, Magri, C, Al-Zahery, N, Benuzzi, G, Quintana-Murci, L, Andellnovic, S, Bullc-Jakus, F, Liu, A, Arslan, A, y Santachiara-Benerecetti, AS (2001). The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from Northern regions of the Black Sea. *Hum Immunol* 62:922-932.
- Penny, D, Steel, M, Waddell, PJ, y Hendy, MD (1995). Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. *Mol Biol Evol* 12:863-882.

- Pereira, L, Prata, MJ, y Amorim, A (2000a). Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet* 64:491-506.
- Pereira, L, Prata, MJ, Jobling, MA, y Amorim, A (2000b). Analysis of the Y-chromosome and mitochondrial DNA pools in Portugal. En: *Archaeogenetics: DNA and the population prehistory of Europe*, Renfrew, C, y Boyle, K (eds). The McDonald Institute for Archaeological Research, Cambridge, pp. 191-195.
- Pérez, J (1991). Los portugueses en Canarias. *Portuguesismos*. Ediciones del Cabildo Insular de Gran Canaria, Las Palmas.
- Pérez, MJ, Hierro, JM, Marrero, MJ, González, AM, y Cabrera, VM (1991). Human enzyme polymorphism on the Canary Islands IV. Eastern Islands. *Hum Hered* 41:385-390.
- Piercy, R, Sullivan, KM, Benson, N, y Gill, P (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Leg Med* 106:85-90.
- Pinto, F, Cabrera, VM, González, AM, Larruga, JM, Noya, A, y Hernández, M (1994). Human enzyme polymorphism on the Canary Islands VI. Northwest African influence. *Hum Hered* 44:156-161.
- Pinto, F, González, AM, Hernández, M, Larruga, JM, y Cabrera, VM (1996a). Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* 60:321-330.
- Pinto, F, González, AM, Hernández, M, Larruga, JM, y Cabrera, VM (1996b). Sub-Saharan influence on the Canary Islands population deduced from G6PD gene sequence analysis. *Hum Biol* 68:517-522.
- Pinto, F, Rando, JC, López, M, Morilla, JM, y Larruga, JM (1996c). Blood group polymorphisms in the Canary Islands. *Gene Geography* 10:171-179.

- Poloni, ES, Semino, O, Passarino, G, Santachiara-Benerecetti, AS, Dupanloup, I, Langaney, A, y Excoffier, L (1997). Human genetic affinities for Y-chromosome p49a,f/*TaqI* haplotypes show strong correspondence with linguistics. *Am J Hum Genet* 61:1015-1035.
- Quintana-Murci, L, Krausz, C, Zerjal, T, Sayar, SH, Hammer, MF, Mehdi, SQ, Ayub, Q, Qamar, R, Mohyuddin, A, Radhakrishna, U, Jobling, MA, Tyler-Smith, C, y McElreavey, K (2001). Y-Chromosome lineages trace diffusion of people and languages in Southern Asia. *Am J Hum Genet* 68:537-542.
- Quintana-Murci, L, Semino, O, Minch, E, Passarino, G, Brega, A, y Santachiara-Benerecetti, AS (1999). Further characteristics of proto-European Y chromosomes. *Eur J Hum Genet* 7:603-608.
- Rando, JC (1998). Composición genética y posible origen de las poblaciones humanas Canarias deducidos del polimorfismo de su ADN mitocondrial. Tesis Doctoral, Universidad de La Laguna, Tenerife.
- Rando, JC, Cabrera, VM, Larruga, JM, Hernández, M, González, AM, Pinto, F, y Bandelt, HJ (1999). Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* 63:413-428.
- Rando, JC, Pinto, F, González, AM, Hernández, M, Larruga, JM, Cabrera, VM, y Bandelt, HJ (1998). Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann Hum Genet* 62:531-550.
- Raymond, M, y Rousset, F (1995). An exact test of population differentiation. *Evolution* 49:1280-1283.
- Renfrew, C (1990). *Arqueología y lenguaje: la cuestión de los orígenes indoeuropeos*. Editorial Crítica, Barcelona.

- Reynolds, J, Weir, BS, y Cockerham, CC (1983). Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767-779.
- Richards, M, Côté-Real, H, Forster, P, Macaulay, V, Wilkinson-Herbots, H, Demaine, A, Papiha, S, Hedges, R, Bandelt, HJ, y Sykes, B (1996). Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59:185-203.
- Richards, MB, Macaulay, V, Bandelt, HJ, y Sykes, BC (1998). Phylogeography of mitochondrial DNA in Western Europe. *Ann Hum Genet* 62:243-262.
- Richards, M, Macaulay, V, Hickey, E, Vega, E, Sykes, B, Guida, V, Rengo, C, Sellito, D, Cruciani, F, Kivisild, T, Villems, R, Thomas, M, Rychkov, S, Rychkov, O, Rychkov, Y, Gölgel, M, Dimitrov, D, Hill, E, Bradley, D, Romano, V, Calì, F, Vona, G, Demaine, A, Papiha, S, Triantaphyllidis, C, Stefanescu, G, Hatina, J, Belledi, M, Di Rienzo, A, Novelletto, A, Oppenheim, A, Nørby, S, Al-Zaheri, N, Santachiara-Benerecetti, S, Scozzari, R, Torroni, A, y Bandelt, HJ (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251-76.
- Roberts, DF, Evans, M, Ikin, EW, y Mourant, AE (1966). Blood groups and the affinities of the Canary Islanders. *Man* 1:512-525.
- Rocha, H, Flores, C, Campos, Y, Arenas, J, Vilarinho, L, Santorelli, FM, y Torroni, A (1999). About the “pathological” role of the mtDNA T3308C mutation... *Am J Hum Genet* 65:1457-1459.
- Rösing, IS (1967). AB0-blutgruppen und Rh-Faktoren auf Teneriffa unter besonderer Berücksichtigung des Vergleichs zwischen vorspanischer und heutiger Bevölkerung. *Homo* 18:96-104.
- Rosser, ZH, Zerjal, T, Hurles, ME, Adojaan, M, Alavantic, D, Amorim, A, Amos, W, Armenteros, M, Arroyo, E, Barbujani, G, Beckman, G, Beckman, L, Bertranpetit, J, Bosch,

- E, Bradley, DG, Brede, G, Cooper, G, Côrte-Real, HBSM, de Knijff, P, Decorte, R, Dubrova, YE, Evgrafov, O, Gilissen, A, Glisic, S, Gölge, M, Hill, EW, Jeziorowska, A, Kalaydjieva, L, Kayser, M, Kivisild, T, Kravchenko, SA, Krumina, A, Kuèinskas, V, Lavinha, J, Livshits, LA, Malaspina, P, Maria, S, McElreavey, K, Meitinger, TA, Mikelsaar, A-V, Mitchell, RJ, Nafa, K, Nicholson, J, Nørby, S, Pandya, A, Parik, J, Patsalis, PC, Pereira, L, Peterlin, B, Pielberg, G, Prata, MJ, Previderé, C, Roewer, L, Rootsi, S, Rubinsztein, DC, Saillard, J, Santos, FR, Stefanescu, G, Sykes, BC, Tolun, A, Villems, R, Tyler-Smith, C, y Jobling, MA (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67:1526-1543.
- Rudbeck, L, y Dissing, J (1998). Rapid, simple alkaline extraction of human genomic DNA from whole blood, buccal epithelial cells, semen and forensic stains for PCR. *BioTechniques* 25:588-590.
- Saitou, N, y Nei, M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
- Salas, A, Comas, D, Lareu, MV, Bertranpetit, J, y Carracedo, A (1998). MtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 6:365-375.
- Sánchez-Mazas, A (2001). African diversity from the HLA point of view: influence of genetic drift, geography, linguistics, and natural selection. *Hum Immunol* 62:937-948.
- Santachiara-Benerecetti, AS, y Semino, O (1996). Y-Chromosome polymorphisms and history of populations. *Cell Pharmacol* 3:199-204.
- Schneider, S, Roessli, D, y Excoffier, L (2000). Arlequin ver. 2000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

- Schwarzfischer, F, y Liebrich, K (1963). Serologische untersuchungen an prähistorische bevölkerungen insbesondere an altkanarischen Mumien. *Homo* 14:129-133.
- Schwidetzky, I (1956). Anthropologische Beobachtungen auf Teneriffe. *Homo* 7:142-52.
- Schwidetzky, I (1963). La población prehispánica de las Islas Canarias Publicaciones del Museo Arqueológico de Tenerife. Nº4. Cabildo Insular de Tenerife.
- Schwidetzky, I (1975). Investigaciones antropológicas en las Islas Canarias: estudio comparativo entre la población actual y la prehispánica. Tenerife: Publicaciones del Museo Arqueológico de Tenerife, Nº10. Cabildo Insular de Tenerife.
- Scozzari, R, Cruciani, F, Pangrazio, A, Santolamazza, P, Vona, G, Moral, P, Latini, V, Varesi, L, Memmi, MM, Romano, V, de Leo, G, Gennarelli, M, Jaruzelska, J, Villems, R, Parik, J, Macaulay, V, y Torroni, A (2001). Human Y-chromosome variation in the Western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 62:871-884.
- Scozzari, R, Cruciani, F, Santolamazza, P, Malaspina, P, Torroni, A, Sellito, D, Arredi, B, Destro-Bisol, G, de Stefano, G, Rickards, O, Martinez-Labarga, C, Modiano, D, Biondi, G, Moral, P, Olckers, A, Wallace, DC, y Novelletto, A (1999). Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet* 65:829-846.
- Seielstad, MT, Hebert, JM, Lin, AA, Underhill, PA, Ibrahim, M, Vollrath, D, y Cavalli-Sforza, LL (1994). Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum Mol Genet* 3:2159-2161.
- Seielstad, MT, Minch, E, y Cavalli-Sforza, LL (1998). Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278-280.

- Semino, O, Passarino, G, Brega, A, Fellous, M, y Santachiara-Benerecetti, AS (1996). A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 59:964-968.
- Semino, O, Passarino, G, Oefner, PJ, Lin, AA, Arbuzova, S, Beckman, LE, de Benedictis, G, Francalacci, P, Kouvatsi, A, Limborska, S, Marcikiae, M, Mika, A, Mika, B, Primorac, D, Santachiara-Benerecetti, AS, Cavalli-Sforza, LL, y Underhill, PA (2000). The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155-1159.
- Semino, O, Santachiara-Benerecetti, AS, Falaschi, F, Cavalli-Sforza, LL, y Underhill, PA (2002). Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* 70:265-268.
- Shen, P, Wang, F, Underhill, P, Franco, C, Yang, W-H, Roxas, A, Sung, R, Lin, AA, Hyman, RW, Vollrath, D, Davis, RW, Cavalli-Sforza, LL, y Oefner, P (2000). Population genetic implications from sequence variation in four Y chromosome genes. *Proc Natl Acad Sci USA* 97:7354-7359.
- Sokal, RR, y Menozzi, P (1982). Spatial autocorrelation of HLA frequencies in Europe supports demic difusión of early farmers. *Am Nat* 119:1-17.
- Spurdle, AB, y Jenkins, T (1996). The origins of the Lemba "Black Jews" of Southern Africa: evidence from p12f2 and other Y-chromosome markers. *Am J Hum Genet* 59:1126-1133.
- Stoneking, M, Fontius, JJ, Clifford, SL, Soodyall, H, Arcot, SS, Saha, N, Jenkins, T, Tahir, MA, Deninger, PL, y Batzer, MA (1997). Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 7:1061-1071.
- Stoneking, M, Jorde, LB, Bhatia, K, y Wilson, AC (1990). Geographic variation of human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717-733.

- Su, B, Xiao, C, Deka, R, Seielstad, MT, Kangwanpong, D, Xiao, J, Lu, D, Underhill, P, Cavalli-Sforza, LL, Chakraborty, R, y Jin, L (2000). Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet* 107:582-590.
- Su, B, Xiao, J, Underhill, P, Deka, R, Zhang, W, Akey, J, Huang, W, Shen, D, Lu, D, Luo, J, Chu, J, Tan, J, Shen, P, Davis, R, Cavalli-Sforza, LL, Chakraborty, R, Xiong, M, Du, R, Oefner, P, Chen, Z, y Jin, L (1999). Y-chromosome evidence for a Northward migration of modern humans into Eastern Asia during the last ice age. *Am J Hum Genet* 65:1718-1724.
- Suárez, JJ, Rodríguez, F, y Quintero, CL (1988). *Historia popular de Canarias. Vol. 2. Conquista y colonización.* Centro de la Cultura Popular Canaria, S/C de Tenerife.
- Sun, C, Skaletsky, H, Rozen, S, Gromoll, J, Nieschlag, E, Oates, R, y Page., DC (2000). Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum Mol Genet* 9:2291-2296.
- Thomas, MG, Parfitt, T, Weiss, DA, Skorecki, K, Wilson, JF, le Roux, M, Bradman, N, y Goldstein, DB (2000). Y chromosomes traveling South: the Cohen modal haplotype and the origins of the Lemba-the “Black Jews of Southern Africa”. *Am J Hum Genet* 66:674-686.
- Thomas, MG, Skorecki, K, Ben-Ami, H, Parfitt, T, Bradman, N, y Goldstein, DB (1998). Origins of Old Testament priests. *Nature* 394:138-140.
- Thomson, R, Pritchard, JK, Shen, PD, Oefner, PJ, y Feldman, MW (2000). Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci USA* 97:7360-7365.
- Tishkoff, SA, Dietzsch, E, Speed, W, Pakstis, AJ, Kidd, JR, Cheung, K, Bonn -Tamir, B, Santachiara-Benerecetti, AS, Moral, P, Krings, M, P  bo, S, Watson, E, Risch, N, Jenkins,

- T, y Kidd, KK (1996). Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 271:1380-1387.
- Torrioni, A, Bandelt, HJ, D'Urbano, L, Lahermo, P, Moral, P, Sellitto, D, Rengo, C, Forster, P, Savantaus, ML, Bonn -Tamir, B, y Scozzari, R (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137-1152.
- Underhill, PA, Jin, L, Lin, AA, Mehdi, SQ, Jenkins, T, Vollrath, D, Davis, RW, Cavalli-Sforza, LL, y Oefner, PJ (1997). Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996-1005.
- Underhill, PA, Passarino, G, Lin, AA, Marzuki, S, Oefner, PJ, Cavalli-Sforza, LL, y Chambers, GK (2001a). Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum Mutat* 17:271-280.
- Underhill, PA, Passarino, G, Lin, AA, Shen, P, Mirazon Lahr, M, Foley, RA, Oefner, PJ, y Cavalli-Sforza, LL (2001b). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43-62.
- Underhill, PA, Shen, P, Lin, AA, Jin, L, Passarino, G, Yang, WH, Kauffman, E, Bonn -Tamir, B, Bertranpetit, J, Francalacci, P, Ibrahim, M, Jenkins, T, Kidd, JR, Mehdi, SQ, Seielstad, MT, Wells, RS, Piazza, A, Davis, RW, Feldman, MW, Cavalli-Sforza, LL, y Oefner, PJ (2000). Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358-361.
- Verneau, R (1887). Rapport d'une mission scientifique dans l'Archipel Canarien. *Archives des Missions Sc. et Litt . 3^{er} s r. XVIII:567-817.*

- Vigilant, L, Pennington, R, Harpending, H, Kocher, TD, y Wilson, AC (1989). Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci USA* 86:9350-9354.
- Vigilant, L, Stoneking, M, Harpending, H, Hawkes, K, y Wilson AC (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253:1503-1507.
- Watkins, WS, Ricker, CE, Bamshad, MJ, Carroll, ML, Nguyen, SV, Batzer, MA, Harpending, HC, Rogers, AR, y Jorde, LB (2001). Patterns of ancestral human diversity: an analysis of *Alu*-insertion and restriction-site polymorphisms. *Am J Hum Genet* 68:738-752.
- Watson, E, Forster, P, Richards, M, y Bandelt, HJ (1997). Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691-704.
- Wells, RS, Yuldasheva, N, Ruzibakiev, R, Underhill, PA, Evseeva, I, Blue-Smith, J, Jin, L, Su, B, Pitchappan, R, Shanmugalakshmi, S, Balakrishnan, K, Read, M, Pearson, NM, Zerjal, T, Webster, MT, Zholoshvili, I, Jamarjashvili, E, Gambarov, S, Nikbin, B, Dostiev, A, Aknazarov, O, Zalloua, P, Tsoy, I, Kitaev, M, Mirrakhimov, M, Chariev, A, y Bodmer, WF (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244-9.
- White, PS, Tatum, OL, Deaven, LL, y Longmire, JL (1999). New, male-specific microsatellite markers from the human Y chromosome. *Genomics* 57:433-437.
- Whitfield, LS, Sulston, JE, y Goodfellow, PN (1995). Sequence variation of the human Y chromosome. *Nature* 378:379-380.
- Wilson, JF, Weiss, DA, Richards, M, Thomas, MG, Bradman, N, y Goldstein, DB (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci USA* 98:5078-5083.
- Wright, S (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395-420.

Zerjal, T, Dashnyam, B, Pandya, A, Kayser, M, Roewer, L, Santos, FR, Schiefenhövel, W, Fretwell, N, Jobling, MA, Harihara, S, Shimizu, K, Semjidmaa, D, Sajantila, A, Salo, P, Crawford, MH, Ginter, EK, Evgrafov, OV, y Tyler-Smith, C (1997). Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 60:1174-1183.

Zilhão, J (2001). Radiocarbon evidence for maritime pioneer colonization at the origins of farming in West Mediterranean Europe. *Proc Natl Acad Sci USA* 98:14180-14185.

APÉNDICE

Combinaciones alélicas de los STR, expresadas como número de repeticiones, detectadas en los haplotipos inspeccionados. Se indica la frecuencia absoluta en las tres áreas geográficas donde se estudiaron: PI, Península Ibérica; C, Canarias; NA, noroeste de África. El recuadro indica la combinación escogida como “linaje raíz” para los cálculos de coalescencia.

Haplotipo	STR					Poblaciones			Total
	<i>DYS391</i>	<i>DYS393</i>	<i>DYS434</i>	<i>DYS466</i>	<i>DYS467</i>	PI	C	NA	
III ₁	9	13	9	7	12	3	3	2	8
	9	13	9	7	13			1	1
	9	13	9	7	15			1	1
	10	13	8	7	13			2	2
	11	13	9	7	11			1	1
	11	13	9	7	13			1	1
	12	12	9	7	12			1	1
III ₃	8	14	10	7	14			1	1
	9	12	9	7	14			1	1
	9	13	8	7	14		1		1
	9	13	9	7	12	2	1	13	16
	9	13	9	7	13	7	12	15	34
	9	13	9	7	14	22	30	52	104
	9	13	9	7	15	1	5		6
	9	14	9	7	13		2	1	3
	9	14	9	7	14		3	1	4
	10	13	9	7	14	1			1
III ₄	9	13	9	7	13			1	1
	10	12	9	7	14		1		1
	10	13	8	7	12		2		2
	10	13	9	7	11	8	9		17

Haplotipo	STR					Poblaciones			Total
	<i>DYS391</i>	<i>DYS393</i>	<i>DYS434</i>	<i>DYS466</i>	<i>DYS467</i>	PI	C	NA	
	10	13	9	7	12	4	2		6
	10	13	9	7	13	1	3		4
	10	13	9	7	14			2	2
	10	13	9	7	15		1		1
	10	14	9	7	11	1	1		2
	10	14	9	7	13		1		1
	10	15	9	7	13		1		1
	11	12	9	7	14			1	1
	11	13	9	7	11		2		2
	11	13	9	7	12	2			2
III ₅									
	9	13	9	7	12		1		1
	10	12	9	7	13	1	1		2
	10	12	9	7	14	1	1		2
III ₆									
	9	13	9	7	14	1	2		3
	9	13	9	7	15	1			1
	9	14	9	7	13		1		1
	9	14	9	7	15	1			1
	9	15	9	6	13		1		1
	10	13	9	7	11	3			3
	10	13	9	7	13	1	2		3
	10	13	9	7	14	1	4		5
	11	13	9	7	14	1	1		2
	11	14	9	7	13		1		1
III ₇									
	10	13	9	7	12		1	1	2

Haplotipo	STR					Poblaciones			Total
	<i>DYS391</i>	<i>DYS393</i>	<i>DYS434</i>	<i>DYS466</i>	<i>DYS467</i>	PI	C	NA	
	10	13	9	7	13			2	2
	10	13	9	8	13			1	1
	10	13	9	7	14	1		2	3
	10	14	9	7	12		1		1
	10	14	9	7	13		1		1
	10	14	9	7	14			2	2
	10	15	8	7	14			1	1
	11	13	9	7	13		2		2
	11	13	9	7	14	1			1
	11	14	9	7	13		1		1
VI ₄	9	13	9	7	13		10		10
	10	13	9	7	11	1			1
	10	13	9	7	12	8	3		11
	10	13	9	7	13	7	3		10
	11	13	9	7	12	1	4		5
	11	13	9	7	13		4		4
VI ₅	9	12	9	7	13			1	1
	10	12	9	6	12		1		1
	10	12	9	6	13		1		1
	10	12	9	7	12	9	11		20
	10	12	9	7	13	4	4		8
	10	12	9	7	14		1		1
	10	12	9	7	15		1		1
	10	13	9	7	14		1		1
	10	13	10	7	12		1		1

Haplotipo	STR					Poblaciones			Total
	<i>DYS391</i>	<i>DYS393</i>	<i>DYS434</i>	<i>DYS466</i>	<i>DYS467</i>	PI	C	NA	
	10	14	9	7	12		2		2
	11	12	9	7	11		2	1	3
	11	12	9	7	12	1	4	5	10
	11	12	10	7	13		1		1
	12	12	9	7	11		1		1
	12	12	9	7	12			4	4
	12	12	9	7	13			2	2
VI ₆	9	12	9	7	11	1			1
	9	12	9	7	12	6	4		10
	9	12	9	7	13		3		3
	9	12	9	7	14		2	1	3
	9	13	9	7	13		2		2
	9	13	9	7	14		2		2
	9	13	9	7	15		1		1
	10	12	8	7	12		2		2
	10	12	8	7	13	1	1		2
	10	12	8	7	14	1	1		2
	10	12	9	7	11	1	1		2
	10	12	9	7	12	4	2		6
	10	12	9	7	13	6	4		10
	10	12	9	7	14	5	4		9
	10	12	9	7	15	2			2
	10	12	9	7	16		1		1
	10	12	10	7	12	1			1
	10	13	9	7	13		1		1
	10	13	9	7	14		2	1	3
	10	13	9	7	15		1		1
	11	12	8	7	14	1			1

Haplotipo	STR					Poblaciones			Total
	<i>DYS391</i>	<i>DYS393</i>	<i>DYS434</i>	<i>DYS466</i>	<i>DYS467</i>	PI	C	NA	
	11	12	9	7	11		1		1
	11	12	9	7	12	2			2
	11	12	9	7	13		1		1
	11	12	9	7	14	1	1		2
VI ₇									
	9	12	9	7	12	1			1
	9	13	9	7	12		2		2
	10	12	9	7	11		1		1
	10	12	9	7	12	4	6		10
	10	12	9	7	13	4	10	2	16
	10	12	9	7	14	1			1
	10	12	9	7	15	2			2
	10	13	9	7	12	1	1		2
	10	13	9	7	13		1		1
	11	12	9	7	12		2	1	3
	11	12	9	7	14	1			1
VIII ₃									
	10	12	9	7	12	1			1
	10	13	9	7	7		1	1	2
	10	13	8	7	11	1	2		3
	10	13	9	7	11	6	2		8
	10	13	9	7	12	3	6		9
	10	13	9	7	13	4			4
	10	13	9	7	14		1		1
	10	14	9	7	7		1		1
	10	14	9	7	11		6		6
	10	14	9	7	12	1			1