



Universitat Autònoma de Barcelona
Unitat d' Antropologia Biològica
Departament de Biologia Animal Biologia Vegetal i Ecologia

Estudio Genético y Biodemográfico del archipiélago de las Azores (Portugal)

Cristina Maria Pereira dos Santos

Tesis Doctoral
2005

Estudio Genético y Biodemográfico del archipiélago de las Azores (Portugal)

Memoria presentada por Cristina Maria Pereira dos Santos para optar al título de Doctor en Biología, programa de doctorado en Biología (opción Antropología) del Departamento de Biología Animal, Biología Vegetal y Ecología de la Universidad Autónoma de Barcelona, codirigida por los profesores

- Dra. M^a Pilar Aluja Paris, Profesora Titular de la Unidad de Antropología del Departamento de Biología Animal, Biología Vegetal y Ecología de la Universidad Autónoma de Barcelona

- Dr. Augusto Manuel Elias Abade, Professor Asociado del Departamento de Antropología de la Universidad de Coimbra

Dra. M^a Pilar Aluja Paris

Dr. Augusto Manuel Elias Abade

Cristina Maria Pereira dos Santos

A minha casa é concha. Como os bichos
Segreguei-a de mim com paciência:
Fechada de marés, a sonhos e a lixos,
O horto e os muros só areia e ausência.

Minha casa sou eu e os meus caprichos.
O orgulho carregado de inocência
Se às vezes dá uma varanda, vence-a
O sal que os santos esboroou nos nichos.

E telhadosa de vidro, e escadarias
Frágeis, cobertas de hera, oh bronze falso!
Lareira aberta pelo vento, as salas frias.

A minha casa... Mas é outra a história:
Sou eu ao vento e à chuva, aqui descalço,
Sentado numa pedra de memória.

“A concha” Vitorino Nemésio

...Claro está que en el fondo somos todos islas
y estamos solos y esa es una gran verdad...

“Viaje Vertical” Enrique Vila-Matas

À minha mãe, pelo exemplo de força e coragem
Ao meu pai, que onde quer que se encontre sempre esteve comigo

A Xavi, por el futuro...

Trabalho financiado pela Fundação para a Ciência e a Tecnologia,
Ministério da Ciência e do Ensino Superior, Portugal
Bolsa de Doutoramento referência: SFRH/BD/723/2000



AGRADECIMIENTOS / AGRADECIMENTOS

Seria imposible referir a todos cuantos me han ayudado y apoyado en la realización de la presente tesis... pero no puedo dejar de agradecer a los que más han contribuido a su concretización.

Seria impossível referir todos aqueles que me ajudaram e apoiaram na realização da presente tese... mas não poderia deixar de agradecer aos que mais contribuíram para a sua concretização.

A la Profesora Dra. Maria Pilar Aluja, de la Unidad de Antropología de la Universidad Autónoma de Barcelona, mi directora de tesis, primero por haber arriesgado en una desconocida y luego por toda su dedicación, apoyo y rigor científico, sin los cuales esta tesis habría sido imposible. Pero te quiero agradecer, Pilar, por tu amistad, por los “cafés” y porque has hecho todo para que me sintiera en casa.

Ao Professor Dr. Augusto Abade, do Departamento de Antropologia da Universidade de Coimbra, que além de ter acompanhado, discutido e incentivado desde inicio este trabalho, foi o iniciador da minha aventura pelos meandros da investigação... por tudo muito obrigado!

À Professora Dra. Manuela Lima, do Departamento de Biologia da Universidade dos Açores, pela sua amizade e pela ajuda incansável nas fases boas e más deste trabalho, sem a sua colaboração esta tese seria absolutamente impossível!

A Rafa por las intensas y productivas discusiones científicas que hemos tenido en estos últimos años, gracias por las enseñanzas y claro, por la amistad, que es lo más importante!

À equipa da secção de Antropologia da Universidade dos Açores pela ajuda incansável e por terem feito as minhas estadias nos Açores mais agradáveis. Um agradecimento especial à São pela amizade e colaboração, sem a sua ajuda durante o período de amostragem uma grande parte deste trabalho não teria sido possível.

A todos los compañeros, amigos, profesores, etc. que he conocido en la Unidad de Antropología, muchas gracias por todo! La verdad es que me habéis hecho sentir siempre muy bien! Un agradecimiento muy especial a Natalia, por haberme enseñado lo que es una pipeta! A Irene que sigue siendo, después de todos estos años, oficialmente, la única que me comprende... A Jordi, Luis y Eli por aguantarme y ayudarme en mis maratones laboratoriales, a Mar que un día se cambiará de sala... a Jessica por la buena disposición y por hacer del laboratorio un sitio más agradable y donde se puede entrar! A Javier el Super Secre, por su disposición de 100%. Un agradecimiento muy muy especial a Blanca por haber "sufrido" como yo estos últimos meses, eres una chica con mucho valor y ha sido un placer trabajar contigo. Al "trío" de paleo y sus Power Analysis y otras discusiones intelectualmente estimulantes.

À Anna Barceló del servicio de secuenciación de la UAB, muchas gracias por atender siempre a un pedido de SOS!

A todos os meus amigos (aos portugueses!) que sempre me apoiaram e entenderam os meus momentos de stress e algumas vezes prolongadas ausências... um beijo enorme para a Cláudia, Francisca, Migas, Campelo, Rui y Zé Miguel que apesar de tudo, sempre tiveram tempo para um jantarzinho entre um avião e o outro!

À a minha família, a todos, em especial aos meus avós pela referência de valores que sempre representaram para mim. À tia Otília, por sempre estar disponível para me "receber e enviar" e ajudar com aquelas malas de peso insuperável!

Aos meus irmãos, Sónia e Melguita, porque gosto muito de vocês dois :-). Um obrigado enorme por terem tentado minimizar as minhas ausências...

À minha mãe, porque uma boa parte do mérito de ter chegado aqui é dela... por tudo o que fez por mim durante toda a vida... obrigado por ter mudado a sua vida por nós, pelo exemplo de coragem e amor, e claro, por sempre me ter incentivado com "todas aquelas fotocópias"!

A Xavi, por todo lo que es imposible escribir... por tu paciencia, comprensión, etc, etc... pero principalmente por todas aquellas cosas que solo pertenecen a nosotros dos...

A todos os dadores de amostras, sem a colaboração dos quais este trabalho seria obviamente impossível. Um agradecimento especial aos “responsáveis” das famílias estudadas pelo empenho incansável que sempre demonstraram.

À Fundação para a Ciência e a Tecnologia que financiou este trabalho com uma bolsa de Doutoramento (ref. SFRH / BD / 723 / 2000).

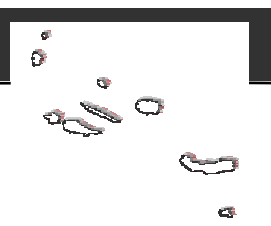
ÍNDICE

	Pág.
I. INTRODUCCIÓN	1
I.1. Evolución y variación genética	3
I.1.1. Fuerzas evolutivas	4
I.1.2. Estructura genética de las poblaciones humanas	5
I.1.2.1. Consecuencias genéticas de la consanguinidad y de la subdivisión poblacional	6
I.1.3. Variabilidad genética humana	7
I.2. El DNA mitocondrial humano	9
I.2.1. Organización	11
I.2.2. Características del DNA mitocondrial humano	13
I.2.3. Tasa de evolución	17
I.2.4. Genealogía y filogeografía del DNA mitocondrial	18
I.2.4.1. Genealogía del DNAMt	18
I.2.4.2. Filogeografía del DNAMt	21
I.2.4.2.1. Los linajes de DNAMt en África	26
I.2.4.2.1. Los linajes de DNAMt fuera de África	27
I.2.4. Aplicaciones del DNAMt humano	31
I.3. La población objeto de estudio: El archipiélago de las Azores (Portugal)	33
I.3.1. Localización geográfica	33
I.3.2. Historia de las Azores	33
I.3.2.1. Descubrimiento y poblamiento	33
I.3.3. Evolución de la población y emigración	37
I.3.4. Antecedentes biodemográficos y genéticos	39
II. JUSTIFICACIÓN Y OBJETIVOS DEL TRABAJO	41
II.1. Objetivos	45
II.1.1. Objetivos generales de la tesis	45
II.1.2. Objetivos de los artículos I y II	45
II.1.3. Objetivos de los artículos III y IV	46
II.1.4. Objetivos del artículo V	47

III. MATERIAL Y MÉTODOS	49
IV. ARTÍCULOS	53
IV.1. Artículo I:	55
Santos, C. , Montiel, R., Angles, N., Lima, M., Francalacci, P., Malgosa, A., Abade, A. y Aluja, M ^a . P. (2004) Determination of Human Caucasian Mitochondrial DNA Haplogroups by Means of a Hierarchical Approach. <i>Human Biology</i> , 76(3): 431-453.	
IV.2. Artículo II:	81
Santos, C. , Lima, M., Montiel, R., Angles, N., Pires, L., Abade, A. y Aluja, M ^a . P. (2003) Genetic structure and origin of peopling in the Azores Islands (Portugal): The view from mtDNA. <i>Annals of Human Genetics</i> , 67: 433-456.	
IV.3. Artículo III:	107
Santos, C. , Lima, M., Abade, A. y Aluja, M ^a P. (2004) Analyse du mouvement marital à l'île de Flores (Azores, Portugal). <i>Antropo</i> , 7: 63-71.	
IV.4. Artículo IV:	119
Santos, C. , Abade, A., Cantons, J., Mayer, F. M., Aluja, M ^a P. y Lima, M. Genetic Structure of Flores Island (Azores, Portugal) in the 19th Century and in the present-day: Evidences from surname analysis. <i>Human Biology (in press)</i> .	
IV.5. Artículo V:	165
Santos, C. , Montiel, R., Sierra, B., Bettencourt, C., Fernandez, E., Alvarez, L., Lima, M., Abade, A. y Aluja, M ^a .P. Understanding differences between phylogenetic and pedigree derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). <i>Molecular Biology and Evolution (in press)</i> .	
V. DISCUSIÓN GENERAL	233
VI. CONCLUSIONES	249
VII. BIBLIOGRAFÍA	255
VIII. APÉNDICE 1	281
VIII.1. Composición poblacional	281
VIII.1.1. Grupo Oriental	281
VIII.1.2. Grupo Central	286
VIII.1.3. Grupo Occidental	289
VIII.2. Índice de masculinidad	292
VIII.2.1. Índice de masculinidad por grupos de edad	292
VIII.2.2. Evolución temporal del índice de masculinidad	295

I.

INTRODUCCIÓN



I. 1. EVOLUCIÓN Y VARIACIÓN GENÉTICA

El concepto de evolución no es algo innato en el hombre, como lo demuestra el hecho de que muchas civilizaciones se hayan basado en conceptos fijistas para explicar el universo. Aunque algunas doctrinas intuitivas han desarrollado la idea evolutiva con sorprendente profundidad, como el caso del budismo, la evolución como teoría científica ha tenido que ser descubierta. La civilización judeocristiana, heredera en buena medida de muchas ideas aristotélico-platónicas, adoptó el fijismo como base conceptual durante más de dieciocho siglos (Fontdevila y Moya, 2003).

Durante el siglo XIX se produjo un cambio radical en la concepción del mundo con el surgimiento y auge del evolucionismo. Mucho antes de la publicación de *El origen de las especies*, de Charles Darwin, en 1859, las escuelas de pensamiento europeo comenzaban a plantearse que el cambio era algo consubstancial a todos los sistemas sociales y naturales. Pero es Darwin, con la teoría de la evolución a través de la selección natural, quien proporciona la prueba que faltaba, es decir, la teoría de cómo y por qué una forma se convierte en otra (Lewontin, 1984). Darwin suponía que la variabilidad seleccionada se heredaba, pero no conocía los mecanismos por los cuales se originaba y pasaba a las siguientes generaciones. La teoría de la herencia se debe a Mendel, que mediante sus famosos cruzamientos con guisantes formalizó las leyes de la herencia. Es Weissman, en 1892, quien relaciona la citología, primero, con la herencia y, después, con la evolución. Según Weissman, la nueva variabilidad sólo podía aparecer en las células germinales por un proceso independiente del ambiente, debido posiblemente a errores (a los que más tarde se llamaría mutaciones) en la duplicación celular (Fontdevila y Moya, 2003).

Darwin propuso, y la moderna biología evolutiva ha confirmado, que la evolución, desde el punto de vista biológico, se basa en la variación de las características de los organismos: diferencias entre individuos de una misma población, entre poblaciones y especies. Así, la piedra angular de la evolución es que exista variación heredable, es decir, variación genética. La evolución, es por tanto, y ante todo, un proceso genético.

La genética de poblaciones proporciona los principios teóricos necesarios para abordar el tema de la evolución del material hereditario de cualquier entidad biológica (Fontdevila y Moya, 1999). Parte del supuesto de que los cambios evolutivos a pequeña escala, o microevolutivos, que ocurren en las poblaciones contienen todos los elementos necesarios para explicar la evolución, ya que la macroevolución, o evolución a gran

escala, no sería más que la extrapolación en el espacio y en el tiempo de los procesos básicos de las poblaciones. Desde el punto de vista de la población, la evolución es, en último término, un cambio acumulativo e irreversible de las frecuencias de las diferentes variantes de los genes, o alelos, en las poblaciones (Relethford, 2001). Así pues, un problema básico de la genética de poblaciones es determinar cómo las frecuencias de los alelos mutantes van cambiando a lo largo del tiempo bajo el efecto de varias fuerzas evolutivas.

I. 1.1. FUERZAS EVOLUTIVAS

Clásicamente son consideradas cuatro las fuerzas evolutivas primarias que pueden producir cambios en la variación de las poblaciones a lo largo del tiempo: la mutación, la selección natural, la migración y la deriva genética (Grant, 1991). Sin embargo, en los eucariotas, una de las fuentes más importantes de variabilidad es la reproducción sexuada ya que esta promueve la formación de nuevas combinaciones de genes por recombinación (Hartl y Clark, 1997).

La función de la mutación es introducir *algo nuevo* en una población, siendo por eso considerada la fuente última de toda variación genética; sin mutación no existiría variación genética y, por lo tanto, la evolución no sería posible. Una nueva variante, surgida por mutación, inicialmente sólo se encuentra en un individuo de la especie. La actuación de las distintas fuerzas evolutivas, hace que, con el transcurso del tiempo, esta variante pueda pasar a tener una frecuencia del 0% (pérdida) o del 100% (fijación) en la especie estudiada, y, en algunos casos, puede permanecer en la población con una frecuencia intermedia (polimorfismo). Muchas de las nuevas mutaciones que surgen en una población reducen la viabilidad o *fitness* de sus portadores, de forma que serán seleccionadas negativamente y muy probablemente eliminadas de la población. Ocasionalmente, pueden surgir mutaciones que confieren ventaja a sus portadores y, en esta circunstancia, su frecuencia se podría incrementar por acción de la selección positiva. Esporádicamente pueden ocurrir mutaciones que presentan la misma viabilidad que el alelo con mayor *fitness* ya presente en la población; estas mutaciones son llamadas mutaciones selectivamente neutras y su destino no depende de la selección, pero sí de la deriva genética, es decir, de la fluctuación al azar de las frecuencias génicas de generación a generación, en una población de tamaño finito (Li, 1997). La deriva genética normalmente actúa eliminando las nuevas mutaciones, ya que estas presentan una frecuencia muy baja en la población en el momento en que se originan, lo que incrementa la probabilidad de pérdida en la nueva generación. Sin embargo, en algunos casos, la frecuencia de un nuevo alelo puede incrementarse al azar por efecto de la deriva, aunque el alelo sea desfavorable (Relethford, 2001).

La migración permite que un nuevo alelo sea transmitido a otra población. La *migración individual*, que se refiere al movimiento de individuos o pequeños grupos o *clusters* de individuos (ej. familias) de una región a otra, es muy frecuente y es normalmente dependiente de la distancia; cuanto más corta es la distancia mayor será la tasa de migración. En lo que respecta a la *migración masiva o colonización*, los fundadores, es decir, los primeros pobladores, son en general pocos, por lo que tendrán un efecto importante en la subsiguiente historia de la población, y lo mismo pasará en el caso de cualquier cuello de botella que pueda ocurrir posteriormente. Este fenómeno es particularmente evidente en la colonización de pequeñas islas, pero también se puede observar en poblaciones, conocidas como *aislados*, que no se mezclan con las poblaciones vecinas por razones sociales y/o geográficas (Cavalli-Sforza, 1973; Cavalli-Sforza *et al.*, 1994).

Las distintas fuerzas evolutivas en algunas ocasiones actúan de forma contrapuesta, mientras que en otros casos actúan en el mismo sentido siendo el destino de un nuevo alelo el resultado de la interacción de la selección natural, deriva genética y migración (Relethford, 2001).

La evolución de la diversidad en una población viene determinada principalmente por un equilibrio entre la tasa de mutación, el efectivo poblacional (que condiciona directamente la deriva genética) y la tasa de migración. Sin embargo, la selección natural puede afectar al nivel de diversidad genética en una población, dependiendo del tipo de selección y de las frecuencias alélicas iniciales. Si hay selección de un alelo respecto a otro, la diversidad disminuye, mientras que si hay selección a favor de los heterocigotos la diversidad puede aumentar. Las fuerzas evolutivas también condicionan la diferenciación entre poblaciones. La deriva genética tiene tendencia a incrementar la diferenciación entre poblaciones, mientras que la migración disminuye la diferenciación. La selección natural puede incrementar o disminuir las diferencias genéticas entre poblaciones dependiendo del tipo de selección y de que las poblaciones se encuentren o no en diferentes ambientes con adaptaciones óptimas distintas (Relethford, 2001).

I. 1.2. ESTRUCTURA GENÉTICA DE LAS POBLACIONES HUMANAS

Las poblaciones constituyen el escenario donde se desarrollan las relaciones entre los individuos, tales como la cooperación o la competición para la obtención de alimentos, o la elección de pareja para reproducirse. Gran parte de los modelos teóricos de la genética de poblaciones están basados en el supuesto de la panmixia de las poblaciones; sin embargo, la realidad de las poblaciones, principalmente de las poblaciones humanas, es bien diferente. Una población no constituye generalmente una unidad panmítica, sino que, por el contrario, se encuentra subdividida por barreras de diversa naturaleza

(geográfica, social, religiosa, lingüística, económica, etc.). Consecuentemente, el *pool* genético está subdividido y estratificado de una forma compleja (Jorde, 1980). Según algunos autores, las subdivisiones geográficas son unas de las más importantes, ya que la distancia es un factor fundamental en la determinación de la selección de pareja (Boyce *et al.*, 1971).

El concepto de estructura de una población, desde un punto de vista genético, se ha definido como el resultado de todos los mecanismos (excluyendo la selección y mutación) que afectan a las frecuencias génicas o genotípicas de una población (Cavalli-Sforza, 1959). Yasuda y Morton (1967) definen la estructura poblacional como la totalidad de las desviaciones a la panmixia o emparejamiento al azar. Así, la estructura poblacional incluye los efectos de la consanguinidad, de los matrimonios organizados, de la deriva y de la migración, y tiene en cuenta factores como el tamaño y distribución de las subpoblaciones (Lasker, 1954; Morton, 1969; Yasuda y Morton, 1967; Harrison y Boyce, 1972; Cannings y Cavalli-Sforza, 1973; Carmelli y Cavalli-Sforza, 1976). La estructura genética de las poblaciones humanas está pues condicionada directamente por la estructura demográfica, de emparejamiento y de parentesco de las mismas (Dyke, 1984), de tal forma que, para poder comprender las causas y los patrones de variación genética de una determinada población, no basta caracterizarla genéticamente, sino que se debe realizar su caracterización biodemográfica (Abade, 1992).

I. 1.2.1. Consecuencias genéticas de la consanguinidad y de la subdivisión poblacional

Las consecuencias genéticas de la ausencia de panmixia son importantes. En primer lugar, cambian las proporciones genotípicas generación tras generación, con lo que ya no se cumple la ley de Hardy-Weinberg. Las desviaciones respecto de la panmixia pueden ser de diferentes tipos, siendo la más frecuente los emparejamientos consanguíneos. Entre los efectos de la consanguinidad el más importante es la redistribución de las frecuencias alélicas de heterocigotos a homocigotos. Las frecuencias genotípicas son alteradas respecto a lo que sería de esperar en equilibrio pero, al contrario de la deriva genética, la consanguinidad por si sola no altera las frecuencias alélicas. En ambos casos (consanguinidad y deriva) existe una disminución de heterocigotos, aunque los procesos son distintos; en la consanguinidad la pérdida de heterocigotos es un proceso determinista intrapoblacional que conduce a un estado final con deficiencia de heterocigotos, y a una frecuencia alélica poblacional igual a la inicial. Sin embargo, en la deriva la pérdida de heterocigotos es un proceso estocástico interpoblacional, es decir, la pérdida paulatina de heterocigosis se da a nivel del conjunto de todas las subpoblaciones, de modo que para una de ellas en particular no se puede predecir cómo será la pérdida de heterocigosis con

el tiempo y, aunque al final todas se fijarán, no se puede predecir qué alelo se fijará en cada una en particular (Futuyma, 1998; Fontdevila y Moya, 1999).

La variabilidad genética total de una población tiene dos componentes: la variabilidad entre subpoblaciones y dentro de cada subpoblación. Esta partición de variabilidad debido a la estructura de la población es fundamental para establecer predicciones sobre la dinámica poblacional. Por esta razón, el desconocimiento de dicha estructura acarrea errores a la hora de aplicar directamente las reglas de la genética de poblaciones no subdivididas. Una de las consecuencias de la subdivisión poblacional es la reducción en la proporción de genotipos heterocigóticos respecto de lo que sería de esperar cuando el emparejamiento ocurre al azar. Esto puede hacer que, cuando se aplica una prueba de Hardy-Weinberg en una población para la cual se desconoce el grado de subdivisión, se encuentren desviaciones al equilibrio. Es posible distinguir las pérdidas de heterocigosis debidas a la subdivisión poblacional y a la consanguinidad. En el primer caso, en un locus multialélico, no todos los heterocigotos están en defecto, algunos pueden presentar un exceso. En el caso de la consanguinidad todos los heterocigotos están en defecto (Hartl y Clark, 1997).

I. 1.3. VARIABILIDAD GENÉTICA HUMANA

A pesar de las semejanzas que reconocemos cuando observamos a los humanos desde una perspectiva filosófica, la experiencia cotidiana demuestra una amplísima gama de diferencias, por ejemplo en la altura, peso, color de la piel, etc. Pero si hay diferencias claras entre individuos de una misma población, estas diferencias son aún más marcadas cuando se comparan individuos de poblaciones muy distintas. Esta variación observable, o variación fenotípica, resulta de miles de años de actuación de mecanismos de selección y adaptación de las poblaciones al medio ambiente en que están establecidas (Chiarelli y Rabino Massa, 2003). Aunque sean claras las diferencias entre los individuos y las poblaciones, los humanos son una especie relativamente homogénea. Se ha visto que la mayoría de la variación (~85%) es intrapoblacional y tan solo un 15% es variación interpoblacional (Lewontin, 1972; Nei y Roychoudhury, 1974; Barbujani *et al.*, 1997; Romualdi *et al.*, 2002).

La frecuencia de polimorfismo o variación, estimada para varios genes humanos (Sunyaev *et al.*, 2000) ha puesto en evidencia que algunos genes son muy polimorficos mientras que otros no presentan prácticamente variación en las poblaciones humanas (Cambien *et al.*, 1999; Cargill *et al.*, 1999; Halushka *et al.*, 1999). Datos recientes han demostrado la existencia de grandes alternancias en la densidad de polimorfismos a lo largo del genoma y se ha postulado la existencia de bloques haplotípicos estables, cada uno con su propia historia evolutiva no perturbada por la recombinación (Jeffreys *et al.*,

2001; Gabriel *et al.*, 2002; Reich *et al.*, 2002; Wang *et al.*, 2002; Kauppi *et al.*, 2003). Así, el estudio de diferentes regiones del genoma puede dar diferentes versiones de la historia de las poblaciones, dependiendo de factores demográficos, pero también de cómo las fuerzas evolutivas han modulado cada una de las zonas del genoma.

El estudio de la variabilidad genética humana, aplicado a las investigaciones antropológicas, remonta al inicio del siglo XX con el descubrimiento del sistema de grupos sanguíneos ABO por Landsteiner en 1901, y la posterior constatación por Hirszfled y Hirszfled (1919) de que las frecuencias de los distintos grupos sanguíneos eran diferentes entre poblaciones. La antropología genética, como la llaman algunos investigadores, ha progresado desde la utilización de grupos sanguíneos (Sanger y Race, 1949; Mourant, 1954; Race y Sanger, 1975; entre otros) y de la electroforesis de proteínas (Pauling *et al.*, 1949; Lewis *et al.*, 1958; Harris, 1966; Mourant *et al.*, 1976; entre otros), pasando a la incorporación de diversos sistemas genéticos como el DNA mitocondrial (DNAMt) (Brown, 1980; Cann *et al.*, 1987; Vigilant *et al.*, 1991; Torroni *et al.*, 1996; Richards *et al.*, 1998; Richards *et al.*, 2000; Ingman *et al.*, 2000; González *et al.*, 2003; Salas *et al.*, 2004; entre otros), el cromosoma Y (Ngo *et al.*, 1986; Hazout y Lucotte, 1986; Hammer y Horai, 1995; Underhill *et al.*, 1997; Underhill *et al.*, 2000; Jobling y Tyler-Smith, 2003; entre otros) los microsatélites (STRs) autosómicos (Di Rienzo *et al.*, 1994; Deka *et al.*, 1995; Rosenberg *et al.*, 2002; entre otros) y las posiciones polimórficas puntuales (SNP) (Stephens *et al.*, 2001; Gabriel *et al.*, 2003; Marth *et al.*, 2003; entre otros). La incorporación de los sistemas genéticos haploides, DNAMt y cromosoma Y en el estudio de las poblaciones humanas, ha revolucionado de forma decisiva la antropología genética. Los dos sistemas se complementan uno al otro, dando información separada, de la genealogía materna y paterna, respectivamente, sin la complicación de la recombinación. Ambos sistemas tienen un efectivo poblacional cuatro veces más bajo que los *loci* autosómicos, lo que hace que sean más sensibles al efecto de la deriva genética y, por lo tanto, más eficaces para detectar la existencia de subestructuración en las poblaciones humanas.

Además de los sistemas genéticos anteriormente mencionados, la variación genética de las poblaciones puede ser estudiada utilizando otro tipo de herramientas, siendo los apellidos una de las más utilizadas. En las sociedades occidentales, en las que la transmisión de apellidos es patrilineal, aunque estos sean caracteres culturales (Cavalli-Sforza y Feldman, 1981), tienen un comportamiento paralelo al de un gen neutro localizado en el cromosoma Y (Yasuda y Morton, 1967; Yasuda y Furusho, 1971; Yasuda *et al.*, 1974; Zei *et al.*, 1983; Zei *et al.*, 2003; entre otros). Esta relación teórica ha sido comprobada por estudios anteriores en que se relaciona el apellido de los individuos con el haplotipo del cromosoma Y (Thomas *et al.*, 1998; Hill *et al.*, 2000; Sykes y Irven, 2000). La utilización de apellidos en el estudio de la variación genética es una opción económica y particularmente atractiva para el estudio de las poblaciones del pasado (Relethford, 1988).

I.2. EI DNA MITOCONDRIAL HUMANO

Las mitocondrias son orgánulos citoplasmáticos que se encuentran presentes en la gran mayoría de las células eucariontes y que tienen una participación fundamental en el metabolismo celular. Además de producir ATP mediante la fosforilación oxidativa, se relacionan con procesos de muerte celular, de envejecimiento, y con diversos procesos patológicos. En general, las mitocondrias presentan una forma ovalada cuyo tamaño, estructura interna y número por célula varía en los diferentes tipos celulares. Están compuestas por dos membranas, una interna y una externa, que engloban una matriz densa que incluye múltiples enzimas del metabolismo intermediario (González-Halphen, 2003).

Las mitocondrias son semiautónomas ya que tienen su propio genoma y la maquinaria necesaria para la replicación, transcripción y síntesis proteica. Sin embargo, su genoma codifica solamente proteínas de la membrana interna, y algunos elementos necesarios para su traducción, como RNAs ribosómicos y de transferencia. Los otros cientos de proteínas que requiere la mitocondria para su funcionamiento se encuentran codificadas en el genoma nuclear y son sintetizadas en el citosol (Saccone *et al.*, 2000; González-Halphen, 2003).

Comúnmente se acepta que la mitocondria descende de eubacterias que en el inicio del proceso de simbiosis tenían su genoma completo. Esta teoría, conocida como *teoría endosimbiótica*, fue propuesta por Linn Margulis (Margulis, 1971) y, en su formulación original, postula el origen de las células eucarióticas en dos pasos: 1) origen del núcleo a partir de una arqueobacteria, y 2) inicio de un proceso de simbiosis con una eubacteria progenitora de las modernas mitocondrias. De acuerdo con esta teoría, los Archaeozoa (eucariotas sin mitocondrias) serían la base del árbol de los eucariotas. Recientemente, se ha propuesto una reformulación de la teoría y la nueva versión, llamada *hipótesis de hidrógeno* (Martin and Muller, 1998; Muller and Martin, 1999), postula la creación simultánea del núcleo eucariótico y de la mitocondria por un proceso de fusión entre una arqueobacteria dependiente de hidrógeno (hospedador) y una alfa-protobacteria productora de hidrógeno (simbionte).

A partir de un conjunto de 1000 genes o más que debieron encontrarse en el genoma del ancestro premitocondrial, la tendencia general de la naturaleza ha sido la de reducir el contenido génico de las mitocondrias. De esta manera, los genomas mitocondriales

actuales representan sólo una fracción de aquellos genes que estaban presentes en el endosimbionte. La mayor parte de esta pérdida de material genético debe haber sucedido durante el establecimiento de la endosimbiosis, y, en principio, puede ser el resultado de la eliminación de genes no esenciales o redundantes, cuya función pudo ser substituida por los genes nucleares del hospedador, y de la exportación de material genético desde la protomitocondria hacia el núcleo celular (Lang *et al.*, 1999; Kurland y Andersson, 2000).

En los últimos años, el estudio de genes que han cambiado su localización del genoma mitocondrial al nuclear ha permitido identificar pasos intermedios del fenómeno de migración de genes. Para que una migración tenga éxito, el gen mitocondrial relocalizado en el genoma nuclear debe activarse y convertirse en un gen funcional. Si por alguna razón no se llevan a cabo las modificaciones necesarias para activarlo, el gen puede desaparecer, o bien permanecer en el genoma nuclear como un pseudogen. En algunos grupos de organismos, la migración de genes al núcleo y su activación funcional parece haber cesado en algún momento de la evolución (González-Halphen, 2003). Tal es el caso de los vertebrados, los cuales muestran un contenido génico prácticamente constante (Boore, 1999). Una de las razones por las cuales la migración de genes mitocondriales al núcleo se ha detenido en los animales y no en las plantas, podría relacionarse con la diferencia que existe entre el código genético de las mitocondrias animales y el código genético universal que se emplea en el núcleo celular, ya que en las plantas no existe tal diferencia (Adams *et al.*, 2000). Es interesante destacar que existen copias de genes mitocondriales que se localizan actualmente en los genomas nucleares de los vertebrados en la forma de pseudogenes, y que podrían sugerir que han existido migraciones sin éxito, en las cuales los genes se integraron, pero nunca se activaron (o bien se activaron pero sus productos proteicos nunca pudieron importarse correctamente al interior de la mitocondria, por lo que eventualmente se inactivaron) (Wallace *et al.*, 1997).

¿Qué favorece la migración de genes mitocondriales al núcleo? La migración de genes al núcleo es el resultado de un proceso indispensable para el establecimiento de simbiosis metabólicas estables. Una de las hipótesis planteadas parte del hecho de que en las mitocondrias suceden reacciones de óxido-reducción que pueden incrementar la tasa de mutación génica, inducida por la presencia de radicales libres y especies reactivas de oxígeno, por lo que la información genética podría conservarse mejor en un ambiente como el nuclear (Allen, 2003). La otra explicación está justificada por un principio fundamental de la genética de poblaciones, el trinquete de Muller o *Muller ratchet*, que postula que en los genomas asexuales, como los genomas de los orgánulos, las mutaciones deletéreas y subletales se acumulan rápidamente, ya que no pueden ser eliminadas por recombinación. De acuerdo con este principio, la acumulación de mutaciones llevaría eventualmente a la extinción del sistema genético mitocondrial.

Así pues, ¿por qué razón el DNAm_t se mantiene en la mayoría de eucariotas y, además, conserva el mismo conjunto de genes en la mayoría de organismos vivos?

Existen varias hipótesis acerca de por qué en los genomas mitocondriales se conserva todavía un conjunto limitado de genes. Una posible explicación ha sido que en las mitocondrias de algunos organismos se utiliza un código genético diferente al que se emplea en el genoma nuclear. Esta heterogeneidad de lenguajes podría evitar la expresión de los genes mitocondriales relocalizados en el núcleo. Sin embargo, esto no explica por qué cierto grupo de genes permaneció en los genomas mitocondriales durante la migración masiva de genes hacia el núcleo, cuando el fenómeno de endosimbiosis apenas comenzaba y no existían diferencias en los códigos genéticos. Otra hipótesis que explica la permanencia de ciertos genes en el DNAm_t toma en consideración la hidrofobicidad de los productos proteicos correspondientes. Las proteínas codificadas en los DNAm_t son generalmente aquellas que se localizan en la membrana interna mitocondrial y que se caracterizan por ser altamente hidrofóbicas. Se ha propuesto que la síntesis en el citosol de este tipo de proteínas podría ocasionar que no se dirigieran al compartimiento celular adecuado (por ejemplo, que se insertaran en las membranas del retículo endoplásmico en lugar de hacerlo en las mitocondrias). Además, al tratarse de polipéptidos hidrófobos podría ocurrir que el transporte a través del sistema de membranas mitocondrial resultara imposible, o bien que los productos proteicos se agregaran irreversiblemente (von Heijne, 1986). También se ha planteado que en la mitocondria permanecen aquellos genes cuya expresión se regula por el estado redox de sus productos o de los portadores de electrones con los cuales interactúan. Esto permitiría a la célula disponer de una localización adecuada del gen que codifica proteínas cuya función demanda un control regulatorio rápido, directo e independiente de otros compartimentos y procesos celulares (Allen, 2003).

I. 2.1. ORGANIZACIÓN

El proceso evolutivo ha conducido a un escenario común en la mayoría de los organismos eucariotas: un genoma mitocondrial de tamaño variable, con información reducida y con un sistema genético que se parece en algunos aspectos a los procariontes y en otros a los eucariontes (Saccone, 1994).

La secuencia completa y la organización del DNA mitocondrial humano han sido descritas en 1981 por Anderson y colaboradores. La secuencia publicada por estos autores, revisada en 1999 por Andrews y colaboradores, ha sido utilizada como secuencia de referencia y es conocida como *Cambridge Reference Sequence (CRS)*. El genoma mitocondrial humano está formado por 16.569 pares de bases (pb), es circular, de doble cadena (la cadena pesada (H) rica en guaninas y la cadena ligera (L) rica en citosinas) y no tiene proteínas asociadas. Está organizado en dos regiones: la región codificante y la región no codificante (Fig. 1).

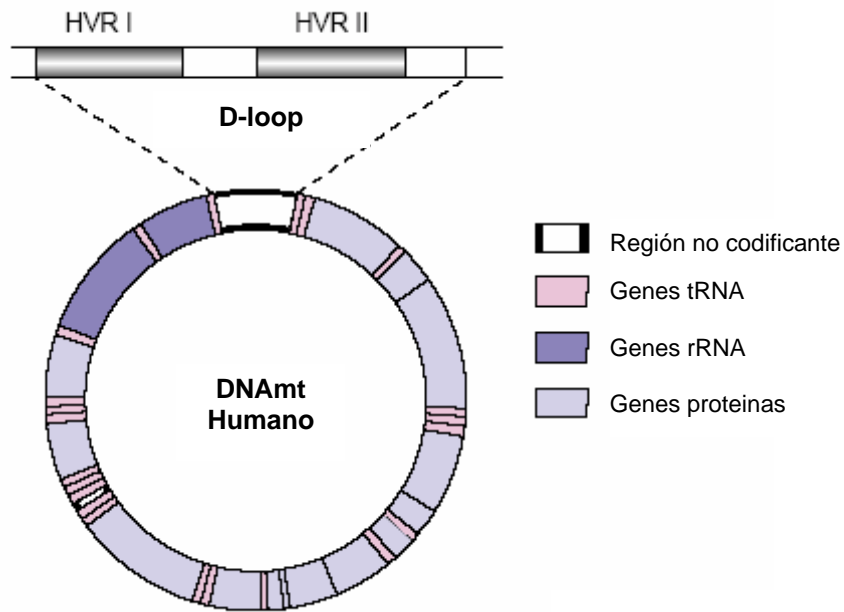


Fig. 1 - Diagrama del genoma mitocondrial Humano en que se puede ver la región codificante, donde se encuentran los genes, y la región no codificante o D-loop (adaptado de Hagelberg, 2003).

Región Codificante

La región codificante representa un 90% del genoma mitocondrial humano y carece de intrones; contiene 37 genes (Fig. 1), de los cuales: 13 codifican para proteínas involucradas en la fosforilación oxidativa, 2 codifican RNAs ribosómicos (12S y 16S rRNA), y 22 codifican RNAs de transferencia (Anderson *et al.*, 1981; Wallace, 1994, 1995).

Región no Codificante

La mayor región no codificante del DNAmT humano comprende 1.121 pb, siendo conocida como región de control o *D-loop* (Fig. 1), pues es en esta región donde se forma el rizo de desplazamiento durante la replicación del DNAmT (Anderson *et al.*, 1981; Wallace, 1994, 1995). La región de control contiene el origen de replicación de la cadena pesada (O_H), las secuencias conservadas relacionadas con el inicio de la replicación de la cadena H (Walberg y Clayton, 1981), las secuencias asociadas con la finalización del rizo de desplazamiento (Chang y Clayton, 1984), las secuencias de control de la transcripción (Cantatore y Attardi, 1980), y los promotores de transcripción (P_H y P_L) (Hixson y Clayton, 1985). Es también en la zona no codificante donde se encuentran dos regiones no funcionales muy variables, la región hipervariable I (HVRI) y la región hipervariable II (HVRII) (Anderson *et al.*, 1981).

I. 2.2. CARACTERÍSTICAS DEL DNA MITOCONDRIAL HUMANO

El DNAm_t humano representa apenas un 0.00006% del genoma nuclear humano, pero por sus características se ha convertido en una de las herramientas más potentes en los estudios de evolución humana y de la genética de poblaciones.

Elevado Número de Copias

El DNAm_t humano está organizado en nucleoides que contienen cada uno 2-8 moléculas de DNAm_t. La naturaleza, número y dinámica de los nucleoides en la mitocondria es prácticamente desconocida; Recientemente, en experimentos de fusión celular, se ha demostrado que estos son móviles y se difunden eficientemente en las mitocondrias (Legros *et al.*, 2004).

Las células somáticas tienen sólo dos copias de cada gen nuclear, sin embargo, pueden tener miles de copias de DNAm_t, ya que en el citoplasma de cada célula hay un elevado número de mitocondrias, y en cada mitocondria se pueden encontrar varias moléculas de DNAm_t (Robin y Wong, 1988). Esta propiedad hace que sea mucho más fácil obtener DNAm_t que DNA nuclear, y convierte el DNAm_t en una herramienta que se puede aplicar, con resultados positivos, en casos en que el material biológico disponible es escaso o está degradado, como sería el caso de los restos forenses y antiguos.

Herencia Materna

En 1980 el grupo de D. Wallace, analizando polimorfismos de sitios de restricción, demostró que la herencia del DNAm_t humano se realiza por vía materna (Giles *et al.*, 1980). Así, los linajes de DNAm_t son clonales o uniparentales y su historia evolutiva puede ser reconstruida sin la complejidad recurrente de la recombinación biparental (Howell, 1997, 1999), siendo las filogenias una representación de las genealogías maternas de los individuos estudiados (Stoneking, 1993). Este “paradigma standard”, como le ha llamado Elson *et al.*, (2001), ha sido aceptado durante mucho tiempo, pero en los últimos años se han establecido algunas dudas y se han generado una serie de discusiones y trabajos en torno a esta cuestión.

En 1988 Spuhler, haciendo referencia al artículo de Wilson y colaboradores (1985), afirma que la mayoría de los modelos de genética evolutiva del DNAm_t en mamíferos asumen que la herencia del mismo es exclusivamente materna. Aunque el espermatozoide de los vertebrados puede contener miles de mitocondrias en la cola, normalmente esta estructura no entra en el huevo fecundado, y si bien que algunas puedan entrar, no sobreviven a la primera división celular (Spuhler, 1988). A propósito de estas afirmaciones, Ankel-Simons y Cummins (1996) escriben: “...esto es claramente incorrecto. Además, el artículo de Wilson y colaboradores (1985) al cual Spuhler se refiere no da información concreta ni ninguna razón para explicar porque la herencia del DNAm_t debería ser exclusivamente materna.”. Los mismos autores advierten que, aunque este hecho haya sido ignorado por

algunas escuelas de antropólogos, el espermatozoide entero penetra completamente en el óvulo durante la fecundación, así que, hasta que no se sepa que ocurre con las mitocondrias del espermatozoide no se debería asumir la herencia estrictamente materna del DNAMt (Ankel-Simons y Cummins, 1996). Trabajos posteriores, en que se investigó la posibilidad de transmisión de DNAMt paterno en casos de fertilización *in vitro* en que ha sido utilizada la técnica de inyección intracitoplasmática de esperma (Torroni *et al.*, 1998a; Danan *et al.*, 1999), no mostraron evidencias de transmisión paterna del DNAMt.

En 2002, Schwartz y Vissing describen el primer (y hasta el momento, el único) caso de contribución paterna del DNAMt en un individuo con una miopatía mitocondrial. Posteriormente, varios investigadores han estudiado esta posibilidad en pacientes con enfermedades mitocondriales y no se ha encontrado ningún otro caso de contribución paterna (Filosto *et al.*, 2003; Taylor *et al.*, 2003; Schwartz y Vissing, 2004).

Con excepción de ciertas familias de bivalvos en que la transmisión del DNAMt se hace por vía paterna, en la mayoría de las especies animales el DNAMt es transmitido por vía materna; aunque pueden existir diferentes grados de contribución paterna, hecho que está relacionado con la forma de eliminación del DNAMt paterno (Birky, 1995, 2001). En los mamíferos parece que hay una destrucción selectiva de las mitocondrias paternas, después de su entrada en el citoplasma del huevo, por la acción proteolítica de la ubiquitina (Sutovsky *et al.*, 2000). La contribución de DNAMt paterno en mamíferos y concretamente en humanos puede ocurrir, pero muy probablemente con una frecuencia muy baja y sólo en aquellos casos en que el sistema de eliminación de mitocondrias paternas no funciona correctamente.

Homoplasmia vs Heteroplasmia

Los estudios iniciales de la variación del DNAMt indicaban que la mayoría de los individuos son homoplásmicos en lo que respecta al DNAMt, es decir, todas sus moléculas de DNAMt son iguales (Monnat y Loeb, 1985; Monnat y Reay, 1986). Hasta no hace muchos años, la heteroplasmia (la condición en que un mismo individuo puede tener más de un tipo de DNAMt) se consideraba bastante frecuente en individuos con enfermedades mitocondriales (Wallace, 1999), y su presencia en posiciones nucleotídicas neutras se continuaba considerando como muy infrecuente, ya que se habían detectado pocos ejemplos (e.g. Howell *et al.*, 1992; Gill *et al.*, 1994; Comas *et al.*, 1995; Ivanov *et al.*, 1996).

Bendall *et al.*, (1996) sugieren, por primera vez, que la heteroplasmia en la región de control es más frecuente de lo que se había supuesto inicialmente, y opinan que la falta de observación de heteroplasmia en estudios previos puede deberse a las dificultades para detectarla mediante secuenciación directa. Con la mejora de los métodos de secuenciación y la aplicación de técnicas más sensibles, como la electroforesis en gradiente desnaturizante (Tully *et al.*, 2000), se han confirmado las sospechas de Bendall *et al.*, (1996) de que la presencia de heteroplasmia no es una excepción, sino que

es un fenómeno bastante común (Tully *et al.*, 2000; Grzybowski *et al.*, 2003; Brandstätter *et al.*, 2004).

En diferentes estudios realizados, se han podido identificar dos tipos de heteroplasmia:

- 1) Heteroplasmia puntual, producida principalmente por sustitución tanto en la región codificante como en la región de control (Bendall *et al.*, 1996; Howell *et al.*, 1996; Parsons *et al.*, 1997; Tully *et al.*, 2000; Sigurðardóttir *et al.*, 2000; Grzybowski *et al.*, 2003; Howell *et al.*, 2003; Brandstätter *et al.*, 2004; entre otros);
- 2) Heteroplasmia de longitud, producida por inserciones y/o deleciones de una misma base en zonas repetitivas, siendo los casos más comunes los observados en dos regiones de policitosinas localizadas en la región HVRI (posiciones 16184-16193, numeración de acuerdo con la CRS) y HVRII (posiciones 303-315, numeración de acuerdo con la CRS) (Hauswirth y Clayton 1985; Bendall y Sykes, 1995; Parson *et al.*, 1998; Sigurðardóttir *et al.*, 2000; Lee *et al.*, 2004; entre otros);

Los estudios de familias en las cuales se ha detectado la presencia de heteroplasmia (Howell *et al.*, 1992; Bendall y Sykes, 1995; Bendall *et al.*, 1996; Howell *et al.*, 1996; Parsons *et al.*, 1997; Lutz *et al.*, 2000; Howell *et al.*, 2003; Brandstätter *et al.*, 2004), han permitido verificar que, aunque se hayan encontrado algunos casos en que los niveles de heteroplasmia son mantenidos más o menos constantes entre generaciones o se registran cambios poco significativos, en la mayoría de casos los niveles de heteroplasmia cambian bastante entre generaciones, existiendo una tendencia rápida hacia la fijación de una de las variantes presentes en el individuo heteroplásmico. Este resultado ha sido explicado considerando que existe una reducción muy grande del número de mitocondrias durante la oogénesis - *bottleneck* o cuello de botella.

Los datos obtenidos hasta el momento sugieren que ocurre un cuello de botella muy importante, una vez que el oocito ha madurado, y que hay muy poca segregación entre la fecundación y el nacimiento (Poulton *et al.*, 1998). Se ha estimado que un oocito contiene ~100.000 moléculas de DNAm_t y que el cuello de botella podría, en algunos casos, reducir a 1 el número de moléculas de DNAm_t. Se han hecho diferentes estimaciones del tamaño del cuello de botella (Howell *et al.*, 1992; Bendall *et al.*, 1996; Marchington *et al.*, 1998) empleando diferentes modelos de selección. Se ha visto que el modelo de selección única (Bendall *et al.*, 1996), que asume un solo momento de muestreo de un pequeño número de DNAm_ts de entre el gran número de moléculas de DNAm_t presentes en el oocito, parece ser más adecuado que el modelo de selecciones repetidas para predecir el tamaño del cuello de botella (Poulton *et al.*, 1998). Aplicando el modelo de selección única, Bendall y colaboradores (1996), utilizando pares de gemelos, han estimado un tamaño de cuello de botella de ~3-30 unidades de segregación. Marchington y colaboradores (1998),

utilizando oocitos han obtenido una estimación similar de 1-31 unidades de segregación, con una moda de 8 y una mediana de 9.

Ausencia de Recombinación

Otro de los postulados que se asume cuando se utiliza el DNAMt, es la ausencia de recombinación. Teóricamente el DNAMt tiene condiciones para poder recombinar; así pues, se ha observado que extractos de proteínas mitocondriales pueden catalizar la recombinación in vitro (Thyagarajan *et al.*, 1996; Lakshmipathy y Campbell, 1999a) y la DNA ligasa III, un enzima clave en la replicación, recombinación y reparación del DNA, está presente en las mitocondrias (Lakshmipathy y Campbell, 1999b).

Se han descrito tres mecanismos potenciales mediante los cuales el DNAMt puede recombinar (Morris y Lightowlers, 2000):

- 1) Recombinación entre diferentes moléculas de DNAMt en un mismo oocito heteroplásmico;
- 2) Recombinación entre el DNAMt y el DNA nuclear;
- 3) Recombinación entre el DNAMt materno y el paterno;

En 1999 se publicó el primer trabajo que muestra evidencias de que el DNAMt humano recombina (Eyre-Walker *et al.*, 1999). Aunque criticado porque existían errores que sesgaban los resultados de forma importante (Macaulay *et al.*, 1999a), trabajos posteriores han conducido a conclusiones semejantes (Awadalla *et al.*, 1999). Eyre-Walker *et al.*, (1999) y Awadalla *et al.*, (1999) sugieren que la elevada tasa de homoplasia observada en las filogenias de DNAMt no serían el resultado de los *hot-spots* mutacionales, pero sí de la recombinación. Posteriormente, varias investigaciones han refutado los resultados de Eyre-Walker *et al.*, (1999) y Awadalla *et al.*, (1999) (Merriwether y Kaestle, 1999; Jorde y Bamshad, 2000; Stoneking, 2000; entre otros) destacando dos trabajos en que se analizan secuencias completas de DNAMt (Ingman *et al.*, 2000; Elson *et al.*, 2001) en las que no se encuentran evidencias de recombinación en el DNAMt humano, apuntando a que la recombinación, si ha ocurrido, no habrá sido un acontecimiento suficientemente frecuente en la evolución humana para no asumir el "paradigma" de la ausencia de recombinación del DNAMt (Elson *et al.*, 2001).

Kraytsberg y colaboradores (2004) han publicado recientemente la evidencia física de recombinación entre el DNAMt materno y paterno en el tejido muscular de un individuo con miopatía (Schwartz y Vissing 2002). Este resultado alerta de la posibilidad de que la recombinación ocurra, siendo apenas posible su observación en individuos con diferentes tipos mitocondriales, ya sean de origen paterno o resultantes de nuevas mutaciones.

I. 2.3. TASA DE EVOLUCIÓN

Según Brown *et al* (1979) cuanto más importante es la función de un gen o proteína, más lentamente actúa el cambio evolutivo. Así pues, sería de esperar un cambio evolutivo lento en el DNAm de las mitocondrias animales, ya que estas son indispensables en la vida celular. Sin embargo, desde los primeros trabajos de Brown, se ha observado que el DNAm evoluciona más rápido que el DNA nuclear. El DNAm tiene una alta tasa de evolución, probablemente debido a la ausencia de un sistema de reparación (Bogehagen, 1999), a la ausencia de histonas protectoras, y al hecho de estar expuesto continuamente a los efectos mutagénicos de los radicales de oxígeno generados en la fosforilación oxidativa (Richter *et al*, 1988).

Es sabido que ocurren muchas más mutaciones de que las que se llegan a fijar, de forma que la tasa de evolución (o sustitución) está en función de la velocidad a que se originan nuevas mutaciones y de la probabilidad de que estas se lleguen a fijar (Kimura, 1983).

La tasa de evolución o sustitución del DNAm puede ser estimada utilizando diferentes métodos. Los primeros que se desarrollaron, y que más se han utilizado, son los llamados métodos filogenéticos (Brown *et al.*, 1979; Cann *et al.*, 1987; Stoneking *et al.*, 1986, 1992; Vigilant *et al.*, 1991; Pesole *et al.*, 1992; Tamura y Nei, 1993; Forster *et al.*, 1996; entre otros). Consisten en valorar la cantidad de diferencia que existe en el DNAm de una, dos o más especies relacionadas, en función del tiempo que ha tardado en producirse esta diferencia, infiriéndose el tiempo, normalmente, a través de evidencias paleontológicas y arqueológicas. Entre los muchos problemas que se apuntan a los métodos filogenéticos, uno de los mayores es la necesidad de utilizar datos externos para obtener una estimación temporal que puede llegar a ser muy imprecisa. Para evitar este problema, se han hecho estimaciones basadas en modelos teóricos, como el de la coalescencia (Lundstrom *et al.*, 1992), pero también en este caso hay que estimar el tamaño efectivo que ha tenido la población a largo plazo, una tarea, en la mayoría de las situaciones, muy complicada. Finalmente, se han utilizado métodos empíricos, que están basados en la observación directa de la mutación en familias y, que no dependen de datos externos (Howell *et al.*, 1996, 2003; Parsons *et al.*, 1997; Sigurðardóttir *et al.*, 2000; Heyer *et al.*, 2001; entre otros, para revisión ver Howell *et al.*, 2003).

Ahora bien, cuando hay que aplicar los métodos filogenéticos a la estimación de la tasa de sustitución del DNAm, existen una serie de problemas metodológicos, en su mayoría inherentes al propio DNAm, hasta ahora imposibles de solucionar. Entre ellos (Howell *et al.*, 2003):

- 1) La heterogeneidad en la tasa de evolución de las diferentes posiciones nucleotídicas del D-loop (Wakeley 1993; Excoffier y Yang, 1999; Meyer *et al.*, 1999; Heyer *et al.*, 2001; Meyer y van Haeseler, 2003);

- 2) La elevada tasa de homoplasia evidenciada en varios estudios, tanto de la región codificante como no codificante (i.e. Richards *et al.*, 1998; Howell *et al.*, 2004).
- 3) El no cumplimiento de las premisas de la evolución según el reloj molecular (Ingman *et al.*, 2000; Torroni *et al.*, 2001a; Howell *et al.*, 2004);
- 4) El no cumplimiento de los supuestos de neutralidad (Gerber *et al.*, 2001; Rand, 2001; Mishmar *et al.*, 2003; Nielsen y Yang, 2003; Elson *et al.*, 2004; Ruiz-Pesini *et al.*, 2004);
- 5) La dificultad de la mayoría de métodos para tener en cuenta los efectos de la dinámica poblacional, mestizaje y migración (Rannala y Bertorelle 2001);

Intuitivamente, los métodos empíricos permitirían la estimación de la tasa de sustitución sin la complicación resultante de todos los factores anteriormente mencionados. Esta ha sido la razón que ha motivado a Howell y colaboradores, en 1996, a estimar directamente la tasa de sustitución en familias. Utilizando esta metodología, los autores estimaron una tasa de sustitución de la región de control de 1.9 mutaciones/base/millón de años, un valor ~10 veces más alto que el obtenido en los estudios filogenéticos. Estudios posteriores en familias han confirmado esta discrepancia (revisión en Howell *et al.*, 2003), lo que ha promovido un intenso debate sobre las causas de esta diferencia (Pääbo, 1996; Parsons *et al.*, 1997; Macaulay *et al.*, 1997; Sigurðardóttir *et al.*, 2000; Heyer *et al.*, 2001; Howell *et al.*, 2003; Hagelberg, 2003). Actualmente está claro que esta discrepancia no puede ser causada por un único factor. La deriva genética, la selección, la posibilidad de recombinación, entre otros, son algunos factores que pueden estar involucrados. Además, cuando se trata de estimar la tasa de sustitución del DNAm hay que tener presente su organización jerárquica dado que los factores evolutivos pueden actuar a diferentes niveles de organización (la mitocondria, la célula, el individuo y la población), condicionando en cada uno de los diferentes niveles el destino de una nueva mutación (Rand, 2001).

Independientemente de las controversias sobre las formas de estimar la tasa de sustitución, los datos obtenidos hasta el momento indican que la tasa de sustitución del DNAm es bastante más alta que la del promedio del DNA nuclear, y la tasa de la región de control es más alta que la de la región codificante.

I. 2.4. GENEALOGÍA Y FILOGEOGRAFIA DEL DNA MITOCONDRIAL

I. 2.4.1. Genealogía del DNAm

La genealogía de un gen se puede obtener utilizando diferentes metodologías. En el caso concreto del DNAm, hasta mediados de la década de los 90, su genealogía se ha inferido utilizando métodos de máxima parsimonia (Cann *et al.*, 1987; Vigilant *et al.*, 1991; Penny *et al.*, 1995; entre otros), de máxima verosimilitud (Templeton, 1993; Hasegawa y

Horai, 1991; entre otros) y basados en distancias genéticas (Bertranpetit *et al.*, 1995; Zischler *et al.*, 1995; entre otros). Sin embargo, los métodos antes referenciados fueron desarrollados para determinar la relación entre especies y pueden no ser adecuados para analizar la variabilidad intraespecífica del DNAMt (Goldman y Barton, 1992; Bandelt *et al.*, 1995). Bandelt y colaboradores (1995) discuten las limitaciones de estas metodologías aplicadas al DNAMt y consideran que estos métodos no son los más apropiados para la reconstrucción filogenética, ya que el número de posiciones/caracteres utilizados es normalmente pequeño y, principalmente, porque es frecuente la presencia de homoplasia (mutaciones paralelas y retromutaciones).

En base a los argumentos anteriores, Bandelt y colaboradores (1995) proponen la utilización de un método que presenta todas las relaciones de un determinado grupo de datos en forma de una red o *network*, siendo así posible visualizar todos los árboles de máxima parsimonia que se pueden deducir. Esta aproximación identifica las substituciones conflictivas, mostrándolas como puntos reticulares, lo que facilita la interpretación en términos de homoplasia, recombinación, errores de secuenciación o superposición de secuencias. Sin embargo, Montiel (2001) utilizando secuencias de DNAMt para las cuales se conocía el haplogrupo, ha verificado que el análisis de secuencias de DNAMt de la región HVRI, utilizando un método basado en la distancia de Tamura y Nei (1993) y posterior construcción del árbol con el algoritmo de Neighbour-Joining (Saitou y Nei, 1987), es tan adecuado como el uso del método de las redes medianas. Según Montiel (2001), lo importante es elegir correctamente el modelo de distancia y tener en cuenta la variación de la tasa de substitución entre nucleótidos, es decir, aplicar la corrección gamma con una estimación adecuada del parámetro α .

Los pioneros en la utilización del DNAMt como marcador molecular fueron Wesley Brown y Douglas Wallace, a finales de la década de los 70. Los dos primeros trabajos se basaban en la digestión del DNAMt con enzimas de restricción. En uno de ellos se utilizó una única enzima en una muestra con muchos individuos (Denaro *et al.*, 1981) y en el otro se utilizaron varias enzimas y pocos individuos (Brown, 1980).

Los estudios subsecuentes se han basado en dos metodologías distintas que tenían por objetivo común diferenciar los tipos mitocondriales o haplotipos del DNAMt: 1) el análisis con enzimas de restricción de todo el genoma (análisis de RFLPs) y 2) la secuenciación. Los análisis con enzimas de restricción pueden ser de baja resolución utilizando 5 o 6 enzimas de restricción (Johnson *et al.*, 1983; Santachiara-Benerecetti *et al.*, 1988; Scozzari *et al.*, 1988), o de alta resolución utilizando 12 o 14 enzimas de restricción (Cann *et al.*, 1987; Torroni *et al.*, 1992). La secuenciación es una técnica de alta resolución, utilizada por primera vez en el estudio del DNAMt por Vigilant *et al.*, (1991).

Utilizando independientemente el análisis de RFLPs (Torroni *et al.*, 1994b; Chen *et al.*, 1995) y la secuenciación (Richards *et al.*, 1996; Watson *et al.*, 1997), se ha visto que en ambos casos los haplotipos formaban grupos de DNAs mitocondriales relacionados

filogenéticamente y específicos de una determinada región geográfica. A estos grupos se les ha llamado haplogrupos, cuando resultaban del análisis de RFLPs, y clusters, cuando resultaban del análisis de secuencias. Estos resultados, y el hecho de que la región codificante y la región de control se heredan juntas y por lo tanto están completamente ligadas, ha llevado a algunos investigadores a buscar una correlación entre los haplogrupos definidos por los polimorfismos de restricción y las sustituciones nucleotídicas en las secuencias del D-loop (Graven *et al.*, 1995; Torroni *et al.*, 1996). Se ha verificado que cuando se analizaban las secuencias del D-loop con el método de las redes medianas (Bandelt *et al.*, 1995), los clusters encontrados tanto en poblaciones europeas (Richards *et al.*, 1996) como africanas (Watson *et al.*, 1997), se correspondían con los haplogrupos determinados con enzimas de restricción (Torroni *et al.*, 1994b; Chen *et al.*, 1995). Esto ha conducido a la unificación de la nomenclatura (Richards *et al.*, 1998), adoptando el sistema de haplogrupos iniciado por Torroni *et al.*, (1992).

La región de control, como se ha visto anteriormente, tiene una alta tasa de mutación. Esta característica es muy útil cuando se analiza la variación intraespecífica de una especie tan reciente como es el *Homo sapiens*, pero, en contrapartida, la alta tasa de mutación también implica la presencia de mutaciones recurrentes y sustituciones paralelas, que pueden dificultar el proceso de reconstrucción filogenética. Así pues, una vez se ha reconocido que la región control y la región codificante evolucionan juntas, se han iniciado trabajos en los cuales se combina la secuenciación de la región de control con el análisis de RFLPs de la región codificante (de alta resolución o por PCR-RFLP) para definir mejor la filogenia del DNAm_t en diferentes regiones (Macaulay *et al.*, 1999b). La mayoría de trabajos de caracterización poblacional combinan la secuenciación de la región control con el análisis por PCR-RFLP de posiciones de la región codificante con relevancia en la determinación del haplogrupo (Pereira *et al.*, 2001; Malyarchuk *et al.*, 2002, 2003; González *et al.*, 2003; Stevanovitch *et al.*, 2004; Dubut *et al.*, 2004; Rosa *et al.*, 2004; entre otros).

En el 2001 se publicó el primer trabajo (Finnilä *et al.*, 2001) con secuencias completas de DNAm_t que tiene por objetivo directo aclarar la filogenia de los haplogrupos. Finnilä y colaboradores (2001) utilizan secuencias completas de 192 individuos fineses, población que tiene una variación de DNAm_t claramente europea, obteniendo dos *network*, una de la región de control y otra de la región codificante, que representan la genealogía del DNAm_t europeo. Estos *networks*, y en particular el de la región codificante, han permitido seleccionar nuevos marcadores para la diferenciación de haplogrupos y subhaplogrupos encontrados en población europea.

En el 2002, Herrnstadt y colaboradores (2002), utilizando 560 secuencias completas de la región codificante presentan tres redes de haplogrupos, una para los haplogrupos que anteriormente habían sido definidos como típicamente europeos, otra para haplogrupos africanos y otra para haplogrupos asiáticos/nativos americanos.

Los trabajos de Kong y colaboradores (2003), Tanaka y colaboradores (2004) y Palanichamy y colaboradores (2004), en los que se analizaron secuencias completas del este, oeste y sur de Asia, han permitido definir con detalle la filogenia del DNAMt en Asia. Estos trabajos, conjuntamente con el trabajo de Quintana-Murci y colaboradores (2004), en el que se ha analizado con técnicas de alta resolución el DNAMt de 23 poblaciones, en su mayoría del sudoeste asiático (desde el Cáucaso hasta la India), permiten, en estos momentos, establecer una filogenia robusta del DNAMt de Europa y Asia, teniendo en cuenta su continuidad geográfica.

Por lo que respecta al continente africano, Kivisild y colaboradores (2004) han publicado recientemente un trabajo, en el que, aunque no utilicen secuencias completas para inferir la filogenia del DNAMt, utilizan secuencias de la región de control y 184 posiciones informativas de la región codificante que han sido seleccionadas teniendo como base la información acumulada en trabajos anteriores (Ingman *et al.*, 2000; Torroni *et al.*, 2001a; Herrnstadt *et al.*, 2002; Maca-Meyer *et al.*, 2003), lo que les ha permitido obtener una filogenia con una elevada resolución.

Finalmente, cabe destacar el estudio detallado de la filogenia del haplogrupo X (Reidla *et al.*, 2003), del haplogrupo H (Loogväli *et al.*, 2004; Achilli *et al.*, 2004) y del haplogrupo U (Achilli *et al.*, 2005) por medio de secuencias de DNAMt completas o utilizando un gran número de posiciones informativas de la región codificante, seleccionadas teniendo como base las filogenias de secuencias completas publicadas anteriormente.

Combinando la información de los diferentes trabajos publicados recientemente, se ha esquematizado la filogenia del DNAMt en la figura 2. La clasificación actual de los linajes de DNAMt reconoce el haplogrupo L0 de origen Africano, como la ramificación más antigua del árbol de DNAMt (Mishmar *et al.*, 2003; Kivisild *et al.*, 2004). Además del haplogrupo L0, se han descrito como originarios en África los haplogrupos L1, L2, L3, L4, L5 y L6. Del haplogrupo L3 han derivado en el este africano el haplogrupo L4 y dos grandes grupos, M y N, que, con la salida de África del hombre moderno, han dado origen a los linajes de DNAMt presentes en Asia y Europa (Fig. 2).

I. 2.4.2. Filogeografía del DNAMt

Con el conjunto de los primeros trabajos para población nativa americana (Ward *et al.*, 1991; Horai *et al.*, 1993; Torroni *et al.*, 1993a, 1994a, d), asiática (Ballinger *et al.*, 1992; Torroni *et al.*, 1993b; Torroni *et al.*, 1994c), europea (Torroni *et al.*, 1994b) y africana (Chen *et al.*, 1995), se ha constatado que los haplogrupos de DNAMt son específicos de un determinado continente. Dicha constatación ha potenciado la utilización de la filogeografía (Awise *et al.*, 1987) para inferir la historia de las poblaciones humanas utilizando el DNAMt, siendo particularmente interesante su utilización en el estudio de los movimientos migratorios, tanto a corta como larga escala (como es el caso de la

colonización de zonas deshabitadas). Con la definición cada vez más clara de la filogenia del DNAMt, se puede, en estos momentos, saber con relativa precisión el origen y dispersión de los distintos haplogrupos del DNAMt.

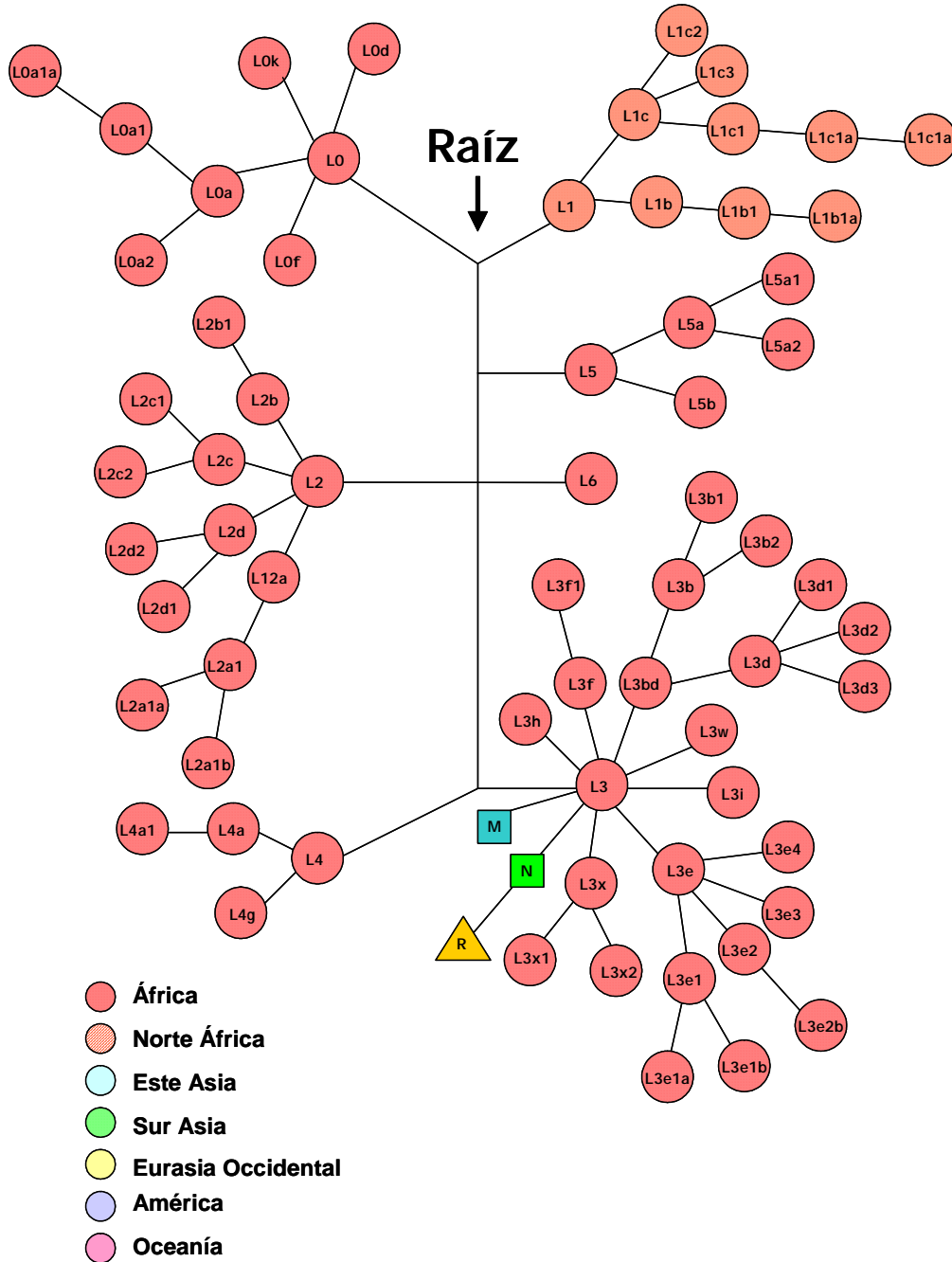
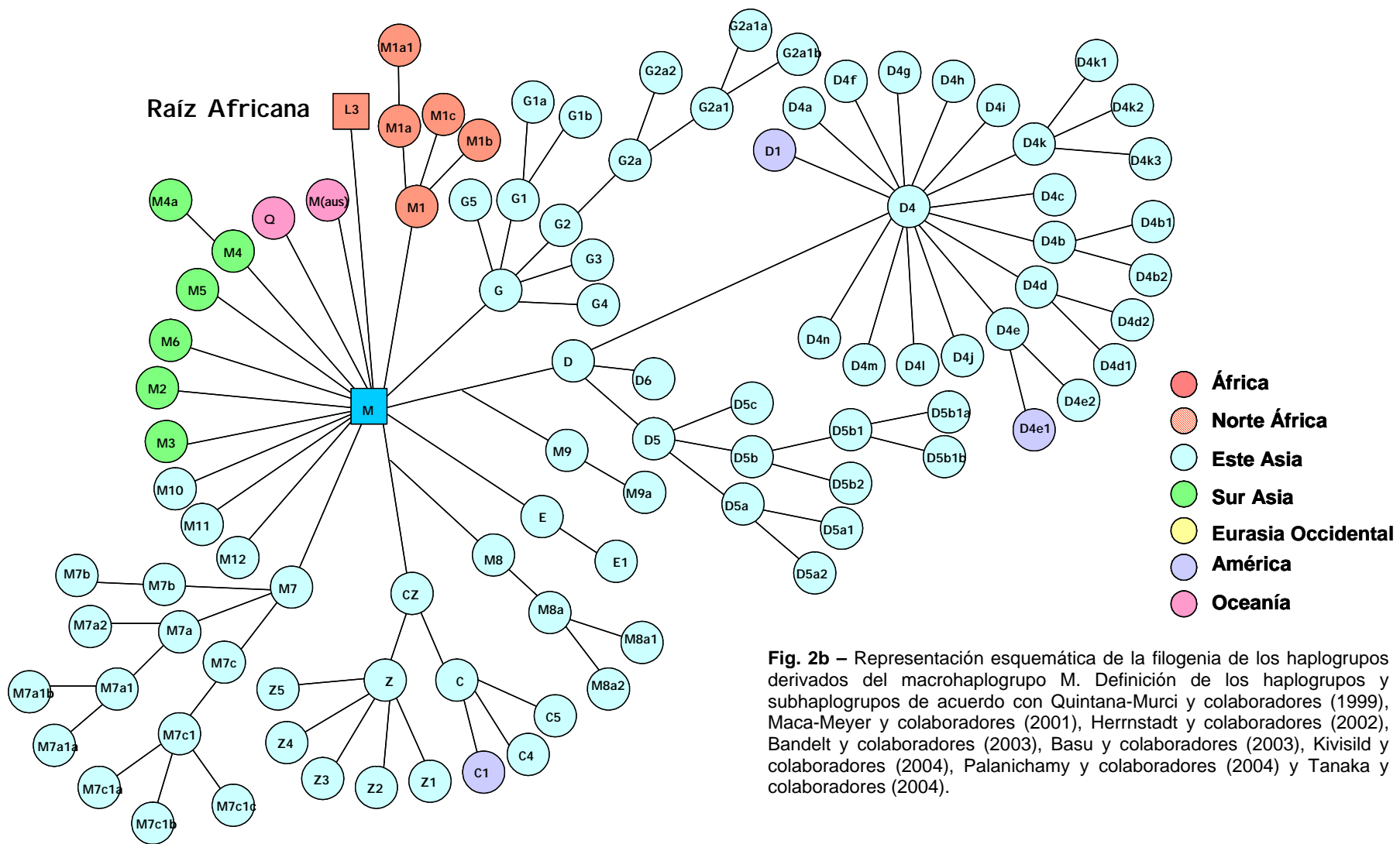


Fig. 2a – Representación esquemática de la filogenia de los haplogrupos L0, L1, L2, L3, L4, L5 y L6. Definición de los haplogrupos y subhaplogrupos de acuerdo con Torroni y colaboradores (2001a), Salas y colaboradores (2004), Shen y colaboradores (2004) y Kivisild y colaboradores (2004).



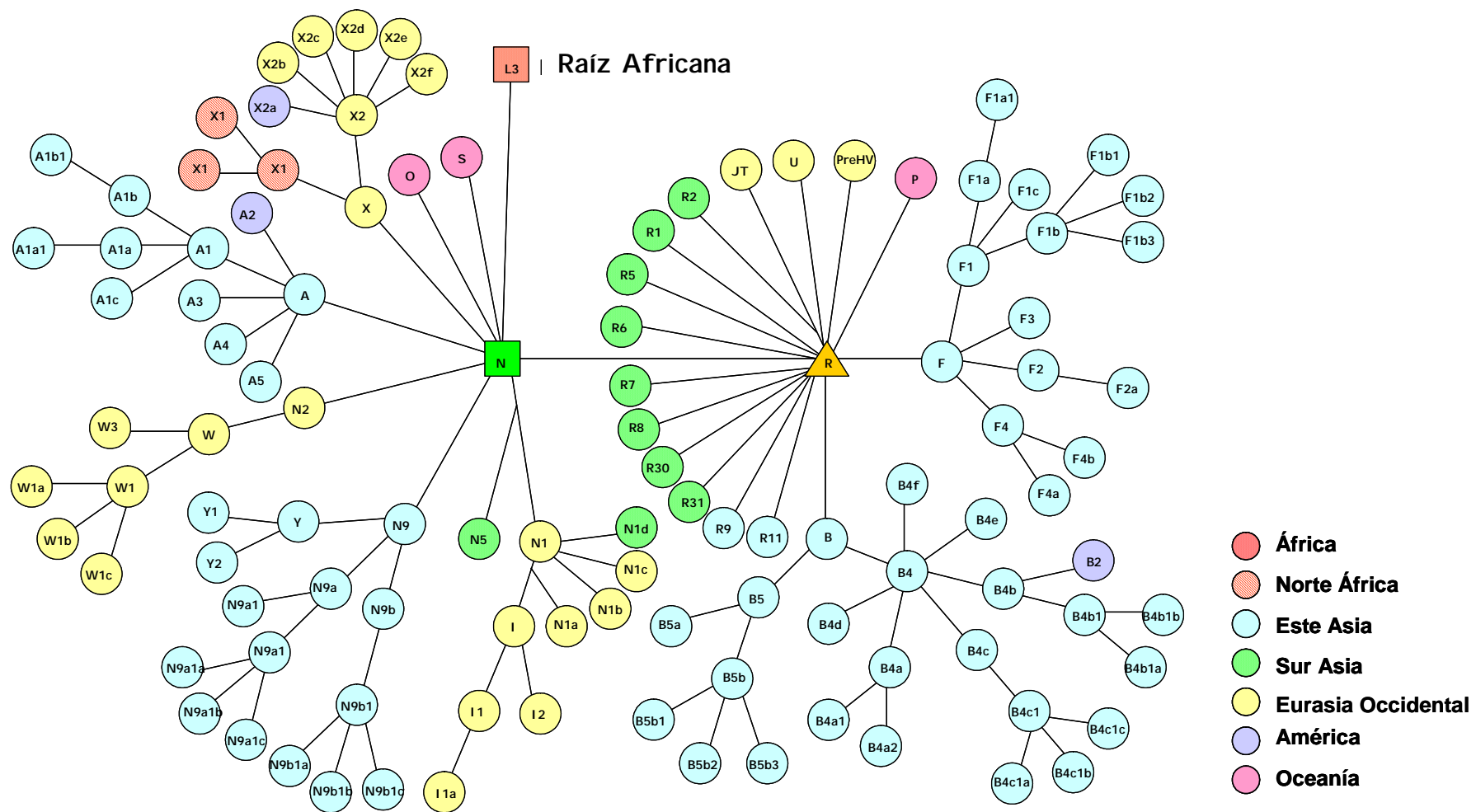


Fig. 2c – Representación esquemática de la filogenia de los haplogrupos derivados de los macrohaplogrupos N y R. Definición de los haplogrupos y subhaplogrupos de acuerdo con Forster y colaboradores (2001) Reidla y colaboradores (2003), Bandelt y colaboradores (2003), Basu y colaboradores (2003), Palanichamy y colaboradores (2004), Tanaka y colaboradores (2004) y Quintana-Murci y colaboradores (2004).

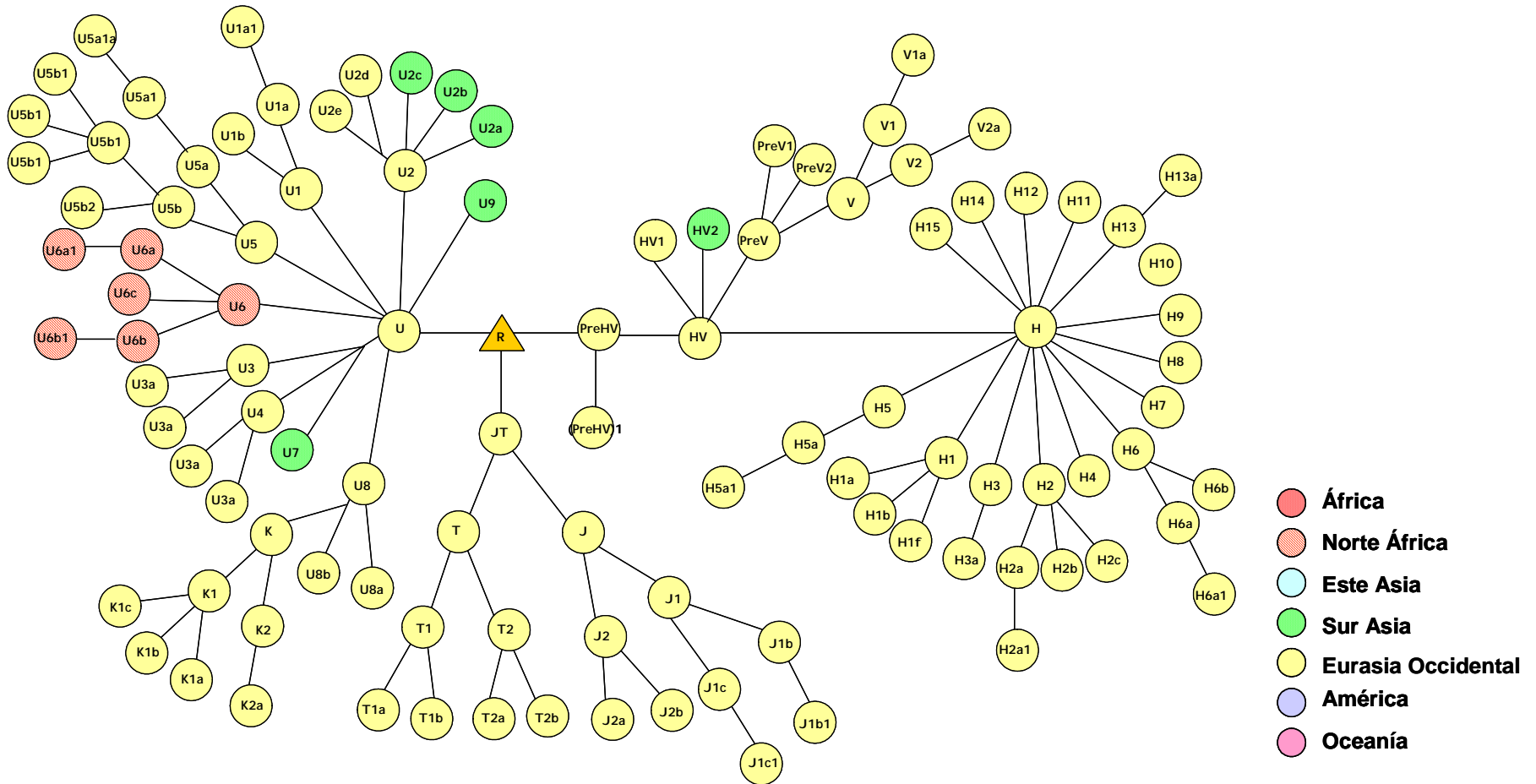


Fig. 2d – Representación esquemática de la filogenia de los subgrupos U, JT y PreHV del macrogrupo R. Definición de los haplogrupos y subhaplogrupos de acuerdo con Richards y colaboradores (2000), Maca-Meyer y colaboradores (2003), Achilli y colaboradores (2004), Basu y colaboradores (2004), Palanichamy y colaboradores (2004), Quintana-Murci y colaboradores (2004) y Achilli y colaboradores (2005).

I. 2.4.2.1. Los linajes de DNAMt en África

En África subsahariana, prácticamente la totalidad del DNAMt pertenece a los haplogrupos con origen en dicho continente L0, L1, L2, L3, L4, L5 y M1 (Fig. 2a, b) (Chen *et al.*, 1995, 2000; Watson *et al.*, 1997; Quintana-Murci *et al.*, 1999; Salas *et al.*, 2002; Rosa *et al.*, 2004; Shen *et al.*, 2004; Kivisild *et al.*, 2004).

El haplogrupo L0, con origen en el este africano (Salas *et al.*, 2002; Kivisild *et al.*, 2004), incluye 4 subhaplogrupos, L0a, L0d, L0f y L0k, que fueron previamente clasificados como subgrupos de L1 (Watson *et al.*, 1997; Salas *et al.*, 2002), de los cuales el L0a es el más frecuente. El haplogrupo L0 no tiene una distribución uniforme, presentando las frecuencias más altas en el sudeste africano (en Mozambique representa un 25% de los linajes de DNAMt) y está prácticamente ausente en el noroeste y sur del continente (Pereira *et al.*, 2001; Salas *et al.*, 2002; Salas *et al.*, 2004; Rosa *et al.*, 2004). Los subhaplogrupos L0d y L0k son específicos del grupo de poblaciones del sur de África que hablan khoisan (Chen *et al.*, 2000; Salas *et al.*, 2002).

El haplogrupo L1, con origen en África central, está subdividido en dos subhaplogrupos, L1b y L1c (Salas *et al.*, 2002; Kivisild *et al.*, 2004). El L1b está concentrado en la zona oeste del continente africano, con alguna dispersión hacia el norte y centro del continente. El L1c presenta las frecuencias más altas en África central y alguna dispersión hacia el oeste y sudeste africano.

El haplogrupo L2 está subdividido en cuatro subhaplogrupos, L2a, L2b, L2c y L2d (Chen *et al.*, 2000; Pereira *et al.*, 2001; Torroni *et al.*, 2001a). El L2a es el más frecuente y ampliamente distribuido de todos los subhaplogrupos Africanos, estando representado en todas las regiones de África con una frecuencia considerable. Los demás subhaplogrupos de L2 están confinados al oeste y centro-oeste del continente africano.

El haplogrupo L3, con origen en el este africano, presenta siete subdivisiones: L3b, L3d, L3e, L3f, L3h, L3i, L3w y L3x (Watson *et al.*, 1997; Salas *et al.*, 2002; Kivisild *et al.*, 2004). Está distribuido por todo el continente pero presenta las mayores frecuencias en el este africano, donde aproximadamente la mitad de los DNAMt son L3 (Salas *et al.*, 2002). El subhaplogrupo L3e es el más frecuente y el que presenta la mayor dispersión geográfica, representando un tercio de los L3 de la África subsahariana. Los subhaplogrupos L3b y L3d presentan las frecuencias más altas en el oeste africano y el L3b tiene representación en el norte de África y Próximo Oriente. Los demás clados del haplogrupo L3 y los haplogrupos L5 y L6, están prácticamente restringidos al este africano, concretamente a la región que va desde Egipto hasta Tanzania (Kivisild *et al.*, 2004).

El haplogrupo M1 es el único linaje basal del macrohaplogrupo M cuyo origen es atribuido a África, concretamente al este africano (Quintana-Murci *et al.*, 1999). La justificación del origen africano está basada en el hecho de que las mayores frecuencias de este haplogrupo se han encontrado en Etiopía y a lo largo del valle del Nilo (Passarino

et al., 1998; Quintana-Murci *et al.*, 1999; Krings *et al.*, 1999b; Richards *et al.*, 2003; Kivisild *et al.*, 2004). Sin embargo, hay autores que prefieren no excluir la posibilidad de que este haplogrupo se haya originado en Asia y que su presencia en África sea el resultado de una migración de retorno muy antigua (Richards *et al.*, 2003; Kivisild *et al.*, 2004). De los tres clados identificados (M1a, M1b y M1c), el M1a en conjunto con el M1* son los que se encuentran más dispersos, habiendo sido detectados principalmente en el este africano, pero también en frecuencias bajas en el Próximo Oriente, Cáucaso y Europa (Corte-Real *et al.*, 1996; Macaulay *et al.*, 1999b; Richards *et al.*, 2000; Richards *et al.*, 2003). El grupo M1b se restringe al este africano y sólo se ha detectado en Etiopía y Egipto (Quintana-Murci *et al.*, 1999; Kivisild *et al.*, 2004). Finalmente, el grupo M1c no se ha encontrado en Etiopía y abarca la gran parte de la variación de M1 en el noroeste africano, islas Canarias y Próximo Oriente.

En el este africano, aparte de los linajes de DNAMt antes presentados, existe una proporción importante de linajes pertenecientes al macrohaplogrupo N (Richards *et al.*, 2003; Kivisild *et al.*, 2004). En Etiopía (Kivisild *et al.*, 2004) más de 30% de los linajes pertenecen a uno de los haplogrupos del macrohaplogrupo N, destacando la elevada frecuencia de (preHV)1 (~10%) que ha sido interpretada como el resultado de una migración de retorno desde el Próximo Oriente (Richards *et al.*, 2003).

En el norte de África aparte de los haplogrupos de origen africano, hay dos subhaplogrupos que probablemente se han originado en el Próximo Oriente y evolucionado en el norte de África, y que son considerados como específicos del continente africano (Fig. 2c, d). Estos son, los subhaplogrupos U6 (Rando *et al.*, 1998, 1999; Plaza *et al.*, 2003; Maca-Meyer *et al.*, 2003) y X1 (Reidla *et al.*, 2003). El haplogrupo X es poco frecuente, pero el haplogrupo U6, dependiendo de la población, puede alcanzar frecuencias superiores al 20%. En el norte de África, el DNAMt muestra una composición mezclada entre haplogrupos africanos y del oeste de Eurasia (Rando *et al.*, 1998, 1999; Macaulay *et al.*, 1999b; Brakez *et al.*, 2001; Anglés *et al.*, 2003; González *et al.*, 2003; Plaza *et al.*, 2003). La contribución de haplogrupos L, que representaría el flujo génico desde las poblaciones subsaharianas hasta las del norte de África, se ha visto que puede variar entre ~3% en Bereberes de Marruecos y ~43% en mauritanos (Rando *et al.*, 1998; Anglés *et al.*, 2003; Plaza *et al.*, 2003). Así, en esta región pueden encontrarse poblaciones con un pool genético mitocondrial muy similar al de las vecinas poblaciones europeas del mediterráneo y otras con una importante contribución de linajes subsaharianos.

I. 2.4.2.2. Los linajes de DNAMt fuera de África

Como se ha referenciado anteriormente, del haplogrupo L3 africano han derivado dos macrohaplogrupos, M y N, que han evolucionado fuera de África originando la presente variación del DNAMt en Europa, Asia, América y Oceanía. En las figuras 2b, 2c y 2d, se

pueden ver los distintos haplogrupos que han derivado de los macrohaplogrupos M y N, así como su origen más probable.

Asia

De los haplogrupos que conforman el macrohaplogrupo M (Fig. 2b), con excepción del haplogrupo africano M1 y de los dos haplogrupos de Oceanía, todos los restantes tienen un origen asiático. Diez haplogrupos (M7-M12, D, E, G y CZ) tienen su origen en el este asiático y los restantes (M2, M3, M4, M5 y M6), tienen su origen en el sur del continente, probablemente en la India, y su distribución se halla bastante circunscrita a esta región, representando la mayor parte de la variación del DNAm_t resultante del macrohaplogrupo M (Basu *et al.*, 2003; Palanichamy *et al.*, 2004). Los haplogrupos M7 y M9-12 tienen una distribución bastante restringida a poblaciones del este asiático (China, Japón y Corea). El clado monofilético formado por los subhaplogrupos M8a, C y Z presenta una distribución más amplia: el subhaplogrupo M8a tiene su origen probable en el sudeste de China, y, de los tres, es el que tiene una distribución geográfica más restringida; el subhaplogrupo C se habrá originado en el nordeste asiático, a partir de donde se ha difundido; está subdividido en 4 clados, de los cuales el C1 está representado en el continente americano, y el haplogrupo Z, cuyo origen se admite que sea la Asia oriental-central, es el que presenta mayor dispersión, estando referido desde la región de los Saami y Rusia en Europa hasta el extremo del este asiático. El haplogrupo D es el haplogrupo más abundante en el este y centro de Asia; sin embargo, la distribución geográfica de algunos subhaplogrupos es bastante característica; por ejemplo, el D5 es abundante en el sudeste, mientras que el D4a presenta las frecuencias más altas en los Chukchi del nordeste de Siberia. El haplogrupo E tiene una distribución en el extremo sureste asiático, habiendo sido únicamente detectado en la península Malay y algunas islas del Pacífico. El haplogrupo G, a semejanza del D, también tiene una distribución característica; el subgrupo G5 es dominante en el nordeste de Siberia y el G2 es relativamente abundante en el norte de China y en el Asia central (Tanaka *et al.*, 2004).

Por lo que se refiere a los haplogrupos que conforman el macrohaplogrupo N (Fig. 2c, d), pueden identificarse como originarios en el este asiático dos clados derivados directamente del macrohaplogrupo N (A y N9) y cuatro clados derivados del macrohaplogrupo R (R9, R11, B y F). Los haplogrupos A y N9 muestran dispersión en la zona del nordeste del continente asiático, mientras que el B y F tienen una dispersión más al sudeste, aunque la distribución del haplogrupo B sea bastante compleja. Además, pueden identificarse una serie de clados y/o subclados (N5, N1, R1, R2, R5-8, R30, R31, U7, U9, HV2 y U2a-c) que son originarios y/o presentan una distribución geográfica restringida al sur y suroeste asiático (Tanaka *et al.*, 2004; Palanichamy *et al.*, 2004; Quintana-Murci *et al.*, 2004).

De lo que se ha presentado anteriormente, se puede deducir la subdivisión de Asia en dos grandes regiones en base al DNAm, este y sur, pero en que se identifica una contribución equitativa de haplogrupos N y M. Sin embargo, el tamaño del continente asiático, combinado con un complejo proceso de poblamiento, ha dado lugar a diferencias importantes en la composición del DNAm. Un ejemplo extremo es la variación que se observa en Siberia, donde seis de los haplogrupos mencionados anteriormente (A, C, D, G, Z y el subgrupo Y del haplogrupo N9), representan más de un 75% de los DNAs mitocondriales de esta región. También la zona más occidental de Asia muestra diferencias importantes en la composición de su DNAm, presentando una baja frecuencia de haplogrupos del este y sur de Asia, y una composición mitocondrial mucho más cercana a la de las vecinas poblaciones de Europa (Richards *et al.*, 2002; Quintana-Murci *et al.*, 2004).

Eurasia Occidental

En Eurasia occidental el DNAm deriva exclusivamente del macrohaplogrupo N (Fig. 2c, d). Los linajes del macrohaplogrupo N que están presentes en Eurasia occidental son: X, N1 (en que se incluyen los haplogrupos I, N1a, N1b y N1c), N2 (que incluye el haplogrupo W) y los clados JT, U y Pre-HV que derivan del supergrupo R. Los clados JT, U y Pre-HV presentan numerosas subdivisiones, la gran mayoría características de poblaciones europeas y del Próximo Oriente. Con excepción de U5, V y algunos subgrupos del H, que posiblemente se han originado y dispersado a partir del oeste europeo, todos los haplogrupos considerados como típicos de Eurasia occidental tienen un origen probable en el Próximo Oriente (Torroni *et al.*, 1996; Torroni *et al.*, 1998b; Richards *et al.*, 2000; Achilli *et al.*, 2004).

Las poblaciones del Próximo Oriente y de Europa presentan algunas diferencias en lo que respecta a la frecuencia de haplogrupos. En primer lugar, las poblaciones del Próximo Oriente tienen una mayor contribución de haplogrupos típicamente africanos y asiáticos, pero además, pueden encontrarse representantes de todos los haplogrupos considerados típicos de Eurasia occidental. En Europa, con excepción de las poblaciones del Mediterráneo, no se encuentran representantes de linajes africanos. Por otra parte, la composición de linajes de Eurasia occidental es un poco diferente: linajes relativamente frecuentes en el Próximo Oriente, como el N1, N2, U1, U2, U8b y (PreHV)1, son muy poco frecuentes o incluso ausentes en las poblaciones europeas; y linajes bastante frecuentes en Europa, como el H, pre*V y U5, presentan frecuencias más bajas en el Próximo Oriente (Richards *et al.*, 2000; Torroni *et al.*, 2001b; Richards *et al.*, 2002; Richards *et al.*, 2003).

Contrariamente a los estudios con marcadores clásicos (Menozzi *et al.*, 1978; Sokal *et al.*, 1989, 1991; Cavalli-Sforza *et al.*, 1994; entre otros), trabajos previos utilizando frecuencias de haplogrupos del DNAm (Simoni *et al.*, 2000), indicaban que las poblaciones de Eurasia occidental no presentaban una estructuración geográfica

importante. Este resultado ha sido interpretado como la consecuencia de una incorrecta clasificación de algunos haplogrupos (Torroni *et al.*, 2000), y también al hecho de que se han considerado agrupaciones muy amplias que permitían una baja resolución. Trabajos posteriores, en que se han tenido en cuenta, no sólo los haplogrupos, sino también las subdivisiones dentro de los haplogrupos, han demostrado que la distribución de haplogrupos mitocondriales presenta un gradiente sudeste-noroeste, compatible con lo anteriormente observado para los marcadores clásicos (Richards *et al.*, 2002).

Ahora bien, los avances recientes en la resolución de la filogenia del DNAMt en Europa, concretamente del haplogrupo H (Loogväli *et al.*, 2004; Achilli *et al.*, 2004), que representa un ~40% de los linajes mitocondriales en Europa y más de un 25% en el Próximo Oriente, va a permitir, en el futuro, un nuevo análisis de las poblaciones de Eurasia occidental de forma más concluyente, como lo han demostrado ya los trabajos de Loogväli y colaboradores (2004) y Achilli y colaboradores (2004).

América

Actualmente no quedan dudas de que la población ancestral de nativos americanos migró de Asia a América a través del estrecho de Bering durante el Pleistoceno (Cavalli-Sforza *et al.*, 1994). Sin embargo, el número de migraciones y la fecha permanecen poco claros (Silva *et al.*, 2002; Bandelt *et al.*, 2003). Entre los nativos americanos se encontraron los cuatro haplogrupos asiáticos, A, B, C y D (Ward *et al.*, 1991; Horai *et al.*, 1993; Torroni *et al.*, 1993a, 1994a, d), y el haplogrupo euroasiático X (Ward *et al.*, 1991; Forster *et al.*, 1996; Scozzari *et al.*, 1997; Brown *et al.*, 1998; Smith *et al.*, 1999; Malhi *et al.*, 2001).

Con el reciente avance en la secuenciación de todo el DNAMt, se ha podido determinar que subclados de los haplogrupos A, B, C, D y X corresponden a los nativos americanos (Bandelt *et al.*, 2003). Como se puede observar en las figura 2b y 2c, los linajes del haplogrupo A que se encuentran en los nativos americanos son representantes del clado A2 (Forster *et al.*, 1996; Bandelt *et al.*, 2003), los del haplogrupo B pertenecen al clado B2 (Herrnstadt *et al.*, 2002; Bandelt *et al.*, 2003), los del C corresponden al clado C1 (Bandelt *et al.*, 2003), los del haplogrupo D pertenecen al clado asiático D4, concretamente a los grupos D1 (Forster *et al.*, 1996; Derbeneva *et al.*, 2002) y D4e1 (Tanaka *et al.*, 2004), anteriormente definido como D2 (Derbeneva *et al.*, 2002; Bandelt *et al.*, 2003), y los del haplogrupo X pertenecen al grupo X2a (Bandelt *et al.*, 2003).

Oceanía

A semejanza de América, donde el DNAMt presenta una continuidad con el norte asiático, también en Oceanía se pueden encontrar un gran número de linajes de DNAMt típicamente asiáticos, en su mayoría pertenecientes a los haplogrupos B, F y E. Aparte de estos linajes asiáticos, se han identificado dos haplogrupos derivados del

macrohaplogrupo M y tres derivados del macroaplogrupo N (Fig. 2b, c), que son específicos de Oceanía (Forster *et al.*, 2001; Ingman *et al.*, 2000; Kivisild *et al.*, 2002; Palanichamy *et al.*, 2004).

I. 2.5. APLICACIONES DEL DNAMt HUMANO

Dadas sus características, el DNAMt, desde sus aplicaciones iniciales en los años setenta, se ha utilizando extensivamente en las más diversas áreas de investigación. En el contexto de las poblaciones humanas cabe destacar sus aplicaciones en Antropología Física o Biológica. La caracterización del DNAMt en poblaciones humanas, tanto actuales como antiguas, en otros primates, y en fósiles del género Homo, ha contribuido, en las últimas décadas, a aclarar cuestiones centrales en la evolución humana (Cavalli-Sforza y Feldman, 2003), por ejemplo:

- La relación entre el hombre y los homínidos africanos (Horai *et al.*, 1995; Gagneux *et al.*, 1999; Noda *et al.*, 2001; Vigilant y Bradley, 2004; entre otros);
- El origen y dispersión del hombre moderno (Cann *et al.*, 1987; Jorde *et al.*, 1995; Krings *et al.*, 1997; Quintana-Murci *et al.*, 1999; Ingman *et al.*, 2000; Maca-Meyer *et al.*, 2001; Bulayeva *et al.*, 2003; entre otros);
- La relación entre el hombre anatómicamente moderno y el Neandertal (Krings *et al.*, 1997, 1999a; Gutierrez *et al.*, 2002; Caramelli *et al.*, 2003; Serre *et al.*, 2004; Cooper *et al.*, 2004; Currat y Excoffier, 2004; entre otros).

Asimismo, el DNAMt destaca como uno de los sistemas genéticos más utilizados en la caracterización genética de las poblaciones humanas, habiendo sido utilizado con los más diversos objetivos y propósitos:

- Como herramienta filogenética (Richards *et al.*, 1998; Richards *et al.*, 2000; Finnilä *et al.*, 2001; Herrnstadt *et al.*, 2002; Kong *et al.*, 2003; entre otros): La variación en el DNAMt puede ser usada para construir árboles filogenéticos, o varios árboles alternativos dispuestos en una red (Bandelt *et al.*, 1995; Bandelt *et al.*, 2000), que describen la relación entre secuencias individuales. La estructura de este árbol contiene información demográfica que, junto con una tasa de mutación calibrada para la secuencia en estudio, puede ser utilizada para estimar la datación de eventos prehistóricos. Además, utilizando lo que es conocido por aproximación filogeográfica, ya anteriormente mencionada, la distribución de los linajes en un árbol o red puede ser usada para detectar movimientos de una región a otra (Richards *et al.*, 1998; Macaulay *et al.*, 1999b; Richards y Macaulay, 2001; Maca-Meyer *et al.*, 2003; entre otros).
- En la caracterización genética e inferencia del origen e historia demográfica de numerosas poblaciones antiguas y modernas de los diferentes continentes

(Mountain *et al.*, 1995; Bertranpetit *et al.*, 1995; Côté-Real *et al.*, 1996; Bertorelle *et al.*, 1996; Comas *et al.*, 1997; Stone y Stoneking, 1998; Pereira *et al.*, 2000; Moreli *et al.*, 2000; Chen *et al.*, 2000; Montiel, 2001; Salas *et al.*, 2002; Malyarchuk *et al.*, 2003; Anglés *et al.*, 2003; entre otros).

- Para documentar migraciones prehistóricas y recientes (Brown *et al.*, 1998; Torroni *et al.*, 1998b; Krings *et al.*, 1999b; Torroni *et al.*, 2001b; Quintana-Murci *et al.*, 2004; entre otros).
- Para documentar procesos de colonización prehistóricos y relativamente recientes (Pinto *et al.*, 1996; Soodyall *et al.*, 1997; Mateu *et al.*, 1997; Rando *et al.*, 1999; Lum y Cann, 2000; Helgason *et al.*, 2000; Helgason *et al.*, 2001; Endicott *et al.*, 2003; Eshleman *et al.*, 2004; Artículo II de la presente tesis; entre otros).
- Para inferir el mestizaje entre poblaciones (Bortolini *et al.*, 1997; Batista dos Santos *et al.*, 1999; Alves-Silva *et al.*, 2000; Mesa *et al.*, 2000; Brehm *et al.*, 2003; Rodas *et al.*, 2003; entre otros).

I. 3. LA POBLACIÓN OBJETO DE ESTUDIO: EL ARCHIPIÉLAGO DE LAS AZORES (PORTUGAL)

I. 3.1. LOCALIZACIÓN GEOGRÁFICA

El Archipiélago de las Azores (Portugal), con una área total de 2329 Km² (SREA, 2001a), está formado por nueve islas de origen volcánico, situadas en el océano Atlántico (Fig. 3), entre los paralelos 36° 55'N y 39° 45'N, y los meridianos 24° 45'W y 31° 17'W (SREA, 2001b).



Fig. 3 – Localización geográfica del archipiélago de las Azores (Portugal).

Las nueve islas del archipiélago definen tres grupos distintos, teniendo en cuenta su proximidad geográfica: El grupo Oriental (islas de São Miguel y Santa Maria), el grupo Central (islas Terceira, Faial, Pico, Graciosa y São Jorge) y el grupo Occidental (islas de Flores y Corvo) (Fig. 3).

I. 3.2. HISTORIA DE LAS AZORES

I. 3.2.1. Descubrimiento y Poblamiento

Contexto general

En la historia de los descubrimientos portugueses, el de las Azores es, sin duda, uno de los temas más discutibles, permaneciendo todavía por esclarecer. Los elementos disponibles para el estudio del descubrimiento y poblamiento del archipiélago son escasos, siendo la información muchas veces contradictoria y no concluyente (Matos, 1989a). Hasta el siglo XIX, se aceptó la versión de Gaspar Frutuoso, cronista del siglo XVI, basada en información oral. Según el cronista, las Azores fueron descubiertas por Gonçalo Velho Cabral en 1432 (Mendonça, 1996). En cambio, a partir del siglo XIX, se

propusieron varias hipótesis sobre su descubrimiento, pero ninguna de ellas definitiva. Actualmente dos teorías dividen los historiadores:

- 1) El archipiélago de las Azores habría sido descubierto por casualidad durante el reinado de D. Afonso IV (siglo XIV), y los viajes del siglo XV habrían servido esencialmente para poblar las islas (Mendonça, 1996; Pires, 1983). Los defensores de esta idea se basan principalmente en que varias islas a occidente de Portugal continental aparecen ya representadas en mapas de finales del siglo XIV (Mendonça, 1996).
- 2) El descubrimiento de las islas de los grupos Oriental y Central del archipiélago, por Diogo de Silves, marinero del infante D. Henrique, habría ocurrido en 1427 (Mendonça, 1996; Matos, 1989a). Para los investigadores que apoyan esta teoría, las islas representadas en los mapas del siglo XIV no son más que islas imaginarias (Mendonça, 1996).

El descubrimiento de las islas del grupo Occidental, Flores y Corvo, es también un poco dudoso; sin embargo, es sabido que es posterior al de las demás islas, probablemente tuvo lugar en 1452, por Diogo de Teive y su hijo (Mendonça, 1996; Matos, 1989a).

Después del descubrimiento de las Azores, el rey portugués dio permiso en 1439 al infante D. Henrique para poblar el archipiélago. Por delegación del infante, Gonçalo Velho inicia el poblamiento de las islas del grupo Oriental (Matos, 1989a). Todos los datos indican que la ocupación humana habría sido lenta y llena de dificultades. Aunque la pequeña isla de Santa Maria se pobló fácilmente, no pasó lo mismo con São Miguel (Matos, 1989a). Así, consta que los primeros pobladores llegaron a vivir en cafúas de paja y heno, estando constantemente atemorizadas por temblores de tierra (Mendonça, 1996). En las islas del grupo Central y Occidental, la ocupación humana fue igualmente desigual y lenta (Matos, 1989a). No es de extrañar pues, que el infante D. Henrique recurriera a la ayuda extranjera, mayoritariamente de Flandes, para promover el poblamiento de las islas (Mendonça, 1996), de forma que en 1450, la capitania de la isla Terceira fue entregada a Jácome de Bruges, y algún tiempo después la de Faial a José Van Hurtere, ambos flamencos (Mendonça, 1996; Matos, 1989a).

Al igual que casi todo lo que se relaciona con las Azores, también el origen de los pobladores ha sido objeto de especulación. En cualquier caso, es posible adelantar alguna información de carácter general acerca de los primeros pobladores de las Azores (Mendonça, 1996; Matos, 1989a; Merelim, 1966; Gomes, 1997):

- 1) Procedían mayoritariamente de las diversas provincias de Portugal continental, principalmente de Algarve, Alentejo y Minho, aunque su participación se cree desigual y variable de isla a isla;

- 2) Exceptuando los dirigentes, los niveles sociales altos de la sociedad portuguesa no participaron en el poblamiento de las islas, pero sí personas con menos recursos y deseosas de mejorar su condición de vida;
- 3) En el poblamiento habrían contribuido también, en gran medida, madeirenses;
- 4) El archipiélago habrá sido utilizado como refugio por un número importante de judíos y conversos;
- 5) La presencia de *esclavos negros y moriscos* está documentada históricamente, sobre todo en São Miguel, Flores y Corvo.
- 6) Muchos extranjeros también habrían ido a las islas de las Azores en busca de mejor suerte, entre ellos: franceses, alemanes, españoles, ingleses, italianos, y principalmente flamencos, que tuvieron una contribución decisiva en el poblamiento de las islas del grupo Central.

Grupo Oriental: Santa María y São Miguel

La isla de Santa María fue la primera del archipiélago en ser poblada. En 1439, Gonçalo Velho, su primer capitán-donatario, acompañado por dos sobrinos y un grupo de colonos, en su mayoría algarvios, se instalaron en la costa este de la isla de Santa María (Matos, 1989a). Más tarde, João Soares de Albergaria, sobrino de Gonçalo Velho, da un nuevo impulso al poblamiento de la isla al llevar familias del continente, sobre todo algarvias, que se instalan en Santa María (Matos, 1987; Matos, 1989a, 1989b).

El inicio del poblamiento de la isla de São Miguel en 1444, es más o menos contemporáneo al de Santa María. Del grupo de pobladores iniciales de la isla de São Miguel, destacan una vez más los algarvios, a los que acompañan *esclavos negros y moriscos* (Matos, 1989a). A propósito de la presencia de esclavos en la isla de São Miguel, Gaspar Frutuoso, historiador de la época, escribe: "... no había hombre de esta isla que yendo allá no comprase uno, dos, tres, cuatro, según los medios y hacienda de cada uno..." (Frutuoso *in* Matos, 1989b). Algunos autores, basándose en datos toponímicos, refieren la influencia de individuos oriundos de la Bretaña francesa en la isla de São Miguel (Matos, 1989a).

Las frecuentes relaciones marítimas entre la región del Algarve y las Azores, especialmente con São Miguel, aparte de permitir intercambios comerciales, fomentaban un cambio de residencia y un estrechamiento de relaciones entre las poblaciones del continente y de las islas. En general es aceptado, por parte de los historiadores, etnógrafos y geógrafos, que en las islas de poblamiento más antiguo, Santa María y São Miguel, hay una influencia clara del sur de Portugal, del Alentejo y sobre todo del Algarve (Matos, 1987; Matos, 1989a, 1989b). Con la muerte de Gonçalo Velho, la capitania de São Miguel es vendida a un miembro de la familia donataria de Madeira, iniciándose (en 1474) el flujo de familias madeirenses para las Azores (Matos, 1989a)

Grupo Central: Terceira, Faial, Pico, São Jorge y Graciosa

Las buenas relaciones políticas entre Portugal y Flandes (reforzadas por uniones matrimoniales entre la familia real portuguesa y el ducado de Borgoña), habrían propiciado la participación flamenca en el poblamiento azoreano, de tal forma, que en todas las islas del grupo Central se deja sentir su influencia. Pero es, sin duda, en la isla de Faial donde esa influencia es más marcada (Matos, 1989a). El flamenco José Van Hurtere, en su primer viaje a Faial, se hizo acompañar de unos 15 hombres, compatriotas que se instalaron ahí (Matos, 1989a). Muchos flamencos continuaron llegando a esta isla después de la atribución de la capitanía de Faial, en 1468, a José Van Hurtere, habiendo servido la isla tanto de colonia penal, como de refugio al ambiente social y político que se vivía en Flandes (Pires, 1983; Matos, 1989a).

Los primeros pobladores de la isla del Pico, que inicialmente fue usada para pastorear el ganado, provienen de la cercana isla del Faial y fueron también flamencos (Matos, 1989a).

El poblamiento de la isla Terceira se inicia a mediados del siglo XV con la concesión de la capitanía al flamenco Jácome de Bruges (Matos, 1989a). Éste, con ocasión de su primera ida a Terceira se hizo acompañar por algunos flamencos. Además de los flamencos, los primeros pobladores de la isla Terceira eran oriundos de Portugal continental y Madeira. Algunos de ellos eran “nobles hidalgos” del continente, y otros hijos y nietos de la nobleza madeirense. De este grupo social terceirense acabaron por salir algunos pobladores para otras islas, sobre todo Flores y Corvo (Matos, 1989a).

A mediados del siglo XV, la isla Graciosa ya tenía pobladores, habiendo sido Vasco Gil Sodré, natural de Montemor-o-Velho (Centro de Portugal continental), acompañado por su familia y criados, el pionero en el poblamiento de la isla (Lima, 1997). La influencia de la isla Terceira parece haber sido decisiva en el poblamiento y desarrollo de la isla Graciosa, pues existen autores que aceptan que la isla Graciosa fue poblada a partir de la isla Terceira (Ferreira, 1987).

Dado la clara influencia flamenca en el poblamiento de las islas, ¿por qué razón la lengua Flamenca desapareció tan rápido de las Azores, y sobretodo de Faial? Sin duda esta pregunta es pertinente. Ya se ha dicho que los flamencos, por circunstancias políticas y sociales menos favorables, habrían ido a las Azores; pero a partir de cierto momento el rey portugués decidió, por razones estratégicas, limitar ese movimiento migratorio. Así, debido a este condicionamiento, y a que los flamencos solteros (que eran muchos) contraían matrimonio con mujeres portuguesas, empezó una generación que aprendía más portugués con las madres, que alemán de Flandes con los padres (Pires, 1983), lo que comporta una completa asimilación del elemento extranjero por el portugués (Matos, 1989a).

Grupo Occidental: Flores y Corvo

La primera tentativa de poblamiento de la isla de las Flores, hecha por Guilherme da Silveira, no fue exitosa. El poblamiento definitivo habría tenido lugar a finales de la primera década del siglo XVI, promovido por Antão Vaz, que llegó al extremo del archipiélago azoreano con un grupo de pobladores oriundos de Terceira y Madeira. Más tarde llegaron colonos provenientes de las restantes islas (principalmente São Miguel) y del continente portugués (Matos, 1989a, 1989b). La presencia de esclavos en la isla de las Flores está bastante bien documentada, siendo posible cuantificar su número en base a los registros parroquiales (Matos, 1989a; Gomes, 1997).

Por lo que respecta al islote de Corvo, no se consigue una ocupación humana con éxito hasta 1548. El capitán Gonçalo de Sousa envía a la isla esclavos de su confianza, que tenían por misión cultivar la tierra y cuidar del ganado (Matos, 1989a). Mientras tanto, fomentado por la proximidad geográfica, existe un estrecho contacto entre la isla de las Flores y del Corvo, de tal forma que pasan individuos de una isla hacia la otra (Matos, 1989a).

El archipiélago, y particularmente las islas de Flores y Corvo, tenía una posición estratégica durante el período de expansión marítima, funcionando como una frontera en medio del Atlántico. Todas las embarcaciones venidas de la India, África y América hacían parada en las islas de Flores y Corvo. Según los historiadores, en las aguas de esta zona había constantemente embarcaciones de piratas y corsarios que esperaban a las embarcaciones portuguesas y, más tarde también a las españolas, de los viajes de retorno de India, África y América (Riley, 1991). Consta que algunas de las batallas marítimas más impresionantes han tenido lugar a lo largo de estas dos pequeñas islas.

I. 3.3. EVOLUCIÓN DE LA POBLACIÓN Y EMIGRACIÓN

Como se puede ver en la tabla 1, de finales del siglo XVI hasta el tercer cuarto del siglo XIX, la población azoreana aumentó a un ritmo considerable, siendo las islas de São Miguel y Terceira las que presentaron el mayor efectivo poblacional. A partir de entonces y hasta ~1930, se registra un ligero descenso en la población, que corresponde a la corriente migratoria que se intensifica en el siglo XIX.

La primera noticia de salida de azoreanos en busca de mejor suerte en otros parajes está datada hacia la primera mitad del siglo XVII (Mendonça, 1996). A lo largo del siglo XVIII, la emigración hacia Brasil se vuelve más regular, pero es durante el siglo XIX que el fenómeno migratorio alcanza proporciones anteriormente desconocidas, implicando entre 1881 y 1885 un total de 22.397 individuos. Los emigrantes se dirigen principalmente a Brasil y América del Norte. La justificación de esta corriente migratoria parece haber sido la

Tabla 1 – Evolución de la población en las varias islas del archipiélago desde finales del siglo XVI hasta 2001 (datos extraídos de: Mendonça, 1996; Madeira, 1999; INE, 2001, 2004).

Isla	Año	Final siglo XVI	1695	1747	1820	1849	1864	1878	1890	1900	1911	1920	1930	1940	1950	1960	1970	1980	1991	2001
S. Maria	2600	4235	5023	3810	5647	5863	6338	6247	6383	6247	6457	7158	8067	11839	13233	9765	6500	5922	5578	
São Miguel	27132	34251	54670	69000	91638	105407	119933	118511	121726	116957	111745	127758	147948	165166	168687	149000	131908	125915	131609	
Terceira	21560	21078	26433	32700	41545	45895	45030	46637	48920	47824	46277	48716	53233	60372	71610	65500	53570	55706	55833	
Graciosa	2708	6656	8037	9230	9075	8718	8321	8449	8394	7603	7477	8470	9193	9517	8669	7180	5377	5189	4780	
São Jorge	2676	6716	13995	15600	17808	17998	18272	17065	16138	14277	13362	13980	15798	16507	15895	12970	10361	10219	9674	
Faial	4048	13287	16669	21340	24755	26204	24962	23630	22385	20214	18917	21510	23280	23923	20281	16375	15489	14920	15063	
Pico	3508	10259	20639	24940	31246	27736	26396	25411	24125	21853	19927	20204	21423	22557	21837	18125	15483	15202	14806	
Flores	632	3235	5583	7700	10840	10508	9662	8838	8141	7220	6720	6992	7447	7850	6583	5630	4352	4329	3995	
Corvo	80	478	523	910	810	883	880	806	805	746	661	676	691	728	681	470	370	393	425	
Total	64944	100195	151573	185230	223409	248028	259800	255594	256291	243002	231543	255464	287091	318558	327480	285015	243410	237795	241763	

inestabilidad política y del sistema administrativo vigente en la época, que potenciaba la distribución desigual de propiedades, creando una acentuada separación entre ricos y pobres. Este fue el factor determinante para que las clases menos favorecidas emigrasen con el objetivo de reducir sus problemas económicos.

Desde 1930 (tabla 1), hay un aumento considerable de la población, que alcanza su máximo en 1960. A partir de este momento la población inicia una acentuada disminución, una vez más a causa del movimiento migratorio, sobre todo hacia los Estados Unidos, Canadá y Bermudas. En efecto, entre 1960 y 1975 emigraron 127537 azoreanos fundamentalmente para los Estados Unidos (Mendonça, 1996). En poco más de un siglo algunas islas, destacando Flores, han reducido su población a menos de la mitad.

I. 3.4. ANTECEDENTES BIODEMOGRÁFICOS Y GENÉTICOS

Las nueve islas del archipiélago de las Azores presentan un enorme potencial para las investigaciones antropológicas y evolutivas (ya sean macro o micro evolutivas), dado su aislamiento geográfico, reducido efectivo poblacional y su descubrimiento y poblamiento relativamente recientes. Sin embargo, hasta no hace mucho se habían realizado pocos estudios de caracterización biodemográfica y genética de las poblaciones azoreanas.

El primer trabajo **biodemográfico** realizado en las distintas islas del archipiélago de las Azores (Cunha, 1987) presenta los valores de consanguinidad, de la segunda mitad del siglo XIX. Los niveles de consanguinidad obtenidos son bastante similares a los encontrados en poblaciones de Portugal continental (Abade, 1986, 1992; Mota, 1983) y poblaciones europeas (Bittles, 1998). Posteriormente, se han publicado dos estudios centrados en la isla de Flores (Cunha et al., 1989; Smith et al., 1992) y los resultados de ambos indican que el patrón de evolución de la consanguinidad en Flores, desde 1861 hasta 1980, es similar al encontrado en poblaciones del sur de Europa en el mismo período (Calderon, 1983). Posteriormente son analizadas algunas poblaciones concretas de la isla de São Miguel que presentan características sociales muy particulares (Lima, 1990, 1991, 1991/1992, 1992/1993) y en las que los valores de consanguinidad observados son ligeramente más elevados que los obtenidos considerando el total de la isla. En estos momentos existe una base de datos genealógica de la isla de Flores (Lima, 1998, 1999) y se halla prácticamente terminada la de Pico (NEPS, 2003) lo que ha permitido que se hayan realizado varios trabajos para estas dos islas utilizando diferentes tipos de fuentes (Alves *et al.*, 2000; Alves, 2000; Santos y Lima, 2000; Cruz *et al.*, 2002; Cruz, 2003; Amorim y Lima, 1986; Amorim 1987, 1988a, 1988b, 1992, 1993, 2003, 2004).

Por lo que se refiere a los estudios de variación **genética**, los primeros fueron limitados a la isla de Flores (Amorim *et al.*, 1979) y de São Miguel (Santos *et al.*, 1992; 1994; Santos y Amorim, 1994). En estos trabajos los autores utilizaron marcadores

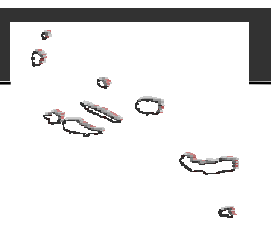
clásicos con el propósito de investigar el efecto de la deriva genética, más fuerte en poblaciones pequeñas. Dichos estudios revelaron un comportamiento de las islas estudiadas bastante similar al observado en otras poblaciones continentales. Los estudios posteriores se han centrado en dos objetivos principales:

- 1) La construcción de una base de datos de microsatélites para utilización forense (Corte-Real *et al.*, 1999; Prata *et al.*, 1998; Anjos *et al.*, 2000; Velosa *et al.*, 2002);
- 2) El estudio de la historia del gen de la enfermedad de Machado-Joseph, enfermedad neurodegenerativa, con un modo de transmisión autosómico dominante, y que presenta una elevada prevalencia en las Azores (Lima, 1996; Lima *et al.*, 1997; Lima *et al.*, 1998; Gaspar *et al.*, 2001).

A su vez, Bruges-Armas y colaboradores (1999), utilizando las moléculas de clase I y II del sistema HLA, y Pires (1999), utilizando la región hipervariable I (HVRI) del DNA mitocondrial (DNAm_t), presentaron los primeros trabajos en que se realiza la caracterización genética de la población de las Azores a la vez que se efectúan inferencias sobre sus orígenes. Este es un abordaje particularmente interesante, teniendo en cuenta el proceso de poblamiento del archipiélago; sin embargo, aunque originales en su abordaje, en ambos trabajos las muestras utilizadas son en su mayoría de individuos de la isla Terceira por lo que podrían no ser representativos de la población de las Azores en general.

Más recientemente, se han realizado una serie de trabajos (artículos de la presente tesis; Brehm *et al.*, 2003; Carvalho *et al.*, 2003a; Carvalho *et al.*, 2003b; Fernandes y Brehm, 2003, Branco y Mota-Vieira, 2003; Montiel *et al.*, 2005; Pacheco *et al.*, 2005; Fernando *et al.*, *in press*; Silva *et al.*, *in press*) que serán discutidos en el apartado V, con el objetivo de profundizar en la estructura genética de la población de las Azores y hacer inferencias sobre sus orígenes, entre los cuales figuran las aportaciones realizadas en la presente tesis, que resultan de la colaboración entre la Universidad de las Azores y la Universidad Autónoma de Barcelona.

II. **JUSTIFICACIÓN** **Y OBJETIVOS**



II. JUSTIFICACIÓN Y OBJETIVOS DEL TRABAJO

Las poblaciones de las islas de las Azores son potencialmente muy interesantes en el contexto antropológico y evolutivo, dado su aislamiento geográfico, reducido efectivo poblacional y su descubrimiento y poblamiento relativamente recientes.

Los datos históricos que relatan el poblamiento y caracterizan la población original de las islas del archipiélago, indican la existencia de alguna diferencia en el origen de los pobladores de las distintas islas; sin embargo, las fuentes históricas raramente dan idea del número y sexo de los individuos que llegaban a las islas y además, tienden a referirse a una pequeña minoría de los pobladores, privilegiando a los miembros de las clases sociales más elevadas y al clero. Por ello, es de interés ampliar el conocimiento sobre el origen y composición de la población original de las islas.

El análisis genético de las poblaciones humanas ha demostrado ser uno de los enfoques más potentes para investigar el origen, estructura y dinámica de las poblaciones. Aunque los estudios genéticos no substituyen la necesidad de efectuar las investigaciones históricas y biodemográficas, que siempre deben acompañar una investigación genética, los datos genéticos de las poblaciones actuales dan información sobre aquellas características que estarían en la población original y que siguen estando representadas en la población actual. De esta forma, el análisis genético de las poblaciones azoreanas puede ser utilizado para hacer inferencias sobre el origen y composición de la población original de las islas y, oportunamente, conocer la estructura genética de la población actual.

Por otra parte, las oscilaciones demográficas registradas a lo largo del tiempo, incluyendo fuertes reducciones poblacionales recientes producidas por movimientos migratorios, hacen de la población azoreana un modelo excelente para evaluar el efecto de las fluctuaciones demográficas en la diversidad y microdiferenciación de las poblaciones humanas.

Teniendo en cuenta lo expuesto, se ha planteado centrar una parte importante de la investigación en el DNAMt; la elección de éste marcador se ha basado en un conjunto de criterios, unos inherentes a la población y otros al propio DNAMt:

- La historia que relata el poblamiento no se refiere, en general, al sexo de los pobladores y, probablemente, representa mejor la realidad masculina que la femenina;

- La información histórica da indicaciones de que el poblamiento no se ha hecho de forma igual en todo el archipiélago, siendo clara la separación de los tres grupos de islas (Oriental, Central y Occidental);
- La filogenia y filogeografía del DNAMt estaban ya bastante bien establecidas (Macaulay et al., 1999b; Richards et al., 2000) aunque no estuviera propuesta una metodología de análisis jerárquico que combinara la secuenciación de la región de control y el análisis de PCR-RFLPs de la región codificante;
- No existía la caracterización poblacional a nivel de las Azores basada en el DNAMt, sólo se había realizado un estudio preliminar de 45 individuos (Pires, 1999) centrado en la región hipervariable I;
- El DNAMt de las poblaciones parentales, es decir, las poblaciones que según la información histórica han aportado pobladores, estaba en su mayoría caracterizado, lo que permitía un buen análisis filogeográfico;
- Mayor sensibilidad al efecto de la deriva genética y, por lo tanto, más eficacia para detectar la existencia de subestructuración en las poblaciones humanas.

Dado que ha sido reportada, en diferentes estudios posteriores, una discrepancia importante en el valor de la tasa de mutación obtenido por métodos filogenéticos y por métodos empíricos, y aprovechando la oportunidad de poder muestrear familias de origen Azoreano, se ha planteado realizar la investigación, de forma integrada, de algunas de las características inherentes al propio DNAMt; entre ellas, la tasa de mutación, transmisión materna, ausencia de recombinación y homoplasia.

Además, a partir del propósito original de estudiar la estructura genética de la población de las islas de las Azores nos hemos planteado realizar un estudio de la estructura genética y biodemográfica de la isla de Flores (grupo Occidental) a lo largo del tiempo. Este estudio parecía particularmente interesante ya que la isla de Flores:

- Es una de las más pequeñas del archipiélago y una de las más aisladas geográficamente;
- Ha sido, junto con el islote de Corvo, la última en poblarse, pero según la información histórica tiene una posición estratégica importantísima, pudiendo haber funcionado como una frontera en medio del Atlántico;
- De todas las islas del archipiélago es la que ha estado más afectada por la emigración, habiéndose reducido drásticamente su población desde mediados del siglo XIX;
- Ninguno de los trabajos biodemográficos ya realizados abarca todas las parroquias de la isla de Flores.

II. 1. OBJETIVOS

II. 1.1. Objetivos Generales de la Tesis

1. Establecer una metodología jerárquica para el análisis de los haplogrupos, discutiendo la misma teniendo en cuenta los recientes avances en la definición de la filogenia del DNAMt;
2. Presentar una perspectiva de la estructura biodemográfica de las islas de las Azores, teniendo en cuenta los resultados obtenidos en esta tesis y en otros trabajos;
3. Dibujar la estructura genética de las islas de las Azores, teniendo en cuenta tanto los resultados obtenidos en esta tesis como otros trabajos;
4. Inferir el origen de los pobladores de las islas de las Azores teniendo en cuenta los resultados obtenidos, además de otros datos existentes;
5. Investigar de forma integrada las características inherentes al DNAMt: transmisión materna, ausencia de recombinación, homoplasmia y tasa de mutación.

II. 1.2. Objetivos de los Artículos I y II

Artículo I: **Santos, C.**, Montiel, R., Angles, N., Lima, M., Francalacci, P., Malgosa, A., Abade, A. y Aluja, M^a. P. (2004) Determination of Human Caucasian Mitochondrial DNA Haplogroups by Means of a Hierarchical Approach. *Human Biology*, 76(3): 431-453.

Artículo II: **Santos, C.**, Lima, M., Montiel, R., Angles, N., Pires, L., Abade, A. y Aluja, M^a. P. (2003) Genetic structure and origin of peopling in the Azores Islands (Portugal): The view from mtDNA. *Annals of Human Genetics*, 67: 433-456.

Considerando los antecedentes histórico-poblacionales del archipiélago de las Azores era previsible la detección de haplogrupos no estrictamente europeos en las mencionadas islas. Partiendo de este presupuesto, nos propusimos diseñar una metodología jerárquica que posibilitara el análisis del DNAMt en poblaciones caucasoides. Esta metodología, descrita detalladamente en el artículo I, se iba perfeccionando a medida que avanzaba el estudio poblacional presentado en el artículo II.

Objetivo general del artículo I: *Proponer una metodología jerárquica para analizar los haplogrupos del DNAMt de poblaciones caucasoides (objetivo general de la tesis: 1).*

Objetivos particulares:

- 1) Proponer una metodología de análisis jerárquico que permita la determinación de haplogrupos del DNAMt en poblaciones caucasoides;
- 2) Describir las condiciones de análisis (primers, condiciones de PCR y de digestión o secuenciación) para cada una de las posiciones seleccionadas para la determinación de haplogrupos.

Objetivo general del artículo II: *Contribuir a la caracterización de la estructura genética de la población actual de las Azores, y realizar inferencias sobre el origen de los pobladores del sexo femenino (objetivos generales de la tesis: 3, 4).*

Con el análisis del DNAMt de los tres grupos de islas del archipiélago de las Azores nos hemos propuesto cumplir los siguientes objetivos particulares:

- 1) Describir la variación del DNAMt en la población azoreana en general y por grupos de islas;
- 2) Contribuir a la caracterización de la estructura genética de la población de las Azores, concretamente:
 - a. Verificar si existen diferencias en la composición de DNAMt de los tres grupos de islas;
 - b. Evaluar y comparar los niveles de diversidad observados en el archipiélago (general y por grupos de islas) con los presentados por otras poblaciones;
 - c. Evaluar el efecto de la deriva genética;
- 3) A través del análisis filogeográfico, hacer inferencias sobre el origen de los pobladores del sexo femenino de los diferentes grupos de islas;
- 4) Deducir la proporción de mestizaje en la población actual de mujeres;
- 5) Situar filogenéticamente las poblaciones de las Azores en el contexto de poblaciones europeas, africanas y del Próximo Oriente.

II. 1.3. Objetivos de los Artículos III y IV

Artículo III: **Santos, C.**, Lima, M., Abade, A. y Aluja, M^a P. (2004) Analyse du mouvement marital à l'île de Flores (Azores, Portugal). *Antropo*, 7: 63-71.

Artículo IV: **Santos, C.**, Abade, A., Cantons, J., Mayer, F. M., Aluja, M^a P. y Lima, M. Genetic Structure of Flores Island (Azores, Portugal) in the 19th Century and in the present-day: Evidences from surname analysis. *Human Biology (in press)*.

Objetivo general de los artículos III y IV: *Contribuir a la caracterización biodemográfica y genética de la isla de Flores desde el siglo XVIII hasta la actualidad (objetivos generales de la tesis: 2, 3).*

Objetivos particulares del Artículo III:

- 1) Hacer el análisis del movimiento marital en cuatro parroquias de Flores situadas en puntos extremos de la isla (Norte, Sur, Este, Oeste), concretamente:
 - a. Analizar y comparar entre parroquias los valores de endogamia y de los diferentes tipos de exogamia;
 - b. Analizar y comparar entre parroquias la evolución temporal de la endogamia, exogamia y inmigración efectiva;
- 2) Calcular e interpretar los valores de consanguinidad obtenidos a partir de las dispensas eclesiásticas;
- 3) Analizar la evolución de la diversidad y microdiferenciación en la isla de Flores, utilizando como herramienta los apellidos.

Objetivos particulares del Artículo IV:

- 1) Analizar y comparar los valores de endogamia y de diferentes tipos de exogamia entre todas las parroquias existentes en la isla en la segunda mitad del siglo XIX;
- 2) Caracterizar genéticamente, teniendo por base la distribución de apellidos, todas las parroquias de Flores en el siglo XIX y en la actualidad;
- 3) Discutir la influencia del efectivo poblacional y de la distancia geográfica en la estructura genética de las poblaciones de Flores;
- 4) Documentar los posibles cambios en la estructura genética de la isla de Flores en los últimos dos siglos;
- 5) Investigar el impacto de la fuerte reducción de población que ha ocurrido en Flores, desde mediados del siglo XIX hasta la actualidad, en la estructura genética de la isla.

II. 1.5. Objetivos del Artículo V

*Artículo V: Santos, C., Montiel, R., Sierra, B., Bettencourt, C., Fernandez, E., Alvarez, L., Lima, M., Abade, A. y Aluja, M^a.P. Understanding differences between phylogenetic and pedigree derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). *Molecular Biology and Evolution* (in press).*

Objetivo general del artículo 5: *Estimar la tasa de mutación de la región de control del DNAm_t en familias de origen azoreano, teniendo en cuenta las posibilidades de contribución paterna y recombinación (objetivo general de la tesis: 5).*

Los objetivos particulares son los siguientes:

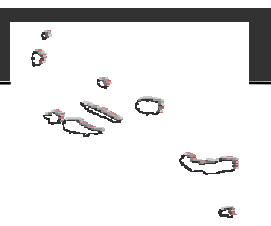
- 1) Contribuir a la comprensión del proceso evolutivo del DNAm_t, concretamente de los factores que controlan los niveles y el destino de la heteroplasmia;

- 2) Evaluar la posibilidad de contribución de DNAm_t paterno a niveles detectables en el individuo joven/adulto;
- 3) Testar la posibilidad de recombinación entre moléculas de DNAm_t heterólogas;
- 4) Estimar empíricamente la tasa de mutación del DNAm_t;

III.

MATERIAL Y

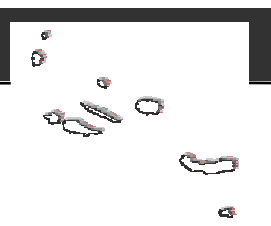
MÉTODOS



III. MATERIAL Y MÉTODOS

El material de base, así como la metodología de laboratorio y de análisis de datos, utilizados en cada una de las publicaciones de las que consta la presente tesis doctoral, está descrito en los correspondientes artículos I-V. Sin embargo, para complementar el estudio de la estructura demográfica de la población azoreana se han construido las pirámides poblacionales y calculado el índice de masculinidad [$IM = (N^{\circ} \text{ hombres} / N^{\circ} \text{ mujeres}) \times 100$] en los años 1864, 1890, 1911, 1930, 1950, 1970 y 2000 (Apéndice 1). Para tal, se han utilizado los datos poblacionales organizados por edad y sexo, extraídos de los censos 1864-2001 (INE, 2003, 2004), de cada municipio del archipiélago de las Azores, y los mismos han sido posteriormente agrupados para cada isla y por grupos de islas.

IV. ARTÍCULOS



IV. 1

Artículo I

Determination of Human Caucasian Mitochondrial DNA
Haplogroups by Means of a Hierarchical Approach

Santos *et al.* (2004) ***Human Biology***, 76 (3): 431-453

Determination of Human Caucasian Mitochondrial DNA Haplogroups by Means of a Hierarchical Approach

CRISTINA SANTOS,^{1,2} RAFAEL MONTIEL,^{1,3} NATÀLIA ANGLÉS,¹ MANUELA LIMA,³ PAOLO FRANCALACCI,⁴ ASSUMPCIÓ MALGOSA,¹ AUGUSTO ABADE,² AND M. PILAR ALUJA¹

Abstract In this paper we propose a hierarchical approach that allows the screening of mitochondrial DNA (mtDNA) haplogroups in populations that have essentially West Eurasian mtDNA backgrounds but that could have some non-West Eurasian contributions. To develop and validate this scheme, we used data on 18 coding region polymorphisms (17 analyzed by RFLP analysis and 1 by sequencing) and sequences of hypervariable segment I (HVSI) of the mtDNA control region from the Azores Islands (Portugal) population. The proposed scheme allows the characterization of almost all West Eurasian and African major clusters by means of RFLPs. Furthermore, the scheme includes information on situations in which sequencing is pertinent to defining a particular haplogroup. The validity of the scheme is ensured by (1) using relatively stable polymorphic positions, (2) screening more than one position to define a specific haplogroup, and (3) typing confirmatory positions. Dubious samples can be resolved by sequencing. The robustness of this approach was assessed by sequencing all samples for HVSI, taking advantage of the previously established relationships between RFLPs and control region sequence polymorphisms. The use of this hierarchical approach avoids the screening of unnecessary control region polymorphisms and therefore results in a more rapid and cost-efficient screening than one in which all polymorphic positions are analyzed. Even if this approach leads to a lower level of phylogeographic resolution than the sequencing of all samples, it allows us to define population movements on a continental level and can be applied, unlike sequencing all samples, with a low cost in any laboratory.

¹Anthropology Unit, Department BABVE, Faculty of Sciences, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain.

²Department of Anthropology, University of Coimbra, 3000 Coimbra, Portugal.

³CIRN and Department of Biology, University of the Azores, 9500 Ponta Delgada, S. Miguel, Azores, Portugal.

⁴Department of Zoology and Biological Anthropology, University of Sassari, 07100 Sassari, Italy.

Human Biology, June 2004, v. 76, no. 3, pp. 431–453.

Copyright © 2004 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: MTDNA, HAPLOGROUPS, RFLP, AZORES, HIERARCHICAL APPROACH TO HAPLOGROUP SCREENING.

Because human mitochondrial DNA (mtDNA) presents particular characteristics, such as maternal inheritance, fast evolutionary rate, and lack of recombination, it has been used extensively in studies from several disciplines, including biological anthropology, forensics, and medical genetics. Particularly in biological anthropology, many questions related to human evolution and population genetics have been addressed, namely, those related to the origin of modern humans [e.g., Cann et al. (1987) and Quintana-Murci et al. (1999)] and the peopling of several regions, for example, Europe [e.g., Richards et al. (1996), Comas et al. (1997), and Simoni et al. (2000)], the Americas [e.g., Torroni et al. (1992), Bailliet et al. (1994), and Silva et al. (2002)], and the Pacific Islands [e.g., Lum et al. (1994), Hagelberg (1997), and Lum and Cann (2000)].

Variation in mtDNA has been assessed mainly by using two different approaches: high-resolution analysis of the whole genome using restriction enzymes (restriction fragment length polymorphisms, RFLPs) or sequencing of the hypervariable control region. Used independently, both methods have shown that mtDNA evolution produces groups of phylogenetically related haplotypes that have been designated either by haplogroups or by clusters, respectively. More comprehensive studies have included the analysis of RFLPs and control region sequences for the same samples [e.g., Graven et al. (1995), Torroni et al. (1996), and Kolman et al. (1996)]. Previous studies have shown that, as expected, because of the lack of recombination in mtDNA, the groups of mtDNA types defined by each method are correlated. These studies have also shown that the two types of analysis are complementary; their combined use helps to clarify the phylogenetic relationship between mtDNA lineages (Richards et al. 1998; Macaulay et al. 1999) and led to the unification of the nomenclature by adopting the system initiated by Torroni et al. (1992), which is based on RFLPs.

It was possible to assign a geographic origin to each haplogroup, and for this reason the haplogroups can be used to trace back the migration patterns of human populations. In West Eurasia eight haplogroups—H, I, J, T, U (including subhaplogroup K), V, W, and X—almost encompass all mtDNA variability (Torroni et al. 1994; Torroni et al. 1996; Macaulay et al. 1999). In Asia most mtDNA types are members of superhaplogroup M (haplogroups C, D, G, and E) (Ballinger et al. 1992; Torroni, Sukernik et al. 1993; Wallace 1995); the remaining Asian mtDNAs are encompassed by haplogroups A, B, and F (Torroni, Miller et al. 1994). Among Native Americans four Asian haplogroups (A, B, C, and D) were found (Torroni, Schurr et al. 1993). In Africa the mtDNA lineages belong to haplogroups L1, L2, and L3 and subhaplogroup M1 (Chen et al. 1995, 2000; Watson et al. 1997; Quintana-Murci et al. 1999). However, North Africa exhibits a mixed composition of lineages included in West Eurasian and African haplogroups (Rando et al. 1998; Anglés et al. 2003).

High-resolution RFLP analysis involves the amplification of 9 overlapping fragments encompassing the whole mtDNA genome and subsequent digestion with a set of at least 14 endonucleases. Analyses conducted using this method revealed that the haplogroups could be defined by a set of specific restriction sites

(Torrioni, Lott et al. 1994; Macaulay et al. 1999). This promoted the investment in research to devise a simpler method for haplogroup screening. The method is based on PCR amplification of all fragments that include the relevant restriction sites for each sample, followed by digestion with the appropriate endonuclease (Morelli et al. 2000; Montiel et al. 2001). However, this approach can be redundant for many samples, which can be efficiently resolved by simply typing a few polymorphic sites.

In this paper we present a hierarchical approach, based mainly on RFLP analysis, that allows a more complete, rapid, and cost-efficient screening of mtDNA haplogroups than that proposed by Morelli et al. (2000) and Montiel et al. (2001). The specific primers and their reaction conditions are also presented. The proposed scheme was derived from the initial analysis of mtDNA RFLPs from Catalanian (Montiel 2001) and Andorran (Anglés et al. 2003) populations and was refined by using the mtDNA analysis from the Azores Islands (Portugal) population (Santos 2001; Santos et al. 2003), in which a higher number of coding region polymorphisms were analyzed. The robustness of this approach was assessed by sequencing the HVSI region of the mtDNA for all the samples, taking advantage of the previously established relationship between RFLPs and coding region polymorphisms.

Materials and Methods

Sample and Population. We used 146 samples of unrelated individuals from the Azores Islands population, including 42 samples for which Pires (1999) previously analyzed the HVSI region of mtDNA. The samples were obtained by buccal swabs; the existing samples collected by Pires (1999) were obtained by venipuncture. Voluntary donors were asked for informed consent and to fill in an anonymous inquiry concerning the birthplaces of all their known ancestors.

The Azores were discovered uninhabited in the 15th century by Portuguese navigators. According to historical records, most of the settlers who contributed to the peopling of the Azores were Europeans; however, Jews and Moorish and African slaves also contributed to the peopling of the archipelago (Mendonça 1996).

DNA Extraction. Total DNA from buccal cells was extracted using Instagene matrix (BioRad), according to the manufacturer's specifications. DNA from blood samples was previously extracted by Pires (1999) using standard protocols.

Analysis of Coding Region Polymorphisms

DNA Amplification. Relevant polymorphic sites for mtDNA haplogroup determination (Table 1) were selected from those proposed by Torrioni, Lott et al. (1994), Torrioni et al. (1996), Torrioni and Wallace (1994), Chen et al. (1995), Lamminen et al. (1997), Macaulay et al. (1999), Quintana-Murci et al. (1999), and Saillard et al. (2000).

Table 1. Coding Region Positions Analyzed to Define Each Haplogroup, According to the Hierarchical Scheme Presented in Figure 1^a

Haplogroup	Position																	
	1715	3592	4216	4577	7025	8994	9052	10032	10394	10397	10871	11719	12308	13704	14465	14766	15606	12705
	<i>DdeI</i>	<i>HpaI</i>	<i>NlaIII</i>	<i>NlaIII</i>	<i>AluI</i>	<i>HaeIII</i>	<i>HaeII</i>	<i>AluI</i>	<i>DdeI</i>	<i>AluI</i>	<i>MniI</i>	<i>SmaI</i>	<i>HinfI</i>	<i>Bsf0I</i>	<i>AccI</i>	<i>MseI</i>	<i>AluI</i>	<i>C/T</i>
Pre-HV	-	-	-	+	+	+	-	-	-	-	+	+	-	-	-	+	-	-
HV or pre*V	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
V	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
U	-	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-
K	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
J	+	+	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
T	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
I	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-
X	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-
W	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
M	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-
L1 or L2	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
L3 (?)	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
R*	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-
N*	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-

a. All the positions were analyzed by RFLP analysis, except position 12705, which was analyzed by sequencing. Plus signs indicate the presence of cleavage; minus signs indicate the absence of cleavage.

Primers and PCR conditions for the analysis of each polymorphic site are reported in Table 2. Primers for the analysis of positions 4577 *Nla*III (Izagirre 1998), 9052 *Hae*II (Montiel 2001; Montiel et al. 2001), 11719 *Sma*I (Saillard et al. 2000), 12308 *Hinf*I (Izagirre 1998), 13704 *Bst*0I (Montiel 2001; Montiel et al. 2001), and 15606 *Alu*I (Montiel 2001; Montiel et al. 2001) were available from the literature. Primers for the analysis of positions 7025 *Alu*I and 14465 *Acc*I were provided by P. Francalacci and A. Torroni, respectively. The rest of the primers (for the analysis of positions 1715 *Dde*I, 3592 *Hpa*I, 4216 *Nla*III, 8994 *Hae*III, 10032 *Alu*I, 10394 *Dde*I, 10397 *Alu*I, 10871 *Mnl*I, 14766 *Mse*I, and 12705 C/T) were designed in our laboratory (Table 2).

The PCR mix for each sample consisted of 1 U of *Taq* DNA polymerase, 50 pmol of each primer, 200 μ M of each dNTP, 2 mM MgCl₂, 67 mM Tris-HCl, 160 mM (NH₄)₂SO₄, 0.1% Tween-20, and 100 ng DNA in a final volume of 25 μ l. The reaction program involved an initial denaturation step (5 min) followed by 40 cycles of PCR (50 s at 94°C, 1 min at annealing temperature, and 1 min at 72°C) and a final extension step of 5 min at 72°C.

The detection of coding region polymorphisms (Table 2) was performed using restriction enzyme analysis, except for the polymorphic position 12705 C/T, which was sequenced.

Restriction Enzyme Analysis. Six microliters of PCR product were used in restriction enzyme digestions (Table 2), and the enzymatic cleavage was performed according to the manufacturer's specifications. The results of the restriction analysis were resolved by electrophoresis in agarose gels (2%).

Sequencing of Coding Region Fragments. Position 12705 C/T was analyzed by sequencing the region that encompasses position 12705, using primers described in Table 2.

PCR products were purified with the QIAquick PCR purification kit (Qiagen). Sequence reactions were carried out using the sequencing kit BigDye Terminator v. 3 (Applied Biosystems) and were run in an ABI Prism 3100 sequencer.

HVSI Amplification and Sequencing. Sequencing was performed for all samples, except for those previously sequenced by Pires (1999). A 443-bp fragment from the HVSI region, located between positions 15978 and 16420, numbered according to Anderson et al. (1981), was amplified using the primers L15997 (5'-CACCATTAGCACCCAAAGCT-3') and H16401 (5'-TGATTTACGGAGGATGGTG-3') and the PCR conditions described by Pereira et al. (2000). Purification and sequence reactions were performed as described for the coding region fragments.

Data Analysis. Sequences were aligned and manually checked using BioEdit (Hall 1999), and all polymorphic positions were confirmed in chromatograms.

Table 2. Primers, Annealing Temperatures, PCR Fragment Sizes, and Digestion Fragments for Each Polymorphic Site

Polymorphic Site	PCR		Digestion			
	Primers	Annealing Temperature (°C)	Fragment Size (bp)	Enzyme		
1715 <i>DdeI</i> (1719G)	L1671: 5'-ACTTAACTTGACCGCTCTGAG-3' H1837: 5'-CAITATGCAGAAAGGTATAGGGG-3'	52	208	<i>DdeI</i>	37	143 bp; 65 bp
3592 <i>HpaI</i> (3594C)	L3388: 5'-CTAGGCTATATACAACTACGC-3' H3717: 5'-GGCTACTGCTCGCAGTG-3'	50.9	330	<i>HpaI</i>	37	207 bp; 123 bp
4216 <i>NlaIII</i>	L4153: 5'-ATACCCCGATTCCGGCTACG-3' H4396: 5'-ATGGGGTGTGATAGGTGGC-3'	53.5	266	<i>NlaIII</i>	37	177 bp; 89 bp
4577 <i>NlaIII</i> (4580A)	L4538: 5'-CACTCATCACAGCGCTAAGC-3' H4621: 5'-TGGCAGCTTCTGTGGAAC-3' (Zagierre 1998)	51	120	<i>NlaIII</i>	37	62 bp; 58 bp
7025 <i>AluI</i> (7028C)	L6949: 5'-CCGTAGGTGGCCCTGACTGGC-3' H7052: 5'-TGATGGCAAATACAGCTCCT-3' (Francalacci, personal communication, 1993)	56	123	<i>AluI</i>	37	Nonpolymorphic: 108 bp; 15 bp Polymorphic: 78 bp; 30 bp; 15bp
8994 <i>HaeIII</i>	L8880: 5'-ACAGTGATTATAGGCTTTCGC-3' H9042: 5'-GTGGCGCTTCCAATTAGG-3'	51.5	200	<i>HaeIII</i>	37	Nonpolymorphic: 167 bp; 33 bp Polymorphic: 136 bp; 33 bp; 31 bp
9052 <i>HaeII</i>	L8997: 5'-AACCAATAGCCCTGGCC-3' H9121: 5'-GCGAATTC TAGGATAGTCAG-3' (Montiel 2001; Montiel et al. 2001)	49	160	<i>HaeII</i>	37	84 bp; 76 bp
10032 <i>AluI</i> (10034C)	L9988: 5'-CTCCATCTATTGATGAGGGTTC-3' H10152: 5'-CTCGTAAAGGGGTGATTTTTC-3'	52	205	<i>AluI</i>	37	139 bp; 66 bp

10394 <i>DdeI</i> (10398G)	L10252: 5'-TTGATCTAGAAAATTGCCCTC-3' HI0527: 5'-GTATTCCTAGAAAGTGAGATG-3'	48.2	276	<i>DdeI</i>	37	Nonpolymorphic: 171 bp; 105 bp Polymorphic: 133 bp; 105 bp; 38 bp 147 bp; 129 bp
10397 <i>AfaI</i> (10400T)	L10252: 5'-TTGATCTAGAAAATTGCCCTC-3'	48.2	276	<i>AfaI</i>	37	
10871 <i>MnlI</i>	HI0527: 5'-GTATTCCTAGAAAGTGAGATG-3' L10727: 5'-CTCAATCTCCAACACATATGGC-3' HI0920: 5'-GGTCGGAGGAAAAGGTTG-3'	51	232	<i>MnlI</i>	37	176 bp; 56 bp
11719 <i>SmaI</i>	HI1718: 5'-CGCAGTCATCTCATAATCGCCCCCGG-3' L11852: 5'-GGGGGTAAGCGGAGGTTAGC-3' (Saillard et al. 2000)	58.5	180	<i>SmaI</i>	25	155 bp; 25 bp
12308 <i>HinfI</i>	L12116: 5'-CACAGAAGCTGCTAACTCATGC-3' HI12338: 5'-ATTACTTTTATTTGGAGTTGCACCAAGATT-3' (Izaguirre 1998)	53	123	<i>HinfI</i>	37	93 bp; 30 bp
13704 <i>Bsr0I</i>	L13627: 5'-TCGAATAATCTTCTCACCC-3' HI13725: 5'-TAGTAATGAGAAATCCTCGG-3' (Montiel 2001; Montiel et al. 2001)	47	137	<i>Bsr0I</i>	60	98 bp; 39 bp
14465 <i>AccI</i>	(Torroni, personal communication, 2001)	54.5	733	<i>AccI</i>	37	529 bp; 204 bp
14766 <i>MseI</i>	L14603: 5'-CTAAACCCCAATAATAGGAG-3' HI14791: 5'-AGGTCGATGAATGAGTGG-3'	50	226	<i>MseI</i>	65	Nonpolymorphic: 201 bp; 21 bp; 4 bp Polymorphic: 184 bp; 21 bp; 17 bp; 4 bp
15606 <i>AfaI</i>	L15503: 5'-AGACCTCCTAGGCGACC-3' HI15620: 5'-GATGGATAGTAATFAGGGCAAG-3' (Montiel 2001; Montiel et al. 2001)	49	154	<i>AfaI</i>	37	121 bp; 33 bp
12705 C/T	L12553: 5'-ACAACCCAGCTCCTCCTAAG-3' HI13127: 5'-TGGAAGCGGATGAGTAAGAA-3'	55	635			

Sequences without ambiguities were obtained between positions 16050 and 16399.

The classification of all the samples, by assigning them to haplogroups, was performed using the information of coding region polymorphic positions and HVSI sequence polymorphisms, both alone and combined. The nomenclature used to classify the samples is summarized by Richards et al. (2000) and is available online at <http://www.stats.ox.ac.uk/~macaulay/founder2000/motif.html> (Macaulay 2000).

Results

Primers and PCR Conditions. Table 2 reports the primer sequences and annealing temperatures for each polymorphic site with which the most efficient PCR reactions were obtained. PCR fragment sizes and digestion fragments are also presented. In some amplified fragments (which include sites 7025 *AluI*, 8994 *HaeIII*, 10394 *DdeI*, and 14766 *MseI*) nonpolymorphic restriction sites exist, producing cleaved fragments after digestion even in the absence of the polymorphic site. Although in these cases the digestion fragments can be difficult to resolve, they provide an internal control for digestion. This is helpful in cases in which the polymorphism relevant for determining a haplogroup consists in the absence of the restriction site, because a failure in the reaction could result in a false-positive result. This is particularly important for position 7025 *AluI*, because this position alone defines haplogroup H, the most frequent haplogroup in European populations. The absence of cleavage in this position defines an individual as haplogroup H; therefore an internal control is essential to warrant the efficiency of the enzymatic reaction and to exclude the possibility of wrong classification of an important number of samples.

The best results from resolved digestion fragments that differ by less than 30 bp relative to the amplified fragment (e.g., 11719 *SmaI*) and the best results between nonpolymorphic and polymorphic restriction fragments (e.g., 8994 *HaeIII*) were archived using NuSieve GTG agarose gels.

Coding Region Analysis of mtDNA. The results of the coding region analysis are presented in Table 3 [adapted from Santos et al. (2003)] and show the positions that were analyzed in each sample and the haplogroup inferred, both by means of coding region polymorphisms and HVSI sequencing. Sequences are available in Santos et al. (2003).

Positions 10394 *DdeI* and 10871 *MnII* were analyzed for all samples with the main purpose of detecting possible non-West Eurasian contributions to the mtDNA gene pool of the study population. The use of both positions was undertaken to double-check this segregation because position 10394 *DdeI* (the first analyzed in our sample) was described as an unstable position (Macaulay et al.

Table 3. Results of the Analysis of Coding Region Polymorphisms and Haplogroup Inferred from Both Coding Region and HVSI Polymorphisms^a

Sample	Haplogroup		Coding Region																	Coding Region Haplogroup	
	HVSI (Santos <i>et al.</i> , 2003)		10871	10394	14766	12308	7025	4216	4577	9052	8994	1715	10032	14465	11719	13705	15606	10397	3592		12705
			<i>MnII</i>	<i>DdeI</i>	<i>MseI</i>	<i>HinfI</i>	<i>AluI</i>	<i>NlaIII</i>	<i>NlaIII</i>	<i>HaeII</i>	<i>HaeIII</i>	<i>DdeI</i>	<i>AluI</i>	<i>AccI</i>	<i>SmaI</i>	<i>BstOI</i>	<i>AluI</i>	<i>AluI</i>	<i>HpaI</i>	<i>C/T</i>	
SM1	M1?	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	M
SM2	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
SM3	U3	+	-	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	U
SM4	H or HV or U or R	+	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-	H
SM5	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
SM6	U5a or U5b	+	-	+	+	+	+	-	+	+	+	+	-	-	-	-	-	-	-	-	U
SM7	J	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	J
SM8	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	H
SM9	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	H
SM10	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
SM11	N1b	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	T	N*
SM12	X	+	-	+	+	-	+	-	+	+	+	+	-	+	-	-	-	-	-	-	X
SM13	T2	+	-	+	-	+	+	+	-	-	-	-	-	-	-	-	+	-	-	-	T
SM14	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
SM15	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
SM16	L2	-	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	L1 or L2
SM17	K	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	K
SM18	K	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	K
SM19	J1	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	J
SM20	U3 or U4	+	-	+	+	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	U
SM21	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
SM22	U6a	+	-	+	+	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	U
SM23	U5 or K	+	-	+	+	+	+	-	-	+	+	+	-	-	-	-	-	-	-	-	U
SM24	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H
SM25	J?	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	J
SM26	T3	+	-	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	T
SM27	U5b	+	-	+	+	+	+	-	-	+	+	+	+	+	-	-	+	-	-	-	U

1999). The other positions were analyzed in an order based mainly on the frequency of haplogroups in European populations. Finally, some confirmatory positions were tested for some samples. This nonoptimized approach leads to the screening of redundant positions in classifying some samples. For example, in European populations haplogroup H is the most frequent (Table 4), and as shown in Table 3, the position 7025 *AluI* (specific to haplogroup H) was tested in almost all the analyzed samples, although this analysis was unnecessary in many cases.

The site 1715 *DdeI* was initially used to define haplogroup X. However, two samples (SM11 and A7) characterized as haplogroup X did not present the corresponding HVSI motif. On the contrary, sample SM12, which already has an HVSI X motif, resulted in +1715 *DdeI* (no X). These results confirm the instability of this polymorphism, as previously noted by Macaulay et al. (1999). Therefore the inclusion of position 14465 *AccI* to correctly type haplogroup X was necessary, as suggested by Macaulay et al. (1999).

Three samples (SM11, A7, and ACR45) out of the 146 analyzed using RFLPs were classified as "other." However, by screening position 12705, we were able to assign two of these samples (SM11 and A7) to superhaplogroup N*. The remaining sample (ACR45) was classified as R*.

Haplogroups HV and pre-*V cannot be resolved by analyzing coding region polymorphisms (Torroni et al. 2001) because the information of HVSI sequencing is necessary to distinguish samples belonging to these clusters. This is the case for samples F21 and F31, which are characterized as pre-*V.

Sample ACR25 was classified as belonging to subhaplogroup K using RFLPs (−9052 *HaeII* and +12308 *HinfI*). However, it presents an HVSI motif (16278T, 16311C) that appears to be typical of haplogroup R1 (Macaulay et al. 1999), although Macaulay et al. (1999) remarked that it is impossible to identify this cluster on the basis of HVSI alone. To exclude the possibility of mischaracterization, we confirmed the results of the RFLP analysis by sequencing the 9052 *HaeII* and 12308 *HinfI* regions for this sample using new primers L8715 (5'-TAAAGGACGAACCTGATCTC-3')/H9372 (5'-GGTATATGGTTAGTGTGTTG-3') and L11978 (5'-TCCCTCTACATATTTACCAC-3')/H12645 (5'-GAATTC-TATGATGGACCATG-3'), respectively. Therefore the absence of 16224C (which with 16311C defines subhaplogroup K) in sample ACR25 is probably due to a retromutation.

The previously described situations led us to develop and validate a hierarchical approach that allows a more rapid, organized and cost-efficient haplogroup screening procedure.

A Hierarchical Approach. In the population analyzed (Santos et al. 2003), all West Eurasian haplogroups (Torroni et al. 1996; Macaulay et al. 1999) were found in similar frequencies to those observed in other European populations (Table 4 and references therein), with the exception of haplogroup H, which is present at a lower frequency in the Azores (Table 4). Most of the mtDNA lineages appear to be West Eurasian; however, a few of the lineages belonging to L1 or

Table 4. mtDNA Haplogroup Distribution in the Azores (Portugal) and in European and North African Populations

Haplotype	Azores ^a (N = 146)	West Mediterranean ^b (N = 217)	Central Mediterranean ^c (N = 302)	East Mediterranean ^d (N = 167)	Southeast Europe ^e (N = 233)	Northwest Europe ^f (N = 456)	North-Central Europe ^g (N = 332)	Northeast Europe ^h (N = 407)	North Africa ⁱ (N = 349)
H	32.2	48.8	42.4	42.5	38.2	48.5	47.9	41.8	31.8
Pre-HV	0.7	1.4	0.3	1.2	1.7	0.2	0.0	0.0	0.0
V, pre-*V, and HV	11.6	6.0	7.0	4.8	6.5	6.3	6.0	4.6	4.9
I	2.1	0.9	2.0	4.2	1.7	3.1	2.1	2.7	0.0
J	10.3	7.4	7.5	12.6	8.2	13.8	9.3	7.8	4.4
T	14.4	6.0	9.9	6.0	11.2	7.3	10.2	7.6	5.2
U	10.9	15.8	15.2	16.2	20.1	9.8	12.6	22.3	22.4
K	5.5	5.5	7.0	3.6	5.2	7.7	6.0	4.7	3.4
W	2.1	2.8	1.3	1.2	3.0	1.1	1.5	4.4	0.6
X	2.7	1.4	3.3	4.2	1.3	0.9	2.1	0.7	1.1
L1 and L2	2.1	0.9	1.0	0.6	0.2	0.2	0.3	0.0	17.5
L3	0.7	1.4	0.6	0.0	0.4	0.2	0.0	0.5	4.3
M	2.7	0.5	0.3	1.2	1.3	0.4	0.0	1.1	2.3
N*, R*, other	2.1	1.4	2.0	1.8	1.2	0.4	1.8	1.4	2.3

- a. This study and Santos et al. (2003).
- b. Crte-Real et al. (1996), Salas et al. (1998), Bertranpetit et al. (1995), and Richards et al. (2000). Data extracted from <http://www.stats.ox.ac.uk/~macaulay/founder2000/>
- c. Francalacci et al. (1996), Torroni et al. (1998), Di Rienzo and Wilson (1991), and Richards et al. (2000). Data extracted from <http://www.stats.ox.ac.uk/~macaulay/founder2000/>
- d. Belleli et al. (2000) and Richards et al. (2000). Data extracted from <http://www.stats.ox.ac.uk/~macaulay/founder2000/>
- e. Calafell et al. (1996) and Richards et al. (2000). Data extracted from <http://www.stats.ox.ac.uk/~macaulay/founder2000/>
- f. Piercy et al. (1993) and Richards et al. (1996, 2000). Data extracted from <http://www.stats.ox.ac.uk/~macaulay/founder2000/>
- g. Richards et al. (1996, 2000) and Hofmann et al. (1997). Data extracted from <http://www.stats.ox.ac.uk/~macaulay/founder2000/>
- h. Sajantila et al. (1995, 1996) and Richards et al. (1996, 2000). Data extracted from <http://www.stats.ox.ac.uk/~macaulay/founder2000/>
- i. Crte-Real et al. (1996), Macaulay et al. (1999), Rando et al. (1998), Watson et al. (1997), and Larruga [personal communication; cited in Gonzlez et al. (2003)]. Data extracted from Gonzlez et al. (2003).

L2, L3, and M* clusters were also detected in the Azores, giving evidence of a mixed composition of mtDNAs in the archipelago.

In general, in all European populations studied so far, almost all mtDNAs belong to West Eurasian haplogroups. However, a small number of African lineages were detected in the different regions of Europe (Table 4). In African white populations the mtDNA presents a mixed composition of West Eurasian and African haplogroups, but with African haplogroups represented at a higher frequency than in European populations (Table 4). Because major and minor clusters found in West Eurasian and African Caucasian populations are represented in the Azores, this population appears to be appropriate for developing a hierarchical scheme (Figure 1) to analyze mtDNA haplogroups in these populations.

The first step in the hierarchical scheme, which includes the analysis of the site 14766 *MseI* (Figure 1) in all the samples, allows two major groups of haplogroups to be separated. The remaining positions are tested following the hierarchical scheme proposed in Figure 1, which takes into account the frequencies of haplogroups in European populations. Samples are tested only for the restriction sites required, depending on the results obtained in the previous step.

There are no restriction sites to distinguish between haplogroups HV and pre-*V. Therefore, to discriminate between them, it is necessary to sequence HVSI, because haplogroup pre-*V has a mutation at position 16298 that is absent in haplogroup HV (Torroni et al. 2001).

Position 10394 *DdeI* is used to distinguish haplogroup J from haplogroup T. However, because position 10394 is prone to new mutational events (Macaulay et al. 1999), it is advisable to use confirmatory positions, as suggested in the hierarchical scheme.

Samples classified as "other" at the end of step 7 of the hierarchical scheme represent African, Asian, and minor West Eurasian haplogroups that can be found in European populations. At this point, site 10871 *MnII*, which appears to distinguish clearly between the non-M Eurasian mtDNA types (Quintana-Murci et al. 1999), must be analyzed with the main purpose of differentiating between West Eurasian and non-West Eurasian mtDNAs.

Samples with -10871 *MnII*, which represent non-West Eurasian lineages, are subsequently tested for 3592 *HpaI* with the main purpose of distinguishing haplogroups L1 and L2 from haplogroups L3 and M; samples with -3592 *HpaI* are analyzed for 10397 *AluI*, which distinguishes haplogroup M and potentially haplogroup L3. To better classify non-West Eurasian mtDNAs, it is advisable to sequence the HVSI region. This is particularly useful when a small number of samples are analyzed, allowing faster and less expensive typing compared to the alternative analysis of additional restriction sites.

Samples with +10871 *MnII* belong to macrogroups R* and N* and can be separated by sequencing the region that encompasses position 12705. As suggested for non-West Eurasian lineages, it is also advisable to sequence samples classified as R* and N* to obtain a more accurate classification of these samples.

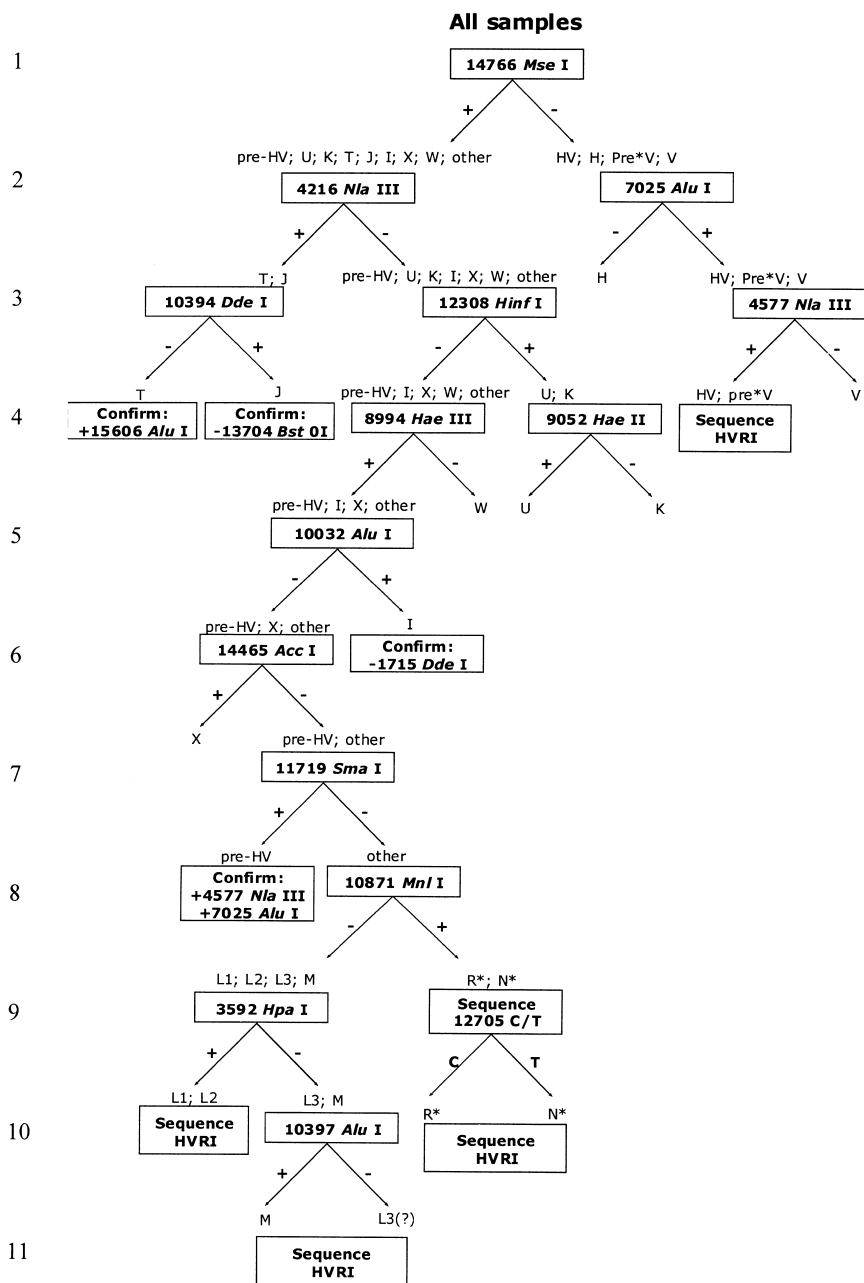


Figure 1. Hierarchical scheme to determine mtDNA haplogroups in Caucasian populations with a low frequency of non-West Eurasian haplogroups.

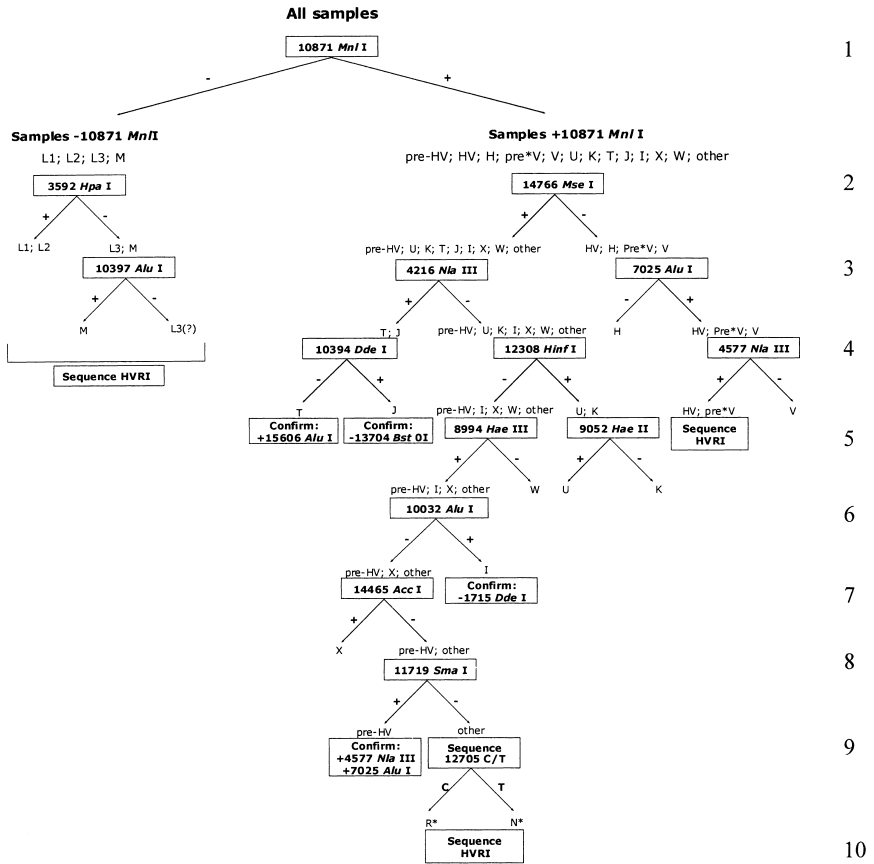


Figure 2. Hierarchical scheme to determine mtDNA haplogroups in Caucasian populations with a high frequency of non-West Eurasian haplogroups.

In Caucasian populations with a high frequency of non-West Eurasian haplogroups (e.g., North African populations), the scheme presented in Figure 2 could be used. In this scheme step 1 (10871 *MnlI*) allows the discrimination between West Eurasian and non-West Eurasian haplogroups.

Discussion and Conclusions

Concordance between HVSI and coding region haplogroups was observed for all the samples analyzed in this study, except for sample ACR25, which was previously discussed. This is in accordance with results published elsewhere [e.g., Torroni, Sukernik et al. (1993), Torroni, Schurr et al. (1993), Graven et al. (1995), Torroni et al. (1996), Kolman et al. (1996), and Macaulay et al. (1999)].

The observed concordance supports the reliability and reproducibility of the determination of mtDNA haplogroups by RFLP analysis.

The observations that mtDNA lineages of sample ACR25 could have suffered a retromutation at position 16224 and that two samples (SM23 and SM28) belonging to haplogroup U (+9052 *Hae*II and +12308 *Hinf*I, confirmed by sequencing) showed an HVSI motif attributed to both haplogroups U5 and K (16224, 16270, 16311) indicate that position 16224 could be an unstable position and probably is not reliable as a diagnostic position to determine HVS haplogroups. This stresses the need to increase the number of samples in which both coding region polymorphisms and sequencing of HVS are performed to better establish the reliability of the correspondence between coding region polymorphisms and HVS sequence haplogroups. The accuracy of this correspondence is important because it allows the use of previous published information for comparisons and could be decisive in authenticating ancient mtDNA (Montiel 2001).

The classification of samples into mtDNA haplogroups is sometimes impossible in the absence of complete information. A clear example is presented in Table 3. When only HVSI information is used, almost 50% (45.89%) of the samples cannot be assigned to only one haplogroup; on the other hand, when 18 coding region polymorphisms are used, 96.55% of the samples can be classified into only one haplogroup. By combining the two types of analysis, 100% of the samples are classified.

In this paper a hierarchical approach to defining mtDNA haplogroups is proposed. This method allows the characterization of almost all major West Eurasian and African clusters by means of RFLPs. Furthermore, the proposed scheme includes information on the situations in which sequencing of coding region fragments or of HVSI is pertinent to define some particular haplogroups.

Our scheme avoids the sequencing of HVSI and the screening of all coding region polymorphisms for all samples, resulting in a more rapid and less expensive analysis. Furthermore, when this scheme is used to characterize mtDNA from West Eurasian and African Caucasian populations, it allows 100% of the samples to be assigned to one haplogroup.

The validity of the scheme is assured by:

1. Using relatively stable polymorphic positions: From the 18 control region polymorphic positions used in the hierarchical scheme, only 2 (1715 *Dde*I and 10394 *Dde*I) were previously described as unstable positions (Macaulay et al. 1999), that is, positions that have suffered homoplasic mutations. Furthermore, in the proposed scheme these positions were never used alone to define a specific haplogroup or to separate groups of haplogroups.
2. Screening more than one position to define a specific haplogroup: Even if a haplogroup can be defined with only one polymorphic position (e.g., haplogroup H), the previous analysis of additional positions decreases the probability of wrong classification.
3. Typing confirmatory positions.

The robustness of the proposed approach was assessed by sequencing all samples for HVSI. The concordance between results obtained by analyzing coding region polymorphisms and sequencing HVSI guarantees an accurate classification of the samples used to develop the hierarchical scheme presented here.

The scheme presented here can also be used to complete the analysis of samples in which only HVS sequencing has been performed, because it is possible to select adequate positions from the scheme to confirm haplogroups.

Even if this approach leads to a lower level of phylogeographic resolution than the sequencing of all samples, the information obtained can be good enough, depending on the goals of the study; for example, the scheme is perfectly applicable to the study of population movements on a continental level or to the quantification of the proportion of admixture of one population. Furthermore, in contrast to automated sequencing, our scheme is a simple method that can be applied with low costs in any laboratory.

Acknowledgments We thank Antonio Torroni for technical support. This study was supported by a grant (ref. SFRH/BD/723/2000) from Fundação para a Ciência e a Tecnologia [Ministério da Ciência e Ensino Superior (MCES), Portugal] with funds from the “Operational Programs Science Technology and Innovation” (POCTI) and the “Society of Information” (POSI).

Received 22 October 2003; revision received 5 January 2004.

Literature Cited

- Anderson, S., A. Bankier, M. Barrell et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–464.
- Anglés, N., R. Montiel, R. Nogués et al. 2003. Haplogrupos mitocondriales en tres poblaciones caucasoides. In *Antropología y Biodiversidad*, M.P. Aluja, A. Malgosa, and R.M. Nogués, eds. Barcelona, Spain: Edicions Bellaterra, v. 2, 26–35.
- Bailliet, G., F. Rothhammer, F.R. Carnese et al. 1994. Founder mitochondrial haplotypes in Amerindian populations. *Am. J. Hum. Genet.* 55:27–33.
- Ballinger, S.W., T.G. Schurr, A. Torroni et al. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* 130:139–152.
- Belledi, M., E.S. Poloni, R. Casalotti et al. 2000. Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. *Eur. J. Hum. Genet.* 8:480–486.
- Bertranpetit, J., J. Sala, F. Calafell et al. 1995. Human mitochondrial DNA variation and the origin of Basques. *Ann. Hum. Genet.* 59:63–81.
- Calafell, F., P. Underhill, A. Tolun et al. 1996. From Asia to Europe: Mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann. Hum. Genet.* 60:35–49.
- Cann, R., M. Stoneking, and A. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.
- Chen, Y.S., A. Olckers, T. Schurr et al. 2000. MtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am. J. Hum. Genet.* 66:1,362–1,383.

- Chen, Y.S., A. Torroni, L. Excoffier et al. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* 57:133–149.
- Comas, D., F. Calafell, E. Mateu et al. 1997. Mitochondrial DNA variation and the origin of the Europeans. *Hum. Genet.* 99:443–449.
- Côrte-Real, H., V. Macaulay, M. Richards et al. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann. Hum. Genet.* 60:331–350.
- Di Rienzo, A., and A.C. Wilson. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 88:1,597–1,601.
- Francalacci, P., J. Bertranpetit, F. Calafell et al. 1996. Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am. J. Phys. Anthropol.* 100(4):443–460.
- González, A., A. Brehm, J. Pérez et al. 2003. Mitochondrial DNA affinities at the Atlantic finger of Europe. *Am. J. Phys. Anthropol.* 120:391–404.
- Graven, L., G. Passarino, O. Semino et al. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka population. *Mol. Biol. Evol.* 12:334–345.
- Hagelberg, E. 1997. Ancient and modern mitochondrial DNA sequences and the colonization of the Pacific. *Electrophoresis* 18:1,529–1,533.
- Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41:95–98.
- Hofmann, S., M. Jaksch, R. Bezold et al. 1997. Population genetics and disease susceptibility: Characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants, and association with disease. *Hum. Mol. Genet.* 6(11):1,835–1,846.
- Izagirre, N. 1998. *Aplicación de la biología molecular en el estudio antropológico de las poblaciones humanas del pasado: Análisis del ADN mitocondrial*. Ph.D. dissertation, Basque Country University, Basque Country, Spain.
- Kolman, C., N. Sambuughin, and E. Bermingham. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1,321–1,334.
- Lamminen, T., K. Huoponen, P. Sistonen et al. 1997. mtDNA haplotype analysis in Finnish families with leber hereditary optic neuropathy (LHON). *Eur. J. Hum. Genet.* 5:271–279.
- Lum, J., and R. Cann. 2000. mtDNA lineage analyses: Origins and migrations of Micronesians and Polynesians. *Am. J. Phys. Anthropol.* 113:151–168.
- Lum, J., O. Richards, C. Ching et al. 1994. Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum. Biol.* 66:567–590.
- Macaulay, V. 2000. *Haplogroup Motifs*. Retrieved October 8, 2001, from <http://www.stats.ox.ac.uk/~macaulay/founder2000/motif.html>
- Macaulay, V., M. Richards, E. Hickey et al. 1999. The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* 64:232–249.
- Mendonça, L. 1996. *História dos Açores: Visão geral (sécs. XV–XIX)*. Ponta Delgada, Azores, Portugal: Centro de Apoio Tecnológico à Educação.
- Montiel, R. 2001. *Estudio diacrónico de la variabilidad del DNA mitocondrial en población Catalana*. Ph.D. dissertation, Autonomous University of Barcelona, Barcelona, Spain. Also available from <http://www.tdcat.cesca.es/TDCat-0726101-095837>
- Montiel, R., A. Malgosa, and P. Francalacci. 2001. Authenticating ancient human mitochondrial DNA. *Hum. Biol.* 73(5):689–713.
- Morelli, L., M. Grosso, G. Vona et al. 2000. Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. *Hum. Biol.* 72(4):585–595.
- Pereira, L., M.J. Prata, and A. Amorim. 2000. Diversity of mtDNA lineages in Portugal: Not a genetic edge of European variation. *Ann. Hum. Genet.* 64:491–506.

- Piercy, R., K. Sullivan, N. Benson et al. 1993. The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int. J. Legal Med.* 106:85–90.
- Pires, L. 1999. *A Origem da população dos Açores com base em estudos de DNA mitocondrial*. B.Sc. thesis, University of Coimbra, Coimbra, Portugal.
- Quintana-Murci, L., O. Semino, H.-J. Bandelt et al. 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nature Genet.* 23:437–447.
- Rando, J.C., F. Pinto, A.M. González et al. 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.
- Richards, M., H. Côrte-Real, P. Forster et al. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 59:185–203.
- Richards, M., V. Macaulay, H.J. Bandelt et al. 1998. Phylogeography of mitochondrial DNA in Western Europe. *Ann. Hum. Genet.* 62:241–260.
- Richards, M., V. Macaulay, E. Hickey et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67:1,251–1,276.
- Saillard, J., P.J. Magalhaes, M. Schwartz et al. 2000. Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. *Hum. Biol.* 72(6):1,065–1,068.
- Sajantila, J., P. Lahermo, T. Anttinen et al. 1995. Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Res.* 5:42–52.
- Sajantila, J., A.H. Salem, P. Savolainen et al. 1996. Paternal and maternal mtDNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc. Natl. Acad. Sci. USA* 93:12,035–12,039.
- Salas, A., D. Comas, M. Laureu et al. 1998. mtDNA analysis of the Galician population: A genetic edge of European variation. *Eur. J. Hum. Genet.* 6:365–375.
- Santos, C. 2001. *Haplogrupos del DNAm t en la población de las Azores (Portugal)*. M.Sc. thesis, University Autónoma of Barcelona, Barcelona, Spain.
- Santos, C., M. Lima, R. Montiel et al. 2003. Genetic structure and origin of peopling in the Azores Islands (Portugal): The view from mtDNA. *Ann. Hum. Genet.* 67:433–456.
- Silva, W., S. Bonatto, A. Holanda et al. 2002. Mitochondrial genome diversity of Native Americans supports a single early entry of founders into America. *Am. J. Hum. Genet.* 71:187–192.
- Simoni, L., F. Calafell, D. Pettener et al. 2000. Geographic patterns of mtDNA diversity in Europe. *Am. J. Hum. Genet.* 66:262–278.
- Torroni, A., H.-J. Bandelt, L. D'Urbano et al. 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am. J. Hum. Genet.* 62:1,137–1,152.
- Torroni, A., H.-J. Bandelt, V. Macaulay et al. 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am. J. Hum. Genet.* 69:844–852.
- Torroni, A., K. Huoponen, P. Francalacci et al. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1,835–1,850.
- Torroni, A., M. Lott, M. Cabell et al. 1994. mtDNA and the origin of Caucasians: Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am. J. Hum. Genet.* 55:760–776.
- Torroni, A., J. Miller, L. Moore et al. 1994. Mitochondrial DNA analysis in Tibet: Implications for the origin of the Tibetan populations and its adaptation to high altitude. *Am. J. Phys. Anthropol.* 93:189–199.
- Torroni, A., T. Schurr, M. Cabell et al. 1993. Asian affinities and continental radiation of the four founding Native Americans mtDNAs. *Am. J. Hum. Genet.* 53:563–590.
- Torroni, A., T. Schurr, C. Yang et al. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153–162.
- Torroni, A., R. Sukernik, T. Schurr et al. 1993. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* 53:591–608.

- Torrioni, A., and D. Wallace. 1994. Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *J. Bioenerg. Biomembr.* 26:261–271.
- Wallace, D. 1995. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am. J. Hum. Genet.* 57:201–223.
- Watson, E., P. Foster, M. Richards et al. 1997. Mitochondrial footprints of human expansion in Africa. *Am. J. Hum. Genet.* 61:691–704.

IV. 2

Artículo II

Genetic Structure and Origin of Peopling in The Azores Islands (Portugal): The View from mtDNA

Santos *et al.* (2003) ***Annals of Human Genetics***,
67: 433-456

Genetic Structure and Origin of Peopling in The Azores Islands (Portugal): The View from mtDNA

C. Santos^{1,2,*}, M. Lima³, R. Montiel^{3,1}, N. Angles¹, L. Pires², A. Abade² and Ma. P. Aluja¹

¹Unity of Anthropology, Department BABVE, Faculty of Sciences, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain

²Department of Anthropology, University of Coimbra, 3000 Coimbra, Portugal

³CIRN and Department of Biology, University of the Azores, 9500 Ponta Delgada, S. Miguel, Azores, Portugal

Summary

The Azores islands (Portugal), uninhabited when discovered by Portuguese navigators in the fifteenth century, are located in the Atlantic Ocean 1500 km from the European mainland. The archipelago is formed by nine islands of volcanic origin that define three geographical groups: Eastern (S. Miguel and Sta. Maria), Central (Terceira, Faial, Pico, Graciosa and S. Jorge) and Western (Flores and Corvo). To improve the genetic characterisation of the Azorean population, and to clarify some aspects related to the history of settlement, a study of mtDNA was conducted in the population of the archipelago. The HVRI region was sequenced and specific RFLPs were screened in 146 samples obtained from unrelated individuals with Azorean ancestry (50 from the Eastern group, 60 from the Central group, and 37 from the Western group). Samples were classified into haplogroups based on the information obtained from both sequencing and RFLP analysis.

All the analyses performed support the idea that, in the whole group of islands, the majority of mtDNA lineages originated from the Iberian Peninsula, mainly from Portugal (mainland). However contributions from other European populations, especially from Northern Europe, cannot be disregarded. The values obtained for the various diversity parameters in the Azores archipelago indicate that the Azorean population, as a whole, does not exhibit the typical characteristics of an isolated population. The analysis of genetic data by groups of islands showed that the Western group exhibited particular features. The distribution of haplogroups in the Western group is very atypical, being significantly different from what is observed in the Eastern and Central groups. Furthermore, the diversity values are, in general, lower than those observed in other populations used for comparison. African haplogroups were found in all the groups of islands. Therefore the presence of *Moorish* and *African* slaves on the islands, as reported in historical sources, is supported by the mtDNA genetic data, especially in the Eastern group. The presence of Jews in the Central group is also supported by the mtDNA data. Neither historical nor genetic data (phylogeography of mtDNA) supports the idea of a differential settlement history for the Western group; however, it is represented in the phylogenies as an isolated branch. The effect of genetic drift, induced by the reduced population size since peopling occurred, has led to a very atypical distribution of haplogroups/haplotypes in this group of islands.

We cannot ignore the influence of biodemographic and genetic processes, namely founder effect, genetic drift, migration, and even recent mutational events in the mtDNA lineages of the Azorean populations. Nevertheless, a great part of the variation in the Azorean mtDNA can be explained by the settlement history.

*Corresponding author: Cristina Santos, Unity of Anthropology, Department BABVE, Faculty of Sciences, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain, Fax: +34 935811321, e-mail: Cristina.Santos@uab.es

Introduction

The Azores archipelago (Portugal), located in the Atlantic Ocean, 1500 km from the European mainland, is formed by nine islands of volcanic origin whose relative positions define three geographical groups: Eastern (S. Miguel and Sta. Maria), Central (Terceira, Faial, Pico, Graciosa and S. Jorge) and Western (Flores and Corvo). With an area of 2344 km², the Azores presently has a total population of 237315 inhabitants, distributed in a very asymmetric way among the three groups of islands (134885 in the Eastern group, 98101 in the Central group, and only 4329 in the Western group) (INE, 2001).

Historical data indicates that the islands were uninhabited when discovered by Portuguese navigators in the fifteenth century. The peopling of the islands was a slow and difficult process, and was initiated in 1439 in the islands of Sta. Maria and S. Miguel; the small islands of Flores and Corvo were only inhabited towards the end of 16th century (Matos, 1989; Mendonça, 1996). According to historical records, the first settlers came mainly from various regions of mainland Portugal and from Madeira Island. However, people of different origin, such as Spanish, French, Italian, English, German and Flemish (whose important presence is always referred to in accounts of the peopling of the Central group), also made up part of the early settlers (Matos, 1989; Mendonça, 1996). There is clear evidence that Jews also contributed to the peopling of the archipelago. Sephardim Jews expelled from the Iberian Peninsula may have used the Azores as a refuge. The presence of *African* and *Moorish* slaves in the islands is also referred to in historical documents (Matos, 1989; Mendonça, 1996), being well-documented especially for the Western group (Gomes, 1997). However, the specific contribution of these settlers to the different islands is not clear, due to the lack of both historical and genetic data.

The complex process of peopling, the evolution of island populations, always affected by migration along with geographic isolation, has been the main stimulus for carrying out biodemographic and genetic studies on the Azorean population. To improve the genetic characterisation of this population and enhance our knowledge about several aspects related to its origin and dynamics, a research project that includes the analysis of nuclear

autosomal markers, Y chromosomal and mitochondrial DNA (mtDNA), is presently being undertaken by our group. Here we present the results of the sequencing of HVRI, and specific RFLP screening of mtDNA. Our main goals were: (i) to describe the variation of mtDNA in the Azorean population; (ii) to gather data in order to contribute to the establishment of the genetic profile of the archipelago; (iii) to shed more light on some aspects related to the history and dynamics of the settlement of the Azores.

Material and Methods

Samples

One hundred and forty-six unrelated individuals, born in the Azores and with Azorean ancestry, were sampled, including 42 previously tested by Pires (1999). The birthplace of the earliest known maternal ancestor was chosen as the sample location. Fifty individuals were originally from the Eastern Group, 60 from the Central Group, and 36 from the Western Group. The samples were taken by buccal swabs and existing samples collected by Pires (1999) by venupuncture. Voluntary donors were asked for informed consent and to fill in an anonymous inquiry concerning the birthplaces of all their known ancestors. The comparative data set consisted of 1493 mtDNA HVRI sequences from the populations listed in Table 1. Populations were selected, when possible, in order to include groups that are mentioned in historical documents as sources of settlers for the Azores islands.

DNA Extraction

Total DNA from buccal cells was extracted using Instagene matrix (BioRad) according to the manufacturer's specifications. DNA from blood samples was previously extracted by Pires (1999) using standard protocols.

MtDNA Analysis

RFLP Analysis

Relevant restriction sites for mtDNA haplogroup determination (Table 2) were selected from those

Table 1 Population (corresponding codes in Figure 2), sample size, references and source of the information used in this study

Population	N	Reference	Source
Azores – Eastern Group	50	This work	This study
Azores – Central Group	60	This work, Pires, 1999	This study, Provided by the author
Azores – Western Group	36	This work	This study
Mainland North Portugal (Pt)	100	Pereira <i>et al.</i> 2000	GenBank
Mainland Central Portugal (Pt)	82	Pereira <i>et al.</i> 2000	GenBank
Mainland South Portugal (Pt)	59	Pereira <i>et al.</i> 2000	GenBank
Spain (Mainland) (Sp)	89	Côrte-Real <i>et al.</i> 1996; Pinto <i>et al.</i> 1996	HVRbase (Handt <i>et al.</i> 1998) ¹
Galicia (Sp)	92	Salas <i>et al.</i> 1998	HVRbase (Handt <i>et al.</i> 1998) ¹
Basque Country (Sp)	106	Bertranpetit <i>et al.</i> 1995; Côrte-Real <i>et al.</i> 1996	HVRbase (Handt <i>et al.</i> 1998) ¹
France (Fr)	50	Rousselet & Mangin, 1998	HVRbase (Handt <i>et al.</i> 1998) ¹
UK (UK)	100	Piercy <i>et al.</i> 1993	HVRbase (Handt <i>et al.</i> 1998) ¹
Germany (Ger)	67	Hofmann <i>et al.</i> 1997	HVRbase (Handt <i>et al.</i> 1998) ¹
Austria (Aus)	101	Parson <i>et al.</i> 1998	HVRbase (Handt <i>et al.</i> 1998) ¹
Italy (It)	49	Francalacci <i>et al.</i> 1996	HVRbase (Handt <i>et al.</i> 1998) ¹
Israel Arabs/Palestinians (Isr)	117	Di Rienzo & Wilson, 1991; Richards <i>et al.</i> 2000	Database of Richards <i>et al.</i> 2000 ²
Algeria (Alg)	85	Côrte-Real <i>et al.</i> 1996	HVRbase (Handt <i>et al.</i> 1998) ¹
Moroccan Jews (MJ)	115	Thomas <i>et al.</i> 2002	Provided by the author
Morocco (Berber) (MBer)	60	Rando <i>et al.</i> 1998	Provided by the author
Morocco (Moroccans) (Mor)	32	Rando <i>et al.</i> 1998	Provided by the author
Senegal (Sen)	119	Graven <i>et al.</i> 1995 Rando <i>et al.</i> 1998	HVRbase (Handt <i>et al.</i> 1998) ¹ Provided by the author
Mauritania (Maur)	30	Rando <i>et al.</i> 1998	Provided by the author
S. Tome (ST)	50	Mateu <i>et al.</i> 1997	Article
Mozambique (Moz)	109	Pereira <i>et al.</i> 2001	Article

¹<http://www.hvrbase.de>²<http://www.stats.ox.ac.uk/~macaulay/founders2000/tableA.html>

proposed by Torroni *et al.* (1993, 1994, 1996), Torroni & Wallace (1994), Chen *et al.* (1995), Macaulay *et al.* (1999), Quintana-Murci *et al.* (1999) and Saillard *et al.* (2000). Some of the specific primers and PCR conditions for each polymorphic site were available from the literature, but others were designed in our laboratory (Montiel, 2001; Santos, 2001).

Six μ l of PCR product were used in restriction enzyme digestions and the enzymatic cleavage was performed according to the manufacturer's specifications. Results of restriction analyses were resolved through electrophoresis in agarose gels (2%).

HVRI Amplification and Sequencing

Sequencing was performed for all samples, except those previously sequenced by Pires (1999). A fragment of 443 bp from the HVRI, located between positions

15978 and 16420 (numbered according to Anderson *et al.* 1981), was amplified using primers and PCR conditions described by Pereira *et al.* (2000). PCR products were purified using the QIAquick PCR purification kit (Quiagen). Sequence reactions were carried out using the kit ABI PRISM BigDye Terminator version 3 and run in an ABI3100 sequencer (Perkin Elmer).

Data Analysis

Sequences were aligned and manually checked using BioEdit (Hall, 1999), and all polymorphic positions were confirmed in chromatograms. Sequences without ambiguities were obtained between positions 16050 and 16399. The classification of all the samples by assigning them to haplogroups was performed

Results and Discussion

Sequencing and RFLP Results

Data obtained for each sample from HVRI sequencing and RFLP analysis, as well as the haplogroup determined from the RFLP and HVRI motifs, are presented in the appendix. It was possible, using the combined information provided for both systems of mtDNA analysis, to classify all samples into one of the haplogroups described in the literature. This work confirms the previously established correlation between RFLPs and HVRI data (Torroni *et al.* 1996; Francalacci *et al.* 1999; Macaulay *et al.* 1999), and supports the idea that a correct and complete classification of mtDNA is only possible when both methods of analysis are used simultaneously.

Haplogroup Frequencies

The haplogroup frequencies observed in the total sample of the Azores, as well as the partial frequencies observed in the Eastern, Central and Western groups, are summarised in Table 3. The 146 samples were distributed among 31 different haplogroups/sub-haplogroups. Distinct numbers of clusters were found in the Eastern (21), Central (19) and Western (11) groups. The comparison of the haplogroup distribution in the three groups of islands, using the exact test of population differentiation, revealed that there is a similarity between the Eastern and Central groups ($p = 0.29955$). The Western group showed significant differences when compared with the other two groups (Western/Eastern: $p < 0.00001$; Western/Central: $p = 0.00135$).

A comparison of the frequencies of each haplogroup in the three groups revealed that haplogroups J, U and V contribute to significantly differentiate the three groups of islands, and largely to differentiate the Western group (significant differences: J-Western/Central/Eastern: $p = 0.04084$, Western/Eastern: $p = 0.038$, Western/Central: $p = 0.023$; U-Western/Central/Eastern: $p = 0.00056$, Eastern/Central: $p < 0.0001$; V-Western/Central/Eastern: $p < 0.0001$, Western/Eastern: $p < 0.0001$, Western/Central: $p < 0.0001$).

In the total sample of the Azores, all European haplogroups (Torroni *et al.* 1996) were found, as well as

the African clusters U6, M1, L1, L2 and L3 (Rando *et al.* 1998; Quintana-Murci *et al.* 1999; Chen *et al.* 1995, 2000; Watson *et al.* 1997) and the Near Eastern N1b cluster (Richards *et al.* 2000). This result provides evidence for a mixed composition of mtDNAs in the Azores islands, that is supported by historical data referring to the contribution of individuals from multiple origins to the peopling of the archipelago.

The analysis by groups (Table 3) revealed that in the Central group, all the main European haplogroups (Torroni *et al.* 1996) were found. A very low frequency of haplogroup U was observed when compared to values observed in European populations. The frequency of J is closer to that observed in North European populations. The frequencies of the remaining haplogroups were similar to those observed in European populations. African (M1, L1) and Near Eastern (N1b) sequences were found in low frequencies in the Central group.

In the Eastern Group, and when considering European haplogroups, only I is absent. The most interesting features are the low frequency of haplogroup H (30%), with respect to European populations, and the higher number of sub-clusters of haplogroup U relative to the Western and Central groups of the archipelago. In the Eastern group, high frequencies of non-European sequences were observed, with members of the U6, M1, L2 and L3 African clusters and N1b Near Eastern cluster detected.

The Western group exhibited a very atypical haplogroup distribution, where the I, J and W haplogroups are absent. Haplogroup U (excluding K) is under-represented, and the few samples belonging to this cluster are members of the U2 and U6b sub-haplogroups, which are uncommon in European populations but more frequent in Near Eastern and North African populations, respectively (Macaulay *et al.* 1999). The most frequent haplogroup is not H, as it is in European populations, but V, the frequency of which was higher than 33%. In all the results published so far, frequencies of V higher than 10% were only found in Catalonia (12.5%) (Montiel *et al.* 2001), the Basque Country (12.4%) and the Skolt Saami (52%) (Torroni *et al.* 2001). Members of the East African M1 cluster were also found in this group of islands.

Table 3 mtDNA haplogroup distribution (number of individuals and percentage) in the total sample of the Azores, and in Eastern, Central and Western groups

Haplogroup	Azores (N = 146)		Eastern Group (N = 50)		Central Group (N = 60)		Western Group (N = 36)	
	f _i	%	f _i	%	f _i	%	f _i	%
H	47	32.19	15	30	24	40	8	22.22
I	3	2.05			3	5		
J	8	5.48	5	10	3	5		
J1	1	0.68	1	2				
J1a	2	1.37	1	2	1	1.67		
J1b1	3	2.05			3	5		
J2	1	0.68			1	1.67		
L1c	1	0.68			1	1.67		
L2	2	1.37	2	4				
L3e2b	1	0.68	1	2				
M1	4	2.74	1	2	1	1.67	2	5.56
N1b	2	1.37	1	2	1	1.67		
T	3	2.05			3	5		
T1	3	2.05	1	2			2	5.56
T2	9	6.16	3	6	4	6.67	2	5.56
T3	6	4.11	1	2	2	3.33	3	8.33
U								
U2	1	0.68					1	2.78
U3	2	1.37	2	4				
U5	2	1.37	2	4				
U5a/b	3	2.05	3	6				
U5a1a	2	1.37	1	2	1	1.67		
U5b	1	0.68	1	2				
U6a	3	2.05	3	6				
U6b	2	1.37					2	5.56
K	8	5.48	3	6	4	6.67	1	2.78
pre-HV	1	0.68			1	1.67		
pre*V	2	1.37					2	5.56
V	15	10.27	1	2	2	3.33	12	33.33
W	3	2.05	1	2	2	3.33		
X	4	2.74	1	2	2	3.33	1	2.78
R*	1	0.68			1	1.67		

HVRI Diversity

Table 4 presents diversity parameters for the Azorean population and populations cited in Table 1, obtained using HVRI sequences between positions 16069–16370. In the total Azorean sample, 80 different HVRI lineages, characterised by 74 polymorphic sites, were observed. The highest values of diversity estimators were obtained for the Central and Eastern group, with similar values, and the lowest values were observed in the Western group. Compared to other populations, the values for mean numbers of pairwise differences (θ_{π}) and nucleotide diversity (π) were higher than the

majority of European populations in the Azores sample (total or by groups of islands). A median value of gene diversity was observed in the Azores. However, the value registered for the Western group appears to be one of the lowest when the three groups of islands were analysed separately. The values of diversity decreased in the three groups of islands when African sequences (L1, L2, L3 and M1) were excluded (Table 4).

The analysis of diversity by haplogroups in the Azorean sample compared to mainland Portugal (Table 5) showed high values of diversity for the haplogroups U and J, while the diversity of the remaining

Table 4 mtDNA HVRI diversity (from nucleotide positions 16069 to 16370) in the Azorean populations including and excluding African sequences (L1, L2, L3 and M1), and in populations used for comparison

Population	N	K (% K)	S (% S)	transi/transv/ indel	$\hat{H} \pm$ sd	$\theta_{\pi} \pm$ sd	$\pi \pm$ sd	θ_k	Private lineages (%)
Azores	146	80 (54.79)	74 (24.50)	64/13/2	0.9603 ± 0.0107	5.4339 ± 2.6322	0.0179 ± 0.0096	71.7723	41 (51.25)
Eastern Group	50	33 (66.00)	50 (16.56)	41/8/2	0.9486 ± 0.0236	5.3168 ± 2.6102	0.0175 ± 0.0095	41.1428	17 (51.52)
Central Group	60	40 (66.67)	52 (17.22)	47/6/0	0.9565 ± 0.0190	5.4937 ± 2.6799	0.0182 ± 0.0098	51.2801	16 (40.00)
Western Group	36	16 (44.44)	34 (11.26)	30/4/1	0.9159 ± 0.0276	5.1745 ± 2.5649	0.0171 ± 0.0094	10.4725	6 (37.50)
Mainland Portugal	241	142 (58.92)	94 (31.13)	89/12/0	0.9588 ± 0.0096	5.1369 ± 2.4982	0.0170 ± 0.0092	144.2037	84 (59.15)
North Portugal	100	67 (67.00)	69 (22.85)	64/7/0	0.9533 ± 0.0162	5.1406 ± 2.5120	0.0170 ± 0.0092	87.7000	35 (52.22)
Central Portugal	82	61 (74.39)	63 (20.86)	59/7/0	0.9765 ± 0.0107	5.2758 ± 2.5753	0.0175 ± 0.0094	106.3182	26 (42.62)
South Portugal	59	41 (69.49)	54 (17.88)	50/4/0	0.9433 ± 0.0246	4.9621 ± 2.4488	0.0164 ± 0.0090	58.2382	18 (43.90)
Spain	89	71 (79.78)	68 (22.52)	65/5/0	0.9837 ± 0.0081	5.4974 ± 2.6693	0.0182 ± 0.0098	160.6716	37 (52.11)
Galicia	92	51 (55.43)	55 (18.21)	53/3/1	0.9233 ± 0.0239	3.3215 ± 1.7218	0.0110 ± 0.0063	46.3011	21 (41.18)
Basque Country	106	53 (50.00)	53 (17.55)	49/5/0	0.9364 ± 0.0181	3.1821 ± 1.6587	0.0105 ± 0.0060	41.5205	23 (43.40)
France	50	42 (84.00)	52 (17.22)	47/7/0	0.9878 ± 0.0086	4.85720 ± 2.4092	0.0161 ± 0.0089	121.0537	22 (52.38)
UK	100	68 (68.00)	67 (22.19)	62/5/2	0.9752 ± 0.0092	4.7785 ± 2.3549	0.0157 ± 0.0086	92.2527	36 (52.94)
Germany	67	52 (77.61)	56 (18.54)	52/3/1	0.9724 ± 0.0139	4.1728 ± 2.1004	0.0138 ± 0.0077	104.8412	30 (57.69)
Austria	101	64 (63.37)	72 (23.84)	68/7/0	0.9505 ± 0.0170	4.9565 ± 2.4320	0.0164 ± 0.0089	74.0609	36 (56.25)
Italy	49	40 (81.63)	54 (17.88)	52/2/0	0.9685 ± 0.0189	5.5298 ± 2.7042	0.0183 ± 0.0099	99.3853	28 (70.00)
Israel	117	101 (86.32)	93 (30.79)	90/12/3	0.9945 ± 0.0031	6.6033 ± 3.1412	0.0219 ± 0.0115	348.3041	66 (56.35)
Algeria	85	30 (35.29)	37 (12.25)	37/0/0	0.9431 ± 0.0103	5.2725 ± 2.5729	0.0175 ± 0.0094	16.0942	15 (50.00)
Moroccan Jews	115	45 (39.13)	51 (16.89)	49/3/0	0.9077 ± 0.0224	4.0584 ± 2.0398	0.0134 ± 0.0075	26.7323	25 (55.56)
Morocco (Berber)	60	38 (63.33)	50 (16.56)	46/5/1	0.9633 ± 0.0147	5.0957 ± 2.5064	0.0169 ± 0.0092	43.5081	18 (47.37)
Morocco (Moroccans)	32	29 (90.63)	47 (15.56)	44/3/0	0.9879 ± 0.0141	6.8612 ± 3.3158	0.0227 ± 0.0122	144.6747	12 (41.38)
Mauritania	30	23 (76.67)	31 (10.26)	31/0/0	0.9747 ± 0.0172	6.8674 ± 3.3245	0.0227 ± 0.0123	43.3056	8 (34.78)
Senegal	240	124 (51.67)	80 (26.49)	77/6/1	0.9858 ± 0.0025	7.3792 ± 3.4633	0.0249 ± 0.0129	102.4841	104 (83.87)
S. Tome	50	32 (64.00)	50 (16.56)	46/6/0	0.9731 ± 0.0110	9.3350 ± 4.3606	0.0309 ± 0.0160	37.2763	21 (65.62)
Mozambique	109	50 (45.87)	62 (20.53)	58/7/2	0.9611 ± 0.0080	9.8708 ± 4.5518	0.0327 ± 0.0167	35.1729	43 (86.00)
Excluding African sequences (L1, L2, L3 and M1)									
Azores	138	73 (52.90)	67 (22.19)	59/10/2	0.9557 ± 0.0012	4.9439 ± 2.4211	0.0163 ± 0.0088	62.0844	33 (45.21)
Eastern Group	46	29 (60.04)	43 (14.24)	36/5/2	0.9391 ± 0.0274	4.9383 ± 2.4483	0.0162 ± 0.0089	32.6527	13 (44.83)
Central Group	58	38 (65.52)	45 (14.90)	43/3/0	0.9534 ± 0.0202	4.9690 ± 2.4524	0.0165 ± 0.0090	46.7723	14 (36.84)
Western Group	34	15 (44.12)	32 (10.60)	28/4/1	0.9073 ± 0.0304	4.5447 ± 2.2907	0.0150 ± 0.0084	9.7030	5 (33.33)

N-sample size; K-number of different haplotypes; S-number of polymorphic sites; \hat{H} -gene diversity; θ_{π} -mean number of pairwise differences; π -nucleotide diversity; θ_k -theta estimator based on the number of different haplotypes; Private lineages are defined as those found only in one of the populations included in the present study

Table 5 Diversity within major haplogroups in the Azores (Az) and Mainland Portugal (Pt)

Haplogroups	N		K		$\hat{H} \pm \text{sd}$		$\theta_\pi \pm \text{sd}$		$\pi \pm \text{sd}$		ρ	
	Az	Pt	Az	Pt	Az	Pt	Az	Pt	Az	Pt	Az	Pt
U (including K)	24	54	19	39	0.9819 ± 0.0164	0.9762 ± 0.0113	5.5696 ± 2.7719	5.3857 ± 2.6370	0.018382 ± 0.010198	0.017834 ± 0.009687	27/24 = 1.125	59/54 = 1.093
J	15	17	10	10	0.9333 ± 0.0449	0.9265 ± 0.0417	3.9738 ± 2.1058	3.3984 ± 1.8290	0.013158 ± 0.007818	0.011253 ± 0.006778	14/15 = 0.933	15/17 = 0.882
T	21	26	11	16	0.9286 ± 0.0306	0.9323 ± 0.0337	4.3704 ± 2.2489	4.6208 ± 2.3424	0.014472 ± 0.008312	0.015301 ± 0.008641	12/21 = 0.571	22/26 = 0.846
H	47	99	19	44	0.6966 ± 0.0768	0.8023 ± 0.0431	1.4868 ± 0.9145	1.9230 ± 1.1031	0.004891 ± 0.003339	0.006368 ± 0.004046	20/47 = 0.426	43/99 = 0.434
V	15	17	3	10	0.5905 ± 0.0771	0.7941 ± 0.1035	0.6632 ± 0.5394	1.6799 ± 1.0353	0.002196 ± 0.002003	0.005562 ± 0.003836	2/15 = 0.133	12/17 = 0.706
L1+L2+L3+M1	8	18	7	17	0.9643 ± 0.0772	0.9935 ± 0.0210	8.5883 ± 4.4445	9.1597 ± 4.4202	0.028344 ± 0.016709	0.030330 ± 0.016367	17/8 = 2.125	38/18 = 2.110

N-sample size; \hat{H} -gene diversity; θ_π -mean number of pairwise differences; π -nucleotide diversity. Major haplogroups are defined as those with a total frequency higher than 10%; ρ -average number of sites differing between a set of sequences and a specified common ancestor (between positions 16051–16399)

haplogroups is lower in the Azores. In the Azorean sample, H and V haplogroups present the lowest values of diversity. The African haplogroups have high values of θ_π , π and ρ , followed by the U haplogroup. However, in the Azores, the highest value for gene diversity is displayed by haplogroup U.

Relating gene diversity and nucleotide diversity in the three groups of islands (Table 4) to intra-haplogroup variation (Table 5), it is possible to infer that the higher nucleotide diversity observed in the Azores, as compared to European populations, is mainly the result of the presence of African sequences in the three groups of islands. Nevertheless, the influence of the variability of some specific haplogroups, especially U, also has some influence on the increase in both nucleotide and gene diversity.

According to Helgason *et al.* (2000), the theta estimator based on the number of different haplotypes (θ_k) is likely to provide a more reliable evaluation of current and historical female effective-population size than θ_π . Of all the populations studied, the Western group presents the lowest value of θ_k , and the Central and Eastern groups also presented relatively low values (Table 4). This indicates that the Azores, and especially the Western group, have recently had a relatively small female effective-population size.

In general, the proportion of private lineages sampled from geographical proximate populations should increase as a function of θ_k , as this parameter reflects the probability of new lineages arising by mutation (Helgason *et al.* 2001). High values of private lineages associated with low values of θ_k , found in some populations used in the comparison such as Mozambique, are the result of the non-inclusion of a sufficient number of geographically related populations in the comparative database to detect gene flow. Considering the analysis of European populations alone, a significant positive correlation is observed between the percentage of private lineages and the values of θ_k ($r_{sp} = 0.596$; $p = 0.009$). The relationship between the percentage of private lineages and the values of θ_k is presented in Figure 1, and varies greatly among the three groups of islands. The Eastern group presents an excess of private lineages in contrast to the Central group, which tends to a scarcity of private lineages. The western group, and the total sample of the Azores, have a

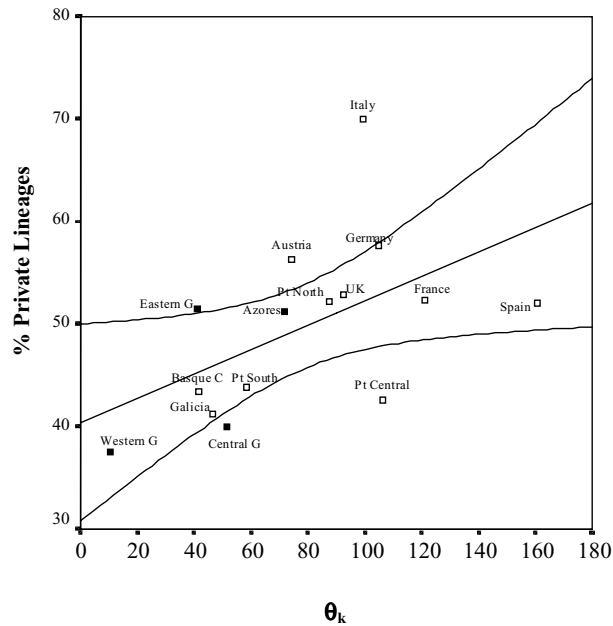


Figure 1 Scatterplot of θ_k values and percentage of private lineages. The least-squares regression line is shown. The curved lines show the 95% confidence region around the regression line.

proportion of private lineages that can be expected by their θ_k .

The excess of private lineages observed in the Eastern group could have different causes. An intuitive explanation is that the isolation of the islands can hinder the migratory flow of new lineages to and from neighbouring populations. However, six of the private lineages found in the Eastern group belong to African clusters (U6a, M1, L2 and L3). Consequently, it is more likely that a sufficient number of African populations was not included in the comparative database.

Within the Azores, the Central group shares the highest number of lineages with the other two groups of islands. Therefore, the relative scarcity of private lineages observed in the Central group could be indicative of either a very high level of emigration (where few lineages remain private for long, because of rapid outward gene flow to neighbouring populations), or of immigration (where new lineages arriving into the population would increase θ_k but not the proportion of private lineages).

The values of diversity parameters, reported for the archipelago, allow us to infer that the Azores population as a whole does not present the typical characteris-

tics of an isolated population. Thus, the values recorded are similar to those observed in the majority of European populations used in the comparative analysis, and higher than those observed in isolated populations like the Basque country and Galicia. Furthermore, neither an excess nor a scarcity of private lineages is observed, in relation to the respective θ_k value, suggesting that values of diversity in the Azores are adjusted to the effective population size of females.

These results are in accordance with studies of autosomal STRs (Lima, personal communication), classical markers (Amorim *et al.* 1979) and those reported in biodemographic studies, which indicate that the Azorean archipelago, when considered as a whole, does not exhibit levels of endogamy and consanguinity higher than those reported for some populations from mainland Portugal (Cunha, 1987; Lima, 1991; Lima & Soares, 1992; Smith *et al.* 1992).

When analyses were carried out by groups of islands, the Western group appears to have different features. The haplogroup distribution is very atypical, diversity values are in general lower than those observed in the other groups of islands, and the value of gene diversity is one of the lowest registered in all of the populations used for comparison. However, the relationship between the percentage of private lineages and values of θ_k indicate that this group of islands does not undergo a higher level of isolation than that observed in the other groups of islands. This leads to the conclusion that the observed haplogroup distribution and lower diversity are strictly related to the reduced number of females present in the island since peopling, and makes the roles of founder effect and genetic drift in small populations evident.

Phylogeography of Azores mtDNA

Figure 2 (a, b, c, d, e) represents phylogenetic networks of haplotypes found in the Azores and populations used for comparison. To simplify networks, haplotypes absent in the Azores were only represented if they were phylogenetically related to haplotypes present in the Azorean population. Examining the networks, from the 80 different haplotypes (based on HVRI information) found in the Azores, only six are present in more than one group of islands. Two haplotypes are present in the three groups, three are common to the Central and Eastern

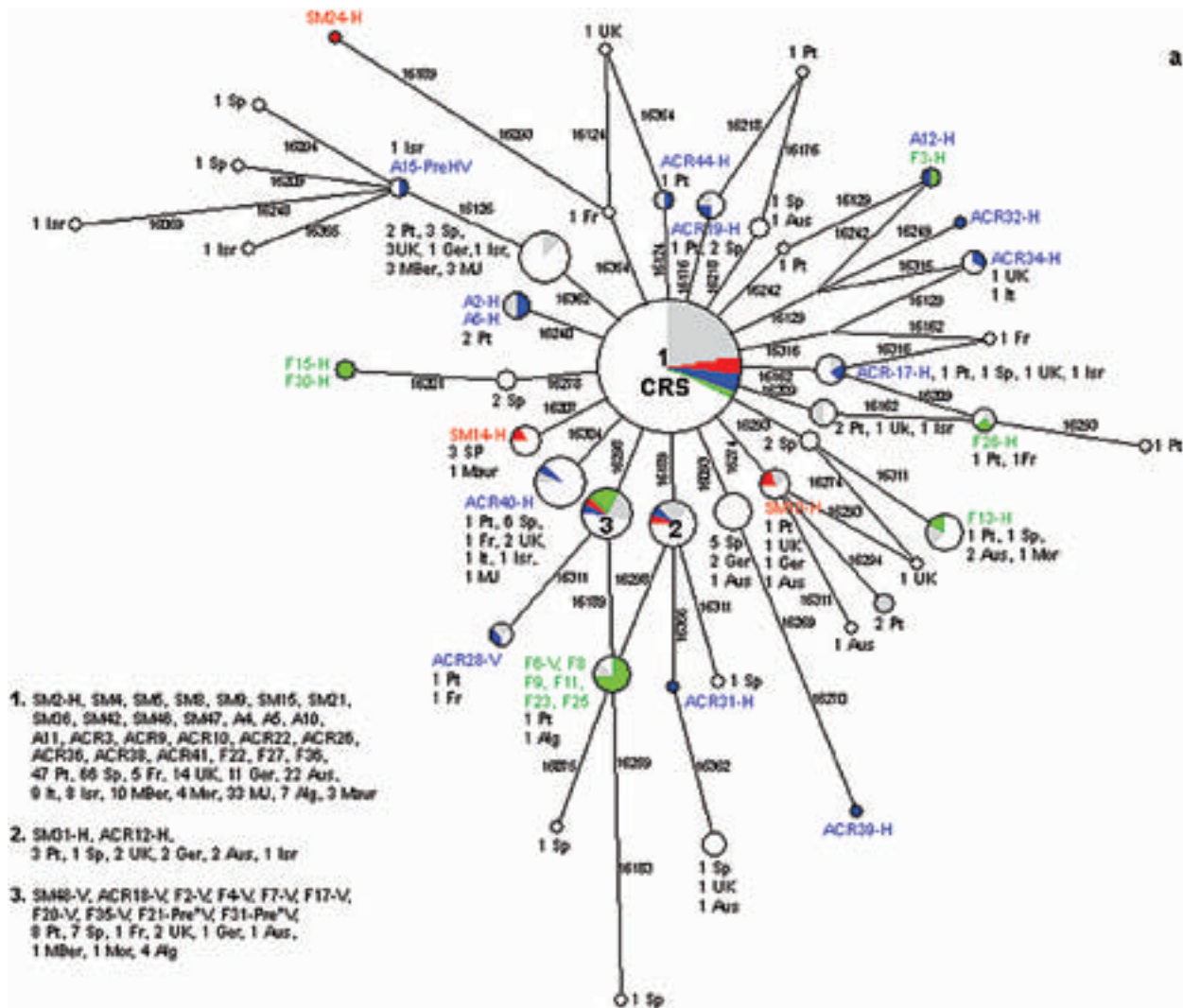


Figure 2 Phylogenetic Networks of (a) H, V, HV, Pre-HV, Pre*V; (b) U (including K); (c) J, T, JT; (d) N, I, W, X; (e) L1, L2, L3, M. Circle sizes are proportional to the haplotype frequency. (Red - Eastern group; Blue - Central group; Green - Western group; Gray - Mainland Portugal; White - other populations).

groups, and one to the Central and Western groups. The Central group is the one that shares the highest number of haplotypes with the other two. This result indicates limited migratory movements among the three groups of islands, until at least the beginning of the 19th century (since the geographical location of the sample was based on the origin of the great-grand-parents of the donor).

Forty-one haplotypes (represented in the networks as single colour circles: red to the Eastern group, blue to the Central group and green to the Western group) appear to be exclusive to the Azores, and thirty-nine are shared with other populations. By analysing the shared

haplotypes, thirty-three were found to be common to the Iberian Peninsula (mainly to mainland Portugal). However, only eleven are shared exclusively with the Iberian Peninsula, the remaining twenty-two also being present in one or more of the other populations. Specifically, sixteen are found in other European populations, ten in Israel, five in Moroccan Jews, eleven in North Africa and four in West Africa. Considering that these haplotypes belong to European haplogroups (H, J, T U, K, V and X) (Torroni *et al.* 1996) their most probable origins are the Iberian Peninsula or other European populations. However, a possible North African origin cannot be excluded for some sequences. The work of

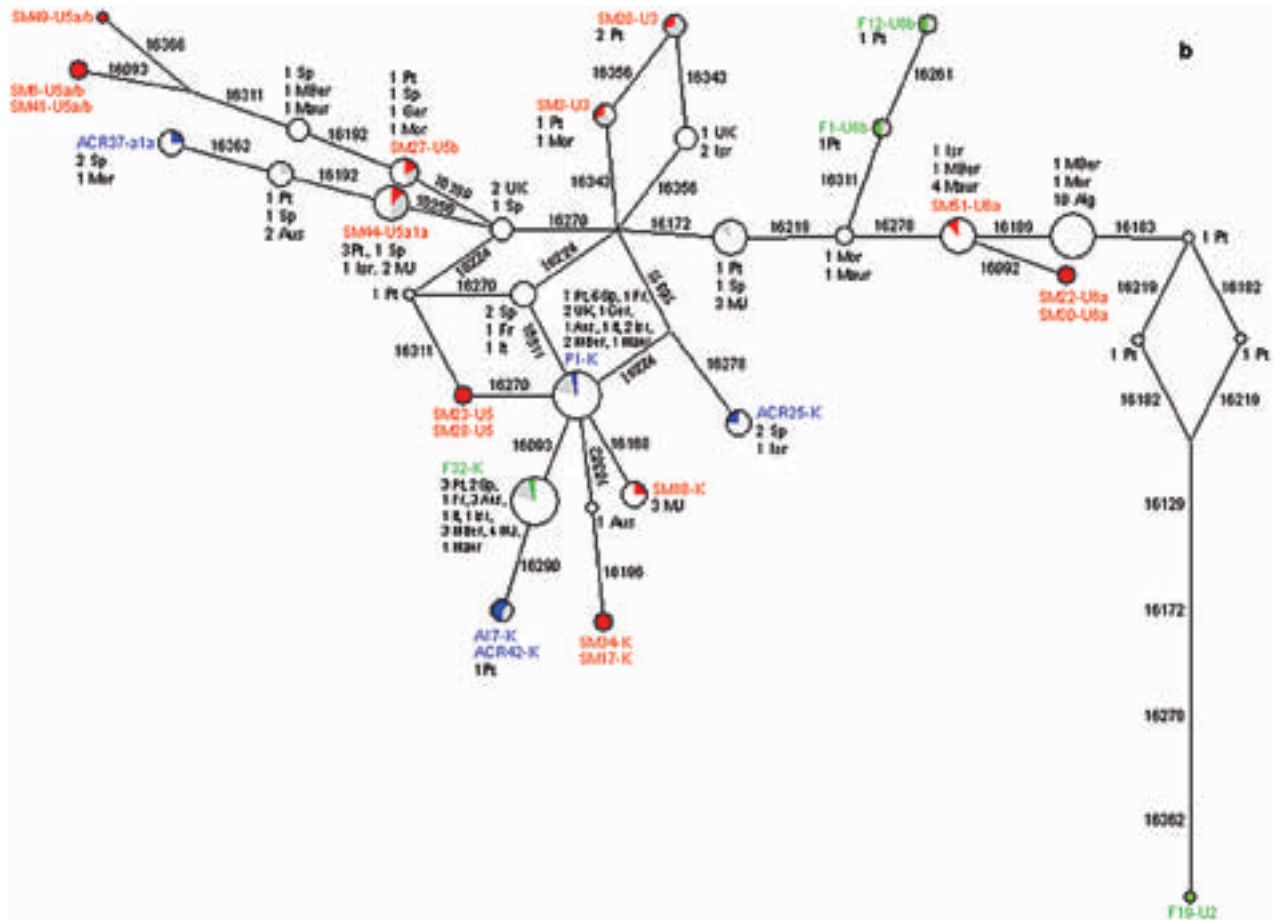


Figure 2 (continued)

Rando *et al.* (1998) attests to the fact that there is strong evidence for some degree of European genetic input into North Africa. There is, for example, evidence that haplogroups U5 (Richards *et al.* 1998) and V (Torroni *et al.* 1998) are represented in North Africa. Only one haplotype (ACR34-H) is shared exclusively with non-Iberian European populations (Fig. 2a).

Haplotypes represented by A15-pre-HV (Fig. 2a) and SM11-N1b (Fig. 2d) are shared exclusively with the Israeli population, and SM18-K (Fig. 2b) with Moroccan Jews. Taking into account the fact that the Israel Arab/Palestinian sample represents an extant population living in the region to which Jews trace their origin (Thomas *et al.* 2002), and that Moroccan Jews probably included descendants of the Sephardim Jews that were expelled from the Iberian Peninsula, the three haplotypes (represented by A15, SM11 and SM18) previously mentioned could indicate the presence of Sephardim Jews in the Azores. The haplotype represented by SM51

(Fig. 2b), belonging to sub-haplogroup U6a, is shared with Israeli and North African populations. This haplotype is probably originally from North Africa, since the U6 haplogroup has been described as a North African cluster (Rando *et al.* 1998).

U6 sequences are represented in both the Eastern and Western groups; however, in the Western group (samples F1, F12) they belong to sub-cluster U6b, which, according to Plaza *et al.* (2001), is restricted to the Iberian Peninsula. In the Eastern group (samples SM22, SM30 and SM51) they are assigned to U6a, found in both the Iberian Peninsula and North Africa (Plaza *et al.* 2001; Rando *et al.* 1998). Thus, only lineages found in the Eastern group could plausibly constitute evidence of the influence of Moorish slaves, which is in accordance with historical data.

The haplotype represented by A16 (Fig. 2d), a member of the European haplogroup I, is present in non-Iberian European populations. In S. Tome, the same

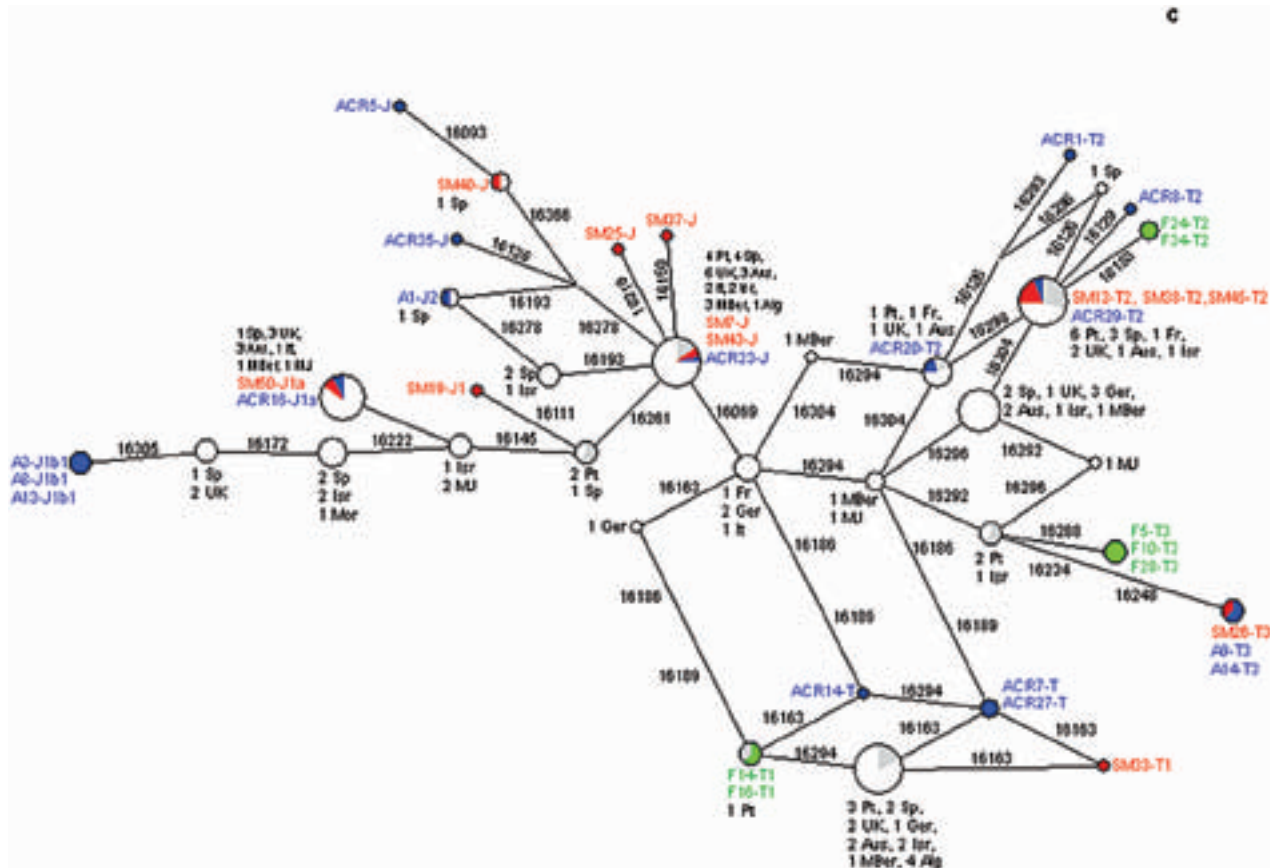


Figure 2 (continued)

haplotype was found (Mateu *et al.* 1997); however that sample probably belongs to haplogroup L3 (Calafell, Comas & Mateu, personal communication).

In summary, the analysis of shared haplotypes indicates that the majority of haplotypes is shared with Iberian Peninsula populations, principally with mainland Portugal. A small influence from other European (especially North European), Near East/Jewish and North African populations can also be confirmed.

The 41 private lineages found could have been introduced into the Azores both by European and by African individuals belonging to populations that are not included in the comparison, for example, populations from Flanders and Madeira Island, for which data are not available. Alternatively, as suggested by Helgason *et al.* (2000), they may arise from populations included in the analysis in which these lineages have not yet been sampled, or have been lost since the inhabitation of the Azores.

On the other hand, the hypothesis of the origin of some private lineages in recent mutational events occurring after the settlement of the islands cannot be disregarded.

The possible origin of the non-shared haplotypes, to be discussed later, was inferred using the partial networks presented in Figure 2 and two mtDNA databases, namely the database of Richards *et al.* (2000) and the HVRbase (Handt *et al.* 1998), both available online.

A worldwide search for the 41 private haplotypes found in the Azores was performed in mtDNA databases, and eight matches with samples that were not included in the comparison were found. Haplotypes represented by ACR31-H (Fig. 2a), F24-T (Fig. 2c), ACR30-W (Fig. 2d) and SM49-U5a/b (Fig. 2b) are shared with Northern European populations that were not included in the analysis; ACR8-T2 (Fig. 2c), SM12-X (Fig. 2d) and ACR7-T (Fig. 2c) match with one sequence each, found, respectively, in Rome (Italy), Romany and Armenia; SM22 (Fig. 2b), which

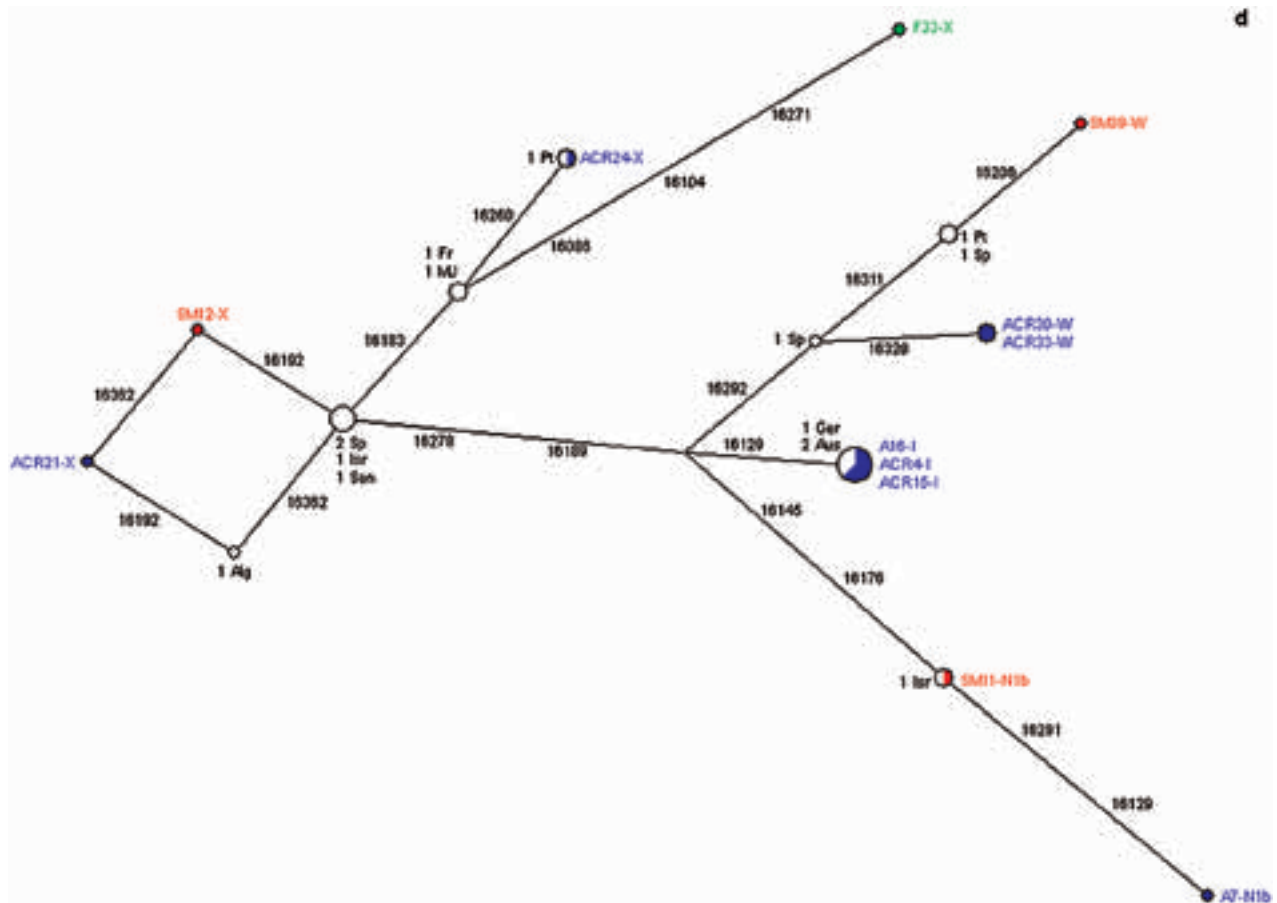


Figure 2 (continued)

belongs to the North African cluster U6a (Rando *et al.* 1998), is not shared with populations used for comparison, but was found in the Canary Islands (Rando *et al.* 1999) and in Moroccan Berbers (Anglés, personal communication).

Haplotypes classified as J can provide valuable information for the inference of the origin of the Azorean population. This haplogroup is present at high frequencies in Northern European populations, and some specific clusters, such as J1b1, can be assigned to a very restricted geographic area. Haplogroup J is absent in the Western group, but its frequency is close to 14% in both the Eastern and Central groups, higher than that observed in the Iberian Peninsula and similar to that observed in Northern European populations. Among the ten J haplotypes found in the Azores, six are private lineages (Fig. 2c).

The haplotype represented by ACR5 could have been derived from the haplotype 16069, 16126, 16278,

16366, found in mainland Portugal and in the Eastern group (SM40). SM19 could have been derived from the haplotype 16069, 16126, 16261, found in mainland Portugal. Haplotypes represented by SM25 and SM37 could be derived from haplotype 16069, 16126 found in the Azores (SM7, SM43 and ACR23) and at high frequencies in Europe, especially in United Kingdom. Haplotype A3 was possibly derived from the haplotype 16069, 16126, 16145, 16172, 16222, 16261, found in the United Kingdom and Spain.

Nevertheless, other origins can be postulated for the previously mentioned J haplotypes. Importantly, the influence of the population from the region of Flanders that was not included in the analysis is always mentioned in the peopling of the Central group. It is expected that this population, like the surrounding Northern European populations, would have high frequencies of haplogroup J, and some of the non shared haplotypes found in the Azores could match with those which,

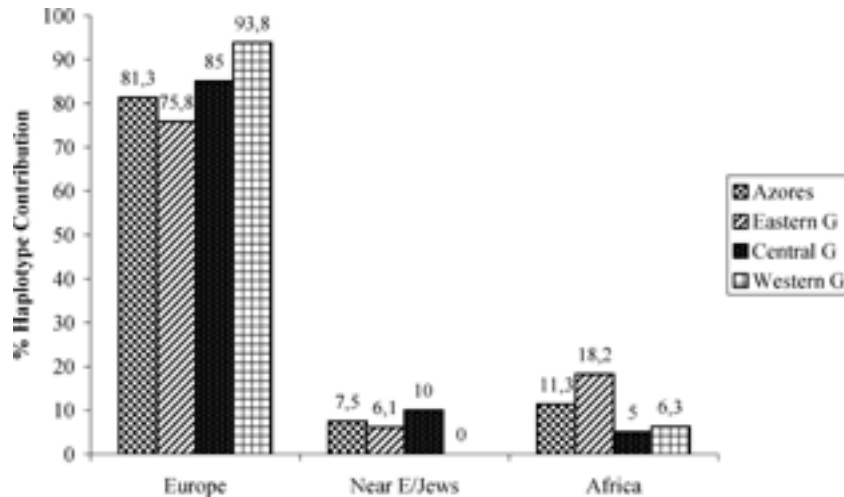


Figure 3 Bar chart showing the pattern of haplotype contribution from parental populations to the Azores.

Whenever a haplotype is shared with a European population, it is admitted that its most probable origin is Europe. Thus, it is possible to infer that probably: a minimum of 11.25% of the haplotypes found in the Azores are originally from African populations; a minimum of 7.5% are derived from Near Eastern and Jewish populations; and a maximum of 81.25% are from European populations.

Figure 3 illustrates the observation that mtDNA gene pools of all the groups of islands have their major contribution from European populations. However, there are some differences among the three groups of islands. The Eastern group exhibits the highest contribution from non-European populations (close to 25%), mainly African (18.2%), revealing the influence of slaves in the peopling of this group of islands.

In the Central group, the non-European influence is approximately 15%, 10% from Near Eastern/Jewish populations, and 5% from African populations. The Western group exhibits the lowest contribution of non-European lineages (6.25%), which is exclusively African. According to historical information the contribution of slaves to the peopling of the islands, mainly to the Eastern and Western groups, was very important, being well-documented for the Western group (for review see Gomes, 1997). One of the headmen of Flores island was also leader of Santo Antão island (Cabo Verde), and a great number of the slaves who arrived at the Western group were originally from Cabo Verde. Surprisingly,

no L1, L2 or L3 mtDNA lineages were found in the Western group. This indicates that factors such as differential mortality may have affected the fitness of slaves, limiting the transmission of their genes until now.

Genetic Distances and Phylogenetic Reconstruction

To examine the relationship between all pairs of populations used in the comparison, genetic distances according to Reynolds *et al.* (1983) were calculated, and used to build a phylogenetic tree (not shown). In contrast to European populations, African populations present higher genetic distances among themselves. This justifies the existence of an identifiable gradient observed in the N-J tree from sub-Saharan to North African populations, and the aggregation of the European populations in one extreme of the tree. It is impossible to establish any relation among European populations. Furthermore, the three groups of islands are located in this assemblage.

To try to clarify the position of the Azores with respect to other populations, various N-J trees were built. The position of the various populations was very similar in all the trees. In Figure 4, a tree that excludes Algerian and Western and Eastern African populations is shown, where the African populations that show less genetic distance relative to the Azores are maintained.

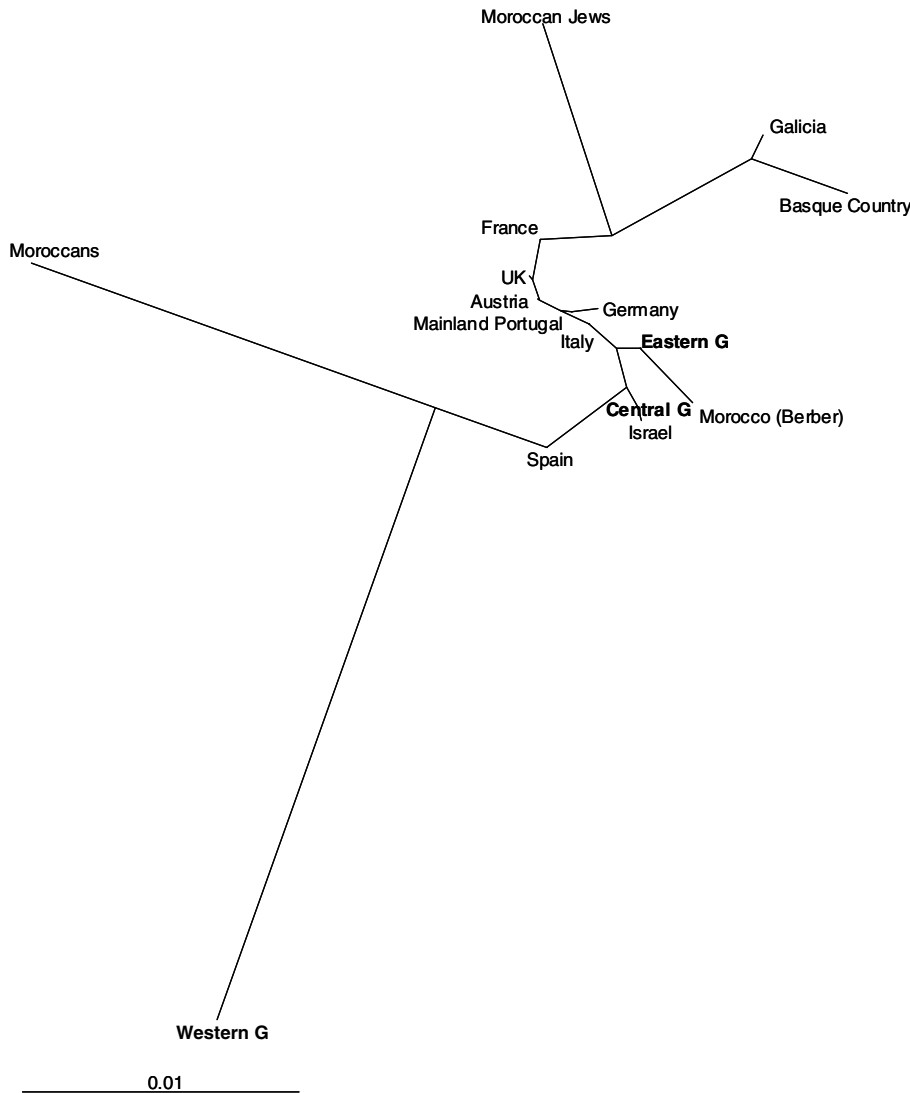


Figure 4 Neighbour-Joining tree (Saitou & Nei, 1987) constructed from Reynolds *et al.* (1983) genetic distances.

The analysis of Figure 4 reveals that the geographically and/or culturally isolated populations, such as those of the Basque Country, Galicia and Moroccan Jews, are separated from the remaining European populations. Moroccans also appear to be separated, because this population is the most differentiated relative to European populations. The three groups of islands are located in different branches of the tree that emerge from a group of European populations. However, the Western group appears in a very profound and isolated branch, reinforcing the previous results that indicate a differentiation of this group of islands. The Central group is located in the same branch as Israel, while the Eastern group is posi-

tioned in the same branch as the Moroccan Berbers. The position of the Eastern group corroborates previous results, indicating some affinities with African populations. In a similar manner the position of the Central group indicates that there is some relationship with Near Eastern populations.

The analysis of phylogeography of Azores mtDNA, estimates of admixture, genetic distances and phylogenetic reconstruction allow us to infer a possible origin for the mtDNA gene pool of the Azores. All the analyses support the idea that, in all the groups of islands, the majority of mtDNA lineages come from the Iberian Peninsula, mainly from Portugal, but the contribution

of other European populations, especially from Northern Europe, cannot be disregarded. The presence of *Moorish* and *African* slaves in the islands is also supported by mtDNA genetic data, especially in the Eastern group, as well as the presence of Jews, mainly in the Central group.

Neither historical nor genetic data supports the idea of a differential settlement for the Western group; however, it appears in the phylogenies in an isolated branch. With 4329 inhabitants (INE, 2001) today, the Western group has reduced its population size by half in less than one century. Thus, this “bottleneck”, combined with a small population size since peopling, has led to a very atypical distribution of haplogroups/haplotypes that influences all the results obtained. Future genetic and biodemographic investigations in the Western group will certainly help to clarify its position relative to the rest of the archipelago, and to other populations.

Conclusions

This work confirms that studies of large population samples in restricted geographical contexts can produce valuable insights concerning specific genetic and demographic features that would remain undetectable in broad scale surveys.

The distinct analyses allow us to infer that the Azorean population does not present the typical characteristics of an isolated population, as is usually postulated for islands. However, the Azores archipelago has a recent demographic history, was peopled by individuals of multiple origins, which naturally leads to a population with high diversity, and it is probable that the effect of isolation, if it exists, remains undetectable at this time.

Separate analyses of the three groups of islands appeared to be valuable for detecting the specific features of the Western group, which presents some differentiation with respect to the other groups of islands. This differentiation should be the result of genetic drift processes, induced by the small population size since peo-

pling, and eventually the result of the recent reduction in population size.

The analysis of lineage sharing supports the idea that mainland Portugal and Northern European populations could have contributed the majority of the mtDNA lineages that are currently observed in the Azorean mtDNA gene pool. However, many lineages are still private to the Azores, and their origins are sometimes difficult to infer, since historically related populations such as the population of Flanders are not used for comparison, because they remain uncharacterised for mtDNA.

Contrary to studies using autosomal markers (Lima *et al.* personal communication), it was possible to infer the influence of African and Near East/Jewish individuals in the islands, pointing out the effectiveness of non-recombining markers to reconstruct the history of populations. It was even possible to detect differential contributions of non-European populations to the peopling of the three groups of islands, mainly the contribution of Africans to the Eastern group, and the contribution of Near East/Jewish to the Central group, both in accordance with historical data.

Part of the variation in the Azorean mtDNA can be explained by the settlement history of the archipelago. Furthermore, it has been necessary to bear in mind that the variation of mtDNA was affected during the last 500 years, by biodemographic and genetic processes like founder effect, genetic drift, migration, and even recent mutational events. Thus, the results presented here arise from the combination of the history of peopling and the evolution of populations, with all their biodemographic and genetic interactions.

Acknowledgments

This work was supported by a grant (ref. SFRH/BD/723/2000) from Fundação para a Ciência e a Tecnologia - Operational programs Science Technology and Innovation (POCTI) and Society of Information (POSI) financed by UE funds and national funds of MCES.

Appendix HVRI region: sequences and RFLP specific screening in the Azorean population

Sample Name	HVRI (16050–16399)	Hapl. Seq.	MspI	DdeI	MseI	HinfI	AflII	NlaIII	NlaIII	HaeII	HaeIII	DdeI	AflII	AclI	SmaI	Bso01	AflII	AflII	HpaI	12705 C/T	Hapl. RFLP	Hapl. Seq./RFLP	MI?
Eastern group																							
SM1	129, 183C, 189, 189msC(1), 249, 311	MI?	-	+	-	+	+	-	-	-	-	-	-	-	-	-	-	+	-	-	M		
SM2	GRS	H or HV or U or R	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM3	343, 390	U3	+	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	U	U3	
SM4	GRS	H or HV or U or R	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM5	GRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM6	93, 189, 189msC(1), 192, 270, 311	U5a or U5b	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U5a/b	
SM7	69, 126	J	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	J	J	
SM8	GRS	H or HV or U or R	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM9	GRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM10	274	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM11	145, 176G, 223, 390	N1b	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Other	N1b	
SM12	189, 192, 223, 278	X	+	-	+	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	X	X	
SM13	126, 294, 296, 304	T2	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	T	T2	
SM14	207	H or HV or U or R	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	H	H	
SM15	GRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM16	223, 260, 278, 390, 399	L2	-	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	+	-	L1 or L2	L2	
SM17	196, 224, 311, 362	K	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	K	K	
SM18	168, 224, 311	K	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	K	K	
SM19	69, 111A, 126, 261	J1	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	J	J1	
SM20	343, 356, 390	U3 or U4	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U3	
SM21	GRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM22	92, 172, 219, 278	U6a	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U6a	
SM23	224, 270, 311	U5 or K	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U5	
SM24	93, 189, 354	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM25	126, 218A	J?	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	J	J	
SM26	126, 234, 248, 292, 294	T3	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	T	T3	
SM27	189, 270	U5b	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U5b	
SM28	224, 270, 311	U5 or K	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U5	
SM30	92, 172, 219, 278	U6a	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U6a	
SM31	189, 189msC(2)	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM32	93A, 223, 278, 294, 309, 390	L2	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	L1 or L2	L2	
SM33	126, 163A, 186, 189, 294	T1	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	T	T1	
SM34	196, 224, 311, 362	K	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	K	K	
SM36	GRS	H or HV or U or R	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	H	H	
SM37	69, 126, 150	J	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	J	J	
SM38	126, 294, 296, 304	T2	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	T	T2	
SM39	206C, 223, 292, 311	W	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	W	W	
SM40	69, 126, 278, 366	J	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	J	J	
SM41	93, 189, 189msC(1), 192, 270, 311	U5a or U5b	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U5a/b	
SM42	GRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM43	69, 126	J	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	J	J	
SM44	256, 270, 399	U5a1a	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U5a1a	
SM45	126, 294, 296, 304	T2	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	T	T2	
SM46	GRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM47	GRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	H	
SM48	298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	V	
SM49	189, 192, 270, 311, 336	U5a or U5b	+	-	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	U	U5a/b	

Appendix (continued)

Sample Name	HVR1 (16050-16399)	Hapl. Seq.	MnII	DdeI	MseI	HinfI	AflII	NdeIII	NlaIII	HaeIII	HaeIII	DdeI	AflII	Acl	SnaI	BstOI	AflII	AflII	HpaI	12705 C/T	RFLP	Hapl. Seq./RFLP
SM50	69, 126, 145, 231, 261	J1a	+	+	+	+	+	+													J	J1a
SM51	172, 219, 278	U6a	+	-	+	+	+	-													U	U6a
SM52	172, 182C, 183C, 189, 223, 320, 390C	L3e2b	-	+			+			+											L3?	L3e2b
Central group																						
A1	69, 126, 193, 278	J2	+	+	+	+	+	+													J	J2
A2	240	H or HV or U or R	+	-	+	+	+	+													H	H
A3	69, 126, 145, 172, 222, 261, 305T	J1b1	+	+	-	+	+	+													J	J1b1
A4	GRS	H or HV or U or R	+	-	-	-	-	-													H	H
A5	GRS	H or HV or U or R	+	-	-	-	-	-												C	H	H
A6	240	H or HV or U or R	+	-	-	-	-	-													H	H
A7	129, 145, 176G, 223, 291, 390	N1b	+	-	+	-	+	-												T	N*	N1b
A8	69, 126, 145, 172, 222, 261, 305T	J1b1	+	+	+	+	+	+													J	J1b1
A9	126, 234, 248, 292, 294	T3	+	-	+	+	+	+									+				T	T3
A10	GRS	H or HV or U or R	+	-	-	-	-	-													H	H
A11	GRS	H or HV or U or R	+	-	-	-	-	-													H	H
A12	129, 242	H or HV or U or R	+	-	-	-	-	-													H	H
A13	69, 126, 145, 172, 222, 261, 305T	J1b1	+	+	+	+	+	+													J	J1b1
A14	126, 234, 248, 292, 294	T3	+	-	+	+	+	+									+				T	T3
A15	126, 362	pre-HV	+	-	+	+	+	+													pre-HV	pre-HV
A16	129, 223, 391	I	+	+	+	+	+	-													I	I
A17	93, 224, 290, 311	K	+	+	+	+	+	-													K	K
P1	224, 311	K	+	+	+	+	+	-													K	K
ACR1	293, 294, 304	T2?	+	-	+	+	+	+													+	T2?
ACR3	GRS	H or HV or U or R	+	-	-	-	-	-													H	H
ACR4	129, 223, 391	I	+	+	+	+	+	-													I	I
ACR5	69, 93, 126, 278, 366	J	+	+	+	+	+	+													J	J
ACR6	129, 187, 189, 223, 265C, 278, 286A, 292, 294, 311, 360	L1c	-	+	+	+	+	+													+	L1 or L2
ACR7	126, 186, 189, 294	T	+	-	+	+	+	+													+	T
ACR8	126, 129, 294, 296, 304	T2	+	+	+	+	+	+													+	T2
ACR9	GRS	H or HV or U or R	+	-	-	-	-	-													H	H
ACR10	GRS	H or HV or U or R	+	-	-	-	-	-													H	H
ACR11	129, 182C, 183C, 189, 223, 230, 249, 311, 359	M1	-	+	+	+	+	+													+	M1
ACR12	189	H or HV or U or R	+	-	-	-	-	-														H
ACR14	126, 186, 189	T?	+	-	+	+	+	+													+	H
ACR15	129, 223, 391	I	+	+	+	+	+	+													+	T
ACR16	69, 126, 145, 231, 261	J1a	+	+	+	+	+	+													+	I
ACR17	162	H or HV or U or R	+	+	+	+	+	+														J
ACR18	298	Pre*V or V	+	-	-	-	-	-														V
ACR19	176	H or HV or U or R	+	-	-	-	-	-														H
ACR20	126, 294, 304	T2	+	-	+	+	+	+													+	T2
ACR21	189, 192, 223, 278, 362	X	+	-	+	+	+	+														X
ACR22	GRS	H or HV or U or R	+	-	-	-	-	-														H
ACR23	69, 126	J	+	+	+	+	+	+													+	J

Appendix (continued)

Sample Name	HVRI (16050–16399)	Hapl. Seq.	MntII 10871	DdeI 10394	MseI 14766	HinfI 12308	AluI 7025	NlaIII 4216	NlaIII 4577	HaeIII 9052	HaeIII 8994	DdeI 1715	AluI 10032	AclI 14465	SmaI 11719	Bst01 13705	AluI 15606	AluI 10397	HpaI 3592	12705 C/T	Hapl. RFLP	Hapl. Seq/RFLP
ACR24	183C, 189, 223, 260, 278	X	+	-	+	-	+	-	-	+	+	-	-	+	-	-	-	-	X		X	X
ACR25	278, 311	R1	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	K		K	K
ACR26	CRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR27	126, 186, 189, 294	T	+	-	+	+	+	-	-	-	-	-	-	-	-	-	+	-	T		T	T
ACR28	298, 311	Pre*V or V	+	-	+	+	+	-	-	-	-	+	-	-	-	-	-	-	V		V	V
ACR29	126, 294, 296, 304	T2	+	-	+	+	+	-	-	-	-	+	-	-	-	-	-	-	T		T	T2
ACR30	223, 292, 320	W	+	-	+	+	+	-	-	-	-	+	-	-	-	-	-	-	W		W	W
ACR31	189, 356	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR32	129, 249	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR33	223, 292, 320	W	+	-	+	+	+	-	-	-	-	+	-	-	-	-	-	-	W		W	W
ACR34	129, 316	H or HV or U or R	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR35	69, 278	J	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	J		J	J
ACR36	CRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR37	192, 256, 270, 362, 399	U5ala	+	-	+	+	+	-	-	+	+	+	-	-	-	-	-	-	U		U	U5ala
ACR38	CRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR39	93, 269, 270	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR40	304	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR41	CRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR42	93, 224, 290, 311	K	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	K		K	K
ACR44	124	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
ACR45	288, 311	H or HV or U or R	+	-	+	+	+	-	-	+	+	+	-	-	-	-	-	-	R*		R*	R*
Western group																						
F1	51, 172, 219, 311	U6b	+	-	+	+	+	-	-	+	+	+	-	-	-	-	-	-	U		U	U6b
F2	298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F3	129, 242	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
F4	298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F5	126, 288A, 292, 294	T3	+	-	+	+	+	-	-	-	-	-	-	-	-	-	+	-	T		T	T3
F6	189, 298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F7	298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F8	189, 298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F9	189, 298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F10	126, 288A, 292, 294	T3	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	T		T	T3
F11	189, 298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F12	172, 219, 261, 311	U6b	+	-	+	+	+	-	-	+	+	+	-	-	-	-	-	-	U		U	U6b
F13	293, 311	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
F14	126, 163, 186, 189	T?	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	T		T	T
F15	201, 278	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
F16	163, 186, 189	T?	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	T		T	T
F17	298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F19	51, 129C, 182C, 183C, 189, 189msC(1), 362, 399	U2	+	-	+	+	+	-	-	+	+	+	-	-	-	-	-	-	U		U	U2
F20	298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F21	298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Pre*V or HV		Pre*V or HV	pre*V
F22	CRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
F23	189, 298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F24	126, 193, 294, 296, 304	T	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	T		T	T
F25	189, 298	Pre*V or V	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V		V	V
F26	162, 209	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H
F27	CRS	H or HV or U or R	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H		H	H

Appendix (continued)

Sample Name	HVRI (16050-16399)	Hapl. Seq.	MnlI	DdeI	MseI	HinfI	AluI	NlaIII	NlaIII	HaeIII	DdeI	AluI	AccI	SnaI	BcoI	AluI	AluI	HpaI	12705 C/T	Hapl. RFLP	Hapl. Seq./RFLP
F28	126, 288A, 292, 294	T3	+	-	+	-	+	+	+	8994	1715	10032	14465	11719	13705	15606	10397	3592	-	T	T3
F29	129, 182C, 183C, 189, 223, 249, 311, 359	M1	-	+	-	-	+	-	-	9052	4577	4216	7025	7025	7025	7025	7025	-	-	M	M1
F30	201, 278	H or HV or U or R	+	-	-	-	+	-	-	8994	1715	10032	14465	11719	13705	15606	10397	3592	+	H	H
F31	298	Pre*V or V	+	-	-	+	+	+	+	9052	4577	4216	7025	7025	7025	7025	7025	-	-	Pre*V or HV	pre*V
F32	93, 224, 311	K	+	+	+	+	+	+	+	9052	4577	4216	7025	7025	7025	7025	7025	-	-	K	K
F33	86, 104, 183C, 189, 189insC(1), 223, 271, 278	X	+	-	+	-	+	-	-	9052	4577	4216	7025	7025	7025	7025	7025	-	-	X	X
F34	126, 193, 294, 296, 304	T2	+	-	+	-	+	-	-	8994	1715	10032	14465	11719	13705	15606	10397	3592	+	T	T2
F35	298	Pre*V or V	+	-	-	-	+	-	-	9052	4577	4216	7025	7025	7025	7025	7025	-	-	V	V
F36	CRS	H or HV or U or R	+	-	-	-	+	-	-	8994	1715	10032	14465	11719	13705	15606	10397	3592	+	H	H
F37	129, 182C, 183C, 189, 223, 249, 311, 359	M1	-	+	-	-	+	-	-	9052	4577	4216	7025	7025	7025	7025	7025	-	-	M	M1

Variant positions from the CRS are shown between 16050 and 16399 in HVRI (minus 16000). Substitutions are transitions unless the base change is explicitly indicated. Insertions of one and two cytosines are shown by 'insC(1)' and 'insC(2)' respectively. The results of specific RFLP screening are also presented. + and - indicates, respectively, the presence and absence of the restriction site. Classification into haplogroups is presented using individual and combined information of sequences and RFLPs. Haplogroups presenting an ambiguous classification are assigned with a question mark (?).

References

Amorim, A., Faria, A. S. N. & Pereira, E. L. (1979) Red Cell Acid Phosphatase (ACP₁) Polymorphism in Flores Island (Azores). *Trabalhos do Instituto de Antropologia "Dr. Mendes Corrêa"*. **35**, 1-4.

Anderson, S., Bankier, A., Barrell, M., Bruijn, M. H. L., Coulson, A. R., Drouin, J., Epcron, I. C., Nicrlich, D. P., Roc, B. A., Sanger, F., Chreier, P. H., Smith, A. J. H., Standen, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-464.

Bandelt, H.-J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37-48.

Bertranpetit, J., Sala, J., Calafell, F., Underhill, P. A., Moral, P. & Comas, D. (1995) Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* **59**, 63-81.

Chen, Y. S., Olckers, A., Schurr, T., Kogelnik, A. M., Huoponen, K. & Wallace, D. C. (2000) MtDNA variation in the south African Kung and Khwe and their genetic relationships to other African populations. *Am J Hum Genet* **66**, 1362-1383.

Chen, Y. S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S. & Wallace, D. C. (1995) Analysis of mtDNA variation in African population reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* **57**, 133-149.

Côrte-Real, H., Macaulay, V., Richards, M., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. & Sykes, B. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**, 331-350.

Cunha, E. (1987) Consanguinity of the Azores Islands (1979-1985). *Antropologia Portuguesa* **4/5**, 67-79.

Di Rienzo, A. & Wilson, A. C. (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* **88**, 1597-1601.

Ewens, W. J. (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* **3**, 87-112.

Felsenstein, J. (1995) *PHYLIP (Phylogeny Inference Package) ver 3.57c*. Department of Genetics. University of Washington, Seattle.

Forster, P., Harding, R., Torroni, A. & Bandelt, H.-J. (1996) Origin and evolution of Native Americans mtDNA variation: a reappraisal. *Am J Hum Genet* **59**, 935-945.

Francalacci, P., Bertranpetit, J., Calafell, F. & Underhill, P. A. (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthropol* **100**(4), 443-460.

Francalacci, P., Montiel, R. & Malgosa, A. (1999) A mitochondrial DNA database - applications to problems of nomenclature and population genetics. In: *Genomic*

- Diversity: Applications in Human population genetics* (eds. S. S. Papiha R. Deka & R. Chakraborty), pp. 103–119. Kluwer Academic/Plenum Publishers, New York.
- Gomes, F. (1997) A ilha das Flores: da redescoberta à actualidade (subsídios para a sua história). Câmara Municipal de Lajes das Flores, Azores, Portugal.
- Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A. & Excoffier, L. (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka population. *Mol Biol Evol* **12**, 334–345.
- Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**, 95–98.
- Handt, O., Meyer, S. & Haeseler, A. (1998) Compilation of Human mtDNA control region sequences. *Nucleic Acids Res* **26**(1), 126–129.
- Helgason, A., Hickey, E., Coodacre, S., Vidar, B., Stefánsson, K., Ward, R. & Sykes, B. (2001) mtDNA and the islands of north Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* **68**, 723–737.
- Helgason, A., Siguroardóttir, S., Gulcher, J., Ward, R., Stefánsson, K. (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet*. **66**, 999–1016.
- Hofmann, S., Jaksch, M., Bezold, R., Mertens, S., Aholt, S., Paprotta, A. & Gerbitz, K. D. (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. *Hum Mol Genet* **6**(11), 1835–1846.
- INE (2001) *Estimativas de população residente 2001 aferidas para os resultados provisórios dos Censos 2001* Instituto Nacional de Estatística, Portugal.
- Lima, M. (1991) *Contribuição para o estudo biodemográfico da freguesia de Rabo de Peixe* Provas de aptidão pedagógica e capacidade científica. University of the Azores, Portugal.
- Lima, M. & Soares, A. O. (1992) Níveis de consanguinidade e endogamia da população de Furnas. *Antropologia Portuguesa* **9–10**, 33–42.
- Macaulay, V. (2000) *Haplogroup motifs*. Retrieved October 8, 2001 from the World Wide Web: <http://www.stats.ox.ac.uk/~macaulay/founder2000/motif.html>
- Macaulay, V., Richards, M., Hickey, E., Veja, E., Cruciani, F., Guida, V., Scozzari, R., Bonnè-Tamir, B., Sykes, B. & Torroni, A. (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–249.
- Mateu, E., Comas, D., Calafell, F., Pérez-Lazaun, A., Abade, A. & Bertranptit, J. (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann Hum Genet* **61**, 507–518.
- Matos, A. (1989) Povoamento e Colonização dos Açores in *Portugal no Mundo* Publicações Alfa, Lisboa, Portugal
- Mendonça, L. (1996) *História dos Açores – Visão geral (sécs. XV–XIX)* Centro de Apoio Tecnológico à Educação, Ponta Delgada, Azores, Portugal
- Montiel, R. (2001) *Estudio diacrónico de la variabilidad del DNA mitocondrial en población Catalana* PhD dissertation. Autonomous University of Barcelona, Barcelona, Spain. (<http://www.tdcat.cesca.es/TDCat-0726101-095837>)
- Montiel, R., Malgosa, A. & Francalacci, P. (2001) Authenticating ancient human mitochondrial DNA. *Hum Biol* **73**(5), 689–713.
- Nei, M. (1987) *Molecular evolutionary genetics* Columbia University Press, New York.
- Page, R. (1996) Treeview: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.
- Parson, W., Parsons, T., Scheithauer, R. & Holland, M. M. (1998) Population data for 101 Austrian caucasian mitochondrial DNA d-loop sequences: Application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* **111**, 124–132.
- Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M. J. & Amorim, A. (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* **65**, 439–458.
- Pereira, L., Prata, M. J. & Amorim, A. (2000). Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet* **64**, 491–506.
- Piercy, R., Sullivan, K., Benson, N. & 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., González, A., Hernández, M., Larruga, J. M. & Cabrera, V. M. (1996) Genetic relationship between the Canary islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* **60**, 321–330.
- Pires, L. (1999) *A Origem da população dos Açores com base em estudos de DNA mitocondrial* BSc dissertation. University of Coimbra, Coimbra, Portugal
- Plaza, S., Calafell, F., Bertranpetit, J. & Comas, D. (2001) *Mitochondrial genome analysis in Human populations from the western Mediterranean* Poster presented in the Euroconference “Inherited Disorders and their Genes in Different European Populations, Exploring Human Genetic Variability.” Sant Feliu de Guíxols, Spain, 15–19 November.
- Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavery, K. & Santachiara-Benerecetti, A. S. (1999) Genetic evidence of an early exit of *Homo sapiens*

- sapiens* from Africa through eastern Africa. *Nature Genetics* **23**, 437–447.
- Rando, J. C., Cabrera, V. M., Larruga, J. M., Hernández, M., González, A. M., Pinto, F. & Bandelt, H.-J. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary islands. *Ann Hum Genet* **63**, 413–428.
- Rando, J. C., Pinto, F., González, A. M., Hernández, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H.-J. (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. & Rousset, F. (1995) An exact test for population differentiation. *Evolution* **49**, 1280–1283.
- Reynolds, J., Weir, B. & Cockerham, C. (1983) Estimation of the coancestry coefficient: basis for short-term genetic distance. *Genetics* **105**, 767–779.
- Richards, M. & Macaulay, V. (2000) Genetic data and the colonization of Europe: genealogies and founders. In: *Archaeogenetics: DNA and the population prehistory of Europe* (eds C. Renfrew & K. Boyle), pp. 139–151. McDonald Institute for Archaeological Research, Cambridge
- Richards, M., Macaulay, V., Bandelt, H. J. & Sykes, B. C. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* **62**, 241–260.
- Richards, M., Macaulay, V., Hickey, E., Veja, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A. & Bandelt, H.-J. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276.
- Rousselet, F. & Mangin, P. (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasians individuals and application to forensic casework. *Int J Legal Med* **111**(6), 292–298.
- Saillard, J., Magalhaes, P. J., Schwartz, M., Rosenberg, T. & Norby, S. (2000) Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. *Hum Biol* **72**(6), 1065–1068.
- Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**(4), 406–425.
- Salas, A., Comas, D., Laureu, M., Bertranpetit, J. & Carracedo, A. (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* **6**, 365–375.
- Santos, C. (2001) *Haplogrupos del DNAmT en la población de las Azores (Portugal)* MSc dissertation. University Autónoma de Barcelona, Barcelona, Spain.
- Schneider, S., Roessli, D. & Excoffier, L. (2000) *ARLEQUIN ver 2.000: A software for population genetics data analysis* Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Smith, M., Abade, A. & Cunha, E. (1992) Genetic structure of the Azores: marriage and inbreeding in Flores. *Ann Hum Biol* **19**(6), 595–601.
- SPSS Inc., 1989–1999. *SPSS for Windows – Release 9.0.0* Chicago.
- Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.
- Thomas, H. (1998) *The slave trade – the history of the Atlantic slave trade 1440–1870*. London: Macmillan Publishers Ltd.
- Thomas, M., Weale, M., Jones, A., Richards, M., Smith, A., Redhead, N., Toroni, A., Scozzari, R., Gratrix, F., Tarekgn, A., Wilson, J., Capelli, C., Bradman, N. & Goldstein, D. (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* **70**, 1411–1420.
- Torroni, A. & Wallace, D. C. (1994) Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *J Bioenerg Biomembr* **26**, 261–271.
- Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. & Wallace, D. C. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850.
- Torroni, A., Bandelt, H.-J., d’Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonne-Tamir, B. & Scozzari, R. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* **62**, 1137–1152.
- Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H. V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards, O., Savontaus, M. L., Huoponen, K., Laitinen, V., Koivumaki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zaheri, N., Santachiara-Benerecetti, A. S., Semino, O. & Scozzari, R. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69**, 844–852.
- Torroni, A., Lott, M., Cabell, M., Chen, Y. S., Lavergne, L. & Wallace, D. C. (1994) mtDNA and the origin of Caucasians. Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet* **55**, 760–776.

Torroni, A., Schurr, T., Cabell, M., Brown, M. D., Neel, J. V., Larsen, M., Smith, D. G., Vullo, C. M. & Wallace, D. C. (1993) Asian affinities and continental radiation of the four founding Native Americans mtDNAs. *Am J Hum Genet* **53**, 563–590.

Watson, E., Foster, P., Richards, M. & Bandelt, H.-J. (1997) Mitochondrial footprints of human expansion in Africa. *Am J Hum Genet* **61**, 691–704.

Received: 2 October 2002

Accepted: 13 December 2002

IV. 3

Artículo III

Analyse du Mouvement Marital à l'île de Flores
(Azores, Portugal)

Santos *et al.* (2004) ***Antropo***, 7: 63-71



Antropo



**Groupement des
Anthropologistes
de Langue Française**

Colloques du Groupement des Anthropologistes de Langue Française (GALF)

Analyse du mouvement marital à l'île de Flores (Azores, Portugal)

Analysis of marital movement in the island of Flores (Azores, Portugal)

Cristina Santos^{1,2}, Manuela Lima³, Augusto Abade², M^a Pilar Aluja¹

1. Unitat Antropologia, Universitat Autònoma de Barcelona, Espagne
2. Departamento de Antropologia, Universidade de Coimbra, Portugal
3. CIRD & Departamento de Biologia, Universidade dos Açores, Portugal

Correspondance: Santos, Cristina. Unitat Antropologia, Dep. BABVE. Universitat Autònoma de Barcelona. 08193 Bellaterra, Barcelona. Espagne. E-mail: Cristina.Santos@uab.es

Mots-clés: Açores, endogamie, exogamie, consanguinité, noms de famille.

Key Words: Azores, endogamy, exogamy, consanguinity, surnames.

Résumé

L'archipel des Açores (Portugal) est situé dans l'Océan Atlantique à 1500 Km du Portugal continental. Il est formé par neuf îles d'origine volcanique qui forment trois groupes: Oriental, Central et Occidental. L'île de Flores, la plus occidentale et isolée de l'archipel, présente des caractéristiques géographiques et démographiques uniques et elle est idéale pour réaliser des études génétiques et biodémographiques. Dans ce travail, on a analysé l'information extraite de 6081 mariages, qui ont eu lieu dans quatre paroisses de Flores dans la période de 1675 à 1880, pour décrire le mouvement marital à Flores et analyser la structure génétique de l'île en utilisant les noms de famille. Les résultats obtenus, relatifs aux analyses d'origine des fiancés, à l'endogamie / exogamie et de l'immigration, semblent indiquer que l'île de Flores a des régions, qui ont eu un haut niveau d'endogamie et une basse immigration, et aussi des régions qui ont eu des valeurs d'endogamie et immigration moyennes. Les résultats de la diversité, ainsi que ceux de la microdifférentiation, indiquent que les paroisses sont génétiquement homogènes. Mais l'effet de l'endogamie au long des générations, conjugué avec les dimensions limitées de la population, influence de manière diverse les fréquences génétiques de chaque paroisse.

Abstract

The Azores archipelago (Portugal) is located in the Atlantic Ocean 1500 Km from the mainland Portugal. It is formed by nine islands of volcanic origin that define three geographical groups: Eastern, Central and Western. Flores Island, the most western and isolated of the archipelago, has unique geographic and demographic characteristics and appears to be ideal to perform genetic and biodemographic studies. In this work, we have analysed the information extracted from 6081 marriages that were celebrated from 1675 to 1880 in four parishes of Flores. The main propose is describing the marital movement in Flores and analyse the genetic structure of the island using surnames. The results obtained for the analysis of the origin of couples, endogamy/exogamy and immigration, seem to indicate that in Flores Island there are regions that have had high levels of endogamy and low levels of immigration, and others with mean values of endogamy and immigration. The results of diversity and microdifferentiation indicate that these parishes are genetically homogenous. On the other hand, the effect of the endogamy across generations combined with the low number of inhabitants could change differentially genetic frequencies in the different parishes.

Introduction

L'archipel des Açores (Portugal), découvert inhabité au XVIème siècle, est situé dans l'Océan Atlantique à 1500 Km du Portugal continental. Il est composé par neuf îles d'origine volcanique qui forment trois groupes: Oriental (S. Miguel et Sta. Maria), Central (Terceira, Faial, Pico, Graciosa et S. Jorge) et Occidental (Flores et Corvo). L'île de Flores est la plus occidentale et isolée de l'archipel, cette île présente des caractéristiques géographiques et démographiques uniques et elle est idéale pour réaliser des études génétiques et biodémographiques. Cette raison a mené la section d'Anthropologie de l'Université des Açores à informatiser tous les registres paroissiaux (Baptêmes, Mariages et Décès) de l'île de Flores dans une base de données généalogique.

Les registres paroissiaux, et concrètement les registres de mariage, sont des sources d'information précieux pour la caractérisation biodémographique et génétique des populations humaines. L'information extraite de ce type de registres permet l'analyse du mouvement marital, de la consanguinité, et aussi des études sur les noms de famille. Les noms de famille sont des caractéristiques culturelles qui se transmettent d'une manière similaire aux caractères génétiques monoparentales.

Nous présentons dans cet article l'analyse de l'information extraite des actes de mariages qui ont eu lieu dans quatre paroisses de Flores depuis 1675 jusqu'à 1875. Les objectifs de ce travail sont les suivants: (i) analyser le mouvement marital dans quatre paroisses de Flores et (ii) utiliser les noms de famille comme indicateurs de la structure génétique de l'île de Flores.

Matériel et méthodes

Données

On a extrait l'information de 6081 actes de mariage, des paroisses de Ponta Delgada, Santa Cruz, Lajes et Fajãzinha (Fig. 1), de la base de données généalogiques de l'île de Flores (Lima, 1998). Cette base de données utilise le software ANALYPOP, développé par l'Equipe de Recherches sur la Dynamique des Populations Humaines (EDYPH) de l'Université du Québec.

Les paroisses considérées se localisent en quatre points distincts de l'île (Fig. 1) et ont un effectif très différent, bien que leur évolution a été similaire (Fig. 2)

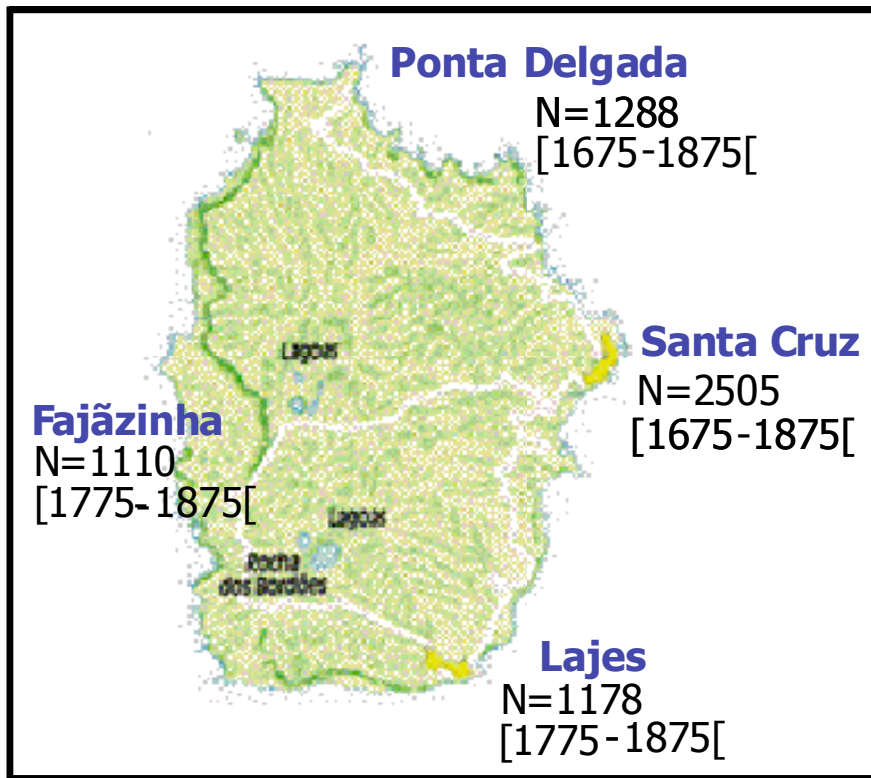


Figure 1. Carte de l'île de Flores. Localisation des paroisses étudiées, nombre de mariages et période analysée.
Figure 1. Map of Flores Island. Localization of studied parishes, number of marriages and analysed period.

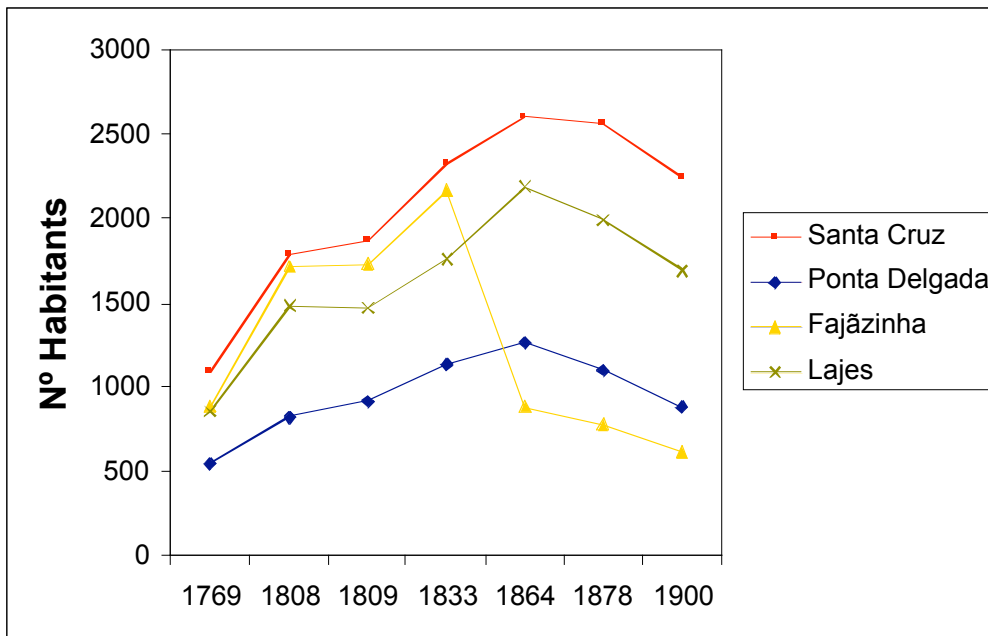


Figure 2. Evolution de la population dans les quatre paroisses étudiées.
Figure 2. Evolution of population in the four studied parishes.

Distribution temporelle des mariages

Les actes de mariage ont été distribués par périodes de 25 ans et l'uniformité de la distribution a été testée avec le test de χ^2 disponible au programme SPSS v. 11.5.1 (SPSS Inc., 2001).

Origine des fiancés

Pour chaque paroisse, les fréquences relatives des lieux de naissance des fiancés (paroisse de mariage, autre paroisse de Flores, Corvo, îles du groupe Central, îles du groupe Oriental et Portugal Continental) sont calculées.

Endogamie/exogamie

Sont considérés comme mariages endogamiques, ceux qui se réalisent entre fiancés nés dans la même paroisse ou la même île, quand les deux sont nés dans une autre île de l'archipel, et exogamiques ceux où au moins un des fiancés n'est pas né dans une paroisse de mariage. Les mariages endogamiques et exogamiques sont classifiés en différents types :

Types de mariages endogamiques:

- end* : les deux fiancés sont de la paroisse de mariage;
- end (HM autre Flores)*: les deux fiancés sont d'autres paroisses de Flores;
- end (HM autre île)* : les deux fiancés sont d'une autre île;

Types de mariages exogamiques:

- exo (H autre Flores)* : la femme est née à la paroisse de mariage et l'homme est né à une autre paroisse de Flores;
- exo (F autre Flores)* : l'homme est né à la paroisse de mariage et la femme est née à une autre paroisse de Flores;
- exo (H autre île)* : la Femme est née à la paroisse de mariage ou à une autre paroisse des Flores et l'homme est né dans une autre île;
- exo (F autre île)* : l'homme est né à la paroisse de mariage ou à une autre paroisse des Flores et la femme est née dans une autre île;
- exo (HM autre île)* : les deux fiancés sont nés dans une autre île; mixtes exogamique

Immigration effective

L'immigration effective a été obtenue avec la formule (Esparza *et al*, 1999):

$$\frac{N^{\circ} \text{ Mariages Exogamiques} + \frac{N^{\circ} \text{ Mariages Mixtes}}{2}}{N^{\circ} \text{ Total Mariages}} \times 100$$

Consanguinité

Dans l'étude de la consanguinité sont utilisés les dispenses ecclésiastiques reportées aux actes de mariage. Le coefficient de consanguinité (α) a été calculé selon Wright (1922).

Noms de famille et structure génétique

Afin d'évaluer la diversité des noms, on a calculé l'index de diversité de Shannon-Weber – H (Shannon-Weber, 1949), le valeur de la redondance de H – R(H) et l' index de Simpson – S (Simpson, 1949). Pour estimer la microdifférentiation a été utilisée la mesure de variance réduite – RST (Relethford, 1988). Ces analyses ont été réalisées avec le programme ISOnymy 3.0 (Abade, 2003).

Résultats et discussion

Distribution temporelle des mariages

La distribution des mariages au long du temps (Fig. 3) montre des différences dans toutes les paroisses (Ponta Delgada: $\chi^2=26,025$; $df=7$; $p<0,001$; Santa Cruz: $\chi^2=66,411$; $df=7$; $p<0,001$; Fajãzinha: $\chi^2=43,571$; $df=3$; $p<0,001$; Lajes: $\chi^2=30,211$; $df=3$; $p<0,001$). Santa Cruz est la paroisse où sont célébrés le plus de mariages; au contraire, Ponta Delgada est la paroisse où le nombre de mariages est le plus bas. La différence entre le nombre de mariages célébrés est liée aux variations d'effectif (Fig. 2).

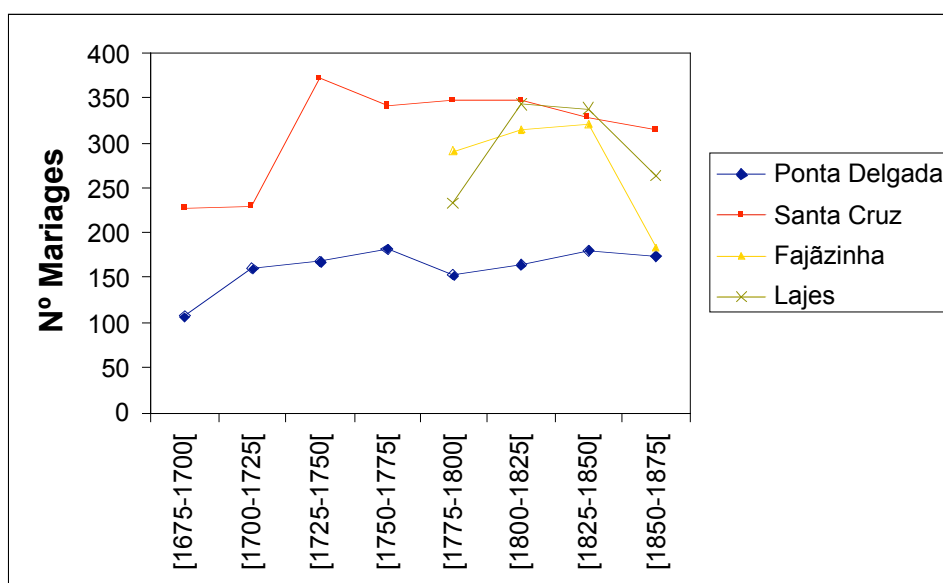


Figure 3. Nombre de mariages au long du temps, par périodes de 25 ans.
 Figure 3. Number of marriages across time, by periods of 25 years.

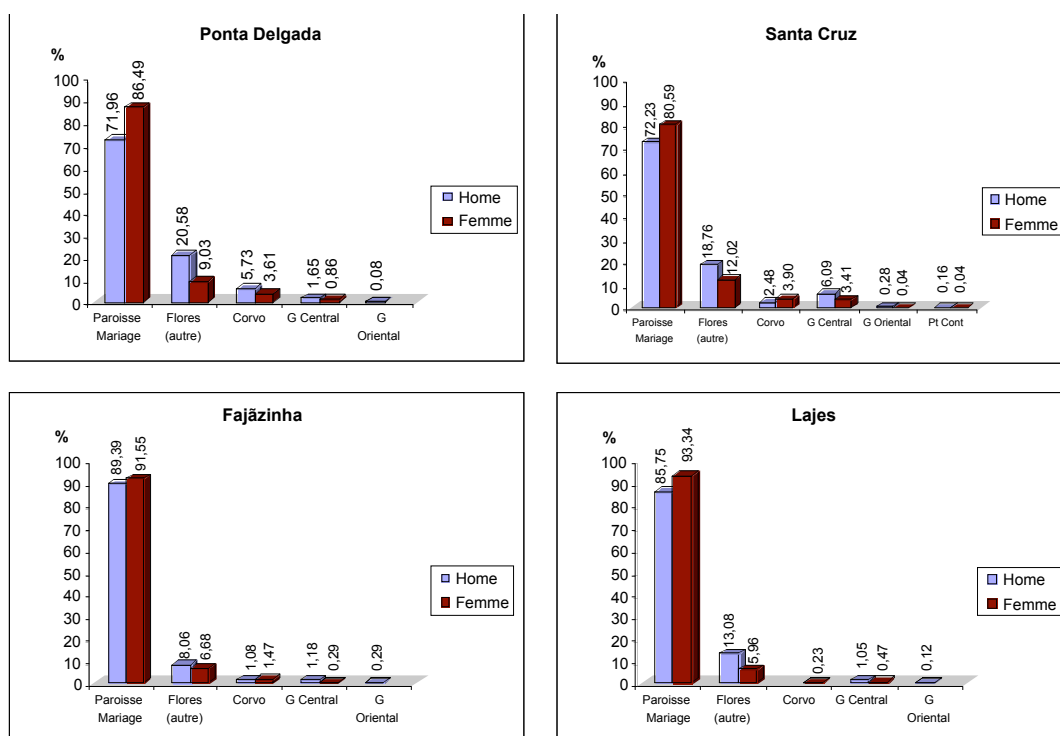


Figure 4. Origine des fiancés dans les quatre paroisses.
 Figure 4. Origin of couples in the four parishes.

Origine des fiancés

Sur la Figure 4 on peut constater que l'origine des fiancés des quatre paroisses montre des tendances similaires.

Comme dans d'autres populations occidentales, la majorité des fiancés sont nés dans la même paroisse où ils se sont mariés. La plupart des individus, qui ne sont pas nés à la paroisse de mariage, sont originaires des autres paroisses de Flores, et une minorité est originaire des autres îles de l'archipel. Les paroisses de Santa Cruz et Ponta Delgada présentent le plus grand nombre de fiancés qui sont nés dans une autre paroisse de Flores, ou sont originaires d'autres îles. La majorité des fiancés qui sont nés dans une autre île sont originaires de l'île voisine de Corvo et du groupe Central. Les individus originaires des îles du groupe Oriental et du Portugal Continental font partie d'une petite minorité.

Endogamie/exogamie

À la Figure 5 on peut observer la distribution des mariages pour chacun des types d'endogamie et exogamie.

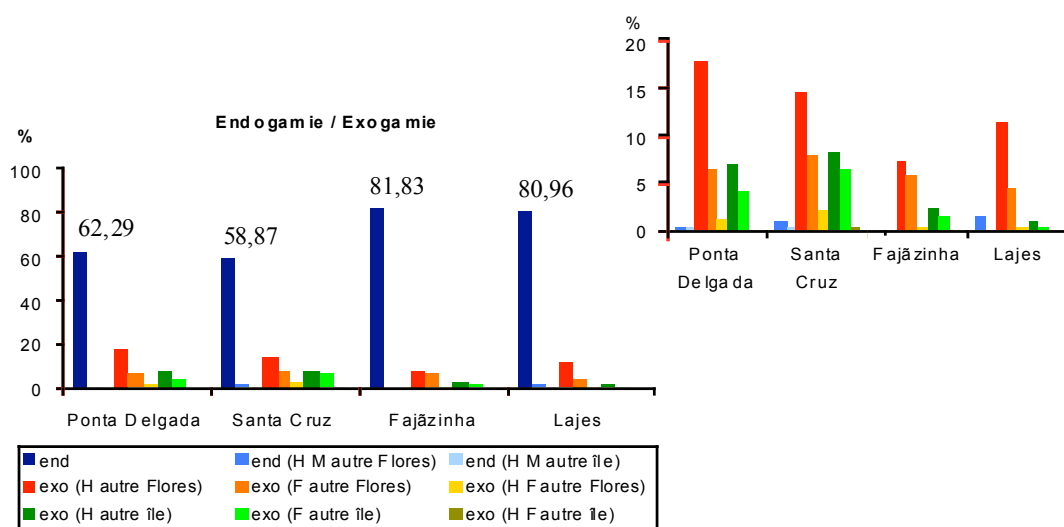


Figure 5. Endogamie et exogamie pendant la période totale.

Figure 5. Endogamy and exogamy at the total period.

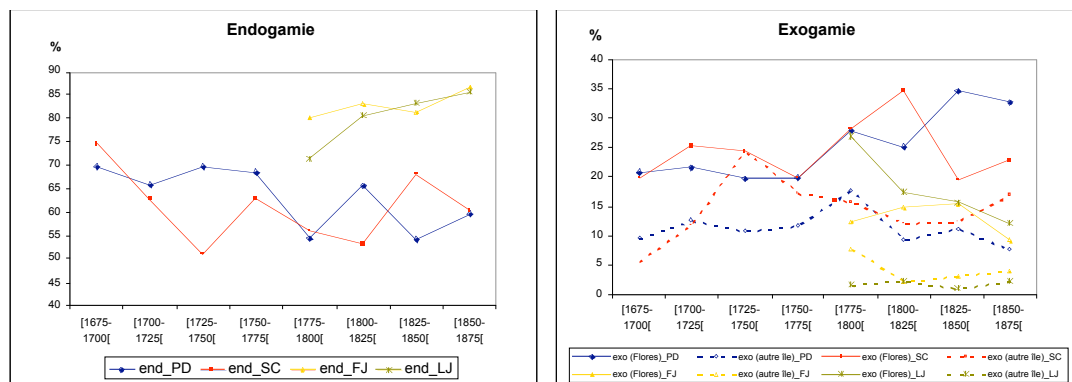


Figure 6. Endogamie et exogamie par périodes de 25 ans (end–mariage endogamique; exo–mariage exogamique; PD–Ponta Delgada; SC–Santa Cruz; FJ–Fajãzinha; LJ–Lajes).

Figure 6. Endogamy and exogamy by periods of 25 years (end–endogamic marriage; exo–exogamic marriage; PD–Ponta Delgada; SC–Santa Cruz; FJ–Fajãzinha; LJ–Lajes).

Santa Cruz a la plus basse fréquence (58,87%) de mariages endogamiques avec les deux fiancés nés à la paroisse de mariage; par contre, Fajãzinha a la plus grande fréquence (81,83%) de ce type de mariages endogamiques. Le pourcentage des autres types de mariages endogamiques est faible

Par rapport à l'exogamie, la plupart des mariages exogamiques sont ceux où un des fiancés est né à une autre paroisse de Flores, mais aux paroisses de Santa Cruz et Ponta Delgada on a

déterminé un important pourcentage de mariages où au moins un des fiancés est né dans une autre île de l'archipel.

Aux paroisses de Santa Cruz et Ponta Delgada l'endogamie a diminué au cours du temps (Fig. 6); par contre, Fajãzinha et Lajes enregistrent une hausse de ces valeurs. À l'exception de Santa Cruz au période [1725-1750], dans toutes les paroisses et périodes les valeurs d'endogamie, donc les fiancés sont nés à Flores, sont supérieures aux valeurs d'exogamie, donc au moins un des fiancés est né dehors de Flores.

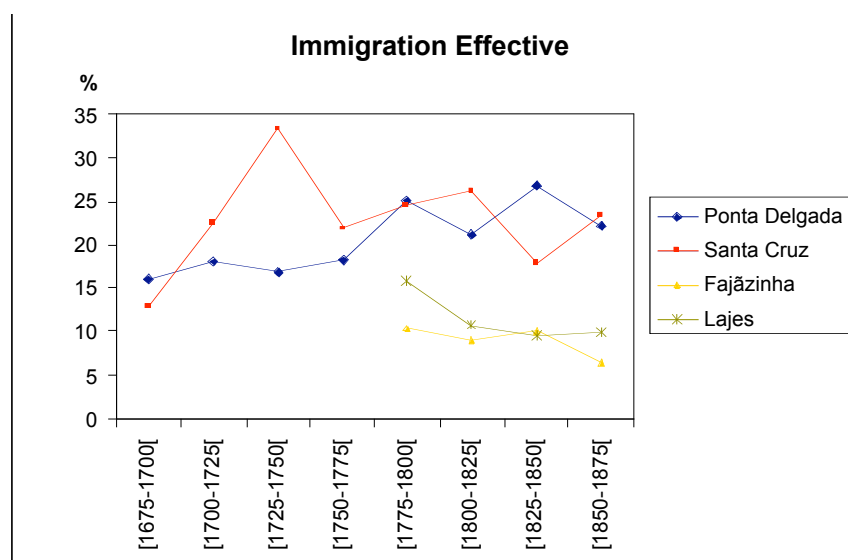


Figure 7. Immigration effective au cours du temps, par périodes de 25 ans.

Figure 7. Effective emigration across time, by periods of 25 years

Immigration effective

Dans la Figure 7, on peut voir que Santa Cruz et Ponta Delgada sont les paroisses les plus ouvertes vers l'extérieur. Depuis 1725 jusqu'à 1749, on a enregistré le maximum d'immigration effective (35%) relative à la paroisse de Santa Cruz. L'immigration en Santa Cruz et Ponta Delgada tend à augmenter tandis que à Fajãzinha et Lajes la tendance est inverse.

Consanguinité (dispense ecclésiastique)

Les pourcentages de mariages consanguins, ainsi que les valeurs du coefficient de consanguinité, sont faibles pendant les deux premières périodes, surtout à Santa Cruz et Lajes (Tableau 1).

	Ponta Delgada		Santa Cruz		Fajãzinha		Lajes	
	% (N)	$\alpha \cdot 10^5$	% (N)	$\alpha \cdot 10^5$	% (N)	$\alpha \cdot 10^5$	% (N)	$\alpha \cdot 10^5$
[1775-1800]	15,1 (152)	272,41	1,7 (346)	25,97	12,7 (292)	196,65	5,6 (234)	70,11
[1800-1825]	8,5 (164)	152,44	1,4 (347)	36,02	12,1 (314)	246,32	3,8 (343)	43,28
[1825-1850]	3,9 (181)	77,69	4,9 (328)	59,55	17,2 (320)	217,29	22,5 (338)	357,11
[1850-1875]	14,4 (174)	338,99	12,1 (314)	226,41	19,6 (184)	467,05	22,8 (263)	337,16
Total	10,3 (671)	207,83	4,9 (1335)	83,98	15,0 (1110)	261,47	13,8 (1178)	204,27

Tableau 1. Pourcentage de mariages consanguins (%), nombre total de mariages (N) et coefficient de consanguinité (α) par périodes de 25 ans.

Table 1. Percentage of consanguineous marriages (%), total number of marriages (N) and coefficient of consanguinity (α) by periods of 25 years.

Au période 1850-1875, les chiffres du coefficient de consanguinité peuvent être considérées médians/hauts, bien qu'inférieurs aux valeurs rapportées à Flores dans les périodes postérieures (Smith *et al*, 1992) et similaires ou inférieurs à ceux rapportés au Portugal continental (Abade, 1986, 1992; Mota, 1983) et sud de l'Europe (Bittles, 1998). Les résultats obtenus dans ce travail,

et dans les travaux précédents (Alves, 2000; Cruz *et al*, 2002), indiquent que à Flores la consanguinité n'était pas enregistrée correctement par les prêtres et qu'elle est certainement sous-estimée, donc il serait plus adéquat d'utiliser des généalogies pour estimer la consanguinité.

Noms de famille et structure génétique

Diversité

Les valeurs de l'indice H témoignent une augmentation de la diversité au cours du temps. Cependant la comparaison statistique montre que seulement les valeurs de H observées dans la première période diffèrent significativement par rapport aux autres périodes en Fajãzinha et à la totalité des quatre paroisses (Tableau 2). Quand on compare les valeurs de H des quatre paroisses, on n'observe pas de différences statistiquement significatives en aucun des périodes considérées. Cependant, quand on étudie la période totale, Santa Cruz présente une valeur de diversité H significativement supérieure à celles obtenus pour les autres paroisses.

Les valeurs de l'index S (probabilité que deux individus de la population présentent le même nom) sont peu élevées dans toutes les paroisses.

Microdifférenciation

Les valeurs de R_{ST} (estimation de la variation entre populations), obtenues par le méthode de Morton, relatifs aux périodes de 25 ans et à la période totale (tableau 2) peuvent être considérées comme peu élevées, si on prend les limites qui ont été suggérées par Relethford (1988): la microdifférenciation est donc peu élevée.

Au cours du temps on observe cependant une tendance à l'augmentation de cette microdifférenciation.

	Ponta Delgada		Santa Cruz		Fajãzinha		Lajes		Total 4 paroisses		R_{ST}
	H	S	H	S	H	S	H	S	H	S	
[1775-1800[4,975	0,048	5,333	0,043	4,616	0,075	5,259	0,038	5,652	0,039	0,0023386
[1800-1825[5,004	0,039	5,929	0,026	5,483	0,042	5,376	0,036	6,205	0,026	0,0017994
[1825-1850[5,120	0,040	6,026	0,021	5,661	0,033	5,301	0,044	6,397	0,022	0,0023606
[1850-1875[5,211	0,039	6,094	0,022	5,439	0,031	5,136	0,050	6,403	0,021	0,0026908
Période Total	5,505	0,035	6,322	0,022	5,652	0,040	5,668	0,036	6,435	0,024	0,0020168

Tableau 2. Valeurs des indices de diversité H et S et valeurs de R_{ST} , par périodes de 25 ans et dans la période totale.

Table 2. Values of diversity indexes H and S, and value s of R_{ST} , by periods of 25 years and in the total period.

Conclusion

Les résultats obtenus relatifs aux analyses d'origines des fiancés, de l'endogamie / exogamie et de l'immigration, semblent indiquer que l'île de Flores a des régions qui ont eu un haut niveau d'endogamie et une basse immigration (Fajãzinha et Lajes), et aussi des régions qui ont eu des valeurs d'endogamie et immigration moyennes (Santa Cruz et Ponta Delgada). Santa Cruz et Ponta Delgada ont été les paroisses les plus ouvertes vers l'extérieur. Les valeurs de consanguinité, qui certainement sont sous-estimées, sont concordantes avec les résultats d'endogamie. Les valeurs d'endogamie et consanguinité les moins élevées s'enregistrent à Santa Cruz tandis que Fajãzinha a les plus hautes.

L'analyse de la diversité des noms de famille indique une grande diversité dans toutes les paroisses, surtout à Santa Cruz. La comparaison des valeurs de H par périodes de temps, indique une tendance à la croissance de la diversité au cours du temps.

Les valeurs de R_{ST} sont basses, donc les quatre paroisses étudiées sont génétiquement homogènes. La croissance de la microdifférenciation au cours du temps peut résulter de la tendance croissante de l'endogamie et de la décroissance de l'immigration effective à Fajãzinha et Lajes, contrairement à Santa Cruz et Ponta Delgada où la tendance de l'endogamie est décroissante et de l'immigration est croissante.

Les résultats obtenus légitiment l'importance de la connaissance des caractéristiques biodémographiques des populations que l'on veut étudier pour des analyses génétiques. Les résultats de la diversité, ainsi que ceux de la microdifférenciation, indiquent que ces paroisses sont homogènes génétiquement. Leur pool de noms ainsi que leur distribution est semblable. D'autre

part, l'effet de l'endogamie au long des générations conjugué avec un effectif réduit les populations, peuvent influencer de manière différente les fréquences génétiques de chaque paroisse. Donc, il faut connaître l'origine des individus échantillonnés et avoir une bonne caractérisation biodémographique des populations quand on veut interpréter les résultats d'analyses génétiques.

Remerciements. Fundação para a Ciência e Tecnologia (SFRH / BD / 723 / 2000), Portugal.

Bibliographie

- Abade, A., 1992, A população inexistente: Estrutura demográfica e genética da população da Lombada, Bragança. Dissertation PhD, Université de Coimbra, Coimbra. pp. 308.
- Abade, A., 1986, Inbreeding as measurement by dispensations and isonymy in Rio de Onor, Portugal. *International Journal of Anthropology*, 1, 225-228.
- Abade, 2003, ISONymy 3. Université de Coimbra, Portugal.
- Alves, C., 2000, Estudo Biodemográfico e Genético de Três Populações de Ilha das Flores (1760-1860). Dissertation BSc, Université des Açores, Portugal. pp. 137
- Bittles, A. H., 1998, Empirical estimates of the global prevalence of consanguineous marriage in contemporary societies. Morrison Institute for Population and Resource studies, Stanford University, California. pp. 69
- Cruz, A.P., Silva, C., Santos, C., Lima, M., 2002, Biodemographic profile of an isolated population – marriage patterns in the parish of Ponta Delgada (Flores Island, Azores, Portugal). *Collegium Antropologicum*, 26 (supplément), 46.
- Esparza, M., Hernandez, M., Susanne, C., 1999, Endogamy, spouse origins and inbreeding in a parish of the Ebro Delta (Spain). *Homo*, 50/3, 211-220.
- Lima, M., 1998, Demografia Histórica e Dinâmica Bio-Cultural da Ilha das Flores (séculos XVII a XIX) – Praxis/PCSH/C/DEM/56/96. Université des Açores, Portugal.
- Mota, P. J. G., 1983, Consanguinidade aparente em S. Vicente da Beira. *Antropologia Portuguesa*, 1, 5-32.
- Relethford, J., 1988, Estimation of kinship and genetic distance from surnames. *Human Biology*, 60, 475-492.
- Shannon, C. E., Weaver, W., 1949, *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.
- Simpson, E. H., 1949, Measurement of diversity, *Nature*, 163, 688.
- Smith, M., Cunha, E., Abade, A., 1992, Genetic structure of the Azores: marriage and inbreeding in Flores. *Annals of Human Biology*, 19(6), 595-602.
- SPSS Inc., 2001, *SPSS for Windows – release 11.5.1*, Chicago.
- Wright, S., 1922, Coefficients of inbreeding and relationship. *The American Naturalist*, 645(56), 330-338

IV. 4

Artículo IV

Genetic Structure of Flores Island (Azores, Portugal) in the 19th Century and in the Present-day: Evidences from Surname Analysis

Santos *et al.* ***Human Biology (in press)***

Genetic Structure of Flores Island (Azores, Portugal) in the 19th Century and in the
Present-day: Evidences from Surname Analysis

Cristina Santos^{1, 2}, Augusto Abade², Jordi Cantons¹, Francine M. Mayer³, M^a Pilar Aluja¹ and Manuela Lima⁴

1. Anthropology Unit, Department BABVE, Faculty of Sciences, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain
2. Department of Anthropology, University of Coimbra, 3000 Coimbra, Portugal
3. EDYPH (Équipe de Recherche sur la Dynamique des Populations Humaines), University of Quebec in Montreal, Montreal, Canada
4. CIRN and Department of Biology, University of the Azores, 9500 Ponta Delgada, S. Miguel, Azores, Portugal

Key-words: Azores Islands, Flores Island, surnames, kinship, genetic diversity, genetic structure

Corresponding author:

Cristina Santos

Anthropology Unit, Department BABVE

Faculty of Sciences

Autonomous University of Barcelona

08193 Bellaterra (Barcelona), Spain

Tel: +34 935812049

Fax: +34 935811321

e-mail: Cristina.Santos@uab.es

Abstract

The island of Flores is the most westerly of the Azores Archipelago (Portugal). Despite its marked geographical isolation and reduced population size, biodemographic and genetic studies conducted so far do not support the idea that its population constitutes a genetic isolate. In this study we conducted a surname analysis of Flores population, in the second half of the 19th century and also in the present-day population. Main purposes were: a) biodemographically and genetically characterise the island taking in care the strong reduction in population, observed from the middle of the 19th century until the present-day; and b) analyse the influence that the effective population size and geographic distance have on the genetic structure of populations. For both periods analysed, all indicators of diversity revealed a high level of surname diversity. Results obtained are in accordance with the diversity estimates obtained both with monoparental genetic markers located in the Y chromosome and using the frequencies of mtDNA haplogroups. Contrarily to what could be expected, considering the strong reduction of population in the last 150 years, a maintenance of diversity and a decrease in microdifferentiation, was observed supporting a higher overture of parishes as a consequence of the increase of the communication routes. From the first to the second period analysed a change in surname composition is evident, although the more frequent surnames in Flores are almost the same for both periods and some of them are reported to be surnames present in the first settlers of Flores. This result testifies the impact of founders on the present-day genetic pool of Flores Island and allows to infer that the genetic characterization of the present-day population of Flores could provide reliable information about the history of peopling.

Introduction

The Azores Islands (Portugal), uninhabited when discovered by Portuguese navigators in the fifteenth century, are located in the Atlantic Ocean, 1500 Km from the European mainland. The archipelago is formed by nine islands of volcanic origin that define three geographical groups: the Eastern group comprising the islands of São Miguel and Santa Maria, the Central group with Terceira, Faial, Pico, Graciosa and São Jorge and finally, the Western group composed of Flores and Corvo. With a total area of 2344 Km², the Azores Islands have presently a total population of 241763 inhabitants, distributed in a very asymmetric way among the three groups of islands (INE, 2001).

The island of Flores is the most westerly island of the Archipelago and one of the smallest, both in area (143.11 Km²) and inhabitants (3995 inhabitants) (INE, 2001). Doing to the rugged volcanic terrain of the island, that limit the population movement by land, the centres of population on Flores were founded in the coast (Fig. 1) and communications across the island were, until recently, performed by the scarce network of roads or by coastal navigation. The island has presently two municipalities, Santa Cruz and Lajes; the municipality of Santa Cruz comprises actually the *freguesias* (Portuguese word for parish, corresponding to a civil administrative area often comprising several settlements, and that is generally equivalent to an ecclesiastic parish) of Santa Cruz, Caveira, Ponta Delgada and Cedros. The municipality of Lajes comprises the parishes of Lajes, Fazenda, Lomba, Lajedo, Mosteiro, Fajãzinha and Fajã Grande (Fig. 1). The oldest parishes of the island, established by the end of the 16th century, are Lajes and Santa

Cruz, whereas the most recently created parishes of the island are Fajã Grande (separated from Fajãzinha in 1862), and Fazenda (detached from Lajes only by 1919).

Flores Island, together with its neighbour Corvo, was the latest island of the Azorean Archipelago to be discovered and peopled. All the islands of the Azores, especially Graciosa and Terceira, as well as Madeira Island, Mainland Portugal and other regions of Europe contributed to the peoplement of Flores Island (Matos 1989; Lages, 2000). An important number of African slaves also made up part of the early settlers (Gomes, 1997). In the beginning of the sixteen century, nearly 50 years after the discovery of the island, the settlement process was initiated (Lages, 2000). By 1587 demographic information, based on parish priests' registers of "souls making confession", referred to only 800 inhabitants (Smith *et al.*, 1992). Subsequently the population grew rapidly and by 1778 the number of inhabitants was already 6275, reaching a maximum in 1856 of 10,937 individuals (Gomes, 1997). After 1856, as result of emigration mainly to North America, the population of Flores began a quick and marked decline, and by 2001 there were only 3995 inhabitants in the Island (INE, 2001).

The peripheral position of Flores Island in the Azorean archipelago and its reduced population size, strongly affected by emigration, favoured the establishment of a population with particular demographic, social and bio-cultural dynamic patterns, that make of Flores Island an appealing case for studies of population and evolutionary genetics. Biodemographic studies, focused on the population of Flores Island (Smith *et al.*, 1992), indicated the presence of levels of inbreeding similar to those observed in other Azorean (Cunha *et al.*, 1987),

mainland Portugal (Abade, 1986, 1992; Mota, 1983) and southern European (Bittles, 1998) populations. Corroborating findings from biodemographic analyses, DNA studies conducted by our team also failed to support the hypothesis that Flores Island constitutes a genetically isolated population, contrary to what could be expected on the basis of its geographic isolation. The study of mtDNA sequences of HVRI, revealed that genetic diversity presented low values in the Western group of islands, when compared with the other groups. However, lower levels of diversity seemed to be strictly related with the reduced number of females since peopling occurred (Santos *et al.*, 2003). Moreover, when mtDNA genetic diversity was calculated taking into account the distribution of mtDNA in major haplogroups (Santos *et al.*, 2001; Montiel *et al.*, *in press*) the value of diversity observed in the Western group was similar to or even higher than the observed in other groups of islands. In regards to the analysis of haplogroups defined using bi-allelic markers of the non-recombining portion of the Y chromosome, the results showed that the Western group presented the highest values of diversity when compared to the other groups of islands (Montiel *et al.*, *in press*).

Parish registers are repositories of precious information that allow to characterize, genetically and biodemographically, human populations. The information extracted from this source provides useful data to calculate important biodemographic indicators like endogamy/exogamy and consanguinity. Surnames are among the more important information that can be extracted from parish records since they can be used as a genetic marker. The main hypothesis has stipulated that people with the same surname all descended from the same common ancestor. In many human populations the surname is inherited in the paternal lineage, and can

therefore be considered a marker for the Y chromosome, and even though surnames are cultural traits (Cavalli-Sforza and Feldman, 1981) and are sometimes polyphyletic, they provide an inexpensive method for obtaining a rough estimation of genetic variation and are very often the only source of data available to study the genetic structure of populations from the past (Relethford, 1988). For these reasons, surnames have been used for long time to study the genetic structure of human populations (see revision of Colantonio *et al.*, 2003). However to the Azores Islands there is only a study of population structure of the island of São Miguel using surnames (Branco and Mota-Vieira, 2003).

In the present research the information extracted from the register of marriages that took place in the second half of the 19th century in Flores Island (the period in which the island rise the high number of inhabitants) as well as the surnames of all individuals with more than eighteen years of age, resident in Flores Island during the electoral census of the year 2003, were collected and analysed in order to: a) characterise biodemographic and genetically the population of Flores Island during the second half of the 19th century and in the present-day; b) discuss the influence of the effective population size and geographic distance in the genetic structure of Flores Island; c) document the potential changes that could have occurred in the genetic structure of Flores Island population in the last two centuries; d) investigate the impact, in the genetic structure of Flores Island, of the strong reduction in population, that occurred from the middle of the 19th century until the present-day.

Material and Methods

Population and Data

In this work we considered the administrative subdivision of Flores Island established in 1862, formed by ten parishes in the Island (Fazenda parish being still considered as a part of Lajes parish) (Fig.1). We created two types of data files extracted from two types of sources (parish records for the 19th century and nominative census list of the year 2003 for the present-day population).

To characterize the population of Flores in the 19th century we used the information contained in parish registers of marriage, computerized in the genealogical database of Flores Island (Lima, 1999). During the building of the database, it was realized that women, in general, did not have any surname; therefore, we decided to work with the surnames of the fathers of both members of the couple. The place of birth of both members of each couple and the surname of their respective fathers were compiled for a total of 1486 marriages registered that took place during the second half of the 19th century (389 in Santa Cruz, 80 in Caveira, 133 in Cedros, 212 in Ponta Delgada, 308 in Lajes, 83 in Lomba, 30 in Lajedo, 56 in Mosteiro, 94 in Fajãzinha and 101 in Fajã Grande). From the 1486 marriages, we would expect to have 2972 individuals each of them with one surname; however, since some of the individuals didn't have family name, the number of individuals considered to build the database of surnames of the 19th century was 2582.

The electoral census list of the year 2003 provided the information to characterize the present-day population of Flores Island. The father's surname of

each of the 3193 individuals listed in the census was recorded, corresponding to the inhabitants of Flores with 18 years of age or older at the time of the census: 1387 for Santa Cruz, 71 for Caveira, 136 for Cedros, 355 for Ponta Delgada, 669 for Lajes, 160 for Lomba, 90 for Lajedo, 52 for Mosteiro, 89 for Fajãzinha and 184 for Fajã Grande.

Surnames of both periods were pooled together in order to detect and uniform all the surnames that were written in different manners but had the same phonetics; for example: Rois (an abbreviation used for Rodrigues), Rodriguez and Rodrigues are distinct ways of writing the same surname, Rodrigues.

Types of endogamous and exogamous marriages

In order to measure the levels of endogamy and exogamy in the 19th century, for each parish, the place of birth of each member of the couple was used to develop categories of marriage. A marriage was defined as endogamous when it was celebrated between individuals born in the same parish. Two types of endogamy were considered: *end1*- both members of the couple were born in the parish of marriage; *end2*- both members of the couple were born in the same parish, although it was not the parish of marriage. Marriages were considered exogamous whenever the place of birth was not the same for the two members of the couple. Four types of exogamy were considered: *exo1*- when the wife was born in the parish of marriage while her husband was born in another parish of Flores; *exo2*- when the husband was born in the parish of marriage, while it is his wife that was born in another parish of Flores; *exo3*- when husband and wife were born in

different parishes of Flores, other than parish of marriage; *exco4*- when at least one of the members of the couple was born outside Flores Island.

Surnames and Genetic Structure

In order to evaluate the diversity levels of the different parishes, the number of different surnames by parish and period, the percentage of parish specific surnames, as well as the values of various indicators of diversity, were used. Indicators of diversity, all of them based in the distribution of surnames, included:

a) The diversity index of Shannon-Weaver – H (Shannon-Weber, 1949)

This diversity index was first applied to the information theory (Shannon-Weber, 1949) and latter was used in ecology by Margalef (1957). It is an index of heterogeneity since it groups the concepts of richness (number of surnames) and equitability (uniform distribution of surnames). Being p_i the frequency of a given surname in a group of N individuals with k surnames, the diversity index is:

$$H = -\sum_{i=1}^k p_i \log_2 p_i$$

and the approximate value of variance (Magurran, 1988):

$$\sigma^2 = \frac{\sum_{i=1}^k p_i (\ln p_i)^2 - \left(\sum_{i=1}^k p_i \ln p_i \right)^2}{N} + \frac{\left(\sum_{i=1}^k p_i^2 \right) - 1}{2N^2}$$

The observed diversity of a distribution of surnames is easily understood if compared with its maximum (H_{max}) limiting cases: i) H_{max1} - a group of N individuals all with different surnames, in which $H_{max1} = \log_2 N$ (Barrai et al., 1990), or ii) H_{max2} - a group with k surnames all with the same frequency, in which $H_{max2} = \log_2 k$.

The minimum (H_{min}) of H can also be deduced assuming: i) H_{min1} - all the individuals have the same surname, in which $H_{min1}=0$ (Barrai et al., 1990), or ii) H_{min2} - a group with k surnames in which $k-1$ surnames have the frequency 1 and the remaining surname is present in all the other individuals of the population the maximum value of all with the same frequency, in which the value of H is

$$(DeBenedictis, 1973) H_{min2} = \log_2 N - \frac{N - k + 1}{N} \times \log_2 (N - k + 1).$$

These values describe the maximum and minimum of diversity, H_{max1} and H_{min1} strictly related with the number of surnames and H_{max2} and H_{min2} with equitability, that is, the way by which surnames are distributed in a population.

b) Heterozygosity (or heteronymy, in the case of surnames) (Harpending and Ward, 1982):

$$Het = 1 - \sum_{i=1}^k p_i^2$$

To access the degree of isolation of the populations, the Harpending and Ward (1982) model was applied. According to the referred model, expected heterozygosity, as previously defined, and the intra-population kinship coefficients, calculated as defined by Harpending and Jenkins (1973) – HJ-R_{ii}, can be related to test the effect of different evolutionary forces, such as differences in population size, migration patterns or bottlenecks; given uniform systematic pressure, genetic heterozygosity or heteronymy is negatively correlated with the values of intra-population kinship.

To access the impact of population size in diversity and intra-population kinship, a non-parametric analysis of correlation was performed, and the nonlinear model of regression proposed by Dahlberg (1948), were used to evaluate the

relation between Morton intra-population conditional kinship (R_{ii}) and the number of inhabitants (N_i):

$$R_{ii} = a \left(\frac{1}{N_i} \right)^b$$

The previously described analyses are useful to have an idea of the isolation of parishes, however, they don't tell us anything about the inter-population relationships. To understand the relations between parishes, several analyses were used: 1) the conditional kinship was calculated using the method proposed by Morton *et al.* (1973) based in random isonymy (same surname). The matrices of Morton conditional kinship and the average within-group conditional and a priori kinship (respectively R_{ST} and F_{ST}) were calculated as described in Relethford (1988), using the program ISOnymy v. 3.0.18 (Abade, 2003). 2) For both periods under analysis, a correspondence analysis, using the distribution of surnames with an intra-parish frequency higher than 3% (the limit of 3 % was established since the more frequent surname in Santa Cruz has a frequency around 4%), was performed. 3) To evaluate the interrelationship between surnames and geographic distance (lineal, by roads and coastal navigation), the matrices of Morton conditional kinship were compared with the geographic distribution of populations. Matrix comparisons were performed by MATFIT procedure (Lalouel, 1973), which rotates two matrices to maximum congruence and allows a visual identification of mismatch between pairs of points as an aid to the interpretation of results (Roberts *et al.*, 1981). Correlation between matrices and their significance were calculated by Mantel tests (Mantel, 1967; Smouse and Long, 1992), using the program IBD: Isolation By Distance (Bohonak, 2002). 4) The Monmonier's (1973) maximum-

difference algorithm, implemented in the program Barrier vs. 2.2 (Manni, 2004), was used to identify boundaries between populations. In this analysis the linear geographic distance (Km) between populations and the matrices of Morton conditional kinship were used.

The correspondence analyses, the analyses of non-parametric correlations, the regression analyses and the scatterplots were performed with the statistical package SPSS v. 11.0 (SPSS inc., 2001).

Results and Discussion

Parish endogamy and exogamy in the second half of the 19th Century

The percentages of endogamous and exogamous unions that took place during the second half of the 19th century are presented in table 1. Parishes belonging to the municipality of Santa Cruz presented values of endogamy comprised between 45.1% and 60.8%, respectively in Caveira and Cedros. Values for this municipality are lower than those observed in the parishes of the municipality of Lajes, where the percentages of endogamy ranged from 71.1% to 87.8%, in Mosteiro and Fajãzinha respectively. The high levels of *end2* observed in Caveira and Mosteiro only represent 11 marriages of individuals from Lomba and 9 marriages of individuals from Fajãzinha, respectively. Furthermore, the referred marriages took place across the fifty years under study, and the high percentage observed could only result from the small number of marriages that occurred in Mosteiro and Caveira in the period analysed. The percentage of exogamous marriages where at least one of the members of the couple was born outside Flores island (*exo4*) ranged from 0% in Lajedo to 20.8% in Santa Cruz. The combined analysis of endogamy and exogamy values indicates the existence of two distinct areas in the island, the first (municipality of Santa Cruz) with medium-low levels of endogamy and relatively high levels of immigration from outside the island and the second one (municipality of Lajes) with high levels of endogamy and low levels of immigration from outside the island.

Surnames and Genetic Structure

In table 2 the number of different surnames detected, for each parish and for the whole island, is presented. Considering the island of Flores as a whole, during the second half of the 19th century, out of 2582 individuals considered, 291 different surnames were listed and in the present-day 303 different surnames were recorded in 3193 individuals (table 2). The analysis of the distribution of surnames in all the parishes puts in evidence the high percentage of surnames specific of a determined parish, with more than fifty percent of the surnames detected in Flores being present in only one of the ten parishes (table 2). Moreover, with the exception of Santa Cruz, there is in all the parishes a small variation or a reduction (very pronounced in Cedros and Caveira) in the percentage of parish specific surnames from the 19th century to the present-day (table 2). Even if Santa Cruz presents the highest percentage of specific surnames (table 2), all the other parishes, for both periods, share the highest number of surnames with Santa Cruz (data not shown). This indicates that Santa Cruz, being the parish with the highest number of different surnames, simultaneously shares an important number of surnames, while it maintains an important number of surnames specific to it.

The number of surnames found simultaneously in the two periods as well as the number and percentage of lost and added surnames are also reported in table 2. The percentage of surnames present in the 19th century, but lost since then, can be considered high, ranging from 41.56% in Santa Cruz to 83.72% in Caveira. The ratio of lost/added surnames indicates that there are important differences between parishes, as illustrated by two extreme groups of parishes: a first group, formed by

Caveira, Cedros, Mosteiro and Fajãzinha with a ratio superior to 1.5, indicating that the loss of surnames is much higher than the addition. A second group formed by Santa Cruz and Lajes, presents the lowest values of lost/added ratio, with the highest adding of surnames from the 19th century to the present-day. Considering the island as a whole, half of the surnames present in the actual population, are absent in the second half of the 19th century.

The values of the different indicators of diversity for each parish and period are presented in table 3. The values of the Shannon-Weaver index (H), obtained for each parish and period, are closer to its corresponding maximums than to its minimums (table 3). This reveals a medium/high level of surname diversity in both periods, that results primarily from the high equitability and also from the high number of surnames, taking into account the number of individuals considered for each parish. In both periods the highest value of H was observed in Santa Cruz whereas the lowest was reported in Lajedo. Statistical significant differences in the values of H, obtained for the 19th century, were only observed when Santa Cruz was compared with all the other parishes. In the present-day, Santa Cruz also presents significant high levels of diversity with respect to all the parishes. The value of H reported in Lajes was significantly higher than the observed in all the parishes except Fajã Grande and Ponta Delgada. On the other hand, the value of H observed in Lajedo was smaller than the observed in Santa Cruz, Lajes, Ponta Delgada and Fajã Grande. Comparing the values of H obtained in the 19th century with today populations, significant differences were only detected in Lajes, where the value of H is significantly higher in the present-day population. The analysis of heteronymy are in accordance with the values of H (table 3).

A detailed analysis of the values of Morton intra-population kinship (R_{ii}) in each parish (table 3) puts in evidence, considering the limits proposed by Relethford (1988), the presence of medium-high levels of intra-population kinship ($R_{ii} > 0.004$) in the parishes of Cedros, Lajes, Lajedo, Caveira, Lomba and Fajã Grande, in the 19th century and in all the parishes except Santa Cruz and Lajes, in the present-day. On the other hand, the others parishes present low or very low (Santa Cruz) values R_{ii} . Furthermore, comparing the second half of the 19th century with the present-day, in seven parishes an increase of kinship (that is very pronounced in the case of Lajedo) is observed.

Taking together the results presented in tables 2 and 3, it appears that in both periods, and considering the entire island, the levels of surname diversity can be considered high. However, important differences arise from the comparison of parishes, both in the 19th century and present-day, indicating that some parishes are more inbred than others. Applying the Harpending and Ward (1982) method to the data the most isolated population appears to be Cedros in the second half of the 19th century (Fig. 2a), and Lajedo in the present-day (Fig. 2b). The remaining populations, in both periods, seems to present an adjusted value of heteronomy attending to it intra-population kinship.

One of the factors that can affect the diversity and the intra-population kinship values is the population size. A significant correlation is observed between population size and heteronomy (Rho Spearman correlation: $r=0.704$; $p=0.001$) and between population size and intra-population kinship (Rho Spearman correlation: $r=-0.688$; $p=0.001$). When the non-linear model proposed by Dahlberg (1948) was applied to evaluate the dependence between R_{ii} and the number of inhabitants, a

significant dependence was found ($R^2=0.32$; $F=29.49$; $p<0.001$). The graphical representation of the relation between variables (Fig. 3) allows the identification of the populations in which the values of kinship cannot be explained by their effective population size. In the 19th century the most extreme positive deviations were observed for Cedros and Lajes, indicating that these parishes have a higher value of R_{ii} than predicted attending to its population size; on the other hand, the parishes of Mosteiros, Fajãzinha and Santa Cruz have a lower value of R_{ii} than would be expected attending to population size. In the present-day, Lajedo and Lomba have a higher value of R_{ii} than predicted, whereas all the other parishes have a smaller or similar value of R_{ii} than expected attending to population size. These results indicate that the high levels of isolation previously reported in Cedros (19th century) and Lajedo (present-day), when heteronymy and intra-population kinship were related, cannot be explained by the effective population size, since again these populations appeared as outliers (Fig. 3).

Values of conditional kinship between parishes are reported in table 4. Cedros and Lajedo, in the second half of the 19th century and in the present-day, respectively, present the highest values of inter-population kinship. This result seems contradictory attending to the fact that Cedros in the 19th century and Lajedo in present day present the highest values of R_{ii} and appear as the more isolated parishes in the different analysis performed. Nevertheless, they also present the highest values of kinship with other parishes indicating that many of the surnames present in these parishes are also present in other parishes of the island. The lowest values of inter-population kinship are observed, for both periods, in Santa Cruz. Since all the parishes share with Santa Cruz the highest number of surnames, as it

was previously referred, these results are apparently contradictory. However, they can be explained by the high number of parish specific surnames present in Santa Cruz.

In figures 4 ad 5 the results of the correspondence analysis using surnames that have an intra-parish frequency higher than 3 % are presented. Table 5 lists the surnames used in the correspondence analysis in each period; underlined are surnames that make part of a list of 34 surnames of some of the first settlers of Flores Island (for a detailed list see Lages, 2000, that uses data based on Frutuoso, 1580-1591; Chagas, 1645; Cordeiro, 1717; Silveira, 1960) and could represent signs of the founder population of the Island.

The correspondence analysis for the second half of the 19th century (Fig. 4), shows that the clustering of populations can be in part related with the geographical structure of the Island. In the first dimension a separation of parishes in a North-South gradient can be inferred, whereas the second dimension separates the parishes in an East-West direction. In the negative quadrant of dimension 1 and positive of dimension 2, a cluster formed by all the parishes of the municipal of Santa Cruz can be identified. Fajã Grande is in the negative quadrant of dimension 2, however it is very close of the origin and it is more reliable to include it in the cluster formed by the parishes of the municipal of Santa Cruz. Two other clusters, formed by the remaining parishes of the municipal of Lajes, can be identified: one formed by the parishes of the South (Lajes and Lajedo) and Southeast (Lomba) coasts of the island and a third cluster formed by parishes of the Southwest (Fajãzinha and Mosteiro) coast of the island. Moreover, the identification of founder surnames in each cluster of populations indicates that there were two

principal points of settlement of Flores, one in the North of the Island and another in the South. This result is consistent with historical data that refer to Lajes and Santa Cruz as the oldest villages of the Island.

For the present-day population, the correspondence analysis (Fig. 5) allows to identify three principal clusters of populations, one defined mainly by dimension 1, formed by Fajãzinha and two defined by dimension 2: one positioned in the negative extreme of dimension 2, formed by Lomba and Caveira and another formed by Lajes, Lajedo, Mosteiro, Santa Cruz, Cedros, Ponta Delgada and Fajã Grande. The position of Santa Cruz and Lajes in the present-day, near the centre of coordinates indicates that these parishes does not present important differences relatively to the other parishes of the Island and since they presently present the highest number of inhabitants, they can be considered representative of the genetic variation of the island.

The results of correspondence analyses indicate that the geographic structure of the island can exert some influence on the genetic structure of its populations, especially on what concerns the second half of the 19th century. As previously referred, the centres of population on Flores are in the coast and communication across the island has always been difficult doing to the rugged volcanic terrain. In such a situation, it would be predictable that migration and gene flow would be dependent upon either the scarce network of roads or coastal navigation, rather than linear geographical distance itself (Smith *et al.*, 2000). To test this hypothesis, the correlation between the matrices of conditional kinship, and geographic distances (lineal, by roads and by coastal navigation) was evaluated using a Mantel test. In all cases the correlation was not significant, both in the 19th

century and in the present-day. Furthermore, the congruence of matrices of geographic distances and conditional kinship, evaluated by the MATFIT procedure, indicates that populations are much closer genetically than would be expected by geography (data not showed).

In figure 6 are presented, for both periods, the three boundaries inferred using the matrix of Morton conditional kinship. In the 19th century (Fig. 6a), Cedros appears as the most isolated parish. The high isolation of Cedros, evident by its high intra-population kinship and by the results of the Harpending and Ward (1982) method, is hard to explain based only on its small effective population size, since the value of kinship in Cedros is higher than would be expected attending to its population size. Furthermore, this population presents high values of inter-population kinship, indicating that the differentiation of Cedros neither can be explained by the presence of a different pool of surnames. In our opinion, these results could be the outcome of simply chance operating in small populations, where the number of events was small. Boundaries two and three, define a clear separation between the North-Centre and the South of the island, this separation is also observed in the correspondence analysis, where a north-south gradient can be inferred. The parishes of the south coast of the island, as previously stated, presented the highest values of endogamy and lowest values of input from the outside of the island. Moreover in these parishes there is a small number of surnames that determined differentiation. All together, these results raise the hypothesis of a socio-cultural boundary separating the two extremes of the island, a more open structure in the North-Centre region, with a population more overt to the outside (including to the exterior of the island), contrasting with a closed structure

of the southern part of the island. The fact that Santa Cruz has represented, since the initial settlement that main gate of entrance in Flores (Matos, 1989), and concentrates the main economic activities of the island has probably promoted, over historical times, a differential socio-cultural behaviour.

In the present-day a decrease of isolation can be confirmed from the results (Fig. 6b). The population of Cedros that is, in the 19th century, the most isolated, appears in the present-day integrated with the other populations of the North-Center region of the island. Presently, the most isolated population appears to be Lajedo, a result consistent with the previous observations. As Cedros in the 19th century, the high isolation of Lajedo in the present-day neither can be explained by its small effective population size alone, nor by the presence of a different pool of surnames and can also result from chance. In the present-day all the three barriers computed are located in the south of the island evidencing, first, the high isolation of this region respect to the rest of the island and secondly, the isolation observed between the parishes of this region.

Even if the previous results pointing out some differences between parishes, the values of the average within-group conditional kinship (or reduced variance) - R_{ST} (Relethford, 1988) are, in the 19th century ($R_{ST}=0.00413$) and in the present-day ($R_{ST}=0.00323$), near to the upper limit defined by Relethford (1988) for samples with low values of microdifferentiation (low: 0.0002-0.0039; moderate to high: 0.0087-0.0193), indicating that Flores Island present low levels of microdifferentiation in both periods.

Branco and Mota-Vieira (2003) recently published a surname analysis for one of the 9 Azorean Islands (São Miguel). This island, the largest of the

archipelago, both in surface and in population, has a value of F_{ST} of 0.0016, lower than the value observed in Flores Island in the present-day population ($F_{ST} = 0.00709$). For many reasons the information provided by the values of R_{ST} can be considered more useful than the one provided by F_{ST} values (Relethford, 1988), since the later most often refers to an unknown founding population that, as Crow (1983) pointed out, has a value of kinship equal to zero (this meaning that the founding population is one where each individual has a different surname). Therefore, it is also difficult to compare values of F_{ST} from different populations. The value of F_{ST} can vary considerably depending on the size of subdivision considered (Relethford, 1988); since Flores Island, one of the smallest of the archipelago, has approximately 33 times less inhabitants than São Miguel, and the number of subdivisions considered in both studies is similar, the differences in the values of F_{ST} observed in Flores and São Miguel must be interpreted with caution. However, the available results for the two islands indicate small degree of microdifferentiation in both islands of the archipelago.

Conclusions

In general, for both periods, indicators of diversity revealed a high level of surname diversity. These results are in accordance with diversity estimates obtained using genetic markers located in the Y chromosome (Montiel *et al.*, submitted) and using the frequencies of mtDNA haplogroups (Santos *et al.*, 2001; Montiel *et al.*, submitted). These results, all together, are consistent with the centrality of Flores and Corvo for all the Atlantic routs during Portuguese maritime expansion (Riley, 1991) that could have promote the interrelationship between the population of the island and the foreigners that use Flores as anchorage point.

Contrarily to what could be expected, considering the strong reduction of population in the last 150 years, there has maintenance of diversity from the middle of the 19th century until the present-day. This testifies that the input of new individuals in the island compensated, in terms of genetic diversity, the strong emigration that affected the island during the last century. Moreover, the level of micro-differentiation can be considered low, with a small decrease from the 19th century to the present-day, supporting a higher overture of parishes as a consequence of the increase of the communication routes.

Given the fact that the values of microdifferentiation are low, that Santa Cruz presents the highest number of different surnames and that all the other parishes share the highest number of surnames with Santa Cruz (this being reinforced by the position of the parish in the centre of coordinates in the correspondence analysis) indicates that the genetic variation of this parish could be considered representative of the general diversity of the Island. In this way, genetic

studies that aim to characterise the population of Flores Island as a whole can be based on a sampling where the majority of individuals is from Santa Cruz. Besides allowing an adequate profiling of the population, this strategy of sampling prevents obtaining results affected by high levels of isolation or reduced effective population size.

From the first to the second period analysed a change in surname composition is evident; considering the island as a whole, one half of the surnames present in the actual population are not present in the second half of the 19th century. However, some of the surnames that present high frequencies in the 19th century and in the present-day remained similar. Furthermore, several of these surnames were reported to be surnames present in the first settlers of Flores. This result testifies the impact of founders on the present-day genetic pool of Flores Island and allows to infer that the genetic characterization of the present-day population of Flores can provide reliable information concerning the history of peopling.

Acknowledgements

This work was supported by a grant (ref. SFRH / BD / 723 / 2000) from Fundação para a Ciência e a Tecnologia - Operational programs Science Technology and Innovation ([POCTI](#)) and Society of Information ([POSI](#)) financed by UE funds and national funds of MCES. Genealogical database of Flores Island was framed using funds from a project (ref. POCTI/C/DEM/56/96) funded by FCT-Fundação para a Ciência e a Tecnologia and coordinated by M. Lima

Literature Cited

Abade, A. (1992) *A população inexistente: Estrutura demográfica e genética da população da Lombada, Bragança*. PhD Dissertation, University of Coimbra, Coimbra. pp. 308.

Abade, A. (1986) Inbreeding as measurement by dispensations and isonymy in Rio de Onor, Portugal. *Inter. J. Anthropol.* 1, 225-228.

Abade, A. (2003) *ISONymy 3*. University of Coimbra, Portugal.

Alatalo R. (1981) Problems in the measurement of evenness in ecology. *Oikos* 37:199-204.

Barrai, I., Formica, G., Barale, R. *et al.* (1990) Isonymy in emigrants from Ferrara in 1981-1988. *Ann. Hum. Biol.* 17: 7-18.

Bittles, A. H. (1998) *Empirical estimates of the global prevalence of consanguineous marriage in contemporary societies*. Morrison Institute for Population and Resource studies, Stanford University, California. pp. 69

Bohonak, A. J. (2002) IBD (Isolation By Distance): a program for analyses of isolation by distance. *Jour. Hered.* 93: 153-154.

Branco, C. & Mota-Vieira, L. (2003) Population structure of São Miguel Island, Azores: a surname study. *Hum. Biol.* 75 (6), 929-939.

Cavalli-Sforza L. L. & Feldman M. W. (1981) *Cultural transmission and evolution: a quantitative approach*. Princeton University Press, Princeton. pp. 388.

Colantonio, S. E., Lasker, G. W., Kaplan, B. A. *et al.* (2003) Use of surname models in Human population biology: a review of recent developments. *Hum. Biol.* 75, 785-807.

Crow, J. F. (1983) Discussion: Surnames as biological markers. *Hum. Biol.* 55, 383-397.

Dahlberg, G. (1948) *Mathematical methods for population genetics*. Karger, Bâle, Interscience, New York.

DeBenedictis, P. A. (1973) On the correlation between certain diversity indices. *Am. Natur.* 107: 295-302.

Cunha, E. (1987) Consanguinity of the Azores Islands (1979-1985). *Antropologia Portuguesa*. 4/5, 67-79.

Gomes, F. (1997) *A ilha das Flores: da redescoberta à actualidade (subsídios para a sua história)*. Câmara Municipal de Lajes das Flores, Azores, Portugal.

Harpending, H. C. & Jenkins, T. (1973) Genetic distance among southern African populations. In Crawford M H., Workman, P. L. editors. *Methods and theories of anthropological genetics*. University of New Mexico Press, Albuquerque. pp.177-199.

Harpending, H. C. & Ward, R. (1982) Chemical systematics and human populations. In Nitecki M. editor. *Biochemical aspects of evolutionary biology*. Chicago University Press, Chicago. pp. 213-256.

INE (2001) *Recenseamento da População e da Habitação (R.A. dos Açores) - Censos*. Instituto Nacional de Estatística, Portugal.

Lages, G. (2000) Situação das Flores e do Corvo nos séculos XVI e XVII. *Arquipélago*. IV (2), 29-88.

Lalouel, J. M. (1973) MATFIT. In Morton N. E. editor. *Genetic structure of populations*. University of Hawaii Press, Honolulu. pp. 303-304.

Lima, M. (1999) O contributo da Demografia Histórica para os estudos em Antropologia Genética. *Insulana*. LV: 201-210.

Magurran, A. (1988) *Ecological diversity and its measurements*. Chapman & Hall, London.

Mantel, N. A. (1967) The detection of disease clustering and generalized regression approach. *Cancer Research*. 27, 209-220.

Manni, F., Guérard, E. & Heyer, E. (2004) Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm. *Hum. Biol.* 76 (2), 173–190.

Margalef, R. (1974) *Ecologia*. Editorial Omega, Barcelona, Spain.

Matos, A. (1989) Povoamento e Colonização dos Açores *in Portugal no Mundo*. Publicações Alfa, Lisboa, Portugal.

Monmonier, M. (1973) Maximum-difference barriers: an alternative numerical regionalization method. *Geogr. Anal.* 3, 245-261.

Montiel, R., Bettencourt, C., Silva, C. *et al.* Analysis of Y-chromosome variability and its comparison with mtDNA variability reveals different demographic histories between islands in the Azores Archipelago (Portugal). *Ann. Hum. Genet.* (In press).

Morton, N. E., Klein, D., Hussels, I. E. *et al.* (1973) Genetic structure of Switzerland. *Am. J. Hum. Genet.* 25, 347-361.

Mota, P. J. G. (1983) Consanguinidade aparente em S. Vicente da Beira. *Antropologia Portuguesa*, 1, 5-32.

Relethford, J. (1988) Estimation of kinship and genetic distance from surnames. *Hum. Biol.*, 60, 475-492.

Riley, C. (1991) Uma fronteira Ocidental. As Flores e o Corvo no contexto das rotas Atlânticas. *Boletim do Núcleo Cultural da Horta*. XI, 81-117.

Roberts, D. F., Jorde, L. B. & Mitchell, R. J. (1981) Genetic structure in Cumbria. *J. Biosocial Science*. 13, 317-336.

Santos, C., Lima, M., Abade, A. *et al.* (2001) The Azores Islands (Portugal): What can Mitochondrial DNA tell us. Poster presented in the “EURESCO Conferences: Inherited disorders and their genes in different European populations”. San Feliu de Guixols (Spain).

Santos, C., Lima, M., Montiel, R. *et al.* (2003) Genetic structure and origin of peopling in the Azores island (Portugal): the view from mtDNA. *Ann. Hum. Genet.* 67, 433-456

Shannon, C. E., Weaver, W. (1949) *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

Simpson, E. H. (1949) Measurement of diversity. *Nature*. 163, 688.

Smith, M., Abade, A. & Cunha, E. (1992) Genetic structure of the Azores: marriage and inbreeding in Flores. *Ann. Hum. Biol.* 19 (6), 595-601.

Smith, M., Alves, C., Silva, C. *et al.* (2000) Historical migration and population structure of Flores, Azores. *12th Congress of the EAA - Millennial Perspectives: Past, Present and Future.* p.74

Smouse, P. E. & Long, J. C. (1992) Matrix correlation analysis in anthropology and genetics. *Yrbk. Phys. Anthropol.* 35, 187-213.

SPSS Inc. (2001) *SPSS for Windows* – release 11.5.1, Chicago.

Table 1. Percentage of endogamous and exogamous marriages in each parish of Flores in the second half of the 19th century.

Parish	N° of valid Marriages ¹	Endogamous unions		Exogamous unions			
		end1	end2	exo1	exo2	exo3	exo4
Santa Cruz	385	58.4	0.9	14.9	5.3	0.9	20.8
Ponta Delgada	209	60.3	0	17.8	11.0	1.5	9.8
Caveira	80	31.3	13.8	17.7	21.4	14.1	2.6
Cedros	97	59.8	1.0	16.6	14.5	2.1	6.1
Lajes	271	75.3	5.2	9.1	4.5	2.6	3.4
Lajedo	28	78.6	0	14.3	7.2	0	0
Fajã Grande	85	76.5	1.2	9.6	6.0	0	4.8
Lomba	70	75.7	0	21.4	0	0	2.8
Mosteiro	45	51.1	20.0	2.2	11.0	8.8	6.6
Fajãzinha	90	87.8	0	4.4	3.3	0	4.4

1. N° Marriages with mention to the place of birth of both members of the couple.

end1- both members of the couple were born in the parish of marriage; *end2*- both members of the couple were born in the same parish, although it was not the parish of marriage. *exo1*- when the wife was born in the parish of marriage while her husband was born in another parish of Flores; *exo2*- when the husband was born in the parish of marriage, while it is his wife that was born in another parish of Flores; *exo3*- when husband and wife were born in different parishes of Flores, other than parish of marriage; *exo4*- when at least one of the members of the couple was born outside Flores Island.

Table 2. Number of individuals considered (one surname per individual was used), number and percentage of different surnames, percentage of parish specific surnames, number of surnames present in both periods and lost and added surnames from the 19th century to the present-day (PD).

Parish	N° of individuals		N° different surnames (%)		% parish specific surnames ¹		N° surnames present in both periods	Lost surnames N (%) ²	Add surnames N (%) ³	Lost /Added
	19 th	PD	19 th	PD	19 th	PD				
Santa Cruz	656	1387	154 (23.48)	220 (15.86)	40.91	46.36	90	64 (41.56)	130 (59.09)	0,49
Ponta Delgada	344	355	71 (20.64)	67 (18.87)	19.72	14.93	30	41 (57.75)	37 (55.22)	1,11
Caveira	142	71	43 (30.28)	24 (33.80)	23.26	8.33	7	36 (83.72)	17 (70.83)	2,12
Cedros	224	136	53 (26.66)	38 (27.94)	32.08	10.53	11	42 (79.25)	27 (71.05)	1,56
Lajes	569	669	90 (15.82)	129 (19.28)	26.67	28.68	41	49 (54.44)	88 (68.22)	0,56
Lajedo	57	90	24 (42.11)	21 (23.33)	4.17	4.76	9	15 (62.50)	12 (57.14)	1,25
Fajã Grande	179	184	51 (28.49)	54 (29.35)	21.57	11.11	21	30 (58.82)	33 (61.11)	0,91
Lomba	152	160	50 (32.89)	43 (26.88)	14.00	11.63	17	33 (66.00)	26 (60.47)	1,27
Mosteiro	97	52	43 (44.33)	27 (51.92)	6.98	7.41	12	31 (72.09)	15 (55.56)	2,07
Fajãzinha	162	89	55 (33.95)	30 (33.71)	23.64	16.67	14	41 (74.55)	16 (53.33)	2,56
Total Island	2582	3193	291 (11.27)	303 (9.49)	56.01	57.43	152	139 (47.77)	151 (49.83)	0,92

1. Surnames that were found only in the parish considered
2. with respect to the number of different surnames in the 19th century
3. with respect to the number of different surnames in the present-day

Table 3. Values of diversity indices (H , $H_{\max 1}$, $H_{\max 2}$, $H_{\min 2}$, Heteronymy) and Morton intra-population conditional kinship (R_{ii}) in each parishes and period (19th – 19th century; PD – Present-day). The value of $H_{\min 1}$ is not specified in the table since it is always zero.

Parish	H		$H_{\max 1}$		$H_{\max 2}$		$H_{\min 2}$		Heteronymy		Rii	
	19 th	PD	19 th	PD	19 th	PD	19 th	PD	19 th	PD	19 th	PD
Santa Cruz	6.54730	6.71604	9.35755	10.43775	7.26679	7.78136	2.47626	1.85685	0.98301	0.98500	-0.00112	-0.00029
Ponta Delgada	5.25221	5.25727	8.42626	8.47168	6.14975	6.06609	1.97608	1.81659	0.96191	0.96295	0.00386	0.00473
Caveira	4.71190	4.18604	7.14975	6.14975	5.42627	4.58496	2.47098	2.37400	0.94475	0.93116	0.00719	0.01006
Cedros	4.73934	4.59377	7.80735	7.08746	5.72792	5.24793	2.10504	2.26168	0.93750	0.94064	0.00964	0.00927
Lajes	5.17552	5.97485	9.15228	9.38586	6.49185	7.01123	1.63856	2.04356	0.94851	0.97269	0.00751	0.00261
Lajedo	4.23476	3.67817	5.83289	6.49185	4.58496	4.39232	2.79826	1.72463	0.93444	0.89358	0.00727	0.02031
Fajã Grande	5.09489	5.15593	7.48382	7.52356	5.67243	5.75489	2.43103	2.51607	0.95734	0.95977	0.00436	0.00490
Lomba	5.04930	4.59563	7.24793	7.32193	5.64386	5.42627	2.71694	2.24598	0.95525	0.93797	0.00464	0.01021
Mosteiro	4.98205	4.45949	6.59991	5.70044	5.42627	4.75489	3.32182	3.35022	0.95653	0.94527	0.00341	0.00520
Fajãzinha	5.26475	4.46330	7.33985	6.47573	5.78136	4.90689	2.83659	2.49356	0.96426	0.94281	0.00246	0.00778
Total island	6.60568	6.83969	11.33427	11.64070	8.18488	8.24317	1.42560	1.23079	0.97968	0.98422	-	-

$H_{\min 1}=0$

Table 4. Matrices of Morton inter-population conditional kinship. Second half of the 19th century (19th) below the diagonal; Present-day (PD) above the diagonal.

	SC	Pdel	Cav	Ced	Lajes	Lajedo	FGr	Lom	Mos	Fajáz	Mean PD
SC		-0.000378	-0.000984	-0.000461	-0.000421	-0.000048	-0.001043	-0.000858	-0.001161	-0.001745	-0.000789
Pdel	-0.000935		-0.002498	0.000183	0.000386	-0.000928	-0.000631	-0.000754	-0.001734	0.000249	-0.000678
Cav	-0.001203	0.000312		-0.001985	-0.000749	0.000360	-0.001999	0.000691	-0.000755	-0.002651	-0.001174
Ced	-0.000612	0.003388	0.004783		0.000838	0.004709	0.000340	0.001101	0.000838	0.000659	0.000691
Lajes	-0.002685	-0.001085	0.000741	0.001856		0.003200	-0.000092	0.001956	0.000266	-0.000497	0.000543
Lajedo	-0.003077	-0.001911	-0.000542	-0.000624	0.000148		0.002908	0.006140	0.002071	0.001821	0.002248
FGr	-0.001383	0.000962	0.000613	0.004212	-0.000838	-0.002007		-0.000574	-0.001154	-0.000740	-0.000332
Lom	-0.001305	0.000481	0.003195	0.005297	0.002434	0.000295	0.001462		0.000432	0.000348	0.000942
Mos	-0.002477	-0.001352	-0.000962	0.000018	-0.000978	-0.002149	-0.000497	-0.000411		-0.001062	-0.000251
Fajáz	-0.002570	-0.001468	-0.002194	-0.001266	-0.002005	-0.001992	-0.001451	-0.001196	0.001064		-0.000402
Mean 19 th	-0.001805	-0.000179	0.000527	0.001895	-0.000268	-0.001317	0.000119	0.001139	-0.000860	-0.001453	

Table 5. Surnames with an intra-parish frequency higher than 3%. Surnames underlined and in italic were reported as founder surnames. + refers to surnames present with an intra-parish frequency higher than 3%; - refers to surnames that are absent or present with an intra-parish frequency lower than 3%.

Surname	19 th Century	Present-day
BARCELOS	+	-
BISPO	+	-
BRINDEIRO	+	-
<u>COELHO</u>	+	-
<u>FERNANDES</u>	+	-
FRAGUEIRO	+	-
MACHADO	+	-
<u>NOLA</u>	+	-
NUNES	+	-
PAIVA	+	-
PEREIRA	+	-
ROCHA	+	-
SANTOS	+	-
TEIXEIRA	+	-
TOMAS	+	-
<u>TRIGUEIRO</u>	+	-
VASCONCELO	+	-
VELHO	+	-
ALVES	-	+
AMBROSIO	-	+
ANDRADE	-	+
AZEVEDO	-	+
BALDES	-	+
BORGES	-	+
CABECEIRA	-	+
CAMARA	-	+
CARVALHO	-	+
CORVELO	-	+
<u>COSTA</u>	-	+
CUSTODIO	-	+
DIOGO	-	+
ESTACIO	-	+
FILIPE	-	+
FORTUNA	-	+
FRADE	-	+
GARCIA	-	+
INOCENCIO	-	+
MACIEL	-	+
MEDEIROS	-	+
PACHECO	-	+
PEDRO	-	+
PERPETUA	-	+
RAIMUNDO	-	+
REIS	-	+
SANTO CRISTO	-	+
SERPA	-	+
SILVA	-	+
TAVARES	-	+
TENENTE	-	+
TERRAS	-	+
TEVES	-	+
VASCONCELOS	-	+
AVELAR	+	+
BELO	+	+
DIAS	+	+
EDUARDO	+	+
FAGUNDES	+	+
<u>FRAGA</u>	+	+
FREITAS	+	+
<u>FURTADO</u>	+	+
GOMES	+	+
GONCALVES	+	+
<u>HENRIQUES</u>	+	+
JORGE	+	+
LOPES	+	+
LOURENCO	+	+
MARTINS	+	+
<u>MENDONCA</u>	+	+
<u>PIMENTEL</u>	+	+
RAMOS	+	+
<u>RODRIGUES</u>	+	+
<u>SILVEIRA</u>	+	+
<u>SOUSA</u>	+	+
VALADAO	+	+
<u>VIEIRA</u>	+	+

Fig. 1 – Map of Flores Island

Fig. 2 – Scatterplot of Harpending and Jenkins (1973) intra-population kinship (HJ-Rii) and heteronymy (Harpending and Ward, 1982). The least-squares regression line is shown. a) Second half of 19th century; b) Present-day.

(SC - Santa Cruz; PDel - Ponta Delgada; Cav - Caveira; Ced - Cedros; Lajes - Lajes; Lajedo - Lajedo; FGr - Fajã Grande; Lom - Lomba, Mos - Mosteiro, Fajãz - Fajãzinha)

Fig. 3 – Scatterplot of number of inhabitants and heteronymy and Morton intra-population conditional kinship. The line of predicted values by the model of Dahlberg (1948) is shown.

(SC - Santa Cruz; PDel - Ponta Delgada; Cav - Caveira; Ced - Cedros; Lajes - Lajes; Lajedo - Lajedo; FGr - Fajã Grande; Lom - Lomba, Mos - Mosteiro, Fajãz - Fajãzinha)

Fig. 4 – Correspondence analysis, for the second half of the 19th century, for parishes and surnames. The first dimension accounts for 29% of the inertia, whereas the second accounts for 16.4%.

(SC - Santa Cruz; PDel - Ponta Delgada; Cav - Caveira; Ced - Cedros; Lajes - Lajes; Lajedo - Lajedo; FGr - Fajã Grande; Lom - Lomba, Mos - Mosteiro, Fajãz - Fajãzinha)

Fig. 5 – Correspondence analysis, for the present-day, of parishes and surnames. The first dimension accounts for 20.9% of the inertia, whereas the second accounts for 16.6%.

(SC - Santa Cruz; PDel - Ponta Delgada; Cav - Caveira; Ced - Cedros; Lajes - Lajes; Lajedo - Lajedo; FGr - Fajã Grande; Lom - Lomba, Mos - Mosteiro, Fajãz - Fajãzinha)

Fig.6 – Boundaries inferred using the matrix of Morton conditional kinship. a) 19th century; b) Present-day.

Fig. 1



Fig. 2

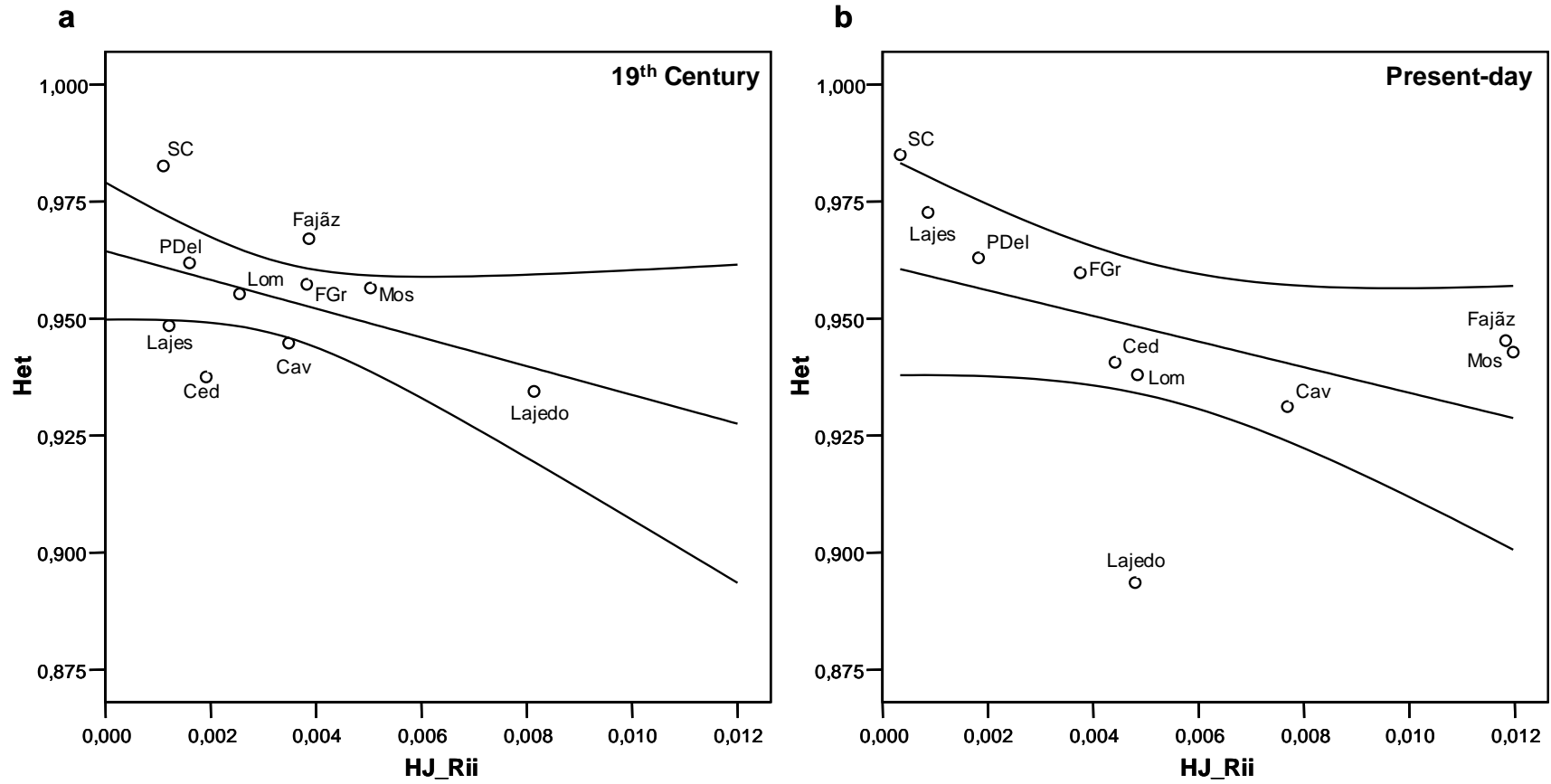


Fig. 3

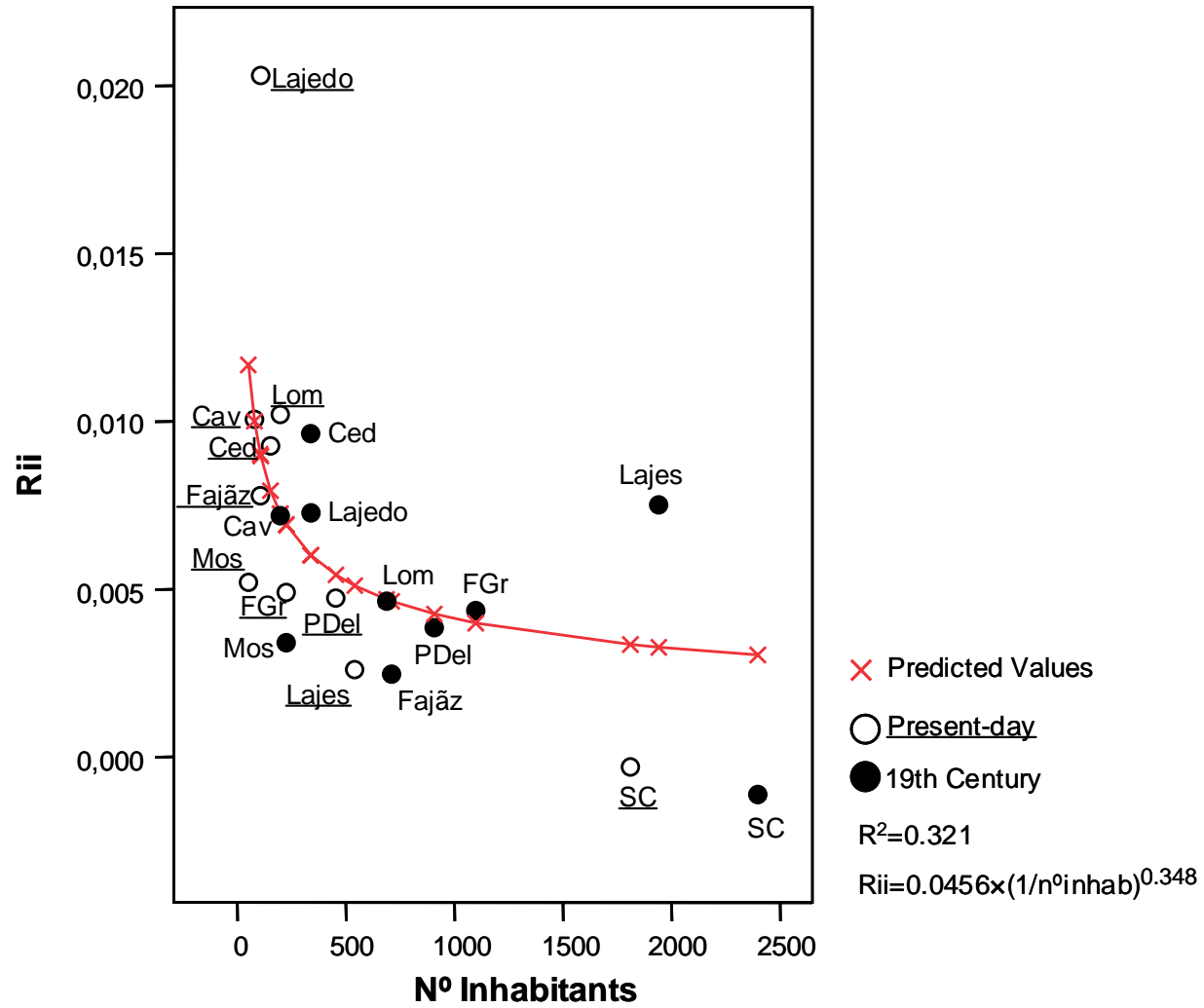


Fig. 4

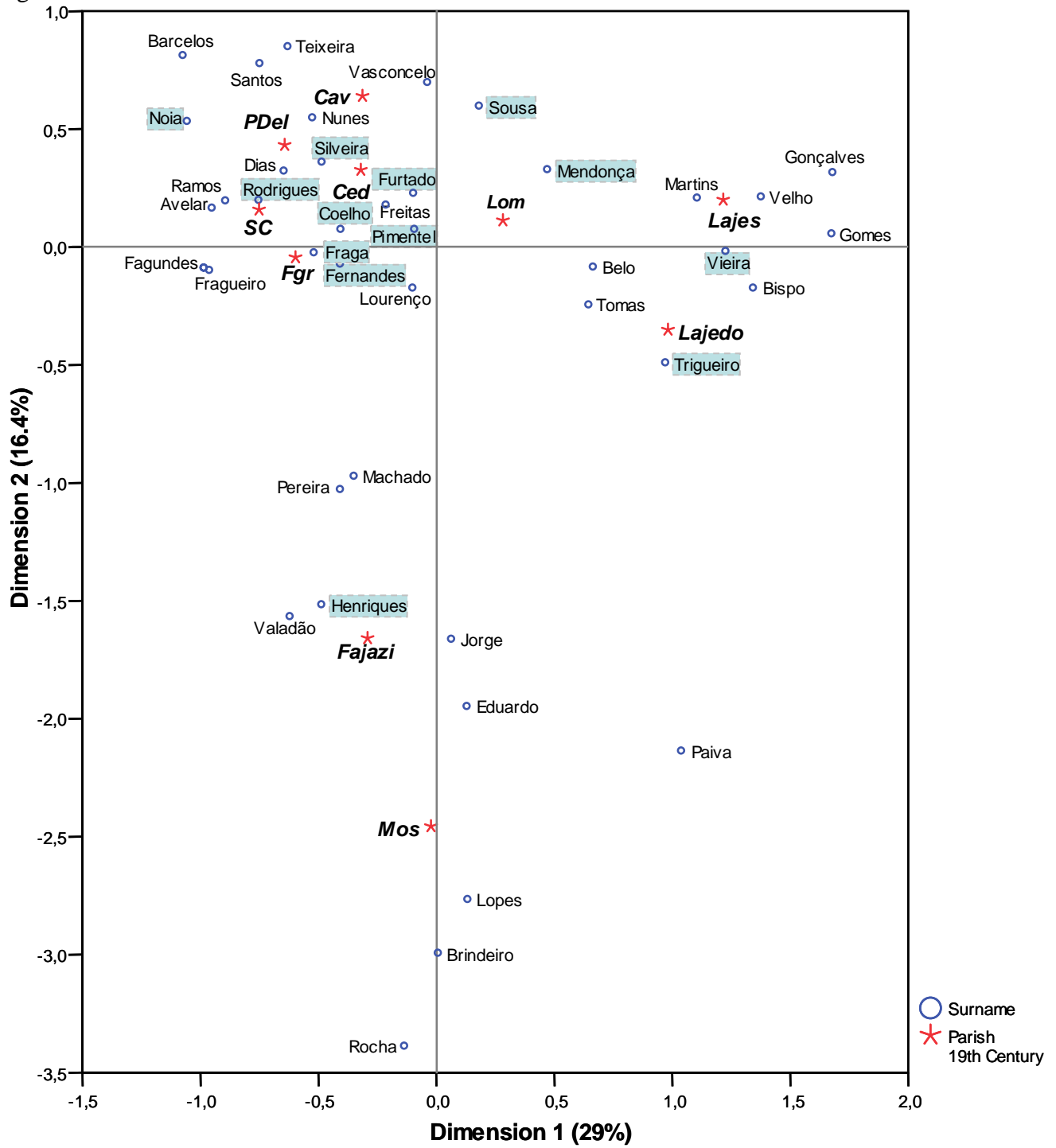


Fig. 5

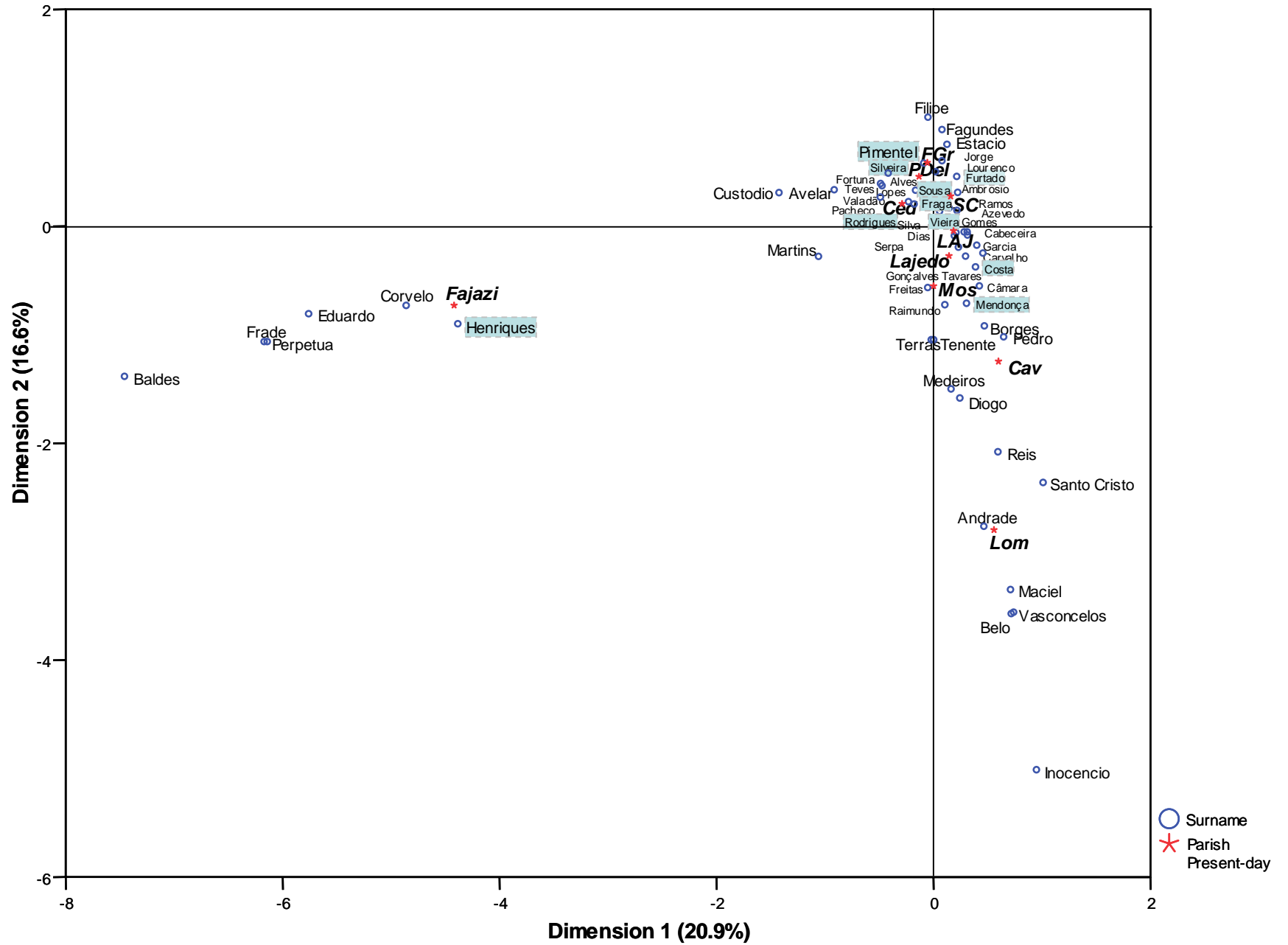
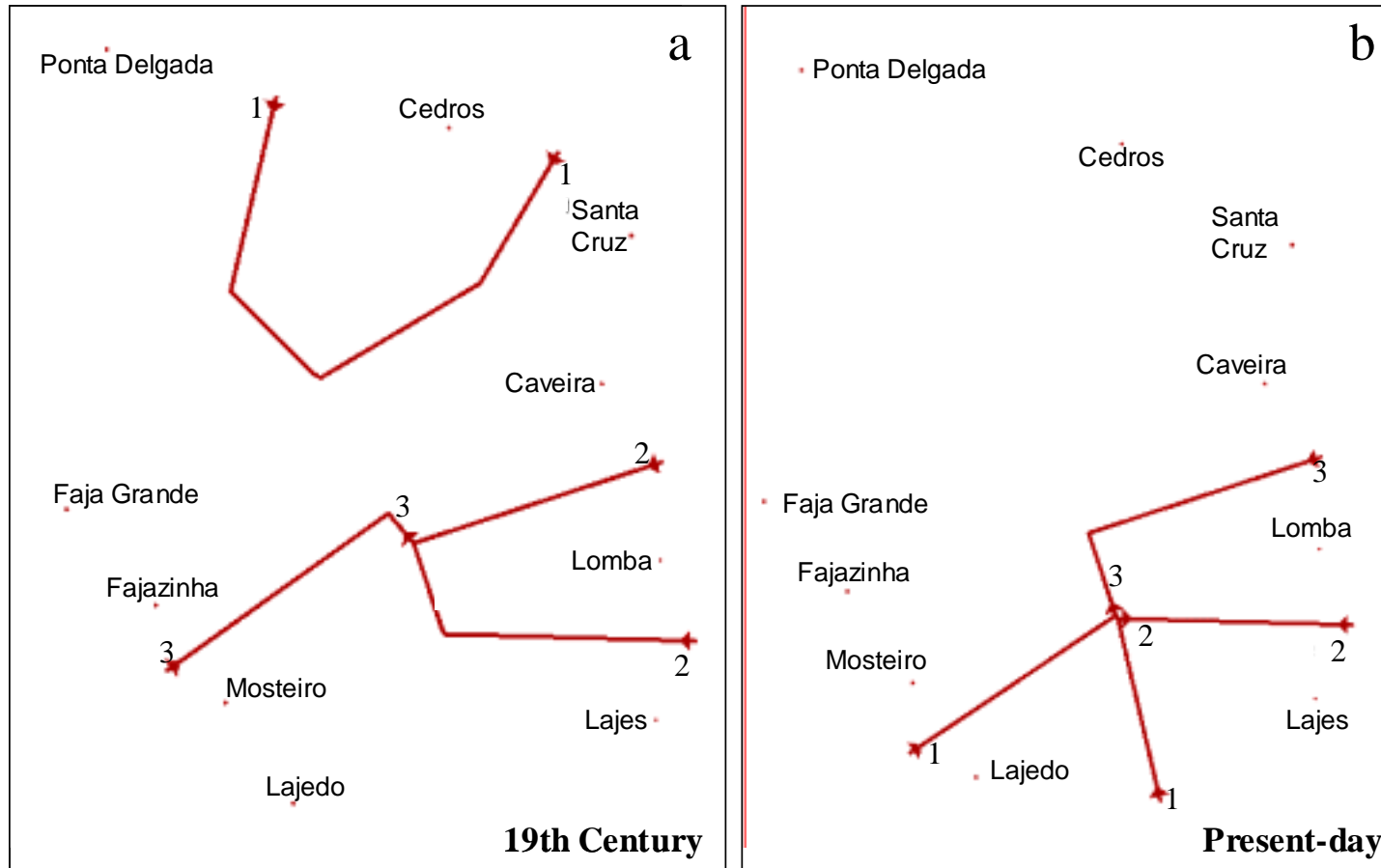


Fig. 6



IV. 5

Artículo V

Understanding Differences Between Phylogenetic and Pedigree Derived mtDNA Mutation Rate: A Model Using Families from the Azores Islands (Portugal)

Santos *et al.* ***Molecular Biology and Evolution***
(*in press*)

Understanding differences between phylogenetic and pedigree derived mtDNA mutation rate: A model using families from the Azores Islands (Portugal)

Research Article

Cristina Santos^{1, 2}, Rafael Montiel³, Blanca Sierra¹, Conceição Bettencourt³, Elisabet Fernandez¹, Luis Alvarez¹, Manuela Lima³, Augusto Abade² and M^a Pilar Aluja¹

1. Anthropology Unit, Department BABVE, Faculty of Sciences, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain
2. Department of Anthropology, University of Coimbra, 3000 Coimbra, Portugal
3. Center for Research in Natural Resources (CIRN), University of the Azores, 9500 Ponta Delgada, S. Miguel, Azores, Portugal

Corresponding author:

Cristina Santos

Anthropology Unit, Department BABVE

Faculty of Sciences

Autonomous University of Barcelona

08193 Bellaterra (Barcelona), Spain

Tel: +34 935812049

Fax: +34 935811321

e-mail: Cristina.Santos@uab.es

Key words: mtDNA mutation rate; heteroplasmy; paternal mtDNA; recombination;
Azores (Portugal)

Running head: mtDNA mutation rate in Azorean families

ABSTRACT

We analysed the control region of the mitochondrial DNA from maternally related individuals originating from the Azores Islands (Portugal) in order to estimate the mutation rate of mitochondrial DNA and to gain insights into the process by which a new mutation arises and segregates into heteroplasmy. Length and/or point heteroplasmy were found at least in one individual of 72% of the studied families. Eleven new point substitutions were found, all of them in heteroplasmy, from which five appear to be somatic mutations and six can be considered germinal, evidencing the high frequency of somatic mutations in mtDNA in healthy young individuals. Different values of the mutation rate according to different assumptions were estimated. When considering all the germinal mutations, the value of mutation rate obtained is one of the highest reported so far in family studies. However, when corrected for gender (assuming that the mutation present in men have the same evolutionary weight of somatic mutations since they will inevitably be lost) and for the probability of intra-individual fixation, the value for the mutation rate obtained for HVRI and HVRII (0.2415 mutations/site/Myr) was in the upper end of the values provided by phylogenetic estimations. These results indicate that the discrepancy, that has been reported previously, between the human mtDNA mutation rates observed along evolutionary timescales and the estimations obtained using family pedigrees can be minimized when corrections for gender proportions in new born individuals and for the probability of intra-individual fixation are introduced. The analyses performed support the hypothesis that: a) in a constant, tight bottleneck genetic drift alone can

explain different patterns of heteroplasmy segregation, and b) in neutral conditions, the destiny of a new mutation is strictly related to the initial proportion of the new variant. Another important point arising from the data obtained is that, even in the absence of a paternal contribution of mtDNA, recombination may occur between mtDNA molecules present in an individual, which is only observable if it occurs between mtDNA types that differ at two or more positions.

INTRODUCTION

The mitochondrial DNA (mtDNA) is one of the most widely studied systems of eukaryotic genomes and has been extensively used in studies of population and evolutionary genetics. In particular, the non-coding control region (CR), or D-loop, of the human mtDNA has been well studied and widely used to address many questions, namely those related with aspects of human evolution (i.e. Cann et al. 1987; Ingman et al. 2000) and the movements of ancient populations (i.e. Richards et al. 1996; Torroni et al. 2001). As stated by Howell et al. (2003), many of these studies rely on phylogenetic analyses of mtDNA CR haplotype trees and phylogenetically derived divergence rates. However, several complications can be acknowledged when using these rates, namely: a) the heterogeneity in the site-specific substitution rate in the D-loop (i.e. Meyer et al. 1999; Meyer and von Haeseler 2003), b) the high rate of homoplasy evidenced in many studies (i.e. Richards et al. 1998; Howell et al. 2004) using median networks (Bandelt et al. 1995), c) the violation of the assumption of clocklike evolution (i.e. Ingman et al. 2000; Torroni et al. 2001; Howell et al. 2004); d) the violation of the neutral evolution assumption (i.e. Mishmar et al. 2003; Elson et al. 2004; Ruiz-Pesini et al. 2004) and e) the failure of most phylogenetic approaches to consider the effects of population dynamics, admixture and migration (Rannala and Bertorelle 2001).

In order to estimate the mtDNA rate of evolution based on minimal assumptions, Howell et al. (1996) developed a novel experimental strategy. The authors analysed the D-loop region in maternally related individuals from families with Leber optic neuropathy and estimated the rate of evolution, counting the number of new

mutations observed in pedigrees by the number of base pairs and generations analysed. Using this approach, Howell et al. (1996) estimated that the rate of evolution in the control region is 1.9 mutations/site/Myr, a value that is 10 times higher than the one derived from the phylogenetic approaches. The work of Howell et al. (1996) was criticized since the authors used families with a mtDNA disease that may have unusually elevated mtDNA mutation rates, however further studies confirmed the discrepancy between the human mtDNA mutation rate observed along evolutionary timescales (i.e. Pesole et al. 1992; Tamura and Nei 1993;) and the corresponding rate estimated using pedigrees [for a summary see Howell et al. (2003) and the references in this work]. A discrepancy with phylogenetic estimations was also observed using long-term series of *Caenorhabditis elegans* mutation accumulation lines (Denver et al. 2000) and comparing mtDNA in ancient and actual Adélie Penguins (Lambert et al., 2002), however the differences are smaller than the previously reported by Howell et al. (1996).

The observation of differences between various estimations of the mutation rate in Human mtDNA has promoted an intense debate (i.e. Pääbo 1996; Macaulay et al. 1997; Sigurðardóttir et al. 2000; Heyer et al. 2001; Howell et al. 2003; Hagelberg 2003) and have been attributed to distinct causes such as: 1) differences in the rate of mutation at different mtDNA positions (it has been suggested that mtDNA mutations, both somatic and germline, occur preferentially at hypervariable sites); 2) the effect of selection and genetic drift, since many of the new mutations observed in pedigree studies are destined to be lost from the population as a whole; 3) the occurrence of somatic mutations during ageing; 4) the unintended sequencing of nuclear mitochondrial pseudogene (Numts); 5) the mutation rates observed in

pedigrees are not due to mutation, but to low levels of leakage of paternal mtDNA and recombination. Several of these causes remain unclear and the processes by which heteroplasmy (the condition in which an individual has a mixture of two or more mtDNA types) is resolved is not well understood, which highlights the need for further data and analyses taking special attention to the complex processes by which mtDNA mutations arise, segregate, and are fixed at the organelle, cell, individual and population levels.

It is recognized that many more mutations arise than those that ultimately become fixed. Furthermore the rate of DNA sequence evolution (i.e. substitution) is a function of the mutation rate and of the probability of fixation (Kimura 1983). This is particularly important in the context of a units-of-evolution view of mtDNA, since the cytoplasmic population constitutes an additional level in the biological hierarchy that can influence the distinction between mutation and substitution (Rand 2001). Heteroplasmy is an obligatory, if transient, phase in the evolution of mtDNA. Therefore, mtDNA evolution can be characterized by a mutation-drift-selection balance in which most cells carry transient levels of heteroplasmy for point mutations, deletions, and rearrangements (Rand 2001). One of the keys to understanding mtDNA evolution is to analyse and to interpret the patterns and segregation of heteroplasmy across generations, and to infer the implications of mitochondrial segregation during the early oocyte development (the mitochondrial genetic bottleneck). A better understanding of heteroplasmy is not only relevant in the context of evolutionary studies, but also has important implications for understanding the mode of inheritance and onset of mitochondrial diseases (Poulton and Marchington 2002).

Recent studies have demonstrated that heteroplasmy in humans is far more frequent than previously believed (Grzybowski et al. 2003). Since research on the transmission of heteroplasmy has been confined to a reduced number of families identified by different authors, a systematic and exhaustive analysis of heteroplasmy transmission in families is still needed.

In this work we present the results of the analysis of the CR of mtDNA from individuals clustered in families from the Azores Islands (Portugal). The main aims are: a) to better understand the mtDNA evolutionary process, including the factors that control the levels and progress of mtDNA heteroplasmy, and b) to provide empirical estimations of the mutation rate of the CR of mtDNA under different assumptions.

This is the first study of mtDNA transmission in families where factors such as paternal contribution of mtDNA and mtDNA recombination between different mtDNA types have been taken into consideration.

MATERIAL AND METHODS

Genealogies and Samples

Genealogical information, provided orally by voluntary participants (nominated primary informants) was recorded for extensive families with Azorean ancestry. Potential sample donors were selected based on these genealogies. Individuals within each family were first contacted through the primary informant. Those willing to participate were subsequently informed by members of the research team about the aims and ethical guidelines of the project. Upon signing an informed consent, the individuals were sampled by mouth scrap. Five hundred samples corresponding to individuals from 26 extensive families with Azorean ancestry were obtained. From this sample bank we selected 422 individuals representing 48 independent mtDNA lineages (Fig.1) and corresponding to 321 mtDNA transmissions.

General Strategy

The analysis of the samples followed 5 main steps:

1) DNA extraction, PCR amplification and sequencing of the D-loop region of the individuals from the most recent generation (individuals are represented by filled black squares and circles in figure 1); 2) Confirmation by a second PCR amplification and sequencing of the individuals that appeared to be heteroplasmic in step 1; 3) DNA extraction, PCR amplification and sequencing of the D-loop of the ancestors of individuals confirmed as heteroplasmic in step 2 (individuals represented by half tone squares and circles in figure 1), whenever heteroplasmy

was not present in the mother, the paternal mtDNA was analysed to verify a possible paternal contribution; 4) To authenticate results an independent DNA extraction, PCR amplification and sequencing were subsequently performed for all individuals showing heteroplasmy; 5) Cloning and sequencing of multiple clones (26-66) of individuals showing heteroplasmy at different levels was subsequently performed.

mtDNA analysis

DNA extraction, D-loop amplification and sequencing

DNA from the buccal cells was extracted using Instagene matrix (BioRad) according to the manufacturer's specifications. The mtDNA region from position 15907 to 580 [according to Anderson et al. (1981)], was amplified using the primers L15907 (5'-ATACACCAGTCTTGTAACC-3') and MTH580 (5'-TTGAGGAGGTAAGCTACATA-3'; Heyer et al. 2001). The PCR reaction program involved an initial 5 min. denaturation step at 95 °C followed by 10 cycles of 94 °C for 1 min., 57°C for 40 s and 72 °C for 2 min., 26 additional cycles at 94 °C for 1 min., 56°C for 40 s and 72 °C for 2 min., and a final extension step of 5 min. at 72 °C. PCR products were purified using the JETQUICK PCR Purification Spin Kit (Genomed).

All samples were fully sequenced between positions 16024-16569 and 1-400. Seven internal primers were used, three in the direct direction [L1_15997 (Pereira et al. 2000) 5'-CACCATTAGCACCCAAAGCT-3', L1_16180 5'-CCCCATGCTTACAAGCAAGT-3' and L2_16485 5'-GAACTGTATCCGACATCTGG-3'] and four in the reverse direction (H1_16166

5'-GGTTTTGATGTGGATTGGGT-3', H1_ 8 5'-
GGTTAATAGGGTGATAGACC-3, H2_288 5'-GGGGTTTGGTGGAAATTTTT-
3' and H2_481 5'-GATTAGTAGTATGGGAGTGG-3'). Sequence reactions were
carried out using the sequencing kit BigDye[®] Terminator v.3 (Applied Biosystems)
according to the manufacturer's specifications and run in an ABI Prism 3100
sequencer.

Cloning

PCR products of some of the samples showing different levels of heteroplasmy were cloned into the pCR[®]4-TOPO[®] vector using the TOPO TA Cloning[®] Kit for Sequencing (Invitrogen) and 26-66 clones for each sample were sequenced using the same primers described above for sequencing.

Data analysis

Alignment and classification of sequence

Sequences were visually verified to detect the presence of heteroplasmy. Subsequently they were aligned in relation to the Cambridge reference sequence – CRS (Anderson et al. 1981), using BioEdit (Hall 1999). All polymorphic positions were confirmed in sequencing electropherograms. Sequences without ambiguities were obtained between positions 16024-16569 and 1-400. Classification of all the samples into haplogroups was performed according to the nomenclature summarized in Richards et al. (2000) and Salas et al. (2004).

Levels of heteroplasmy

To quantify the levels of heteroplasmy two methods were used:

a) *Height of the peaks in the electropherograms*, a strategy similar to that used by Brandstätter et al. (2004), in which the proportion of heteroplasmy was inferred, from direct and reverse electropherograms of a point heteroplasmic sample, measuring the height of the two peaks directly from the electropherogram and calculating the proportion between the height of each peak with respect to the sum of the height of the two peaks. For comparison purposes the obtained numerical proportions were averaged.

b) Cloning of PCR products that encompass the D-loop region and the subsequent sequencing of the 26-66 clones to determine the number of clones that contain each mtDNA variant.

Statistical tests concerning the levels of heteroplasmy were performed with the programs SPSS v.12 (SPSS inc. 1989-2003) and OpenStat2 v1.42 (Miller 2003).

Population Database and Reference Sequences

To calculate the frequency of each variant for a particular nucleotide position, we created an mtDNA population database of European and African individuals for which both the HVRI and HVRII region were available in the HVRbase (Handt et al. 1998). The database consists of 1215 European sequences (Vigilant et al. 1991; Piercy et al. 1993; Calafell et al. 1996; Francalacci et al. 1996; Hofmann et al. 1997; Baasner et al. 1998; Brown et al. 1998; Delghandi et al. 1998; Lutz et al. 1998; Rousselet and Mangin 1998; Parson et al. 1998; Baasner and

Madea 2000; Helgason et al. 2000; Parsons et al. Armed Forces Institute of Pathology) and 288 African sequences (Vigilant et al. 1991; Graven et al. 1995; Soodyall et al. 1996). The database is in SPSS format (SPSS inc. 1989-2003) and each mtDNA position represents a variable that allows the relationship between distinct nucleotide positions to be analysed.

To infer the ancestral state in a given nucleotide position we used two sequences: the consensus sequence proposed by Ingman et al. (2000) - GeneBank accession number NC_001807 - that originates from an African sequence whose ancestor is the most recent common ancestor of the youngest clade containing both African and non-African individuals (GeneBank accession number AF347015); and the sequence of an African San individual - GeneBank accession number AF347008 – that represents the sequence of the mtDNA Eve (Ingman et al. 2000).

Mutation Rate Estimation

The mutation rate was derived from the number of detected mutations per number of “meioses” or transmission events, which is the number of cumulative generations tracing back to the maternal ancestor. When the mutation rate is expressed as mutations per base pair per million years (Myr), the generational time was assumed to be 25 years, because biodemographic studies for the Azorean population in which the marriage age was calculated indicate that the mean age of marriage is around 25 years (Lima 1996). Different values of the mutation rate (considering substitutions only) were estimated according to different assumptions: 1) all the substitutions that were detected were considered for the mutation rate calculation, 2) only the substitutions for which there was evidence of a germinal

origin were considered (applied by the majority of authors, such as Howell et al. 2003), 3) only the substitutions present in women for whom there was evidence of a germinal origin were considered (this takes into account that the mutations present in men have the same evolutionary weight of somatic mutations since they cannot be transmitted to the next generation), and 4) only the substitutions with a germinal origin present in women that would become fixed at the individual level were considered. To perform these estimations we assume that all mutation observed are neutral and the probability of intra-individual fixation of the new variant is equal to its frequency in the heteroplasmic individual.

The 95% confidence intervals for proportions were computed using the program OpenStat2 v1.42 (Miller 2003).

Effect of Random Genetic Drift in Heteroplasmy Resolution

The Wright-Fisher model of random genetic drift (Fisher 1930; Wright 1931) was used to estimate the probability that the mtDNA population within the oocyte drifts from i to j copies of mtDNA type A in one generation (Hartl and Clark 1997; formula 7.2, page 275 but corrected for the present work for haploid genomes):

$$P_{ij} = \binom{N}{j} p^j q^{N-j}$$

The theory of drift is oversimplified and does not take into account the changes of N in time. To solve this problem we computed a population of effective size, N_e , which is the population size that would give the same amount of drift in a reproducing population as that assumed in the simplest drift model (Cavalli-Sforza and Bodmer 1971). Wright (1939) showed that if in different generations the

population sizes are N_1, N_2, \dots, N_i , the effect on the distribution of gene frequencies can be predicted by taking the harmonic mean of the N_i values as an estimate of N . Thus, if there are k values N_i at k different times (Cavalli-Sforza and Bodmer 1971; formula 8.20, page 417) the N_e will be given by:

$$N_e = \frac{k}{\frac{1}{N_1} + \frac{1}{N_2} + \dots + \frac{1}{N_k}}$$

The limited data gathered so far suggests that a significant bottleneck occurs by the time that the oocytes are mature and that there is little segregation between fertilization and birth (Poulton et al. 1998). Furthermore, the single-selection model (Bendall et al. 1996), that assumes a one-time sampling of a small number of mtDNAs from a large pool (an oocyte contains ~100000 mtDNAs), appears to be more suitable than the repeated-selection model for predicting the size of the bottleneck (Poulton et al. 1998). Bendall et al. (1996), using the single-selection model, estimated the most probable bottleneck size in relatives of four twin pairs to be ~3-30. Marchington et al. (1998), examining human oocytes obtained a similar estimate (1-31 segregating units) with the most probable modal being 8 and the median 9.

To calculate the effective population size (N_e) we considered that the oocyte: a) at time zero contains ~100000 mtDNAs (Jenuth et al. 1996), representing the first generation, b) at time 1 the number of mtDNA molecules is reduced by a bottleneck that is considered to be constant ($N=10$) (Poulton et al. 1998), and c) at time 2, after the bottleneck a 50-fold expansion of mtDNA occurs (Chen et al. 1995), which represents the next generation.

RESULTS

Control-Region sequence analysis of Azorean pedigrees

General overview

The mitochondrial control region, between nucleotide positions 16024-16569 and 1-400 [numbering according to Anderson et al. (1981)], was fully sequenced in 232 individuals traceable to 48 “founding mothers” of Azorean ancestry (Fig. 1). The corresponding pedigrees include a total of 321 maternal transmissions. The control region was also sequenced for the fathers of four heteroplasmic individuals. A total of 52 independent mtDNA lineages, 48 from maternally related individuals and 4 corresponding to the fathers of heteroplasmic individuals, were analysed. The results obtained for the control-region sequencing and the haplogroup inferred based on the D-loop haplotype are available as online supplementary material (table I of supplementary material).

About seventy percent of the independent mtDNA lineages display a HVRI haplotype that has been formerly identified in individuals with Azorean ancestry (Santos et al. 2003; Brehm et al. 2003). The haplogroup composition fits the results of the previously reported studies, i.e. that the Azores islands population show a mixed composition of mtDNAs with West-Eurasian and African haplogroups being detected.

From the 48 families analysed only 13 families were fully homoplasmic and in the rest of the families at least one individual displayed length and/or point

heteroplasmy. The families analysed can be clustered into four principal groups (table 1).

As the polycytosine (poly-C) tracts of both hypervariable regions require further analyses we focused our research mainly on point mutational events. Nevertheless, results for length heteroplasmies are also presented.

Length Heteroplasmy

The poly-C tracts of both hypervariable regions I and II (located between positions 16184-16193 and 303-315 respectively) are known to have high insertion/deletion rates, thus originating length heteroplasmy (Hauswirth and Clayton 1985). This type of heteroplasmy is represented by multiple populations of mtDNA containing poly-C stretches of various lengths (Lee et al. 2004).

In this study we observed length heteroplasmy produced by the insertion of cytosine residues in the poly-C tract of HVRI in 11 (22.92%) of the families analysed. Nine families (BM, CAR_III, CV, JA_III, MBP, ML, MN, NP_I, PA_I) show a homoplasmic change in position 16189 T->C associated with length heteroplasmy produced by the insertion of cytosine residues. Family ML is the only one that shows differences between family members in the proportions and type of length variants (see table I of online supplementary material). In the remaining eight families, the length heteroplasmy is present in all the members of each family that were analysed, and the analysis of sequencing electropherograms gave no evidence of variations within the family, this is similar to the results obtained by Bendall and Sykes (1995). However, identifying length variants directly from the sequencing electropherograms when position 16189 changes from T to C is not

easy to perform, this has been previously acknowledged by others (Parson et al. 1998; Lee et al. 2004). Therefore our observations refer only to those variants that present a sufficiently high representation and that can be detected by direct sequencing. Two families (CAR_I and HR_I), that will be described further, are comprised of one and four individuals respectively that are heteroplasmic at nucleotide position 16189 T->C and present length heteroplasmy produced by the insertion of one cytosine residue in the poly-C of HVRI. Confirming the previous results of Lee et al. (2004), no insertion of C residues was observed in individuals that were homoplasmic for 16189T, indicating that the instability of the poly-C tract 16184-16193 is closely related to the presence of 16189C. We detected three within-family length changes for the poly-C tract of HVRI. Our estimate of the overall rate of length change for this region is 3 changes in 321 transmissions or generations, that is 0.009 changes per generation (95% CI: 0 - 0.019). It is likely that this result represents an underestimation, and future analyses will be performed to fully assess the changes between family members.

In the poly-C tract of the HVRII we observed length heteroplasmy produced by the insertion of cytosine residues in 26 (54.17%) of the families analysed. In 14 of the families (AV, BF, CAM_II, CAR_V, CV, JA_I, JL_I, JL_II, JS, MM_II, MN, NP_IV, PA_I, PP), the within family length variants are the same. Furthermore, in the majority of the families all the individuals of a given family presented similar proportions of each length variant, a similar result to that obtained for HVRI poly-C tract. In contrast, in 12 families (CAM_I, CAM_III, CAR_III, CG_II, CG_III, CM, HR_II, JA_II, MBP; CAR_II, ML, PA_II) different length variants were detected in individuals of the same family. In the majority of families

the differences between individuals are not necessarily the result of new mutations and can be explained by differential segregations during oogenesis. Considering the previous results, and assuming one change per family, our estimate of the overall rate of length change for the poly-C tract of HVRII is 12 changes in 321 transmissions or generations, that is 0.037 changes per generation (95% CI: 0.016 - 0.058), a value that is not significantly different to that reported by Sigurðardóttir et al. (2000). We also found length heteroplasmy outside of the poly-C tracts in a mother-son pair from the family MM_I, produced by a guanine deletion in the Poly-G 66-71 [numbering according to the CRS (Anderson et al. 1981)]. In this family, the individual I2 is clearly heteroplasmic presenting mtDNA variants with 6 G (as in the CRS) and 5 G. Her daughter (III1) and her grandchildren (III1 and III2) are completely homoplasmic for the variant with 6 G. Her son (II2) presents low levels of heteroplasmy.

Heteroplasmy Produced by Point Substitutions

Heteroplasmies produced by point substitutions were found in ten (20.83%) of the families analysed, representing 28 individuals with point heteroplasmy; The D-loop of 15 of them was cloned and 26-66 clones of each were subsequently sequenced. The comparison of the mean proportion of variants obtained by the analysis of the height of the peaks in the electropherograms and by counting the clones did not reveal significant differences (Wilcoxon, $Z = -1.396$, $p = 0.163$) indicating that the first methodology can be used to have an idea of the proportion of mtDNA variants. However, it is worth to be noted, that even if a non-parametric

test was used, the absence of differences can be due to the small sample size used for comparison.

In table 2, are presented for each family, the individuals displaying heteroplasmy, the heteroplasmic positions and the levels of heteroplasmy ascertained by the height of peaks and by counting the clones. In table 3, the nucleotide in the homoplasmic individuals of the family in the Cambridge Reference Sequence (Anderson et al. 1981), in the consensus sequence of Ingman et al. (2000) and in an African San individual (Ingman et al. 2000) is presented for each heteroplasmic position. The site-specific rate (Meyer et al. 1999) for each heteroplasmic position and the frequency of each possible variant, calculated from the database of European and African mtDNA sequences, are also presented.

Families showing heteroplasmy produced by point substitutions, described in detail in supplementary material, can be clustered into two groups:

1) Families in which only one individual presented heteroplasmy produced by point substitution (CAR_I, CAR_II, PA_I and JA_I): in these families the reported heteroplasmy may have resulted from somatic mutations that occurred in the individuals themselves, or alternatively, it may have resulted from germline mutations that occurred in the oocyte from which the heteroplasmic individuals developed.

2) Families in which various individuals displayed heteroplasmy produced by point substitution (HR_I, ML, NP_II, NP_III, PA_II and SAO): in these families the mutations present in heteroplasmy segregate at least during two generations. The pattern of heteroplasmy segregation during various generations in each family is described in supplementary material. The majority of individuals present point

heteroplasmy in only one nucleotide position, the exceptions is the mother-son pair formed by the individuals II4 and III2 from family ML. The analysis of the electropherograms of individuals III2 and II4 from family ML indicated that both present high levels of heteroplasmy in two distinct positions: 16086 T/C and 16182 A/C or Del A (table 2). The heteroplasmy observed in position 16182 is probably the result of a deletion of an Adenine, however, it is not possible to exclude the possibility that it represents an A to C transversion. To exactly determine the mtDNA variants found in individuals II4 and III2 the D-loop of these individuals was cloned and 62 and 37 clones were sequenced respectively. In table 4 the mtDNA variants defined by positions 16086 and 16182, and by the number of C residues in the poly-C from HVRI, are presented. Excluding length variation in the poly-C tract, we detected four distinct mtDNA types defined by positions 16086 and 16182 in the clones of the two individuals: the wild-type of the family (16086T-16182A), and types 16086T-16182C/Del A, 16086C-16182A and 16086C-16182C/Del A. These results were replicated using two distinct samples of both individuals and they can be considered authentic (see Montiel et al. 2001). Moreover, additional mutations found in the wild-type (for example a mutation in position 16129) were also present in the other mtDNA variants (table 4), which verify once again the authenticity of these results. In the absence of recombination this pattern can only be explained assuming multiple mutational events that can be justified in part by the instability in this region caused by the presence of 16189C. As previously stated, this mutation is known to produce instability in the poly-C tract and it is plausible that it also influences the number of A's upstream from the poly-C. In fact, in the population database used in this work a significant

association between the variant 16189C and the presence of a C or an A deletion in positions 16182 and 16183 was observed (Fisher exact test, 16189C/16182A: $p < 0.0001$; 16189C/16183A: $p < 0.0001$). However, for the other six families showing the 16183C and 16189C changes, there were no signs of heteroplasmy in position 16182. Thus, our data does not support the hypothesis that two independent changes in the number of As upstream from the poly-C may occur due to the 16189C change. It is likely that this pattern results from recombination between distinct mitotypes present in the same individual.

Overall Analysis of Mutations (excluding positions 16184-16188, 16190-16193 and 303-315)

Mutation Pattern

We detected thirteen new mutations - eleven substitutions, one possible deletion (position 16182) and one deletion (deletion of one G in poly-G 66-71) - that gave rise to heteroplasmy in twelve distinct positions. Five substitutions appear to be somatic mutations and six can be considered germinal, putting in evidence the high frequency of somatic mutations in mtDNA in healthy young individuals (table 3). With the exception of site 16189, that mutated twice, the rest of the sites mutated once. Six mutations were observed at sites that display above average site-specific mutation rates and seven other mutations were detected at sites with slightly or substantially lower than average site-specific mutation rates (Meyer et al. 1999). Moreover, the analysis of the amount of polymorphism at the population level (table 3) has shown that mutations (for example in position 238), that don't

appear to be polymorphic in West-Eurasian and African populations can be found in heteroplasmy.

In relation to the substitutions that occurred in HVRI, we detected five involving pyrimidine transitions and one involving a purine transition, which represents a 5 pyrimidine/purine transition rate. In HVRII three pyrimidine transitions and two purine transitions were identified, representing a 1.5 pyrimidine/purine transition rate. Pooling our data with data obtained by Parsons et al. (1997), Sigurðardóttir et al. (2000); Heyer et al. (2001), Howell et al. (2003) and Brandstätter et al. (2004), that analysed simultaneously the two hypervariable regions, the pyrimidine/purine transition rates are 5.25 for HVRI and 1.1 for HVRII. The transition/transversion rate is 25 in HVRI and 46 if HVRI and HVRII are considered together.

Mutation Rate

In the analysis of the 321 mtDNA transmissions we detected eleven substitutions in the D-loop (more precisely, in 973 bp located between positions 16024-16596 and 1-400), which implies that 0.0343 mutations occur (at a detectable level) in the D-loop in each generation (95% CI: 0.014 - 0.054). If we employ the same definition of mutation used by other authors (for example Howell et al. 2003), only mutations for which there is evidence that they are germinal should be considered (see table 2). This implies that the mutation rate would be reduced almost by half: 6 mutations in 321 mtDNA transmissions, that is 0.0187 mutations/generation for the entire D-loop or 1.921×10^{-5} mutations/site/generation. Assuming that the generational time is 25 years, the mutation rate would be 0.7684

mutations/site/Myr (95% CI: 0.164 - 1.398). In order to make an estimate of the mutation rate for HVRI and HVRII separately we considered HVRI as being delimited by positions 16024-16383 (360 bp) and HVRII as being delimited by positions 57-371 (315 bp) in accordance with Sigurðardottir et al. (2000). This provided an estimation of 1.04 mutations/site/Myr (95% CI: 0 - 2.222) for HVRI and 1.19 mutations/site/Myr (95% CI: 0 - 2.556) for HVRII.

To compare our results with those obtained by others, which have been summarised by Howell et al. (2003) [see table 2 of the article by Howell et al. (2003)], we estimated the divergence rate assuming that the generational time is 20 years and that there are 1,122 bp in the noncoding control region. Under these assumptions, the divergence rate (twice the mutation rate) derived from our data is 1.67 mutations/site/Myr (95% CI: 0.357 - 3.030), which is one of the highest reported so far (Howell et al. 2003). As in other studies, our estimation is much higher than those obtained by phylogenetic methods. The possible causes for this discrepancy will be discussed further below, implicating mainly the process by which heteroplasmy is resolved and the effect of sex proportion, genetic drift and selection at the different nested hierarchies of populations covered by the mtDNA molecule: the organelle, the cell, the tissue, the organism, and so on up through natural populations and species (Rand 2001).

DISCUSSION

In the analysis of the 321 mtDNA transmissions in healthy individuals, eleven substitutions were detected in the D-loop, from these, only 6 can be considered germinal mutations. The mutation rate obtained employing the same definition of mutation used by others, considering only germinal mutations, was 0.7684 mutation/site/Myr (95% CI: 0.164 - 1.398), one of the highest values reported so far. These results indicating that the discrepancy observed in the mutation rate reported in family and phylogenetic studies can't be attributed to the inclusion of somatic mutations in calculations or to the use of families with mtDNA disease. Other explanation that has been given for the differences is the existence of substantial heterogeneity in mutation rates across the D-loop, i.e., mutations occur preferentially at hypervariable sites, from our data, half of the reported mutations displayed above average site-specific mutation rates and the other half were detected at sites with slightly or substantially lower than average site-specific mutation rates (Meyer et al. 1999) indicating that mutations observed in family studies does not seem to occur preferentially in hipervariable sites. The effect of selection and genetic drift, have also be invoked as a causes of discrepancy, since many of the new mutations observed in pedigree studies are destined to be lost from the population as a whole. Some of the mutations observed in this work were not previously detected at population level, moreover, when comparing the values of the pyrimidine/purine transitions observed in our sample, with the estimates made by Meyer et al. (1999) (the authors reported a rate of 3.49 for HVRI and 1.66 for HVRII) slight differences can be seen. Data from family studies [present study and

Parsons et al. (1997), Sigurðardóttir et al. (2000); Heyer et al. (2001), Howell et al. (2003) and Brandstätter et al. (2004)] seem to show a higher value of pyrimidine transitions in HVRI and a higher value of purine transitions in HVRII. However, the available data obtained from family studies are limited, and the small sample number could bias the differences observed. The transition/transversion rate value observed in the family studies is much higher than the value of 15 obtained by phylogenetic estimations for HVRI and HVRII (Vigilant et al. 1991; Tamura and Nei 1993), indicating that a considerable number of transitions are probably eliminated by drift and/or by selection. From the 321 mtDNA transmissions analysed, we did not identify any cases of mtDNA paternal contribution. This result does not support the suggestion made by Hagelberg (2003) that the paternal contribution of mtDNA could be one of the possible causes for the discrepancy observed in the mutation rate reported in family and phylogenetic studies. However, our data put in evidence the possibility that recombination occurs between mtDNA types present in a heteroplasmic individual. Individuals II4 and III2 from family ML display a mtDNA type that could result from a recombination event. However, one of the mutations signalling this possibility occurred in a very unstable region and definitive conclusions cannot be made. Future research is necessary to better address the possibility of recombination in mtDNA of healthy individuals. Kraytsberg and co-authors (2004) have recently reported evidence for recombination between maternal and paternal mtDNA in muscle tissue of a myopathy-affected individual (Schwartz and Vissing 2002). Even excluding mtDNA paternal contribution, recombination may occur in the general population between two or more mtDNA types, originating from new mutations found in

heteroplasmic individuals. If this possibility is confirmed it will be indispensable to estimate the frequency of recombination to correctly assess the mtDNA evolutionary rate. To fulfil a strict concept of mutation (any heritable change in the genetic material) for mtDNA there are other factors, not previously considered, that must be taken into account. One of these factors is the gender of individuals with germinal mutations. In our data, two out (HR_I IV2; ML III2) of the six mutations will never be passed to the next generation because they are carried exclusively by men. This fact automatically reduces the mutation rate that we stated previously to 0.5123 mutations/pb/Myr (95% CI: 0.000 – 1.028). This correction is the equivalent of considering the effective population size for the coalescent estimation of the mutation rate, N , for the mtDNA (Lundstrom et al. 1992). It is worthwhile noting that the proportion of males and females included in our sample that were analysed in the first step (53.7% and 46.3% respectively) is similar to the proportion of newborn males and females in the present day Azorean population [for the year 2002 (INE, 2003): 52% and 48% respectively].

Another point is that for a mutation in mtDNA to reach polymorphism levels in the population (and eventually become fixed) it is first necessary that it passes through heteroplasmy to homoplasmy at the individual level (intra-individual fixation). To estimate the probability of this event many variables should be considered, such as the reduction of segregating units during the bottleneck phase, the initial proportion of each variant and the selective pressure in the mutated site. Since the individual effect of each factor is difficult to assess [see for example the discussion concerning the number of segregating units during the bottleneck in Bendall et al. (1996), Poulton et al. (1998) and Lutz et al. (2000)] the

theoretical estimation of this probability is currently impractical. An empirical estimation of this probability could theoretically be obtained by directly counting the number of new heteroplasmies that resolve to the new variant. Unfortunately at this time this is impossible to perform from our data, since the new variants were found in heteroplasmy and we would need to follow the future generations to access this probability.

If we assume that all the mutations are neutral, then the probability that the individual will eventually become fixed (homoplasmic) for the new variant is approximately the proportion of the new variant. If we take the proportions (based on clone count) of the new variants in women from the most recent generation that present germinal heteroplasmy (NP_II: III1; NP_III: IV7, V2; PA_II: II1, II2; SAO: III2, III3) the approximate number of new mutations that rise intra-individual fixation is 1.308 in all D-loop and 0.213 and 1.095 in HVRI and HVRII respectively. The mutation rate becomes 0.1675 mutations/site/Myr (CI 95%: 0 – 0.452) or 4.1878×10^{-6} mutations/site/generation (CI 95%: 0 – 11.305×10^{-6}) for the D-loop; if we consider only 675 bp from HVRI and HVRII, the mutation rate becomes 0.2415 mutations/site/Myr (CI 95%: 0 – 0.652) or 6.0367×10^{-6} mutations/site/generation (CI 95%: 0 – 16.296×10^{-6}). The mutation rate for HVRI is 0.0737 mutations/site/Myr (CI 95%: 0 – 0.333) and for HVRII is 0.4332 mutations/site/Myr (CI 95%: 0 – 1.270), under these assumptions.

Our estimation of the mutation rate for 675 bp from HVRI and HVRII is in the upper end of the range of values reported by several phylogenetic studies (table 5) and is similar to the value obtained, using the data of family studies where only homoplasmic mutations were found (Heyer et al. 2001), if the correction for sex

(proportion of newborn males and females: ~ 0.5 in the majority of populations) is applied. Considering the data of Heyer et al. (2001) and correcting for gender proportion the mutation rate would be 5.85×10^{-6} mutations/site/generation.

One important observation derived from our data is precisely the low level of heteroplasmies that turned to homoplasmies and the heterogeneity in the process by which heteroplasmy is resolved. This raises the question of what mechanism (or mechanisms) is involved in producing heterogeneity in the resolution of heteroplasmy. Previous studies (Howell et al. 1992; Bendall et al. 1996; Lutz et al. 2000) have focused their interest on this question, paying particular attention to the effect of the number of segregating units during the bottleneck. However other factors, like the proportions of mtDNA variants in the heteroplasmic mothers and the effect of selection at different mtDNA levels of organization (the molecule, the organelle, the cell and the tissue), must also be considered.

In our data there are two heteroplasmic positions (16189 in family HR_I and 152 in family PA_II), in which both variants appear to have a similar fixation probability, displaying a rapid increase of heteroplasmic levels and/or shifting to homoplasmy and high levels of heterogeneity among siblings. In other cases (16086 in family ML) it is not clear if both variants have the same probability of becoming fixed, however, a rapid shift to homoplasmy was still observed. These three cases are compatible with the traditional concept of a tight bottleneck. On the other hand, family NP_III shows low levels of heteroplasmy in position 150 during at least four generations without a significant increase in the minor variant, this pattern is consistent with a wider bottleneck.

We used the Wright-Fisher model of random genetic drift (Fisher 1930; Wright 1931) to estimate the probability of the mtDNA population within the oocyte drifting from i to j copies of mtDNA type A in one generation. The effective population size considered for the Wright-Fisher model was estimated, as described in the data analysis section of material and methods, to be ~ 30 . In figure 2 the probability distributions for proportions of A, in generation one, ranging from 5% to 50% are presented. It can be observed that the shape of the probability distributions is sharper for initial proportions of A inferior to 20%. Therefore for smaller initial proportions of A the probability of wide frequency changes is smaller than for high initial proportions. Thus a tight bottleneck could lead to either rapid fixation or maintenance of the heteroplasmic condition for several generations depending upon the initial proportions of the two variants. This means that whenever a new mutation reaches a high proportion ($\geq 20\%$) in the oocyte, it has a higher probability of increasing its proportion in the next generation and ultimately of becoming fixed in comparison with a new mutation that does not initially reach a high proportion. This leads us to conclude that the moment in which the mutation arises (before, during or after the bottleneck) determines its fate.

The previously proposed model cannot, however, explain the heteroplasmic profile of family SAO. This family displays very low levels of heteroplasmy during generations I and II but in the third generation we observed a significant increase in the heteroplasmy levels in the four descendents. The heteroplasmy profile across generations indicates that the direction of the mutational event is a G to A change in position 215 representing a phylogenetic retromutation. Position 215 presents a low site-specific mutation rate (Meyer et al. 1999) and the 215G variant is very

infrequent in human populations (see table 2). Considering these particularities we put forward the hypothesis that, for selective pressure, the mtDNA variants 215G produced by an A to G mutation in position 215 never totally shift to homoplasmy or if this occurs there is a tendency for retromutation. This observation is supported by the higher proportion of the A variant. Thus, the family SAO raises the possibility that selection could be one of the factors that must be considered when predicting the proportion of heteroplasmic transmissions that lead to intra-individual fixation. As this effect is probably different for each site, with our current knowledge it is impossible to correctly weight it for the estimations of mutation rate. A suitable method would be to empirically estimate the proportion of intra-individual fixation by observing the number of heteroplasmic transmissions that become fixed for the new variant. Then, all the factors affecting the segregation of heteroplasmy would be averaged and therefore implicitly considered. Furthermore, assuming that selection pressure will be similar at the intra-individual level and the population level, the rate of mutations that reach intra-individual fixation will be similar to the rate of population (and species) fixation. This means that only neutral and mildly deleterious mutations will be fixed at an intra-individual level but not deleterious mutations, as is observed in the majority of mitochondrial diseases in which the mutant type never reaches intra-individual fixation. Selection, however, could still modify the population fixation rate in cases in which a change in environment modifies the selective pressure (i.e. Ruiz-Pesini et al. 2004).

In conclusion, our experimental design in which: a) multiple generations were sampled and analysed, b) the sex proportions of the samples analysed were

representative of the proportions in the source population, and c) recombination and paternal contribution of mtDNA were considered, allowed us to take insights in the issue of the previously reported discrepancy between the human mtDNA mutation rate observed along evolutionary timescales and estimations obtained using family pedigrees. Based on our results, these discrepancies can be understood when correcting for the proportion of sexes in newborn individuals and for the probability of intra-individual fixation. Moreover, under neutrality and assuming a constant bottleneck size, genetic drift alone can explain different patterns of heteroplasmy segregation; the destiny of a new mutation being strictly related with the initial proportion of the new variant.

SUPPLEMENTARY MATERIAL

Detailed description of families showing heteroplasmy produced by point substitutions and a table with results (table I) are available as online supplementary material.

ACKNOWLEDGMENTS

This work was supported by the C. Santos PhD grant (ref. SFRH / BD / 723 / 2000) from Fundação para a Ciência e a Tecnologia - Operational programs Science Technology and Innovation (POCTI) and Society of Information (POSI) financed by UE funds and national funds of MCES. R. Montiel is a postdoctoral fellow of the Fundação para a Ciência e a Tecnologia (SFRH/BPD/13256/2003).

Literature Cited

Anderson, S., A. Bankier, M. Barrell et al. (14 co-authors). 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290**:457-464.

Baasner, A., C. Schafer, A. Junge, and B. Madea. 1998. Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. *Forensic Sci Int* **98**:169-178.

Baasner A. and B. Madea. 2000. Sequence polymorphisms of the mitochondrial DNA control region in 100 German Caucasians. *J Forensic Sci* **45**:1343-8.

Bandelt, H. J., P. Forster, B. C. Sykes, and M. B. Richards. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141**:743-753.

Bendall, K. E., and B. C. Sykes. 1995. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet* **57**:248-256.

Bendall, K. E., V. A. Macaulay, J. R. Baker, and B. C. Sykes. 1996. Heteroplasmic point mutations in the human mtDNA control region. *Am J Hum Genet* **59**:1276-1287.

Brandstätter, A., H. Niederstatter, and W. Parson. 2004. Monitoring the inheritance of heteroplasmy by computer-assisted detection of mixed basecalls in the entire human mitochondrial DNA control region. *Int J Legal Med* **118**:47-54.

Brehm, A., L. Pereira, T. Kivisild, and A. Amorim. 2003. Mitochondrial portraits of the Madeira and Acores archipelagos witness different genetic pools of its settlers. *Hum Genet* **114**:77-86.

Brown, M. D., S. H. Hosseini, A. Torroni, H. J. Bandelt, J. C. Allen, T. G. Schurr, R. Scozzari, F. Cruciani, and D. C. Wallace. 1998. mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *Am J Hum Genet* **63**:1852-1861.

Calafell, F., P. Underhill, A. Tolun, D. Angelicheva, and L. Kalaydjieva. 1996. From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* **60 (Pt 1)**:35-49.

Cann R., M. Stoneking, and A. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* **325**:31-36

Cavalli-Sforza L. L., and W. F. Bodmer. 1971. *The genetics of Human Populations*. W. H. Freeman and Company, San Francisco.

Chen, Y. S., A. Torroni, L. Excoffier, A. S. Santachiara-Benerecetti, and D. C. Wallace. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* **57**:133-149.

Delghandi M., E. Utsi, and S. Krauss 1998. Saami mitochondrial DNA reveals deep maternal lineage clusters. *Hum Hered* **48**:108-114

Denver D. R., K. Morris, M. Lynch, L. L. Vassilieva, and W. K. Thomas. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**:2342-2344

Elson, J. L., D. M. Turnbull, and N. Howell. 2004. Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am J Hum Genet* **74**:229-238.

Fisher R. A. 1930. *The genetical theory of Natural Selection*. Clarendon, Oxford.

Forster, P., R. Harding, A. Torroni, and H. J. Bandelt. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* **59**:935-945.

Francalacci P., J. Bertranpetit, F. Calafell, and P. A. Underhill 1996. Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthropol* **100**:443-460.

Graven, L., G. Passarino, O. Semino, P. Boursot, S. Santachiara-Benerecetti, A. Langaney, and L. Excoffier. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* **12**:334-345.

Grzybowski, T., B. A. Malyarchuk, J. Czarny, D. Miscicka-Sliwka, and R. Kotzbach. 2003. High levels of mitochondrial DNA heteroplasmy in single hair roots: reanalysis and revision. *Electrophoresis* **24**:1159-1165.

Hall T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **41**:95-98.

Hagelberg E. 2003. Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. Trends Genet **19**:84-90

Handt O., S. Meyer, and A. Haeseler 1998. Compilation of Human mtDNA control region sequences. Nucleic Acids Res **26**:126-129

Hartl D., and A. Clark 1997. Principles of Population Genetics. 3th edition. Sinauer Associates, Inc., Publishers Sunderland, Massachusetts.

Hasegawa M., and S. Horai 1991. Time of the deepest root for polymorphism in human mitochondrial DNA. J Mol Evol **32**:37-42

Hauswirth W. W., and D. A. Clayton 1985. Length heterogeneity of a conserved displacement loop sequence in human mitochondrial DNA. Nucleic Acids Res **13**:8093–8104

Helgason A., S. Sigurðardóttir, J. Gulcher, R. Ward, and K. Stefánsson 2000. mtDNA and the origin of the Icelanders: deciphering signals of recent population history. Am J Hum Genet **66**:999-1016.

Heyer, E., E. Zietkiewicz, A. Rochowski, V. Yotova, J. Puymirat, and D. Labuda. 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. Am J Hum Genet **69**:1113-1126.

Hofmann, S., M. Jaksch, R. Bezold, S. Mertens, S. Aholt, A. Paprotta, and K. D. Gerbitz. 1997. Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. *Hum Mol Genet* **6**:1835-1846.

Horai, S., K. Hayasaka, R. Kondo, K. Tsugane, and N. Takahata. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci U S A* **92**:532-536.

Howell, N., S. Halvorson, I. Kubacka, D. A. McCullough, L. A. Bindoff, and D. M. Turnbull. 1992. Mitochondrial gene segregation in mammals: is the bottleneck always narrow? *Hum Genet* **90**:117-120.

Howell, N., I. Kubacka, and D. A. Mackey. 1996. How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet* **59**:501-509.

Howell, N., C. B. Smejkal, D. A. Mackey, P. F. Chinnery, D. M. Turnbull, and C. Herrnstadt. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am J Hum Genet* **72**:659-670.

Howell N., J. L. Elson, D. M. Turnbull, and C. Herrnstadt 2004. African Haplogroup L mtDNA Sequences Show Violations of Clock-like Evolution. *Mol Biol Evol.* **21**:1843-1854.

INE – Instituto Nacional de Estatística 2003. Anuário Estatístico de Portugal. Instituto Nacional de Estatística, Portugal.

Ingman, M., H. Kaessmann, S. Pääbo, and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**:708-713.

Jenuth J. P., A. C. Peterson, K. Fu, and E. A. Shoubridge. 1996. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat Genet* **14**:146-151

Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge

Kraytsberg, Y., M. Schwartz, T. A. Brown, K. Ebralidse, W. S. Kunz, D. A. Clayton, J. Vissing, and K. Khrapko. 2004. Recombination of human mitochondrial DNA. *Science* **304**:981.

Lambert, D. M., P. A. Ritchie, C. D. Millar, B. Holland, A. J. Drummond, and C. Baroni. 2002. Rates of evolution in Ancient DNA from Adélie Penguins. *Science* **295**:2270-2273.

Lee, H. Y., U. Chung, J. E. Yoo, M. J. Park, and K. J. Shin. 2004. Quantitative and qualitative profiling of mitochondrial DNA length heteroplasmy. *Electrophoresis* **25**:28-34.

Lima M. M. 1996. *Doença de Machado-Joseph nos Açores: Estudo epidemiológico, biodemográfico e genético*. PhD thesis, University of the Azores, Ponta Delgada, Azores (Portugal).

Lundstrom R., S. Tavaré, and R. H. Ward 1992. Estimating substitution rates from molecular data using the coalescent. *Proc Natl Acad Sci USA* **89**:5961-5965.

Lutz, S., H. J. Weisser, J. Heizmann, and S. Pollak. 1998. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* **111**:67-77.

Lutz, S., H. J. Weisser, J. Heizmann, and S. Pollak. 2000. Mitochondrial heteroplasmy among maternally related individuals. *Int J Legal Med* **113**:155-161.

Macaulay, V. A., M. B. Richards, P. Forster, K. E. Bendall, E. Watson, B. Sykes, and H. J. Bandelt. 1997. mtDNA mutation rates--no need to panic. *Am J Hum Genet* **61**:983-990.

Meyer, S., G. Weiss, and A. von Haeseler. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**:1103-1110.

Meyer, S., and A. von Haeseler. 2003. Identifying site-specific substitution rates. *Mol Biol Evol* **20**:182-189.

Miller W. G. 2003. OpenStat2 (OS2), Version 1.4.4. West DeMoines, Iowa.

Mishmar, D., E. Ruiz-Pesini, P. Golik et al. (13 co-authors). 2003. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* **100**:171-176.

Montiel R. 2001. Estudio diacrónico de la variabilidad del DNA mitocondrial en población Catalana. Ph.D. dissertation, Autonomous University of Barcelona, Barcelona, Spain. Also available from <http://www.tdcat.cesca.es/TDCat-0726101-095837>

Montiel, R., A. Malgosa, and P. Francalacci. 2001. Authenticating ancient human mitochondrial DNA. *Human Biology* **73**:689-713.

Pääbo S. 1996 Mutational hot spots in the mitochondrial microcosm. *Am J Hum Genet* **59**:493-496

Parson, W., T. J. Parsons, R. Scheithauer, and M. M. Holland. 1998. Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* **111**:124-132.

Parsons, T. J., D. S. Muniec, K. Sullivan et al. (11 co-authors). 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* **15**:363-368.

Pereira, L., M. J. Prata, and A. Amorim. 2000. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet* **64**:491-506.

Pesole, G., E. Sbisa, G. Preparata, and C. Saccone. 1992. The evolution of the mitochondrial D-loop region and the origin of modern man. *Mol Biol Evol* **9**:587-598.

Piercy, R., K. M. Sullivan, N. Benson, and P. Gill. 1993. The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* **106**:85-90.

Poulton, J., V. Macaulay, and D. R. Marchington. 1998. Mitochondrial genetics '98 is the bottleneck cracked? *Am J Hum Genet* **62**:752-757.

Poulton, J., and D. R. Marchington. 2002. Segregation of mitochondrial DNA (mtDNA) in human oocytes and in animal models of mtDNA disease: clinical implications. *Reproduction* **123**:751-755.

Rannala B., and G. Bertorelle. 2001. Using linked markers to infer the age of a mutation. *Hum Mutat* **18**:87-100

Rand D. M. 2001. The units of selection on mitochondrial DNA. *Ann Rev Ecol Syst* **32**:415-448

Richards, M., H. Corte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H. J. Bandelt, and B. Sykes. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* **59**:185-203.

Richards, M., V. Macaulay, E. Hickey et al. (37 co-authors). 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**:1251-1276.

Richards, M. B., V. A. Macaulay, H. J. Bandelt, and B. C. Sykes. 1998. Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* **62 (Pt 3)**:241-260.

Rousselet F., and P. Mangin. 1998. Mitochondrial DNA polymorphisms: a study of 50 French Caucasians individuals and application to forensic casework. *Int J Legal Med* **111**:292-298.

Ruiz-Pesini, E., D. Mishmar, M. Brandon, V. Procaccio, and D. C. Wallace. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**:223-226.

Salas, A., M. Richards, M. V. Lareu, R. Scozzari, A. Coppa, A. Torroni, V. Macaulay, and A. Carracedo. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* **74**:454-465.

Santos, C., M. Lima, R. Montiel, N. Angles, L. Pires, A. Abade, and M. P. Aluja. 2003. Genetic structure and origin of peopling in the Azores islands (Portugal): the view from mtDNA. *Ann Hum Genet* **67**:433-456.

Sigurðardóttir S., A. Helgason, J. R. Gulcher, K. Stefansson, and P. Donnelly. 2000. The mutation rate in the human mtDNA control region. *Am J Hum Genet* **66**:1599-1609.

Schwartz M., and J. Vissing. 2002. Paternal inheritance of mitochondrial DNA. *N Engl J Med* **347**:576-580

Soodyall, H., L. Vigilant, A. V. Hill, M. Stoneking, and T. Jenkins. 1996. mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* **58**:595-608.

SPSS Inc. 1989-2003. SPSS 12.0S for Windows. Chicago.

Stoneking, M., S. T. Sherry, A. J. Redd, and L. Vigilant. 1992. New approaches to dating suggest a recent age for the human mtDNA ancestor. *Philos Trans R Soc Lond B Biol Sci* **337**:167-175.

Tamura K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**:512-526

Templeton A. R. 1993. The "Eve" hypotheses: a genetic critique and reanalysis. *Am Anthropolo* **95**:51-72

Torrioni, A., C. Rengo, V. Guida et al. (12 co-authors). 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* **69**:1348-1356.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A. C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science* **253**:1503-1507.

Wright S. 1931. Evolution in Mendelian population. *Genetics* **16**:97-159

Wright S. 1939. Statistical genetics in relation to evolution. Pp 1-64 *in*: G. Teissier, eds.
Exposés de biométrie et de statistique biologique. Vol. 13. Paris, Hermann, pp 1-64

Table 1 – Type and sub-type of families based in the category of heteroplasmies detected.

Family Type	Sub-type	Families
1) All the members analysed present the same mutations in the CR	a) and are fully homoplasmics	13 families: AC_I, AC_II, CAR_IV, CAR_VI, CG_I, ER, MA_I, MA_II, MA_III, NP_V, RUB, TM and ZH
	b) and are heteroplasmic for the number of cytosine residues in the poly-C 16184-16193 tract.	3 families: BM, JA_III, NP_I
	c) and are heteroplasmic for the number of cytosine residues in the poly-C 303-309 tract.	10 families: AV, BF, CAM_II, CAR_V, JL_I, JL_II, JS, MM_II, NP_IV, PP
	d) and are heteroplasmic for the number of cytosine residues in the poly-C tracts 16184-16193 and 303-309.	2 families: CV, MN
2) At least one of the members analysed presents only in the Poly-C tract 303-309 differences respect to the remaining individuals of the family		9 families: CAM_I, CAM_III, CAR_III, CG_II, CG_III, CM, HR_II, JA_II, MBP (all the members of families CAR_III and MBP are heteroplasmic for the number of cytosine residues in the Poly-C tracts 16184-16193)
3) Two of the members analysed present heteroplasmy produced by the deletion of one G in the Poly-G 66-71		1 family: MM_I
4) At last one of the members analysed present differences respect to the remaining individuals of the family, produced at least by a substitution:	a) presents heteroplasmy produced by a substitution.	3 families: NP_II, NP_III, SAO
	b) presents point heteroplasmy and heteroplasmy in the poly-C tract (303-309) similar in all the members of the family	1 family: JA_I
	c) presents point heteroplasmy and the poly-C tract 303-309 is different in some of the members of the family.	2 families: CAR_II, PA_II
	d) presents point heteroplasmy and heteroplasmy in the poly-C tracts 16184-16193 and 303-309 similar in all the members of the family	1 family: PA_I
	e) presents point heteroplasmy at nucleotide position 16189 T/C and heteroplasmy produced by a C insertion in the poly-C tract 16190-16193.	2 families: CAR_I, HR
	f) presents point heteroplasmy in positions 16086 T/C and 16182 A/C or Del A being the poly-C tract 303-309 is different in some of the members of the family.	1 family: ML

Table 2 – For each family analysed, code (according to figure 1), sex and proportion of variants in individuals presenting heteroplasmy produced by substitutions.

Family	Individual	Sex	Position	Proportion height peaks Forward Primer	Proportion height peaks Reverse Primer	Mean proportion height peaks	Proportion clone number (N)
CAR_I	V1	M	16189 T->C	88%T:12%C	92%T:8%C	90%T:10%C	-
CAR_II	II4	M	64 C->T	77%C:23%T	67%C:33%T	72%C:28%T	-
HR_I	IV2	M	16189 T->C	54%T:46%C	60%T:40%C	57%T:43%C	-
	(III5)	F	16189 T->C	87%T:13%C	90%T:10%C	88.5%T:11.5%C	-
	(II3)	F	16189 T->C	95%T:5%C	98%T:2%C	96.5%T:3.5%C	-
	(II2)	F	16189 T->C	95%T:5%C	98%T:2%C	96.5%T:3.5%C	-
JA_I	III10	F	238 A->G	80%A:20%G	70%A:30%G	75%A:25%G	-
ML	III2	M	16086 T->C	61%T:39%C	59%T:41%C	60%T:40%C	73.0%T:27.0%C (37)
			16182 A->C or Del A	61%A:39%C	64%A:36%C	62.5%A:37.5%C	67.6%A:32.4%C (37)
	(II4)	F	16086 T->C	64%T:36%C	66%T:34%C	65%T:35%C	59.7%T:40.3%C (62)
				16182 A->C or Del A	60%A:40%C	66%T:34%C	63%T:37%C
NP_II	III1	F	16150 C<?>T	81%T:19%C	81%T:19%C	81%T:19%C	78.7%T:21.3%C (47)
	(II2)	F	16150 C<?>T	86%T:14%C	89%T:11%C	87.5%T:12.5%C	-
NP_III	IV1	M	16309 A->G	66%A:34%G	73%A:27%G	69.5%A:30.5%G	-
	VII1	M	150 C->T	85%C:15%T	89%C:11%T	87%C:13%T	-
	V2	F	150 C->T	96%C:4%T	95%C:5%T	95.5%C:4.5%T	98%C:2%T (49)
	IV7	F	150 C->T	86%C:14%T	94%C:6%T	90%C:10%T	96%C:4%T (50)
	(V3)	F	150 C->T	96%C:4%T	97%C:3%T	96.5%C:3.5%T	-
	(IV6)	F	150 C->T	85%C:15%T	90%C:10%T	87.5%C:12.5%T	-
	(III5)	F	150 C->T	98%C:2%T	97%C:3%T	97.5%C:2.5%T	98%C:2%T (50)
PA_I	IV4	F	16264 C->T	58%T:42%C	56%T:44%C	57%T:43%C	57.7%T:42.3%C (26)
PA_II	II1	F	152 C<?>T	50%C:50%T	60%C:40%T	55%C:45%T	68.6%C:31.4%T (51)
	II2	F	152 C<?>T	82%C:18%T	85%C:15%T	83.5%C:16.5%T	75%C:25%T (44)
	(I2)	F	152 C<?>T	70%C:30%T	75%C:25%T	72.5%C:27.5%T	-
SAO	III1	M	215 G<?>A	50%G:50%A	62%G:38%A	56%G:44%A	62.1%G:37.9%A (29)
	III2	F	215 G<?>A	76%G:24%A	78%G:22%A	77%G:23%A	66.6%G:33.3%A (27)
	III3	F	215 G<?>A	70%G:30%A	79%G:21%A	74.5%G:25.5%A	86.2%G:13.8%A (29)
	III4	M	215 G<?>A	68%G:32%A	82%G:18%A	75%G:25%A	90.6%G:9.4%A (29)
	(II1)	F	215 G<?>A	94%G:6%A	97%G:3%A	95.5%G:4.5%A	98.5%G:1.5%A (66)
	(I2)	F	215 G<?>A	95%G:5%A	95%G:5%A	95%G:5%A	97.7%G:2.3%A (44)

Table 3 - Positions presenting heteroplasmy (excluding positions 16184-16188, 16190-16193 and 303-315).

Family	Haplogroup	Position	State in homoplasmic individuals of the family	CRS ^a / Consensus ^b / African San ^c	Germline	Paternal Contribution	Site-Specific Rate ^d	Population Database frequency
ML	M1	16086 T->C	T	T / T / T	Yes (2 generations)	-	<1	98.8% T 1.1% C
NP_II	J	16150 C->T	T	C / C / C	Yes (2 generations)	-	<0.5	99.4% C 0.5% T 0.1% A
ML	M1	16182 A->C or Del A	A	A / A / A	Yes (2 generations)	-	<1	99.2% A 0.4 del 0.3% C 0.1 G
CAR_I	H	16189 T->C	T	T / C / C	No	No	~5	77.1% T 22.7% C 0.1% A
HR_I	H	16189 T->C	T	T / C / C	Yes (3 generations)	-	~5	77.1% T 22.7% C 0.1% A
PA_I	U1	16264 C->T	C	C / C / C	No	No	<1	97.4% C 2.6% T
NP_III	K	16309 A->G	A	A / A / A	No	No	4	98.9% A 1.1% G
CAR_II	H	64 C->T	C	C / C / C	No	No	~3	97% C 3% T
MM_I	J1a	Poly-G 66-71 Del G	G	G / G / G	Yes (2 generations)	No	<0.5	100%G
NP_III	K	150 C->T	C	C / T / C	Yes (4 generations)	-	~6	87.5% C 12.5% T
PA_II	H	152 C->T	T	T / T / C	Yes (2 generations)	-	~6	72.3% T 27.6% C
SAO	H	215 G->A	-	A / A / A	Yes (3 generations)	-	~0.5	99.4% A 0.6% G
JA_I	J	238 A->G	A	A / A / A	No	-	0	100% A

a. Anderson et al. (1981)

b. Ingman et al. (2000), GenBank accession number NC_001807

c. Ingman et al. (2000), GenBank accession number AF347008

d. Meyer et al. (1999)

Table 4 – mtDNA variants found by sequencing of multiple clones in individuals ML III2 and ML II4.

1	1	1	1	Poly-C HVRI (16184-16193)	37 clones	62 clones
6	6	6	6		ML III2	ML II4
0	1	1	1		% (N)	% (N)
8	2	8	8			
6	9	2	3			
T	A	A	C	CCCCCCCC	-	1.6 (1)
T	A	A	C	CCCCCCCC	13.5 (5)	6.5 (4)
T	A	A	C	CCCCCCCC	10.8 (4)	11.3 (7)
T	A	A	C	CCCCCCCC	18.9 (7)	19.4 (12)
T	A	A	C	CCCCCCCC	10.8 (4)	12.9 (8)
T	A	A	C	CCCCCCCC	5.4 (2)	1.6 (1)
T	A	A	C	CCCCCCCC	2.7 (1)	1.6 (1)
T	A	A	C	CCCCCCCC	2.7 (1)	-
T	A	-	C	CCCCCCCC	2.7 (1)	-
T	A	-	C	CCCCCCCC	2.7 (1)	-
T	A	-	C	CCCCCCCC	-	3.2 (2)
T	A	-	C	CCCCCCCC	2.7 (1)	1.6 (1)
C	A	A	C	CCCCCCCC	2.7 (1)	-
C	A	A	C	CCCCCCCC	-	1.6 (1)
C	A	-	C	CCCCCCCC	2.7 (1)	1.6 (1)
C	A	-	C	CCCCCCCC	2.7 (1)	11.3 (7)
C	A	-	C	CCCCCCCC	8.1 (3)	11.3 (7)
C	A	-	C	CCCCCCCC	10.8 (4)	6.5 (4)
C	A	-	C	CCCCCCCC	-	4.8 (3)
C	A	-	C	CCCCCCCC	-	3.2 (2)

Table 5 - Evolutionary rate of mtDNA hypervariable regions obtained in this study and using distinct phylogenetic methods (adapted from Montiel 2001).

Reference	Region	Method	Date of divergence ^a	Evolutionary rate as expressed by authors	Standardized Evolutionary rate ^b
Vigilant et al. 1991	HVR I and II	PHY (Maximum parsimony)	4 Myr	11.5%/Myr	0.0575
Vigilant et al. 1991	HVR I and II	PHY (Maximum parsimony)	6 Myr	17.3%/Myr	0.0865
Hasegawa and Horai 1991	HVR I and II	PHY (Markov process)	4 Myr	0.177-0.286/site/Myr	0.1770-0.2860
Stoneking et al. 1992	HVR I and II	PHY (UPGMA)	-	11.81 ± 3.11%/Myr	0.1181 ± 0.0311
Stoneking et al. 1992	HVR I and II	PHY (Jukes-Cantor, distance)	-	11.42 ± 3.33%/Myr	0.1142 ± 0.0333
Pesole et al. 1992	HVR I and II	PHY (Markov process)	5 Myr	12 ± 5/100 site/Myr	0.1200 ± 0.05
Pesole et al. 1992	HVR I and II	PHY (Markov process)	7.5 Myr	18 ± 10/100 site/Myr	0.1800 ± 0.1
Tamura and Nei 1993	HVR I and II	PHY (Tamura-Nei, distance)	4-6 Myr	1.5×10 ⁻⁷ -2.5×10 ⁻⁸ /site/year	0.1500-0.0250
Horai et al. 1995	HVR I and II	PHY (UPGMA)	4.9 Myr	7.00×10 ⁻⁸ / site/year	0.0700
Horai et al. 1995	HVR I	PHY (UPGMA)	4.9 Myr	10.3×10 ⁻⁸ / site/year	0.1030
Forster et al. 1996	HVR I	PHY (Median networks)	-	1 trans/275sites/20,180 years	0.1802 trans
Lambert et al. 2002 ^c	HVRI	Ancient DNA	-	0.8553 substitutions/site/Myr	0.8553
Present study ⁽¹⁾	D-loop (973 bp)	Empirical (families)	-	3.5219×10 ⁻⁵ /site/generation	1.4088
Present study ⁽²⁾	D-loop (973 bp)	Empirical (families)	-	1.9210×10 ⁻⁵ /site/generation	0.7684
Present study ⁽³⁾	D-loop (973 bp)	Empirical (families)	-	1.2807×10 ⁻⁵ /site/generation	0.5123
Present study ⁽⁴⁾	D-loop (973 bp)	Empirical (families)	-	4.1878×10 ⁻⁶ /site/generation	0.1675
Present study ⁽⁴⁾	HVR I and II	Empirical (families)	-	6.0367×10 ⁻⁶ /site/generation	0.2415
Present study ⁽⁴⁾	HVR I	Empirical (families)	-	1.8420×10 ⁻⁶ /site/generation	0.0737
Present study ⁽⁴⁾	HVR II	Empirical (families)	-	10.8300×10 ⁻⁶ /site/generation	0.4332

a. Date of divergence Human/Chimpanze used to calibrate evolutionary rate

b. mutations/site/Myr;

c. mean between three different estimations present by the authors

PHY - Phylogenetic

(1) All the substitutions that were detected were considered for the mutation rate calculation;

(2) Only the substitutions for which there was evidence of a germinal origin were considered;

(3) Only the substitutions present in women for whom there was evidence of a germinal origin were considered;

(4) Only the substitutions with a germinal origin present in women that would become fixed at the individual level were considered;

Figure Legends

Fig. 1 – Forty-eight pedigrees relating sampled individuals. Individuals presenting heteroplasmy (excluding positions 16184-16188, 16190-16193 and 303-315) are signalled with arrows (\rightarrow heteroplasmy produced by substitution; \dashrightarrow heteroplasmy produced by deletion; \dashrightarrow heteroplasmy produced by a possible deletion). Individuals represented by filled black squares and circles were analysed in the first step; individuals represented by half tone squares and circles were analysed after detection of heteroplasmy in individuals analysed in the first step; samples of individuals represented by grille squares and circles were available but were not analysed; samples of individuals represented by white squares and circles were not available.

Fig. 2 – Probability distributions that the variant A drift from i to j copies in one generations. Curves resulted from the application of the Wright-Fisher model of random genetic drift for different initial proportions of the new variant, ranging from 5% to 50%, assuming a constant bottleneck of 30.

Fig 1

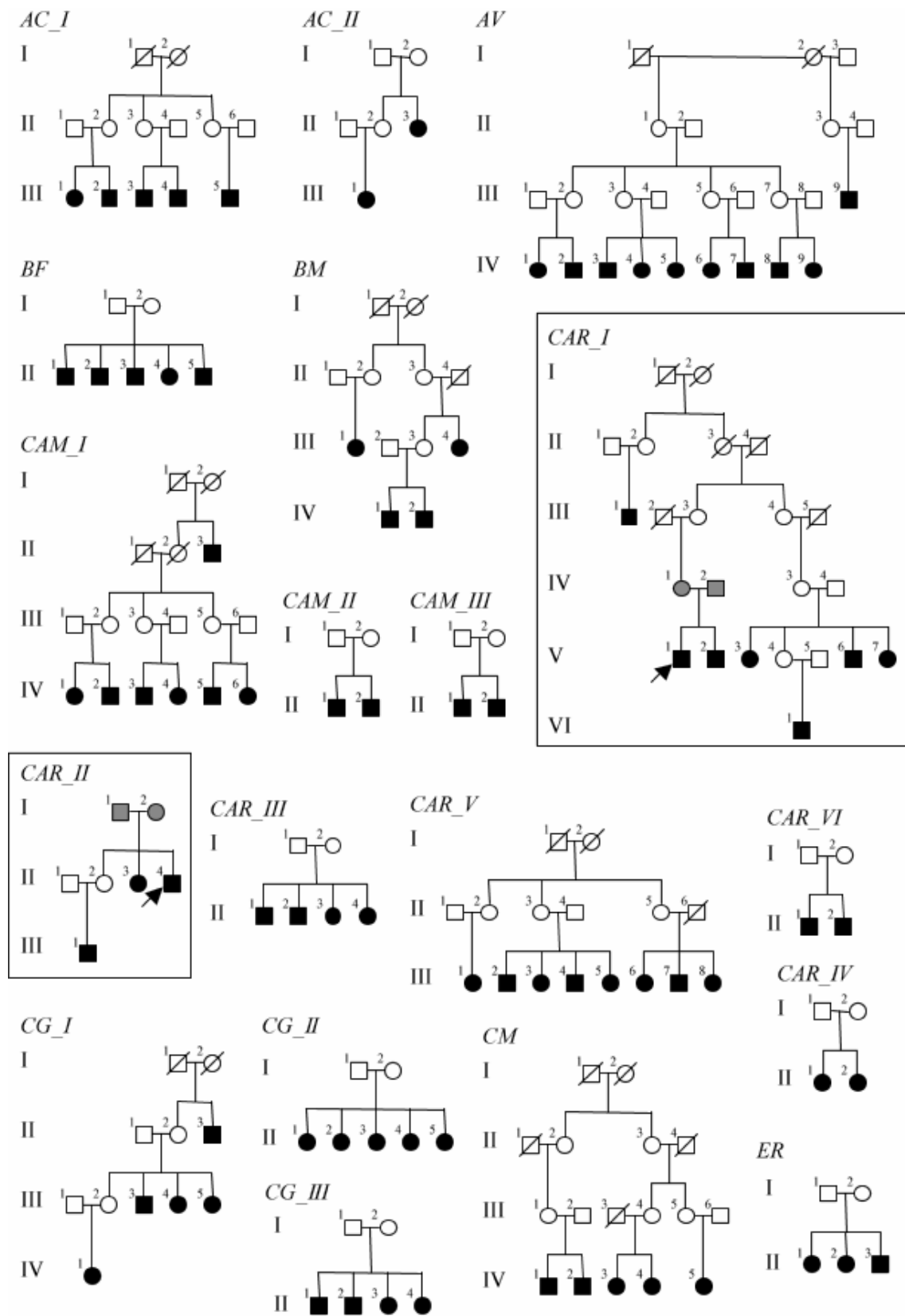


Fig. 1 (cont)

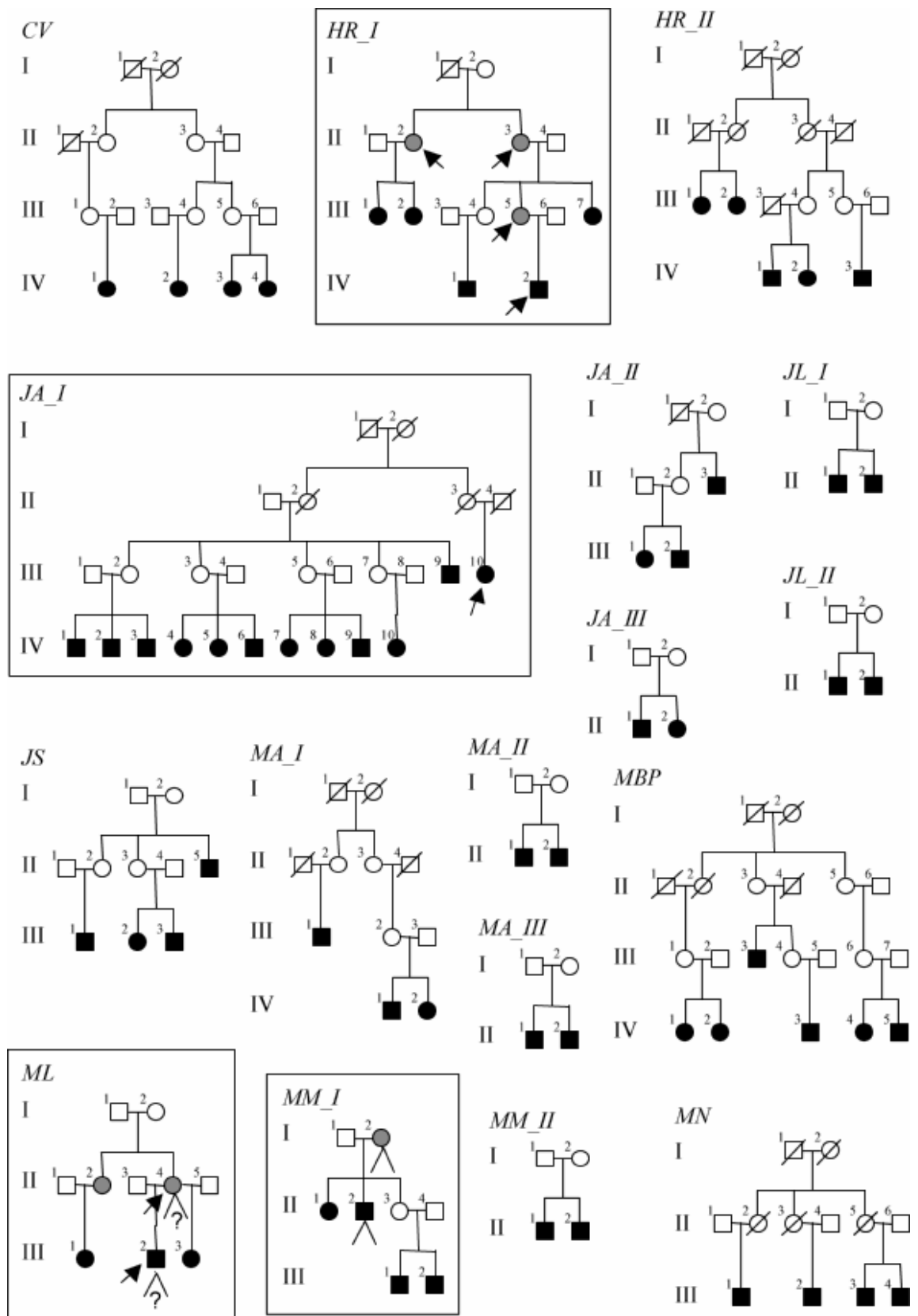


Fig. 1 (cont)

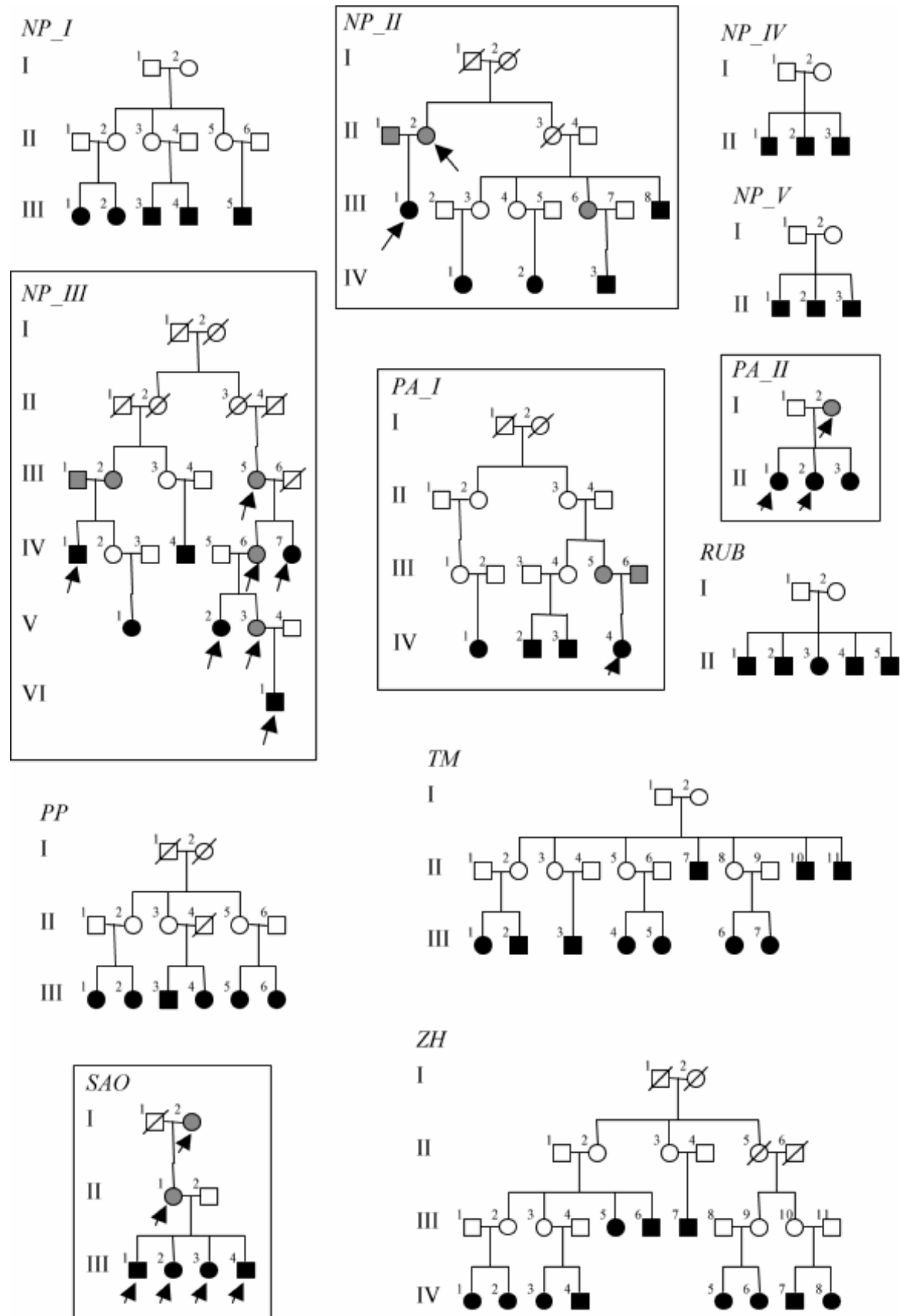
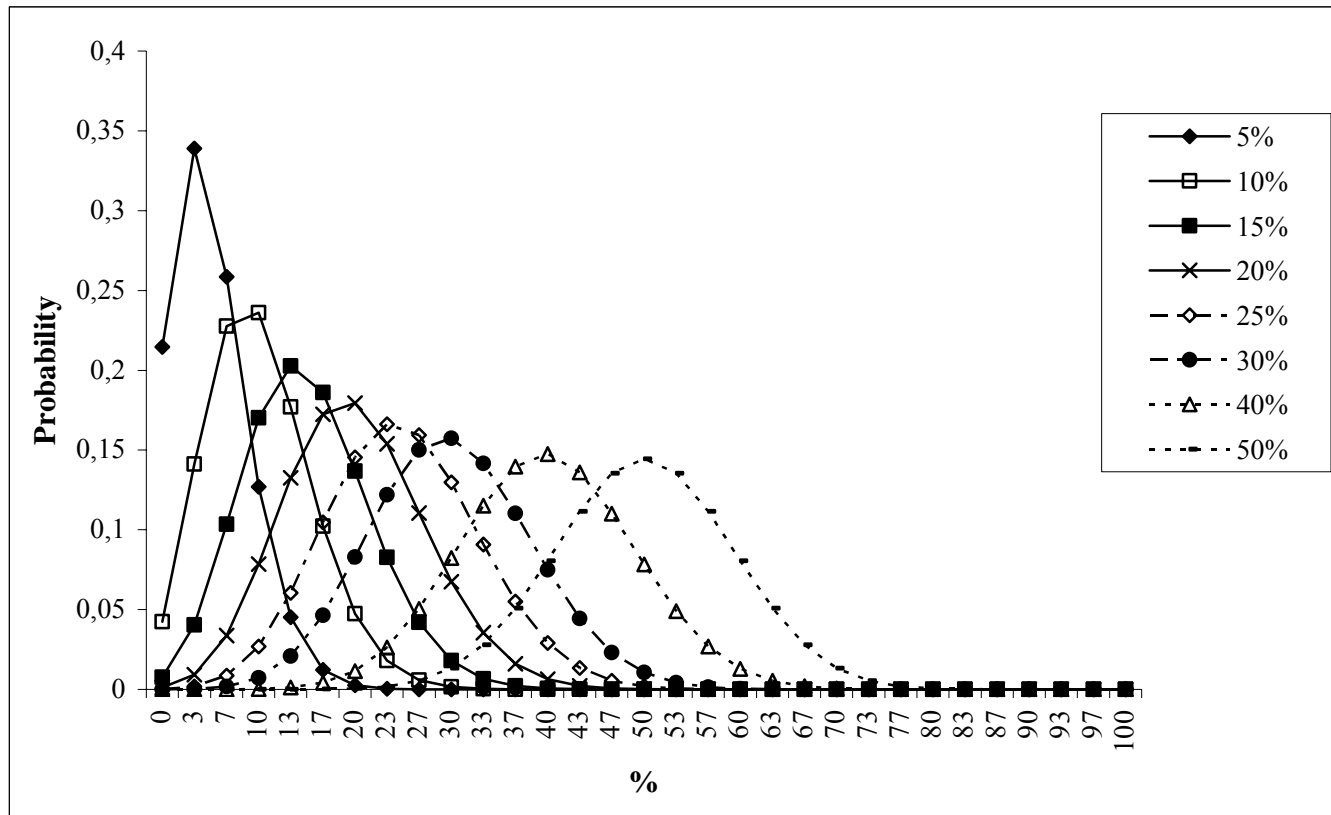


Fig. 2



SUPPLEMENTARY MATERIAL

Detailed description of families presenting point heteroplasmy

Families in which only one individual presented heteroplasmy produced by substitution:

- Family CAR I (Fig.1): in the first step of the analysis we sequenced the D-loop of seven maternally related individuals (5 men and 2 women) from which, one man (V1) displayed low levels of heteroplasmy (90%T:10%C) in position 16189 as the result of a T->C mutation (table 2).

- Family CAR II (Fig. 1): initially we sequenced the D-loop of three maternally related individuals (2 men and 1 woman). From these, one man (II4) displayed median levels (72%C:28%T) of heteroplasmy in position 64, produced by a mutation C->T (table 2).

- Family PA I (Fig.1): the D-loop of four maternally related individuals (2 men and 2 women) was sequenced, revealing that one woman (IV4) displayed high levels (57%T:43%C) of heteroplasmy in position 16264, produced by a mutation C->T (table 2).

In the three families described above, the D-loop of the mother of heteroplasmic individuals was subsequently sequenced and no evidence of the presence of heteroplasmy in the previous generation was found. To verify if the observed heteroplasms could be the result of a paternal contribution of mtDNA, the D-loop of the father of heteroplasmic individuals was also sequenced. However, no evidence of a paternal contribution was found. These results indicate that the

heteroplasmies observed in individuals CAR_I V1, CAR_II II4 and PA_I IV4 must result from somatic mutations that, in this case, occurred in the individuals themselves, or alternatively, it could result from germline mutations that occurred in the oocyte from which these individuals developed.

- Family JA_I (Fig. 1): the D-loop of twelve maternally related individuals (6 men and 6 women) was sequenced. One woman (III10) presented median levels (75%A:25%G) of heteroplasmy in position 238, produced by a mutation A->G (table 2). It was not possible to analyse her mother because she is deceased. Nevertheless, and since no other member of the family presented heteroplasmy, there is no evidence that the mutation is germinal. Thus, the heteroplasmy observed in individual III10 must result from a similar mechanism to the one previously described for families CAR_I, CAR_II and PAI.

Families in which various individuals displayed heteroplasmy produced by substitution:

- Family HR_I (Fig.1): we first sequenced the D-loop of five maternally related individuals (2 men and 3 women). From these, one man (IV2) showed high levels (57%T:43%C) of heteroplasmy in position 16189 produced by a mutation T->C (table 2). The subsequent analysis of his mother (III5), his grandmother (II3) and his grandmother's sister (II2), showed the presence of low levels of heteroplasmy in all of them. Hence, it is likely that the transition from T to C in position 16189 occurred in individual I2 and was transmitted to individual II3 and to one of her daughters. However, this mutation was lost in the descendent line of individual II2. A significant increase of mitotype 16189C is observed across

generations (proportion test, II3/III5: $Z=-2.085$, $p=0.0185$; III5/IV2: $Z=-4.909$, $p<0.0001$).

- Family ML (Fig.1): in the first analysis step we sequenced the D-loop of three maternally related individuals (1 man and 2 women). The analysis of the electropherograms indicated that one man (III2) presented high levels of heteroplasmy in two distinct positions: 16086 T/C and 16182 A/C or Del A (table 2). The heteroplasmy observed in position 16182 is probably the result of a deletion of an Adenine, however, it is not possible to exclude the possibility that it represents an A to C transversion. The subsequent analysis of his mother (II4) also showed high levels of heteroplasmy in the two positions referred to above (table 2). Individual II2 was also analysed but no heteroplasmy was found (excluding the two poly-C). In this family a significant heterogeneity within generations was observed in individuals III2 and III3 presenting very different mtDNA compositions. To determine exactly the mtDNA variants found in individuals II4 and III2 the D-loop of these individuals was cloned and 62 and 37 clones were sequenced respectively. In table 4 the mtDNA variants defined by positions 16086 and 16182 and by the number of C residues in the poly-C from HVRI are presented. Excluding length variation in the poly-C tract, we detected four distinct mtDNA types defined by positions 16086 and 16182 in the two individuals. The wild-type of the family was identified in both the individuals (16086T-16182A), as well as clones 16086T-16182C/Del A, 16086C-16182A and 16086C-16182C/Del A. These results were replicated using two distinct samples of both individuals and they can be considered authentic. Moreover, additional mutations found in the wild-type (for example a

mutation in position 16129) were also present in the other mtDNA variants (table 4), which verifies once again the authenticity of these results.

- Family NP II (Fig.1): the D-loop of five maternally related individuals (2 men and 3 women) was initially sequenced. One woman (III1) presented low levels (81%T:19%C) of heteroplasmy in position 16150 T/C (table 2). The subsequent analysis of her mother (II2) evidenced the presence of heteroplasmy in the previous generation, as the mother presented levels of heteroplasmy (87.5%T:12.5%C) that were not significantly different from those observed in her daughter (proportion test, $Z = 1.157$; $p = 0.1236$).

In this family it is difficult to infer the direction of mutation. The homoplasmic individuals have a thymine residue in position 16150, which represents a change with respect to the original state in the reference sequences (table 3) and in the founder sequence of J haplogroup (Richards et al. 2000). Thus, the heteroplasmy observed in the mother-daughter pair may have two different explanations. The first is that the minority variant 16150C may result from a retromutation T->C in individual II2, an unlikely explanation in our opinion since the site-specific substitution rate of this position (table 3) is inferior to 0.5 (Meyer et al. 1999). The other possibility is that this pair of individuals represents the final steps of the fixation of the variant 16150T.

Family NP III (Fig.1): initially we sequenced the D-loop of six maternally related individuals (3 men and 3 women). One man (IV1) presented high levels of heteroplasmy in position 16309. Low levels of heteroplasmy in position 150 were detected in two women (IV7, V2) and one man (VI1) (table 2). We have not identified the two heteroplasmies segregating in the same individual. The

subsequent analysis of the mother (III2) of individual IV1 [heteroplasmic at position 16309 (69.5%A:30.5%G)] did not give evidence of heteroplasmy in the previous generation. To verify if the heteroplasmy observed could be the result of a paternal contribution, the D-loop of the father (III1) was also sequenced, revealing no evidence of a paternal contribution. The results indicate that the heteroplasmy observed in individual IV1 must result from an adenine to guanine transition that may represent a somatic mutation that occurred in the individual himself or a germline mutation that occurred in the oocyte that gave rise to the individual. Concerning the heteroplasmy in position 150, observed in individuals IV7, V2 and VII, the analysis of the maternally related individuals, III2, III5, IV6 and V3, gave evidence of the presence of low levels of heteroplasmy in individuals V3, IV6 and III5. The presence of heteroplasmy in individual III5, that was almost imperceptible in the electropherograms, was confirmed when fifty clones were sequenced. Forty-nine of these clones harboured 150C and 1 clone had 150T (table 2). The origin and direction of the mutational event seems to be a cytosine to thymine transition in individual III5 or in her mother. These results show the occurrence of a mutation and maintenance of low levels of heteroplasmy during at least four generations without a significant increase in the new variant.

- Family PA II (Fig.1): we initially sequenced the D-loop of three sisters of which, II1 and II2 showed heteroplasmy in position 152 C/T (table 2) and II3 was homoplasmic in 152T. The levels of heteroplasmy in individuals II1 (55%C:45%T) and II2 (83.5%C:16.5%T) were significantly different (proportion test, $Z = 4.281$; $p < 0.0001$). The subsequent analysis of their mother (I2) evidenced the presence of heteroplasmy in the previous generation, as their mother displayed median levels

(72.5%C:27.5%T) of heteroplasmy. Comparing the proportions of heteroplasmy between mother and offspring, significant differences were detected in the two comparisons (proportion test, I2/III1: $Z = -2.497$; $p = 0.0062$; I2/II2: $Z = 1.893$; $p = 0.0292$), showing both a significant increase and decrease of the 152T variant from one generation to the next. Furthermore, a rapid shift to homoplasmy was observed and apparently both variants 152T and 152C could be resolved to homoplasmy. Since there is no available sample from any of individual I2's ancestors, and the two generations analysed are not sufficient to infer the original nucleotide present in position 152 (and therefore to infer the direction of the mutational event), we tried to deduce the original state of position 152 using the data from Ingman et al. (2000). According to the consensus sequence proposed by the authors (GeneBank accession number NC_001807), the original state is 152T. Although this method of inferring the original state is suboptimal considering that this position has a high site-specific substitution rate (table 3) (Meyer et al. 1999), we consider that the original state in this family is in fact 152T. This statement is supported by the observation that H sequences have a high frequency of 152T. Thus, we can assume that the occurrence of a mutation from T to C in individual I2 (or in her mother) generated two distinct mtDNA variants, 152T and 152C. Both variants appear to have an equal probability of fixation, since in the three descendents analysed, one reverted to the original state, one showed more than 80% of mtDNAs from the 152C type and the last one presented equal frequencies of both variants.

- Family SAO (Fig.1): in the first step of the analysis we sequenced the D-loop of four brothers (2 women and 2 men) all presenting heteroplasmy in position 215 G/A (table 2). The level of heteroplasmy in individual III1 is significantly

higher than that observed in individuals III2 (proportion test, III1/III2: $Z = 3.146$; $p = 0.0008$), III3 (III1/III3: $Z = 2.826$; $p = 0.0023$), and III4 (III1/III4: $Z = 2.826$; $p = 0.0023$). The subsequent analysis of their mother (II1) and grandmother (I2) evidenced the presence of heteroplasmy in the previous generations (table 1), as both their mother and grandmother showed very low levels of heteroplasmy that were confirmed by sequencing of multiple clones. Comparing the proportions of heteroplasmy across generations, similar values of heteroplasmy were observed in generations I and II and an increase in the 215A variant was observed from the generations II to III. Furthermore, in three generations none of the individuals resolved to homoplasmy and only variant 215A appears to show a tendency to increase. In this family all the available individuals are heteroplasmic, and once again the nucleotide present in this position before the mutational event that produced heteroplasmy took place cannot be identified with certainty. The profile of heteroplasmy across generations seems to indicate that the direction of the mutational event is a G to A transition in position 215. If we take into account that the original state in the consensus sequence of Ingman et al. (2000) is an adenine, this transition could represent a retromutation in position 215.

Table I - D-loop sequences of the Azorean families analysed.

Family	Samples	Haplogroup	D-loop (16024-16569 and 1-400)
AC_I	III1, 2, 3, 4, 5	H or HV, preHV	16519, 152, 263, 311.1
AC_II	II3, III1	H or HV, preHV	16519, 263, 311.1
AV	III9, IV1, 2, 3, 4, 5, 6, 7, 8, 9	V or pre*V	16291, 16298, 72, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
BF	III,2,3,4,5	J	16069, 16126, 16278, 16366, 16519, 73, 185, 188, 228, 263, 295, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1 [462, 489]
BM	III1, 4, IV1, 2	U6a1	16108, 16111, 16172, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16219, 16278, 16519, 73, 152, 251, 263, 311.1
CAM_I	II3, IV2, 3, 4, 5	U5	16224, 16270, 16T, 73, 150, 199, 263, 279, 311.1, [518T]
	IV1, 6	U5	16224, 16270, 16T, 73, 150, 199, 263, 279, <i>Het Poly-C 303-309 (303.0, 303.1)</i> , 311.1, [518T]
CAM_II	III, 2	U5a1	16114A, 16192, 16256, 16270, 16294, 16526, 73, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
CAM_III	II1	H or HV, preHV	16239G, 16256, 16519, 152, 263, <i>Het Poly-C 303-309 (303.0, 303.1, 303.2, 303.3)</i> , 311.1
	II2	H or HV, preHV	16239G, 16256, 16519, 152, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
CAR_I	III1, V2, 3, 6, 7, VI1	H or HV, preHV	16274, 16294G, 16519, 152, 263, 311.1
	V1	H or HV, preHV	Het 16189 T/C , <i>Het Poly-C 16184-16193 (16190.0, 16190.1)</i> , 16274, 16294G, 16519, 152, 263, 311.1
	IV1	H or HV, preHV	16274, 16294G, 16519, 152, 263, 311.1
	IV2 (Only HVRI)	U5a1a	16086, 16256, 16270, 16342, 16399
CAR_II	II3	H or HV, preHV	16519, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
	III1	H or HV, preHV	16519, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
	II4	H or HV, preHV	16519, Het 64 T/C , 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
	I2	H or HV, preHV	16519, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
	I1 (Only HVRII)	H or HV, preHV	152, 263, 311.1
CAR_III	III, 3, 4	X	16086, 16104, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.1, 16190.2)</i> , 16223, 16271, 16278, 16519, 73, 153, 195, 225, 226, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
	II2	X	16086, 16104, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16223, 16271, 16278, 16519, 73, 153, 195, 225, 226, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
CAR_IV	III, 2	H or HV, preHV	16274, 16294G, 16519, 152, 263, 311.1
CAR_V	III1, 2, 3, 4, 5, 6, 7, 8	J	16069, 16126, 16278, 16366, 16519, 73, 185, 188, 228, 263, 295, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1 [465, 492]
CAR_VI	III, 2	H or HV, preHV	16519, 263, 311.1
CG_I	II3, III3, 4, 5, IV1	L1c2	16129, 16187, 16189, 16223, 16265C, 16278, 16286A, 16292, 16294, 16311, 16360, 16519, 16527, 73, 152, 182, 186A, 189C, 195, 198, 247, 263, 297, 311.1, 316
CG_II	III,2,4	V or pre*V	16298, 72, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
	II3,5	V or pre*V	16298, 72, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
CG_III	III,4	H or HV, preHV	16519, 199, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
	II2,3	H or HV, preHV	16519, 199, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
CM	IV1, 2	H or HV, preHV	16265C, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
	IV3, 5	H or HV, preHV	16265C, 263, <i>Het Poly-C 303-309 (303.0, 303.1, 303.2)</i> , 311.1
	IV4	H or HV, preHV	16265C, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
CV	IV1, 2, 3, 4	L3b	16124, 16171, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16214, 16223, 16278, 16362, 73, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
ER	III, 2, 3	J	16069, 16126, 16221, 73, 185, 228, 263, 295, 303.1, 311.1 [464, 484, 491]

HR_I	III1, 2, 7, IV1	H or HV, preHV	16192, 16260, 16519, 263, 311.1
	IV2	H or HV, preHV	Het 16189 T/C, Het Poly-C 16184-16193 (16190.0, 16190.1) , 16192, 16260, 16519, 263, 311.1
	III5	H or HV, preHV	Het 16189 T/C, Het Poly-C 16184-16193 (16190.0, 16190.1) , 16192, 16260, 16519, 263, 311.1
	II3	H or HV, preHV	Het 16189 T/C, Het Poly-C 16184-16193 (16190.0, 16190.1) , 16192, 16260, 16519, 263, 311.1
	II2	H or HV, preHV	Het 16189 T/C, Het Poly-C 16184-16193 (16190.0, 16190.1) , 16192, 16260, 16519, 263, 311.1
HR_II	III1	H or HV, preHV	16519, 93, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
	III2, IV1, 2, 3	H or HV, preHV	16519, 93, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
JA_I	III9, IV1, 2, 3, 4, 5, 6, 7, 8, 9, 10	J	16069, 16126, 16278, 16366, 16519, 73, 185, 188, 228, 263, 295, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1 [462, 489]
	III10	J	16069, 16126, 16278, 16366, 16519, 73, 185, 188, 228, Het 238 A/G , 263, 295, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1 [462, 489]
JA_II	II3, III2	H or HV, preHV	16293, 16311, 16519, 146, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
	III1	H or HV, preHV	16293, 16311, 16519, 146, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
JA_III	III, 2	H or HV, preHV	16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16519, 187T, 263, 311.1
JL_I	III, 2	H or HV, preHV	16519, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1 [479]
JL_II	III, 2	H or HV, preHV	16519, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
JS	II5, III1, 2, 3	W	16223, 16292, 16320, 16519, 73, 189, 195, 204, 207, 263, <i>Het Poly-C 303-309 (303.0, 303.1)</i> , 311.1
MA_I	III1, IV1, 2	H or HV, preHV	16519, 263, 311.1
MA_II	III, 2	U5	16252, 16270, 16342, 16399, 73, 200, 263, 311.1
MA_III	III, 2	H or HV, preHV	16519, 152, 263, 311.1
MBP	III3, IV1, 5	M1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16223, 16249, 16311, 16399, 16454, 73, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i>
	IV2, 3	M1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.1, 16190.2)</i> , 16223, 16249, 16311, 16399, 16454, 73, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i>
	IV4	M1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.1, 16190.2)</i> , 16223, 16249, 16311, 16399, 16454, 73, 195, 263, <i>Het Poly-C 303-309 (303.0, 303.1, 303.2, 303.3)</i>
ML	III1	M1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16249, 16311, 16519, 73, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1 [489]
	III3	M1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2, 16190.3)</i> , 16249, 16311, 16519, 73, 263, 311.1 [489]
	III2	M1	Het 16086 T/C , 16129, Het 16182 C/A or del A , 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.-2, 16190.-1, 16190.0, 16190.1, 16190.2, 16190.3, 16190.4, 16190.5)*</i> , 16249, 16311, 16519, 73, 263, <i>Het Poly-C 303-309 (303.0, 303.1, 303.2, 303.3, 303.4)</i> , 311.1, [489]
	II4	M1	Het 16086 T/C , 16129, Het 16182 C/A or del A , 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.-2, 16190.-1, 16190.0, 16190.1, 16190.2, 16190.3, 16190.4, 16190.5)*</i> , 16249, 16311, 16519, 73, 263, 303.3, 311.1, <i>Het Poly-C 303-309 (303.0, 303.1, 303.2, 303.3, 303.4)</i> [489]
	II2	M1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.-1, 16190.0, 16190.1, 16190.2, 16190.3)</i> , 16249, 16311, 16519, 73, 263, 303.2, 311.1, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> [489]
MM_I	II1, III1, 2	J1a	16069, 16126, 16145, 16231, 16261, 73, 150, 152, 195, 215, 263, 295, 309.1T, 311.1, 319
	II2	J1a	16069, 16126, 16145, 16231, 16261, Het Poly G 66-71 delG , 73, 150, 152, 195, 215, 263, 295, 309.1T, 311.1, 319
	I2	J1a	16069, 16126, 16145, 16231, 16261, Het Poly G 66-71 delG , 73, 150, 152, 195, 215, 263, 295, 309.1T, 311.1, 319
MM_II	III, 2	H or HV, preHV	16293, 16311, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
MN	III1, 2, 3, 4	V or pre*V	16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2, 16190.3)</i> , 16298, 72, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
NP_I	III1, 2, 3, 4, 5	U6a	16172, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.1, 16190.2)</i> , 16219, 16278, 73, 263, 311.1

NP_II	III8, IV1, 2, 3	J	16069, 16126, 16150, 16519, 73, 185, 188, 228, 263, 295, 311.1 [463, 490]
	III1	J	16069, 16126, 16150 Het T/C , 16519, 73, 185, 188, 228, 263, 295, 311.1 [463, 490]
	II2	J	16069, 16126, 16150 Het T/C , 16519, 73, 185, 188, 228, 263, 295, 311.1 [463, 490]
	III6	J	16069, 16126, 16150, 16519, 73, 185, 188, 228, 263, 295, 311.1 [463, 490]
NP_III	IV4, V1	K	16224, 16311, 16519, 73, 195, 263, 311.1 [498]
	IV1	K	16224, 16309 Het A/G , 16311, 16519, 73, 195, 263, 315 (1C) [498]
	III2	K	16224, 16311, 16519, 73, 195, 263, 315 (1C) [498]
	III1 (Only HVRI)	H or HV, preHV	16519
	IV7, V2, VI1	K	16224, 16311, 16519, 73, 150 Het C/T , 195, 263, 311.1 [498]
	V3	K	16224, 16311, 16519, 73, 150 Het C/T , 195, 263, 311.1 [498]
	IV6	K	16224, 16311, 16519, 73, 150 Het C/T , 195, 263, 311.1 [498]
	III5	K	16224, 16311, 16519, 73, 150 Het C/T , 195, 263, 311.1 [498]
NP_IV	III, 2, 3	H or HV, preHV	16140, 16519, 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
NP_V	III, 2, 3	H or HV, preHV	16304, 185, 263, 311.1 [456]
PA_I	IV1, 2, 3	U1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16249, 16311, 16355, 16519, 73, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
	IV4	U1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16249, 16264 Het T/C , 16311, 16355, 16519, 73, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
	III5	U1	16129, 16183C, 16189, <i>Het Poly-C 16184-16193 (16190.0, 16190.1, 16190.2)</i> , 16249, 16311, 16355, 16519, 73, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
	III6 (Only HVRI)	J	16069, 16126, 16278, 16366, 16519
PA_II	II3	H or HV, preHV	16519, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
	III,2	H or HV, preHV	16519, 152 Het T/C , 263, <i>Het Poly-C 303-309 (303.1, 303.2, 303.3)</i> , 311.1
	I2	H or HV, preHV	16519, 152 Het T/C , 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
PP	III1, 2, 3, 4, 5, 6	L2a	16223, 16278, 16294, 16309, 16390, 73, 143, 152, 195, 263, <i>Het Poly-C 303-309 (303.1, 303.2)</i> , 311.1
RUB	III,2,3,4,5	H or HV, preHV	16519, 263, 311.1
SAO	III1, 2, 3, 4	H or HV, preHV	16201, 16278, 16519, Het 215 G/A , 263, 311.1
	III1	H or HV, preHV	16201, 16278, 16519, Het 215 G/A , 263, 311.1
	I2	H or HV, preHV	16201, 16278, 16519, Het 215 G/A , 263, 311.1
TM	II7, 10, 11, III1, 2, 3, 4, 5, 6, 7	H or HV, preHV	16239, 16519, 263, 311.1
ZH	III5, 6, 7, IV1, 2, 3, 4, 5, 6, 7, 8	H or HV, preHV	16519, 152, 263, 311.1

Variant positions from the CRS are shown between 16024-16569 and 1-400 according to Anderson et al. (1981). Substitutions are transitions unless the base change is explicitly indicated.

* Variants detected by sequencing of multiple clones.

V.

DISCUSIÓN GENERAL



V. DISCUSIÓN GENERAL

Se han presentado y discutido los resultados obtenidos durante el período de investigación que ha conducido a la realización de la presente tesis (artículos I - V). A fin de cumplir los objetivos generales que se han propuesto, se discutirán de forma integrada las aportaciones de los artículos I-V relacionándolos, a su vez, con trabajos realizados por otros investigadores.

V. 1. METODOLOGÍA JERÁRQUICA DE ANÁLISIS DEL DNAm.

Uno de los objetivos de la presente tesis consiste en desarrollar una metodología jerárquica de análisis del DNAm. En el artículo I (*Determination of Human Caucasian Mitochondrial DNA Haplogroups by means of a Hierarchical Approach*), se ha propuesto una metodología jerárquica que permite la clasificación en haplogrupos de los DNAs mitocondriales que pueden encontrarse en poblaciones caucasoides. Se proponen dos abordajes jerárquicos: uno aplicable a poblaciones con un *pool* genético mitocondrial formado casi exclusivamente por haplogrupos típicos del oeste euroasiático, y otro adecuado para analizar poblaciones que, además de presentar un elevado porcentaje de haplogrupos típicos del oeste euroasiático, tienen también una contribución importante de DNAm africano y/o asiático.

La metodología presentada plantea el análisis jerárquico de 18 polimorfismos puntuales filogenéticamente informativos, de la región codificante del DNAm, y la secuenciación automática de la región HVRI en casos puntuales en que es necesaria la información de esta región para definir el haplogrupo, o simplemente, en aquellos casos en que se pretenda una resolución filogeográfica más alta. La validez de la metodología jerárquica presentada se puede asegurar, ya que las posiciones polimórficas utilizadas son estables; para definir cada haplogrupo son comprobadas como mínimo dos posiciones, y, en algunos casos, se sugiere el análisis de posiciones confirmatorias adicionales.

Con la unificación de la nomenclatura de clasificación del DNAm (Richards *et al.*, 1998) y el reconocimiento de que el análisis conjunto de la región de control y de posiciones de la región codificante permitía definir una filogenia más adecuada del DNAm (Macaulay *et al.*, 1999b), la mayoría de grupos de investigación han pasado a analizar de forma rutinaria las dos regiones del DNAm (Pereira *et al.*, 2001; Malyarchuk *et al.*, 2002,

2003; González *et al.*, 2003; Stevanovitch *et al.*, 2004; Dubut *et al.*, 2004; Rosa *et al.*, 2004; entre otros). Sin embargo, aunque algunos trabajos de caracterización poblacional hayan dado una mayor importancia al análisis de posiciones polimórficas de la región codificante (Morelli *et al.*, 2000; Montiel, 2001; Anglés *et al.*, 2003; entre otros), en muchos casos, el análisis de posiciones de la región codificante ha sido utilizado como una metodología complementaria a la secuenciación de la región de control, de forma que se estudiaban sólo algunas posiciones de la región codificante con el objetivo de confirmar y/o complementar la información extraída de la región de control.

Sin duda, la tendencia de priorizar la secuenciación se justifica, en parte, por el hecho de que se han acumulado un gran número de secuencias de la región de control, particularmente de la región HVRI, de innumerables poblaciones, lo que permite la realización de análisis filogeográficos bastante completos. Sin embargo, en otro tipo de estudios, como el realizado por Simoni y colaboradores (2000) en que intentaban identificar un patrón de variación del DNAm_t en Europa y Próximo Oriente, se puede prescindir de la utilización de secuencias de la región de control de todas las muestras, y pueden ser bastante más exitosos si se utiliza una metodología de análisis que permita la realización de una correcta clasificación en haplogrupos, como ha sido discutido por Torroni y colaboradores (2000) y posteriormente demostrado por Richards y colaboradores (2002).

Así, la metodología propuesta en el artículo I es particularmente útil en aquellos casos en que se pretende realizar la clasificación del DNAm_t en haplogrupos, o en estudios en los que se pretende evaluar el grado de mestizaje de las poblaciones. En ambos casos la secuenciación de la región HVRI de todas las muestras es claramente innecesaria y en el caso de poblaciones europeas poco eficiente, ya que casi un 50% de las muestras no se puede clasificar con seguridad utilizando simplemente la información de la región HVRI.

Ahora bien, las ventajas de utilizar una aproximación jerárquica son indiscutibles. Ante todo, permiten el análisis organizado y más rápido que aquellas metodologías que se basan en el análisis de todos los polimorfismos puntuales para todas las muestras. En tal caso, la técnica de detección de los polimorfismos puntuales por PCR-RFLP es la más utilizada y la propuesta en nuestra metodología para analizar 17 de los 18 polimorfismos puntuales sugeridos. A pesar de que se han propuesto otras técnicas más rápidas en que se utilizan PCRs múltiplex y una detección por *SNaPshot* o minisequenciación (Morley *et al.*, 1999; Brandstätter *et al.*, 2003; Plaza *et al.*, 2003; Quintans *et al.*, 2004; entre otros), en estos momentos la utilización de PCR-RFLP sigue presentando ventajas, ya que se trata de una técnica sencilla, fácilmente aplicable en cualquier laboratorio y relativamente económica.

Nuestra propuesta jerárquica sigue siendo válida a pesar de los recientes avances en la definición de la filogenia del DNAm_t. Los estudios de DNAm_t en Eurasia Occidental, seguramente iniciarán una nueva etapa, siendo prioritario realizar la caracterización

poblacional de los subgrupos de los haplogrupos más frecuentes, como los del H (Achilli et al, 2004), y redefinir el patrón de variación del DNAmT en Europa y Próximo Oriente. En este contexto, la metodología presentada en el artículo I puede servir de soporte para esta nueva ola de trabajos, tratándose de incorporar en el esquema jerárquico adecuado las posiciones pertinentes en el seguimiento de los distintos haplogrupos.

V. 2. ESTRUCTURA BIODEMOGRÁFICA DE LA POBLACIÓN DE LAS AZORES

La información sobre la composición poblacional, por sexo y grupos de edad, de las diferentes islas del archipiélago es prácticamente inexistente en el período que va desde el inicio del poblamiento hasta finales del siglo XVIII, ya que únicamente a partir de 1771 empezaron a elaborarse las primeras descripciones poblacionales, en las cuales se ponía en evidencia que entre 1776 y 1796 el índice de masculinidad presenta un valor medio ligeramente superior a 100 en las islas del grupo Occidental, un valor entre 90 y 95 en las islas del grupo Central y un valor de ~83 en las islas del grupo Oriental (Madeira, 1999).

A partir del inicio del siglo XIX, y particularmente a mediados de dicho siglo, el índice de masculinidad disminuye bastante en todos los grupos de islas (Madeira, 1999; Matos, 2004; apéndice I: figuras Ia-b, IIa-b, IIIa-b, IVa-b), presentando valores inferiores a 100 hasta ~1930 (apéndice I: figuras Ia-d, IIa-d, IIIa-d, IVa-d). Los bajos valores del índice de masculinidad son consecuencia del déficit de población masculina, producido por el movimiento migratorio, en los grupos de edad entre 15 y 40 años (apéndice I: figura V) especialmente marcada en los grupo Occidental y Central. Actualmente, el índice de masculinidad es de ~98 y la estructura de la población por edad y sexo indica que se trata de una población que ha superado recientemente la transición demográfica, y en la que se puede detectar la existencia de grupos de edad, de ambos sexos, que están menos representados de lo que se espera (particularmente en el grupo Oriental), signos de la ola de emigración que se ha iniciado en la década de los sesenta y que parece tener características diferentes de las observadas en el siglo XIX ya que la desproporción de sexos no es acentuada (apéndice I: figuras Ig, IIg, IIIg).

La reducción del índice de masculinidad observada a partir del inicio del siglo XIX puede haber tenido consecuencias muy importantes en la estructura biodemográfica de las poblaciones azoreanas, como ha puesto en evidencia el trabajo de Matos (2004) en la parroquia de Ribeira Seca de São Jorge. Dicha reducción constituye una limitación inequívoca a la posibilidad de acceder al matrimonio, y, como consecuencia, puede producir un incremento en la edad al casamiento y una disminución en la natalidad (Matos, 2004), lo que en términos genéticos representaría una disminución de la probabilidad de transmisión de los genes a las siguientes generaciones. Sin embargo, la disminución de la natalidad puede no manifestarse como lo previsto, ya que en un contexto en que

disminuye el índice de masculinidad puede haber un incremento de la ilegitimidad, o sea, del nacimiento de hijos de madres solteras, un fenómeno observado en las Azores como consecuencia de la emigración masculina (Matos, 2004), hecho que también se ha observado en otras poblaciones rurales en las que existían limitaciones al matrimonio de naturaleza económica, relacionadas con el régimen de propiedad de las tierras y las prácticas de herencia (Abade, 1992). Estudios biodemográficos anteriores son concordantes con un incremento, en la isla de Flores, a partir del 1800, del número de nacimientos, tanto de hijos de padre desconocido como de expósitos (Cruz, 2003) y el mismo resultado es observado cuando se analiza la legitimidad en el momento del óbito (Santos y Lima, 2000).

En los distintos trabajos realizados para las islas de las Azores, incluyendo los presentados en esta tesis [artículo III: *Analyse du mouvement marital à l'île de Flores (Azores, Portugal)*, y artículo IV: *Genetic Structure of Flores Island (Azores, Portugal) in the 19th Century and in the Present-day: Evidences from Surname Analysis*], se ha observado que en general los valores de consanguinidad son similares a los obtenidos tanto en Portugal continental como en otras poblaciones del sur de Europa, como ya se había mencionado anteriormente, aunque en algunas poblaciones particulares, como Rabo de Peixe (Lima, 1991; Lima, 1992/1993), se hayan encontrado valores de consanguinidad más altos. Por lo que respecta a los valores de endogamia, estos son en general altos; sin embargo, un análisis detallado por islas, como el que se ha llevado a término para Flores, puede poner de manifiesto algunas diferencias regionales no detectables en estudios aislados de algunas parroquias.

El análisis más detallado de la isla de Flores, realizado en el contexto de este trabajo (artículos III y IV) se ha visto limitado por la calidad de la información extraída de los registros parroquiales de la isla. Los trabajos de Alves y colaboradores (2000) y Alves (2000) ya habían demostrado que en algunas parroquias de Flores existían variables muy importantes, como la edad de matrimonio, la dispensa de consanguinidad y el apellido de la novia, que no eran registradas de forma sistemática en los registros matrimoniales; y lo mismo pasaba con otras variables propias de los registros de defunciones (Santos y Lima, 2000; Cruz, 2003) y de bautizos (Cruz, 2003).

Por lo que respecta a la edad del matrimonio, ésta condiciona de forma directa el tamaño del período reproductivo y, en consecuencia, el número de hijos. Para la isla de Flores no se ha podido estudiar la evolución de la edad de matrimonio a lo largo del período considerado (1675-1875) ya que la misma sólo ha sido registrada de forma más o menos sistemática a partir de mediados del siglo XIX. Entre 1850 y 1875, las edades de matrimonio, tanto en hombres (32.9 +/- 10.759) como en mujeres (27.8 +/- 9.906), son elevadas, presentando valores similares a las observadas, para un período equivalente, en otras poblaciones azoreanas (Matos, 2004) y en poblaciones del Norte de Portugal (Abade, 1992). Los elevados valores de edad de casamiento se han interpretado como el

resultado de la estructura económica y social de las poblaciones; en el caso concreto de Flores, y también de Ribeira Seca (São Jorge) (Matos, 2004), la elevada edad de casamiento puede ser consecuencia de la alta emigración de hombres que, como se ha visto anteriormente, ha tenido un gran impacto en la estructura demográfica de la población azoreana. No obstante, para llegar a conclusiones definitivas sería necesario hacer un estudio longitudinal de la edad de matrimonio en otra isla, ya que éste se ha mostrado imposible en la isla de Flores.

En cuanto a la ausencia de dispensas de consanguinidad, estas no eran una característica exclusiva de los registros parroquiales de Flores puesto que la parroquia de Ribeira Seca (São Jorge) se halla en una situación parecida. Matos (2004) ha comprobado que de las peticiones de dispensa matrimonial por parentesco de los contrayentes en Ribeira Seca, realizadas al Obispado de Angra entre 1800 y 1859, tan sólo algo más de un 10% habían sido anotados en el registro matrimonial. Así, los valores de consanguinidad, presentados en el artículo III y en otros trabajos en que se han utilizado las dispensas eclesiásticas, estarán seguramente subestimados, y sería importante hacer una estimación de los niveles de consanguinidad en el archipiélago utilizando otra metodología distinta a la utilizada hasta el momento.

Por el contrario, por lo que a los valores de endogamia/exogamia e inmigración se refiere, se ha podido hacer un análisis detallado de la isla de Flores (artículos III y IV) ya que en general, el origen de los contrayentes figuraba en los registros matrimoniales. Las parroquias del municipio de Lajes (sur de la isla) destacan por sus elevados valores de endogamia y la tendencia al incremento de los mismos a lo largo del tiempo, con la consecuente disminución de la inmigración. Por otra parte, las parroquias del municipio de Santa Cruz (norte de la isla), presentan valores medios de endogamia y una tendencia decreciente de los mismos a lo largo del tiempo, al igual que se destacan valores de inmigración más altos que los observados en el municipio de Lajes. Estos resultados demuestran la heterogeneidad entre diferentes regiones de la isla de Flores, tanto en relación a los valores de endogamia como a su evolución temporal, confirmando que el estudio de una parroquia aislada, muchas veces motivado por características sociales, geográficas, etc. particulares, no es representativo de la realidad general de una población.

En definitiva, aún reconociendo la imprecisión de las generalizaciones, la población de las Azores se presenta como una población con valores de endogamia medios-altos, que ha sobrevivido a la reducción del número de habitantes, a partir de mediados del siglo XIX, la cual no afecta de igual forma a los individuos de ambos sexos. Además, la ilegitimidad parece surgir como un mecanismo compensatorio de forma que muchas mujeres siguen teniendo oportunidad de reproducirse aunque no puedan acceder al matrimonio debido al bajo índice de masculinidad. En conjunto, estos factores pueden

llevar a que se obtengan patrones genéticos distintos dependiendo del marcador genético empleado, lo que discutiremos en el siguiente apartado.

V. 3. ESTRUCTURA GENÉTICA DE LA POBLACIÓN DE LAS AZORES

En el artículo II [*Genetic Structure and Origin of Peopling in The Azores Islands (Portugal): The View from mtDNA*] se han presentado y discutido los resultados del análisis del DNAmT de la población azoreana de forma global y por grupos de islas. El trabajo presentado en esta tesis presenta cierta originalidad y relevancia, ya que se plantea la posibilidad de que el archipiélago puede no constituir una unidad genética homogénea, posibilidad ignorada por otros trabajos aparecidos casi simultáneamente (Brehm *et al.*, 2003; Carvalho *et al.*, 2003a).

Los resultados obtenidos en el artículo II indican que, cuando se considera el archipiélago en global, el *pool* genético mitocondrial de los azoreanos presenta una elevada diversidad. Se han detectado representantes de todos los haplogrupos originarios en Eurasia Occidental y de algunos de origen africano, como se discutirá más detalladamente en el apartado V4, y los valores de los distintos indicadores de diversidad son más elevados que los observados en la mayoría de poblaciones europeas. Estos resultados son similares a los obtenidos por Brehm y colaboradores (2003) en un análisis de DNAmT de individuos de los archipiélagos de las Azores y Madeira, y son concordantes con los alcanzados utilizando marcadores del cromosoma Y (Carvalho *et al.*, 2003b; Fernandes y Brehm, 2003; Montiel *et al.*, 2005; Pacheco *et al.*, 2005; Fernando *et al.*, *in press*), STRs autosómicos (Velosa *et al.*, 2002; Silva *et al.*, *in press*) y haplotipos del sistema HLA (Bruges-Armas *et al.*, 1999). La elevada diversidad presente en el archipiélago de las Azores puede parecer sorprendente si consideramos su aislamiento geográfico; sin embargo, es fácilmente explicada por el origen reciente y diverso de su población. El efecto del aislamiento geográfico, si realmente es significativo, permanece indetectable en estos momentos.

Cuando se analiza el DNAmT por grupos de islas, puede comprobarse que, por lo menos hasta finales del siglo XIX (puesto que el origen de las muestras analizadas se refiere al origen del ancestro femenino más antiguo recordado), existía un aislamiento importante entre los tres grupos, ya que el número de linajes de DNAmT compartido entre grupos de islas es muy bajo. Por otra parte, se verifica que el archipiélago no es homogéneo genéticamente, ya que el grupo Occidental se diferencia de forma significativa respecto de los otros dos grupos de islas, resultado también obtenido para el cromosoma Y (Montiel *et al.*, 2005). Sin embargo, esta diferenciación del grupo Occidental observada con los dos sistemas, parece ser el resultado de mecanismos distintos, ya que el DNAmT y el cromosoma Y dan resultados totalmente antagónicos en lo que respecta a los valores de diversidad. La diversidad genética (Nei, 1987) del DNAmT presenta, en el grupo

Occidental, el valor más bajo cuando se consideran los haplotipos del DNAm_t encontrados (artículo II) y no presenta diferencias significativas en relación a los observados en los otros grupos de islas cuando los haplotipos de DNAm_t son agrupados en haplogrupos (Montiel *et al.*, 2005); Por el contrario, los haplogrupos del cromosoma Y, presentan, en el grupo Occidental, un valor de diversidad genética (Nei, 1987) significativamente más alto que el observado en los otros grupos de islas (Montiel *et al.*, 2005).

De acuerdo con los resultados del análisis del DNAm_t, la diferenciación del grupo Occidental no parece resultar de un poblamiento diferencial, puesto que la proporción de haplogrupos euroasiáticos y africanos es bastante similar a la observada en el grupo Central. Tampoco se puede atribuir al hecho de que estas islas presenten valores más altos de consanguinidad y/o endogamia, como se puede ver en los artículos III y IV, y ya se ha discutido en el apartado anterior. Por el contrario, sí parece que se trata del reflejo de la actuación de la deriva genética, incrementada por el reducido efectivo poblacional de mujeres de este grupo de islas desde el inicio del poblamiento y por la fuerte reducción poblacional que ha ocurrido desde mediados del siglo XIX hasta la actualidad, como se ha discutido en el artículo II. La elevada diversidad del cromosoma Y en el grupo Occidental concuerda con la gran diversidad de apellidos encontrados en la población de la isla de Flores (artículos III y IV); sin embargo, la diversidad más elevada del grupo Occidental respecto del resto del archipiélago es difícil de explicar, teniendo en cuenta su reducido efectivo poblacional. Montiel y colaboradores (2005) proponen que la composición de hombres durante el proceso de poblamiento podría ser distinta de la de los otros grupos de islas. Esta explicación se apoya en la posición estratégica que ocupaba el grupo Occidental durante el período de expansión marítima, la cual proporcionaba los intercambios tanto de orden económico como seguramente de orden personal entre los tripulantes de las embarcaciones y las isleñas. Además, a parte de las diferencias que puedan haber existido, en lo que al origen de los pobladores se refiere, y que han sido evidenciadas por los datos de Montiel y colaboradores (2005), las informaciones demográficas de finales del siglo XVIII, anteriormente discutidas, indican que existieron diferencias en la composición de sexos de los tres grupos de islas, siendo el grupo Occidental el que muestra un índice de masculinidad mayor, muy probablemente debido a una mayor entrada de hombres que de mujeres en estas islas. Esta diferencia en la proporción de hombres en el grupo Occidental y la posibilidad de que a estas islas hayan llegado hombres de origen muy distinto, podrían ser la base de la diversidad manifestada en el cromosoma Y, la cual sigue presentando los valores más elevados en la actualidad, a pesar de que puede haber sido reducida por acción de la deriva genética.

La estructura genética de las islas de Flores (artículos III y IV) y São Miguel (Branco y Mota-Vieira, 2003) ha sido evaluada utilizando apellidos y se ha puesto en evidencia que existen algunas regiones dentro de cada una de las islas que presentan elevados valores de parentesco intrapoblacional; sin embargo, en general no existe una microdiferenciación

elevada de las subpoblaciones de cada isla. Debido a ello, y a las diferencias manifestadas a través del DNAmT y del cromosoma Y, el archipiélago no puede ser considerado como una unidad mendeliana, es decir, como una población homogénea, sino que pueden existir diferentes jerarquías de subestructuración. Hasta este momento se ha podido comprobar con seguridad que existe subestructuración, como mínimo, entre grupos de islas (artículo II; Montiel *et al.*, 2005) y hay resultados de otros marcadores que hacen pensar que puede existir subestructuración entre diferentes islas e incluso dentro de cada una de ellas (Corte-Real *et al.*, 1999; Silva *et al.*, *in press*; Fernando *et al.*, *in press*). Además, la subestructuración no parece resultar de un único mecanismo, sino que puede reflejar distintos factores, dependiendo del marcador genético considerado. Las investigaciones futuras deberían, pues, centrarse en hacer una caracterización detallada de cada una de las islas haciendo uso de distintas herramientas, ya que ignorar la subestructuración poblacional puede tener consecuencias importantes en la interpretación de resultados tanto de naturaleza poblacional como forense.

Uno de los objetivos particulares del artículo IV ha sido evaluar, utilizando como herramienta los apellidos, el efecto de la reducción poblacional en la diversidad y microdiferenciación de la isla de Flores. Los resultados obtenidos, ya discutidos en el artículo IV, han permitido verificar que en el global de la isla no se observa un cambio significativo en los valores de diversidad comparando la segunda mitad del siglo XIX con la actualidad; sin embargo, en las parroquias más pequeñas hay una reducción (aún que no significativa) de los valores de diversidad y un incremento de los valores de parentesco intrapoblacional. Parece ser, que los nuevos individuos que entran en las parroquias más grandes de la isla – Santa Cruz y Lajes, aportando nuevos apellidos (y por lo tanto diversidad), han compensado la gran pérdida de individuos ocurrida desde mediados del siglo XIX. La microdiferenciación tampoco se incrementa, por el contrario, disminuye ligeramente, lo que tiene que ver con la mejora de las vías de comunicación que se ha registrado en las últimas décadas.

La fuerte reducción poblacional que ocurrió en todo el archipiélago y, particularmente en Flores, parece no haber tenido consecuencias muy significativas en la estructura genética de la población de la isla de Flores. Sin embargo, hay que tener en cuenta que los apellidos presentan algunas limitaciones metodológicas y estas conclusiones pueden estar sesgadas.

V. 4. ORIGEN DE LOS POBLADORES DE LAS ISLAS DE LAS AZORES INFERIDO EN BASE A LOS DATOS GENÉTICOS

En el artículo II de la presente tesis se ha realizado y discutido un exhaustivo análisis filogeográfico de los linajes de DNAmT encontrados en las Azores, utilizando los datos publicados hasta finales del año 2002 para las poblaciones que, según la información

histórica, podrían haber aportado pobladores a las islas de las Azores. En esos momentos se ha visto que cerca de un 50% de los linajes detectados en las Azores no se encontraban en las poblaciones utilizadas en la comparación, y una vez realizada una búsqueda en las bases de datos poblacionales de DNAm, más de un 40% de los linajes seguían siendo exclusivas de las Azores. Estos resultados se han interpretado como consecuencia de que:

- 1) Algunas poblaciones parentales, como la de Flandes y Madeira, no se han incluido en la comparación, ya que no estaban caracterizadas en esos momentos;
- 2) Los linajes pueden estar en las poblaciones utilizadas en la comparación aunque todavía no se han detectado;

La importancia de estos factores se ha podido comprobar con el análisis posterior del DNAm de la población del archipiélago de Madeira (Brehm *et al.*, 2003) y el incremento del número de muestras analizadas para la población de Portugal continental (González *et al.*, 2003). Algunos de los linajes que hemos clasificado como exclusivos de las Azores (artículo II) también están en Madeira y/o en Portugal continental de forma que el porcentaje de linajes no compartidos ha bajado a 35%.

El análisis filogeográfico ha permitido verificar que la gran mayoría de los linajes de DNAm encontrados en las Azores tienen su origen en poblaciones europeas, principalmente en Portugal continental; además, se han podido detectar linajes típicos de poblaciones africanas y de poblaciones del Próximo Oriente, resultados concordantes con los obtenidos por Brehm y colaboradores (2003). El análisis por grupo de islas ha demostrado que en los tres grupos el *pool* genético mitocondrial tiene mayoritariamente un origen en poblaciones europeas, pero la proporción de linajes africanos y del Próximo Oriente difiere entre grupos de islas, siendo los linajes africanos más frecuentes en el grupo Oriental, mientras que los del Próximo Oriente son más frecuentes en el grupo Central.

Del análisis filogeográfico, ya ampliamente discutido en el artículo II, destaca la ausencia de linajes L1, L2 y L3 en el grupo Occidental. Según los datos históricos, para el poblamiento de estas islas han contribuido en gran medida *esclavos negros* (Gomes, 1997), por lo que sería de esperar que se encontrasen linajes típicos de poblaciones del oeste africano. La ausencia de estos linajes de DNAm en las islas del grupo Occidental podría ser explicada por distintas razones:

- 1) Los *esclavos negros* a los que se refieren los documentos históricos serían mayoritariamente hombres;
- 2) Las mujeres *esclavas* no han tenido posibilidad de reproducirse o la viabilidad de sus descendientes es menor;
- 3) Los linajes africanos se han perdido por actuación de la deriva genética;

Para investigar las dos primeras razones se ha realizado un estudio preliminar de la presencia y reproducción de esclavos en la isla de Flores (Lima *et al.*, 2003). Se ha podido verificar que en el siglo XVIII un ~11% de los habitantes de Santa Cruz (Flores) son esclavos, y que la proporción de esclavos hombres y mujeres es semejante. Además, hay descendientes de parejas en que por lo menos uno de los padres es esclavo, siendo la casi totalidad de ellos hijos de madre esclava y de padre desconocido. Estos resultados indican que el hecho de que no se encuentren linajes de DNAMt L1, L2 y L3 en el grupo Occidental no significa que no hubo mujeres esclavas en este grupo de islas, sino que sus linajes se han perdido, posiblemente por efecto de la deriva, o simplemente aún no se han detectado.

Ahora bien, los resultados obtenidos en el artículo II, conjuntamente con los aportados por el análisis del cromosoma Y (Montiel *et al.*, 2005; Pacheco *et al.*, 2005; Fernando *et al.*, in press) y de marcadores autosómicos (Bruges-Armas *et al.*, 1999; Velosa *et al.*, 2002; Silva *et al.*, in press) permiten deducir que la mayor parte del *pool* genético de los azoreanos deriva de poblaciones europeas; sin embargo, se han detectado variantes genéticas características de población africana, del Próximo Oriente y asiática.

La aportación africana, detectada con los distintos marcadores utilizados, seguramente representa un signo de la presencia de esclavos en las islas. Además, contrastando los resultados de DNAMt y cromosoma Y, se puede concluir que en las Azores los esclavos tanto varones como mujeres han podido dejar descendencia, aunque en distinta proporción, al revés de lo que ha ocurrido en Portugal continental, donde sólo se han detectado vestigios de linajes africanos de origen materno (Pereira *et al.*, 2000; González *et al.*, 2003). En lo que respecta a las variantes características de poblaciones del Próximo Oriente, probablemente éstas constituyen marcas de la presencia de judíos sefarditas en las islas. La presencia en las islas de las Azores de esclavos y judíos está documentada históricamente, y la presencia en la población actual de genes derivados de estos grupos indica que se han podido reproducir, de forma que sus genes se han mantenido hasta la actualidad. La presencia de variantes asiáticas detectadas tanto con el análisis del cromosoma Y (Pacheco *et al.*, 2005) como con el análisis de moléculas de clase I y II del sistema HLA (Bruges-Armas *et al.*, 1999), es difícil de explicar, teniendo en cuenta las fuentes históricas que narran el poblamiento; la opción más probable es que estas variantes hayan sido introducidas en las Azores durante el período de expansión marítima en que el archipiélago tenía una posición estratégica en las rutas marítimas.

Finalmente, hay que destacar los resultados obtenidos en el artículo IV, que ponen en evidencia cambios temporales en la composición de apellidos, y, por lo tanto, en la composición genética de la población de la isla de Flores. Se ha verificado que un número importante de apellidos se han perdido desde el siglo XIX hasta la actualidad y, por otra parte, han sido introducidos nuevos apellidos en la población; sin embargo, algunos de los

apellidos de los primeros pobladores de la isla de Flores siguen estando representados en la población actual de la isla. Estos resultados indican que la caracterización genética de la población actual del archipiélago, que ha sido influenciada por procesos complejos de inmigración y emigración, permite inferir el origen de sus primeros pobladores. Sin embargo, hay que tener siempre presente la dinámica de la población, de forma que la ausencia de una determinada variante no implica que ésta no haya estado presente en el pasado.

V. 5. CARACTERÍSTICAS DEL DNA MITOCONDRIAL

El último de los objetivos generales de la tesis es el que recoge el artículo V [*Understanding Differences Between Phylogenetic and Pedigree Derived mtDNA Mutation Rate: A Model Using Families From the Azores Islands (Portugal)*] en el que se presentan y discuten los resultados del estudio de las características inherentes al DNAm: transmisión materna, ausencia de recombinación, homoplasmia y tasa de mutación. Este estudio tiene por objetivo principal el cálculo empírico de la tasa de mutación del DNAm en familias de origen azoreano tomadas al azar de la población, y cuyos miembros no padecen de ninguna enfermedad mitocondrial. Se trata de un trabajo novedoso, ya que para un gran número de familias se analizan los cambios en el DNAm durante varias generaciones, controlando las posibilidades de contribución de DNAm de origen paterno y de recombinación.

Todas las nuevas mutaciones detectadas están presentes en heteroplasmia, existiendo por lo menos un individuo con heteroplasmia de longitud y/o puntual en 73% de las familias analizadas, lo que indica que la heteroplasmia no puede ser considerada como una condición excepcional, como se aceptaba no hace mucho, sino, como evidencian nuestros resultados y los de otros trabajos recientes (Tully *et al.*, 2000; Grzybowski *et al.*, 2003; Brandstätter *et al.*, 2004; entre otros), un fenómeno bastante frecuente.

El análisis de las distintas heteroplasmas encontradas a lo largo de varias generaciones ha permitido comprobar que la resolución de la heteroplasmia a homoplasmia es heterogénea, corroborando los resultados obtenidos en los estudios de heteroplasmia en familias (Howell *et al.*, 1992; Bendall y Sykes, 1995; Bendall *et al.*, 1996; Howell *et al.*, 1996; Parsons *et al.*, 1997; Lutz *et al.*, 2000; Howell *et al.*, 2003; Brandstätter *et al.*, 2004). Según los análisis realizados, la velocidad a que la heteroplasmia evoluciona hacia la homoplasmia es controlada por la deriva genética y depende en gran medida de la frecuencia inicial de la nueva variante, siendo el momento en que ocurre la nueva mutación (antes, durante o después del *bottleneck*) determinante para su destino final; sin embargo, en algunos casos la selección puede jugar un papel importante en el destino de la nueva variante.

En 10 de las 48 familias fueron detectadas 11 nuevas mutaciones, todas presentes en heteroplasmia y producidas por mutaciones puntuales, aunque, en la mitad de los casos, la nueva mutación no puede considerarse germinal, ya que sólo se ha hallado en un individuo de la familia, siendo muy probablemente somática. Las mutaciones somáticas en el DNAm_t se han asociado al proceso de envejecimiento, a enfermedades degenerativas, al cáncer y a la autoinmunidad (Wallace, 1999); sin embargo, los individuos en que se han hallado las mencionadas mutaciones son jóvenes y aparentemente sanos, de forma que nuestros resultados son concordantes con la propuesta de que una gran parte de las mutaciones somáticas tienen un origen bastante temprano, muchas de ellas durante el desarrollo embrionario (Khrapkoo *et al.*, 2003, 2004).

La tasa de mutación del DNAm_t estimada en diferentes trabajos ha puesto en evidencia una discrepancia muy importante entre las estimaciones empíricas y filogenéticas. En el artículo V se ha realizado la estimación empírica de la tasa de mutación en familias de origen azoreano, utilizando diferentes aproximaciones. Siguiendo la propuesta de Howell y colaboradores (2003), se ha calculado la tasa de mutación teniendo en cuenta las nuevas mutaciones germinales, incluyendo aquellas que están presentes en heteroplasmia, lo que ha conducido a la obtención de un valor que se sitúa en el rango superior de los valores de tasa de mutación obtenidos en estudios de familias (ver revisión de Howell *et al.*, 2003). Estos resultados vienen a comprobar que la discrepancia antes reportada, para las estimaciones empíricas y filogenéticas de la tasa de mutación del DNAm_t, no puede ser atribuida al hecho de que algunos de los estudios anteriores se hayan realizado en familias con enfermedades mitocondriales (Howell *et al.*, 1996; Howell *et al.*, 2003), ni a que resulten de la contribución paterna de DNAm_t, como ha sido sugerido por algunos autores (Hagelberg, 2003), ya que los resultados presentados en el artículo V se refieren a familias presuntamente sanas en las que se ha considerado la posibilidad de contribución paterna.

A partir de los resultados obtenidos se ha calculado otro valor de la tasa de mutación basado en dos supuestos:

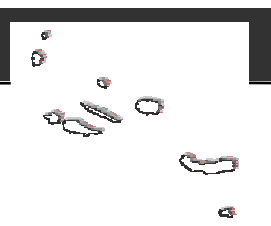
- 1) Las nuevas mutaciones presentes exclusivamente en individuos del sexo masculino tienen el mismo papel evolutivo que las mutaciones somáticas ya que también se pierden;
- 2) Si una mutación está en heteroplasmia tendrá una determinada probabilidad de fijarse a nivel individual que, considerando la mutación neutra, es igual a la frecuencia de la misma en los individuos heteroplásmicos.

Teniendo en cuenta los presupuestos anteriores, el valor de la tasa de mutación para el D-loop sería 4.1878×10^{-6} mutaciones/pb/generación, un valor que está en el rango superior de los valores obtenidos en las estimaciones filogenéticas. Esto permite demostrar que la discrepancia en la tasa de mutación observada por otros autores entre

los estudios familiares y filogenéticos puede ser explicada, en parte, por el hecho de que las nuevas mutaciones, cuando se hallan presentes en individuos del sexo masculino, tienen el mismo peso que las mutaciones somáticas, y por lo tanto no deben ser utilizadas en el cálculo de la tasa de mutación. Por otra parte, una vez se detecta una nueva mutación que esté en heteroplasma, esta tiene una probabilidad X de acabar fijándose a nivel individual, y suponiendo la mutación neutra, la probabilidad se corresponde con su frecuencia inicial.

Finalmente, en un par madre-hijo se han encontrado dos posiciones (16086 y 16182) que mostraban signos de heteroplasma y al secuenciar múltiples clones se han detectado cuatro tipos de DNAm (16086T-16182A, 16086T-16182C/Del A, 16086C-16182A y 16086C-16182C/Del A) presentes en ambos individuos. En ausencia de recombinación, este patrón sólo podría ser explicado por múltiples eventos mutacionales, de forma que hay que tener en cuenta la posibilidad de que resulte de recombinación entre moléculas de DNAm heterólogas. Esto pone en evidencia la posibilidad de que el DNAm recombine aunque no haya contribución paterna del mismo, un hecho que merece futura investigación.

VI. CONCLUSIONES



VI. CONCLUSIONES

Metodología jerárquica de análisis del DNAm

1. El estudio del DNAm de la población azoreana, utilizando la secuenciación de la región HVRI y la determinación de 18 polimorfismos puntuales de la región codificante, ha permitido desarrollar una metodología jerárquica que posibilita la clasificación en haplogrupos de los DNAs mitocondriales de poblaciones caucasoides en las que puede haber distintos grados de aportaciones de DNAm africano y/o asiático. La validez de dicha metodología ha sido comprobada, corroborando los resultados del análisis de polimorfismos de la región codificante con los de la secuenciación de la región HVRI de todas las muestras. Además, se puede asegurar, ya que las posiciones polimórficas utilizadas son estables, para definir cada haplogrupo son testadas como mínimo dos posiciones, y, en algunos casos, es sugerido el análisis de posiciones confirmatorias adicionales.

Estructura biodemográfica de la población de las Azores

2. La población de las Azores se presenta como una población con valores de endogamia medios-altos, y valores de consanguinidad similares a los obtenidos para Portugal continental y en otras poblaciones del sur de Europa.

3. La población azoreana ha sido muy influenciada por la migración y ha sobrevivido a una clara reducción del número de habitantes a partir de mediados del siglo XIX, que afecta particularmente a los individuos de sexo masculino. Además, la ilegitimidad surgió como un mecanismo compensatorio de la baja razón de masculinidad, de forma que muchas mujeres, aunque no tuvieron acceso al matrimonio, siguieron teniendo oportunidad de reproducirse.

4. El análisis biodemográfico más detallado de la isla de Flores se ha visto limitado por la calidad de la información extraída de los registros parroquiales de la isla; sin embargo, el análisis de la endogamia/exogamia e inmigración demuestran la heterogeneidad entre diferentes regiones de la isla de Flores, tanto en relación a sus

valores como a su evolución temporal, confirmando que el estudio de una parroquia aislada, muchas veces motivado por características sociales, geográficas, etc. particulares, no es representativo de la realidad general de una población.

Estructura genética de la población de las Azores

5. Cuando se considera globalmente el archipiélago, el *pool* genético de los azoreanos presenta una elevada diversidad. Dicha diversidad puede parecer sorprendente considerando su aislamiento geográfico; sin embargo, es fácilmente explicada por el origen reciente y muy diverso de su población. El efecto del aislamiento geográfico, si realmente es significativo, puede aún permanecer indetectable en estos momentos.

6. Cuando se analiza el DNAm_t por grupos de islas, puede comprobarse que por lo menos hasta finales del siglo XIX, existía un aislamiento importante entre los tres grupos, ya que el número de linajes de DNAm_t compartidos entre ellos es muy bajo.

7. El análisis del DNAm_t por grupos de islas ha demostrado que el grupo Occidental se diferencia de forma significativa respecto de los otros dos grupos de islas. Dicha diferenciación no parece resultar de un poblamiento diferencial y ni tampoco se puede atribuir al hecho de que estas islas presenten valores más altos de consanguinidad y/o endogamia. Por el contrario, sí parece ser el reflejo de la actuación de la deriva genética incrementada por el reducido efectivo poblacional de mujeres de este grupo de islas desde el inicio del poblamiento.

Efecto de la reducción poblacional en la diversidad y microdiferenciación de la isla de Flores

8. La fuerte reducción poblacional que ocurrió en todo el archipiélago y particularmente en Flores, no parece haber tenido consecuencias muy significativas en la estructura genética de la población de la isla. En el global de la isla de Flores no se observa un cambio significativo en los valores de diversidad, comparando la segunda mitad del siglo XIX con la actualidad, y la microdiferenciación tampoco se incrementa. Sin embargo, en las parroquias más pequeñas hay una reducción (aunque no significativa) de los valores de diversidad y un incremento de los valores de parentesco intrapoblacional. Los nuevos individuos que entran en las parroquias más grandes de la isla, aportando nuevos apellidos (y por lo tanto diversidad), parecen haber compensado la gran pérdida de individuos a la que se ha asistido desde mediados del siglo XIX.

Origen de los pobladores de las islas de las Azores inferido en base a los datos genéticos

9. El análisis de apellidos de la isla de Flores indica que, a pesar de que la población actual del archipiélago de las Azores haya sido influenciada por procesos complejos de inmigración y emigración, la caracterización genética de la misma sigue siendo útil para inferir el origen de sus primeros pobladores, aunque hay que tener siempre presente la dinámica de la población, de forma que la ausencia de una determinada variante no implica que ésta no haya estado presente en el pasado.

10. El análisis filogeográfico del DNAMt ha permitido verificar que la gran mayoría de los linajes de DNAMt encontrados en las Azores tienen su origen en poblaciones europeas, principalmente en Portugal continental. Además, se han podido detectar linajes típicos de poblaciones africanas y de poblaciones del Próximo Oriente. El análisis por grupo de islas ha demostrado que el *pool* genético mitocondrial de los tres grupos de islas tiene un origen mayoritario en poblaciones europeas; sin embargo, la proporción de linajes africanos y del Próximo Oriente presenta alguna variación entre grupos, siendo los linajes africanos más frecuentes en el grupo Oriental y los del Próximo Oriente más frecuentes en el grupo Central. La aportación africana seguramente representa un signo de la presencia de esclavos en las islas, mientras que las variantes características de poblaciones del Próximo Oriente probablemente constituyen marcas de la presencia de judíos sefarditas en las islas.

Inferencias sobre las características del DNA mitocondrial

11. Existe la posibilidad de que el DNAMt recombine aunque no haya contribución paterna del mismo. Por otra parte, la heteroplasmia no puede ser considerada como una condición excepcional, como se aceptaba no hace mucho tiempo, siendo un fenómeno bastante frecuente; su resolución hacia la homoplasmia es heterogénea y la velocidad a que ocurre es controlada por la deriva genética, dependiendo en gran medida de la frecuencia inicial de la nueva variante; sin embargo, en algunos casos, la selección puede jugar un papel importante en el destino de la nueva variante.

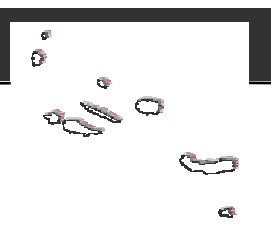
12. La discrepancia en la tasa de mutación observada por otros autores entre los estudios familiares y filogenéticos puede ser explicada, en parte, por el hecho de que, las nuevas mutaciones, cuando presentes en individuos del sexo masculino, tienen el mismo peso que las mutaciones somáticas, no debiendo, por tanto, ser utilizadas en el cálculo de la tasa de mutación. Además, una vez se detecta una nueva mutación que se halle en heteroplasmia, ésta tiene una probabilidad X de acabar fijándose a nivel individual, y

Conclusiones

suponiendo la mutación neutra, dicha probabilidad corresponde a su frecuencia inicial, ya que, teniendo en cuenta los presupuestos anteriores, el valor de la tasa de mutación para el D-loop es de 4.1878×10^{-6} mutaciones/pb/generación, un valor que está en el rango superior de los valores obtenidos en las estimaciones filogenéticas.

VII.

BIBLIOGRAFÍA



VII. BIBLIOGRAFÍA

- Abade, A. (1986) Inbreeding as measurement by dispensations and isonymy in Rio de Onor, Portugal. *Inter J Anthropol*, 1, 225-228.
- Abade, A. (1992) *A população inexistente: Estrutura demográfica e genética da população da Lombada, Bragança*. Tesis Doctoral, Universidad de Coimbra, Coimbra, Portugal.
- Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V. *et al.* (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet*, 75, 910-918.
- Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., Magri, C., Scozzari, R., Babudri, N., Santachiara-Benerecetti, A. S., *et al.* (2005) Saami and Berbers – an unexpected mitochondrial DNA link. *Am J Hum Genet*, 76, 000-000 (electronic publication).
- Adams, K.L., Daley, D.O., Qiu, Y.L., Whelan, J. y Palmer, J.D. (2000) Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature*, 408, 354-357.
- Allen, J.F. (2003) Why chloroplasts and mitochondria contain genomes? *Com Func Genom*, 4, 31-36.
- Alves, C. (2000) *Estudo Biodemográfico e Genético de Três Populações de Ilha das Flores (1760-1860)*. Trabajo realizado para obtención de la Licenciatura en Biología, Universidad de las Azores, Ponta Delgada, Portugal.
- Alves, C., Lima, M., Silva, C., Santos, C., Mayer, F.M., Labelle, E. y Smith, M.T. (2000) Biodemographic Profile of an Isolated Population: Framing of a Genealogical Database and Preliminary Results for Flores Island (Azores, Portugal). In *Investigaciones en Biodiversidad Humana*, T. Varela (Ed.), Universidad de Santiago de Compostela / Sociedad Española de Antropología Biológica, Universidad de Santiago de Compostela, España, 25-27.
- Alves-Silva, J., da Silva Santos, M., Guimaraes, P.E., Ferreira, A.C., Bandelt, H.-J., Pena, S.D. y Prado, V.F. (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet*, 67, 444-461.
- Amorim, A., Faria, M. y Pereira, E. (1979) Red Cell Acid Phosphatase (ACP1) Polymorphism in Flores Island (Azores). *Trabalhos do Instituto de Antropologia "Dr. Mendes Corrêa"*, 35, 1-4.
- Amorim, N. y Lima, L. (1986) Estratégias de sobrevivência em uma paróquia rural açoriana. Perspectiva demográfica (1711-1980). *Boletim do Instituto Histórico da Ilha Terceira*, XLIV.

Bibliografia

- Amorim, N. (1987) Evolucao da populacao e fecundidade legitima numa paroquia acoriana (1700-1980). **Boletim de la Asociacion de Demografia Historica**, 5 (1), 4-28.
- Amorim, N. (1988a) Uma perspectiva de S. Mateus do Pico através dos falecidos no ano de 1760. **Boletim do Instituto Histórico da Ilha Terceira**, XLV.
- Amorim, N. (1988b) Mortalidade e emigração irreversível das gerações nascidas em S. João do Pico entre 1845 e 1884. **Boletim do Instituto Histórico da Ilha Terceira**, XLV.
- Amorim, N. (1992) **Evolução da População em Três Paróquias do Sul do Pico (1680-1980)**. Instituto de Ciências Sociais da Universidade do Minho, Braga.
- Amorim, N. (1993) Nupcialidade e fecundidade diferenciadas. Evolucao de comportamentos nos ultimos tres seculos. O caso do Sul do Pico (Acores). **Boletim de la Asociacion de Demografia Historica**, 11 (2), 55-73.
- Amorim, N. (2003) Maternidade ilegítima e Família. O caso de S. João do Pico em finais do século XIX. **Cadernos do Noroeste**, Série História 3, 523- 553.
- Amorim, N. (2004) **As famílias de S. João nos finais do século XIX**. NEPS e Câmara Municipal das Lajes do Pico, Portugal.
- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F. *et al.* (1981) Sequence and organization of the human mitochondrial genome. **Nature**, 290, 457-465.
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. y Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. **Nat Genet**, 23, 147.
- Anglés, N., Montiel, R., Nogués, R., Backes, C.H., Larrouy, G., Cherkaoui, M., Malgosa, A. y Aluja, M^a P. (2003) Haplogrupos mitocondriales en tres poblaciones caucasoides. In **Antropología y Biodiversidad**, M.P. Aluja, A. Malgosa, y R.M. Nogués (Eds.) Barcelona, España: Ediciones Bellaterra, v. 2, 26–35.
- Anjos, M.J., Carvalho, M., Andrade, L., Corte-Real, F., Vieira, D.N. y Vide, M.C. (2000) Allele frequencies of STR multiplex systems in two Portuguese population samples. **Prog Foren Genet**, 8, 208-211.
- Ankel-Simons, F. y Cummins, J. (1996) Misconceptions about mitochondria and mammalian fertilization: Implications for theories on human evolution. **Proc Natl Acad Sci USA**, 93, 13859-13863.
- Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. y Saunders, N.C. (1987) Intraspecific Phylogeography - the Mitochondrial-DNA Bridge between Population-Genetics and Systematics. **Annu Rev Ecol Syst**, 18, 489-522.
- Awadalla, P., Eyre-Walker, A. y Smith, J.M. (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. **Science**, 286, 2524-2525.
- Ballinger, S.W., Schurr, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K.H. y Wallace, D.C. (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. **Genetics**, 130, 139-152.

Bibliografía

- Bandelt H.-J., Forster P., Sykes B.C. y Richards M.B. (1995) Mitochondrial portraits of human populations using median networks. **Genetics**, 141, 743–753.
- Bandelt, H.-J., Macaulay, V. y Richards, M. (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. **Mol Phylogenet Evol**, 16, 8-28.
- Bandelt, H.-J., Herrnstadt, C., Yao, Y.G., Kong, Q.P., Kivisild, T., Rengo, C., Scozzari, R., Richards, M., Villems, R., Macaulay, V. *et al.* (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. **Ann Hum Genet**, 67, 512-524.
- Barbujani, G., Magagni, A., Minch, E. y Cavalli-Sforza, L.L. (1997) An apportionment of human DNA diversity. **Proc Natl Acad Sci USA**, 94: 4516-4519.
- Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M., Dey, B., Roy, M., Roy, B., Bhattacharyya, N.P. *et al.* (2003) Ethnic India: A genomic view, with special reference to peopling and structure. **Genome Res**, 13, 2277-2290.
- Batista dos Santos, S.E., Rodrigues, J.D., Ribeiro-dos-Santos, A.K. y Zago, M.A. (1999) Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y-DNA. **Am J Phys Anthropol**, 109, 175-80.
- Bendall, K.E. y Sykes, B.C. (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. **Am J Hum Genet**, 57, 248-256.
- Bendall, K.E., Macaulay, V.A., Baker, J.R. y Sykes, B.C. (1996) Heteroplasmic point mutations in the human mtDNA control region. **Am J Hum Genet**, 59, 1276-1287.
- Bertorelle, G., Calafell, F., Francalacci, P., Bertranpetit, J. y Barbujani, G. (1996) Geographic homogeneity and non-equilibrium patterns of mtDNA sequences in Tuscany, Italy. **Hum Genet**, 98, 145-150.
- Bertranpetit, J., Sala, J., Calafell, F., Underhill, P.A., Moral, P. y Comas, D. (1995) Human mitochondrial DNA variation and the origin of Basques. **Ann Hum Genet**, 59, 63-81.
- Birky, C.W.Jr. (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. **Proc Natl Acad Sci USA**, 92, 11331-11338.
- Birky, C.W.Jr. (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. **Ann Rev Genet**, 35, 125-148.
- Bittles, A.H. (1998) **Empirical estimates of the global prevalence of consanguineous marriage in contemporary societies**. Morrison Institute for Population and Resource studies, Stanford University, California, USA.
- Bogenhagen, D.F. (1999) Repair of mtDNA in vertebrates. **Am J Hum Genet**, 64, 1276-1281.
- Boore, J.L. (1999) Animal mitochondrial genomes. **Nucl Acids Res**, 27, 1767-1780.

Bibliografía

- Bortolini, M.C., Salzano, F.M., Zago, M.A., Da Silva Junior, W.A. y Weimer Tde, A. (1997) Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. *Am J Phys Anthropol*, 103, 147-56.
- Boyce, A.J., Kucheman, C.F. y Harrison, G.A. (1971) Population structure and movement patterns. In *Biological aspects of demography*, W. Brass (Ed.). Taylor & Francis Ltd, London, pp. 1-9.
- 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 Sous area. *Ann Hum Biol*, 28, 295-307.
- Branco, C. y Mota-Vieira, L. (2003) Population structure of São Miguel Island, Azores: a surname study. *Hum Biol*, 75, 929-939.
- Brandstätter, A., Parsons, T.J. y Parson, W. (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med*, 117, 291-298.
- Brandstätter, A., Niederstatter, H. y Parson, W. (2004) Monitoring the inheritance of heteroplasmy by computer-assisted detection of mixed basecalls in the entire human mitochondrial DNA control region. *Int J Legal Med*, 118, 47-54.
- Brehm, A., Pereira, L., Kivisild, T. y Amorim, A. (2003) Mitochondrial portraits of the Madeira and Açores archipelagos witness different genetic pools of its settlers. *Hum Genet*, 114, 77-86.
- Brown, M.D., Hosseini, S.H., Torroni, A., Bandelt, H.-J., Allen, J.C., Schurr, T.G., Scozzari, R., Cruciani, F. y Wallace, D.C. (1998) mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *Am J Hum Genet*, 63, 1852-1861.
- Brown, W.M., George, M., Jr. y Wilson, A.C. (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA*, 76, 1967-1971.
- Brown, W.M. (1980). Polymorphism in mitochondrial DNA of human as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci USA*, 77, 3605-3609.
- Bruges-Armas, J., Martinez-Laso, J., Martins, B., Allende, L., Gomez-Casado, E., Longas, J., Varela, P., Castro, M.J. y Arnaiz-Villena, A. (1999) HLA in the Azores Archipelago: possible presence of Mongoloid genes. *Tissue Antigens*, 54, 349-359.
- Bulayeva, K., Jorde, L.B., Ostler, C., Watkins, S., Bulayev, O. y Harpending, H. (2003) Genetics and population history of Caucasus populations. *Hum Biol*, 75, 837-853.
- Calderon, R. (1983) Inbreeding, migration and age at marriage in rural Toledo, Spain. *J Biol Sci*, 15, 47-57.
- Cambien, F., Poirier, O., Nicaud, V., Herrmann, S.M., Mallet, C., Ricard, S., Behague, I., Hallet, V., Blanc, H., Loukaci, V. *et al.* (1999) Sequence diversity in 36 candidate genes for cardiovascular disorders. *Am J Hum Genet*, 65, 183-191.
- Cann, R., Stoneking, M. y Wilson, A. (1987) Mitochondrial DNA and human evolution. *Nature*, 325, 31-36.

Bibliografía

- Cannings, C. y Cavalli-Sforza, L.L. (1973) Human population structure. *Adv Hum Genet*, 4, 105-171.
- Cantatore, P. y Attardi, G. (1980) Mapping of nascent light and heavy strand transcripts on the physical map of HeLa cell mitochondrial DNA. *Nucl Acids Res*, 8, 2605-2625.
- Caramelli, D., Lalueza-Fox, C., Vernesi, C., Lari, M., Casoli, A., Mallegni, F., Chiarelli, B., Dupanloup, I., Bertranpetit, J., Barbujani, G. et al. (2003) Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc Natl Acad Sci USA*, 100, 6593-6597.
- Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N. et al. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet*, 22, 231-238.
- Carmelli, D. y Cavalli-Sforza, L.L. (1976) Some models of population structure and evolution. *Theor Popul Biol*, 9, 329-359.
- Carvalho, M., Mendes, C., Antunes, H., Anjos, M.J., Andrade, L., Lopes, V., Vieira D.N. y Vide, M.C. (2003a) mtDNA analysis in Portuguese populations (Central Portugal and Azores Islands): polymorphic sites in control region sequences. *International Congress Series*, 1239, 535-539.
- Carvalho, M., Anjos, M.J., Andrade, L., Lopes, V., Santos, M.V., Gamero, J.J., Corte Real, F. y Vide, M.C. (2003b) Y-chromosome STR haplotypes in two population samples: Azores Islands and Central Portugal. *Forensic Sci Int*, 134, 29-35.
- Cavalli-Sforza, L.L. (1959) Some data on the genetic structure of human population. *Proceedings of the X International Congress in Genetics*, 1, 389-407.
- Cavalli-Sforza, L.L. (1973) Some current problems in human population genetics. *Am J Hum Genet*, 25, 82-104.
- Cavalli-Sforza L.L. y Feldman M.W. (1981) *Cultural transmission and evolution: a quantitative approach*. Princeton University Press, Princeton.
- Cavalli-Sforza, L.L., Menozzi, P. y Piazza, A. (1994) *The history and geography of human genes*. Princeton University Press, Princeton, New Jersey, USA.
- Cavalli-Sforza, L.L. y Feldman, M.W. (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet*, 33 Suppl, 266-275.
- Chang, D.D. y Clayton, D.A. (1984) Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell*, 36, 635-643.
- Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S. y Wallace, D.C. (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet*, 57, 133-149.
- Chen, Y.S., Olckers, A., Schurr, T.G., Kogelnik, A.M., Huoponen, K. y Wallace, D.C. (2000) mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am J Hum Genet*, 66, 1362-1383.

Bibliografía

- Chiarelli, B. y Rabino Massa, E. (2003) Variabilité de l' Homme actuel: Introduction. In **Anthropologie Biologique**, C. Susanne, E. Rebato y B. Chiarelli (Eds.). De Boeck & Larcier s.a., Bruxelles, pp. 425-428.
- Comas, D.; Pääbo, S. y Bertranpetit, J. (1995) Heteroplasmy in the control region of human mitochondrial DNA. **Genome Res**, 5, 89-90.
- Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E. y Bertranpetit, J. (1997) Mitochondrial DNA variation and the origin of the Europeans. **Hum Genet**, 99, 443-449.
- Cooper, A., Drummond, A.J. y Willerslev, E. (2004) Ancient DNA: would the real Neandertal please stand up? **Curr Biol**, 14, R431-433.
- Corte-Real, F., Souto, L., Anjos, M.J., Carvalho, M., Vieira, D.N., Carracedo, A. y Vide, M.C. (1999) Population study of HUMTH01, HUMVWA31/A, HUMF13A1, and HUMFES/FPS systems in Azores. **J Forensic Sci**, 44, 1261-1264.
- Côrte-Real, H.B., Macaulay, V.A., Richards, M.B., Hariti, G., Issad, M.S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. y Sykes, B.C. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. **Ann Hum Genet**, 60, 331-350.
- Cruz, A.P., Silva, C., Santos, C. y Lima, M. (2002) Biodemographic profile of an isolated population – marriage patterns in the parish of Ponta Delgada (Flores Island, Azores, Portugal). **Coll Antrop**, 26 (suplemento), 46.
- Cruz, A.P. (2003) **Dinâmica Populacional da Ilha das Flores, Açores: Estudo Biodemográfico da população de Ponta Delgada, Flores (1675-1880)**. Trabajo realizado para la obtención de la Licenciatura en Biología, Universidad de las Azores, Ponta Delgada, Portugal.
- Cunha, E. (1987) Consanguinity of the Azores Islands (1979-1985). **Antropologia Portuguesa**, 4/5, 67-79.
- Cunha, E., Abade, A. y Smith, M. (1989) Biodemography of Flores Island (Azores). In **Actas/Proceedings del VI Congreso Español de Antropología Biológica**, E. Rebato y R. Calderón (Eds.). Universidad del País Vasco, Bilbao, pp. 100-107.
- Curat, M. y Excoffier, L. (2004) Modern Humans Did Not Admix with Neanderthals during Their Range Expansion into Europe. **PLoS Biol**, 2, e421.
- Danan, C., Sternberg, D., Van Steirteghem, A., Cazeneuve, C., Duquesnoy, P., Besmond, C., Goossens, M., Lissens, W. y Amselem, S. (1999) Evaluation of parental mitochondrial inheritance in neonates born after intracytoplasmic sperm injection. **Am J Hum Genet**, 65, 463-473.
- Deka, R., Shriver, M. D., Yu, L. M., Ferrell, R. E. y Chakraborty, R. (1995) Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. **Electrophoresis**, 16: 1659-1664.
- Denaro, M., Blanc, H., Johnson, M.J., Chen, K.H., Wilmsen, E., Cavalli-Sforza, L.L. y Wallace, D.C. (1981) Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA. **Proc Natl Acad Sci USA**, 78, 5768-5772.

Bibliografía

- Derbeneva, O.A., Sukernik, R.I., Volodko, N.V., Hosseini, S.H., Lott, M.T. y Wallace, D.C. (2002) Analysis of mitochondrial DNA diversity in the aleuts of the commander islands and its implications for the genetic history of beringia. *Am J Hum Genet*, 71, 415-421.
- Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M. y Freimer, N. B. (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA*, 9, 3166-3170.
- Dubut, V., Chollet, L., Murail, P., Cartault, F., Beraud-Colomb, E., Serre, M. y Mogentale-Profizi, N. (2004) mtDNA polymorphisms in five French groups: importance of regional sampling. *Eur J Hum Genet*, 12, 293-300.
- Dyke (1984) Migration and the structure of small populations. In *Migration and mobility – biosocial aspects of human movement*. A. J. Boyce (Ed.). Taylor & Francis Ltd, London, pp. 69-81.
- Elson, J.L., Andrews, R.M., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. y Howell, N. (2001) Analysis of European mtDNAs for recombination. *Am J Hum Genet*, 68, 145-153.
- Elson, J.L., Turnbull, D.M. y Howell, N. (2004) Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. *Am J Hum Genet*, 74, 229–238.
- Endicott, P., Gilbert, M.T., Stringer, C., Lalueza-Fox, C., Willerslev, E., Hansen, A.J. y Cooper, A. (2003) The genetic origins of the Andaman Islanders. *Am J Hum Genet*, 72, 178-84.
- Eshleman, J.A., Malhi, R.S., Johnson, J.R., Kaestle, F.A., Lorenz, J. y Smith, D.G. (2004) Mitochondrial DNA and prehistoric settlements: native migrations on the western edge of North America. *Hum Biol*, 76, 55-75.
- Excoffier, L. y Yang, Z. (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Mol Biol Evol*, 16, 1357–1368.
- Eyre-Walker, A., Smith, N.H. y Smith, J.M. (1999) How clonal are human mitochondria? *Proc R Soc Lond B Biol Sci*, 266, 477-483.
- Fernandes, A. y Brehm, A. (2003) Y-chromosome STR haplotypes in the Açores Archipelago (Portugal). *Forensic Sci Int*, 135, 239-242.
- Fernando, O., Mota, P., Lima, M., Silva, C., Montiel, R., Amorim, A., Prata, M.J. The peopling of the Azores Islands (Portugal): data from the Y-chromosome. *Hum Biol (in press)*.
- Ferreira, A. (1987) *A Ilha Graciosa*. Livros Horizonte, Lisboa, Portugal.
- Filosto, M., Mancuso, M., Vives-Bauza, C., Vila, M.R., Shanske, S., Hirano, M., Andreu, A. L. y DiMauro, S. (2003) Lack of paternal inheritance of muscle mitochondrial DNA in sporadic mitochondrial myopathies. *Ann Neurol*, 54, 524-526.
- Finnilä, S., Lehtonen, M.S. y Majamaa, K. (2001) Phylogenetic network for European mtDNA. *Am J Hum Genet*, 68, 1475-1484.
- Fontdevila, A. y Moya, A. (1999) *Introducción a la Genética de Poblaciones*. Editorial Síntesis, S. A., Madrid, España.

Bibliografía

- Fontdevila, A. y Moya, A. (2003) ***Evolución: Origen, adaptación y divergencia de las especies***. Editorial Síntesis, S. A., Madrid, España.
- Forster P., Harding R., Torroni A., Bandelt H.-J. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. ***Am J Hum Genet***, 59, 935-945.
- Forster, P., Torroni, A., Renfrew, C. y Rohl, A. (2001) Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. ***Mol Biol Evol***, 18, 1864-1881.
- Futuyma, D. J. (1998) ***Evolutionary Biology***. Sinauer Associates, Inc., Publishers, Massachusetts, USA.
- Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. et al. (2002) The structure of haplotype blocks in the human genome. ***Science***, 296 (5576), 2225-2229.
- Gaspar, C., Lopes-Cendes, I., Hayes, S., Goto, J., Arvidsson, K., Dias, A., Silveira, I., Maciel, P., Coutinho, P., Lima, M. et al. (2001) Ancestral origins of the Machado-Joseph disease mutation: a worldwide haplotype study. ***Am J Hum Genet***, 68, 523-528.
- Gagneux, P., Wills, C., Gerloff, U., Tautz, D., Morin, P.A., Boesch, C., Fruth, B., Hohmann, G., Ryder, O.A. and Woodruff, D.S. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. ***Proc Natl Acad Sci U S A***, 96, 5077-5082.
- Gerber A.S., Loggins R., Kumar S. y Dowling T.E. (2001) Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? ***Ann Rev Genet***, 35, 539-566.
- Giles, R.E., Blanc, H., Cann, H.M. y Wallace, D.C. (1980) Maternal inheritance of human mitochondrial DNA. ***Proc Natl Acad Sci USA***, 77, 6715-6719.
- Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E. y Sullivan, K. (1994) Identification of the remains of the Romanov family by DNA analysis. ***Nat Genet***, 6, 130-135.
- Goldman, N. y Barton, N.H. (1992) Human origins. Genetics and geography. ***Nature***, 357, 440-441.
- Gomes, F. (1997) ***A ilha das Flores: da redescoberta à actualidade (subsídios para a sua história)***. Câmara Municipal de Lajes das Flores, Azores, Portugal.
- González, A.M., Brehm, A., Perez, J.A., Maca-Meyer, N., Flores, C. y Cabrera, V.M. (2003) Mitochondrial DNA affinities at the Atlantic fringe of Europe. ***Am J Phys Anthropol***, 120, 391-404.
- González-Halphen, D., Pérez-Martínez, X., Funes, S., Reyes-Prieto, A. y Santillán-Torres, J. L. (2003) La migración de genes de la mitocondria al núcleo y la evolución de los genomas mitocondriales. In ***Mensaje Bioquímico, Vol XXVII***, O. Flores Herrera, H. Riveros Rosas, A. Sosa Peinado y E. Vázquez Contreras (Eds.). Universidad Nacional Autónoma de México, México, pp. 201-220.
- Grant, V. (1991) *The evolutionary process – A critical study of evolutionary theory*. Columbia University Press, New York, USA.

Bibliografía

- Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A. y Excoffier, L. (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol*, 12, 334-345.
- Grzybowski, T., Malyarchuk, B.A., Czarny, J., Miscicka-Sliwka, D. y Kotzbach, R. (2003) High levels of mitochondrial DNA heteroplasmy in single hair roots: reanalysis and revision. *Electrophoresis*, 24, 1159-65.
- Gutierrez, G., Sanchez, D. y Marin, A. (2002) A reanalysis of the ancient mitochondrial DNA sequences recovered from Neandertal bones. *Mol Biol Evol*, 19, 1359-1366.
- Hagelberg, E. (2003) Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. *Trends Genet*, 19, 84-90.
- Halushka, M.K., Fan, J.B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. y Chakravarti, A. (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet*. 22, 239-247.
- Hammer, M.F. y Horai, S. (1995) Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet*, 56, 951-962.
- Harris, H. (1966) Enzyme polymorphisms in man. *Proc R Soc Lond B*, 164, 298-310.
- Harrison, G.A. y Boyce, A.J. (1972) Migration, exchange, and the genetic structure of populations. In *The Structure of Human Populations*, G.A. Harrison y A.J. Boyce (Eds.). Clarendon Press, Oxford, pp. 128-145.
- Hartl, D. y Clark, A. (1997) *Principles of Population Genetics*. 3th edition. Sinauer Associates, Inc., Publishers Sunderland, Massachusetts, USA.
- Hasegawa, M. y Horai, S. (1991) Time of the deepest root for polymorphism in human mitochondrial DNA. *J Mol Evol*, 32, 37-42.
- Hauswirth, W.W. y Clayton, D.A. (1985) Length heterogeneity of a conserved displacement-loop sequence in human mitochondrial DNA. *Nucleic Acids Res*, 13, 8093-8104.
- Hazout, S. y Lucotte, G. (1986) Towards a genealogy of chromosome Y. *Ann Genet*, 29, 246-252.
- Helgason, A., Sigurðardóttir, S., Nicholson, J., Sykes, B., Hill, E.W., Bradley, D.G., Bosnes, V., Gulcher, J.R., Ward, R. y Stefansson, K. (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet*, 67, 697-717.
- Helgason, A., Hickey, E., Goodacre, S., Bosnes, V., Stefansson, K., Ward, R. y Sykes, B. (2001) mtDna and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet*, 68, 723-737.
- Herrnstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C., Ghosh, S.S., Olefsky, J.M., Beal, M.F., Davis, R.E. et al. (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-

Bibliografia

- region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet*, 70, 1152-1171.
- Heyer, E., Zietkiewicz, E., Rochowski, A., Yotova, V., Puymirat, J. y Labuda, D. (2001) Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet*, 69, 1113-1126.
- Hill, E.W., Jobling, M.A. y Bradley, D.G. (2000) Y-chromosome variation and Irish origins. *Nature*, 404, 351-352.
- Hirszfeld, L. y Hirszfeld, H. (1919) Essai d'application des methods au problème des races. *Anthropologie*, 29, 505-537.
- Hixson, J.E. y Clayton, D.A. (1985) Initiation of transcription from each of the two human mitochondrial promoters requires unique nucleotides at the transcriptional start sites. *Proc Natl Acad Sci USA*, 82, 2660-2664.
- Horai, S., Kondo, R., Nakagawa-Hattori, Y., Hayashi, S., Sonoda, S. y Tajima, K. (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Mol Biol Evol*, 10, 23-47.
- Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. and Takahata, N. (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci U S A*, 92, 532-536.
- Howell, N., Halvorson, S., Kubacka, I., McCullough, D.A., Bindoff, L.A. y Turnbull, D.M. (1992) Mitochondrial gene segregation in mammals: is the bottleneck always narrow? *Hum Genet*, 90, 117-120.
- Howell, N., Kubacka, I. y Mackey, D.A. (1996) How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet*, 59, 501-509.
- Howell, N. (1997) mtDNA recombination: what do in vitro data mean? *Am J Hum Genet*, 61, 19-22.
- Howell, N. (1999) Human mitochondrial disease: answering questions and questions answers. *Int. Rev. Cytol*, 186, 49-116.
- Howell, N., Smejkal, C.B., Mackey, D.A., Chinnery, P.F., Turnbull, D.M. y Herrnstadt, C. (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic y pedigree rates. *Am J Hum Genet*, 72, 659-670.
- Howell N., Elson J.L., Turnbull D.M. y Herrnstadt C. (2004) African Haplogroup L mtDNA sequences show violations of clock-like evolution. *Mol Biol Evol*, 21, 1843-1854.
- INE - Instituto Nacional de Estatística (2001) **Recenseamento da População e da Habitação (R.A. dos Açores) – Censos 2001**. Instituto Nacional de Estatística, Portugal.
- INE - Instituto Nacional de Estatística (2004) **Biblioteca Digital: Recenseamento geral da população e da habitação 1864, 1878, 1890, 1900, 1911, 1920, 1930, 1940, 1950, 1960, 1970, 1981, 1991**. Available at the World Wide Web: <http://www.ine.pt:8080/biblioteca/search.do>
- Ingman, M., Kaessmann, H., Paabo, S. y Gyllensten, U. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature*, 408, 708-713.

Bibliografía

- Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W. y Parsons, T.J. (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. **Nat Genet**, 12, 417-420.
- Jeffreys, A.J., Kauppi, L. y Neumann, R. (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. **Nat Genet**, 29, 217-222.
- Jobling, M.A. y Tyler-Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age. **Nat Rev Genet**, 4, 598-612.
- Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C. y Cavalli-Sforza, L.L. (1983) Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. **J Mol Evol**, 19, 255-271.
- Jorde, L. (1980) The Genetic structure of subdivided human populations. In **Current developments in anthropological genetics**, Vol. 1, J. H. Mielke y M. H. Crawford (Eds.). Plenum Press, New York, pp. 135-208.
- Jorde, L.B., Bamshad, M.J., Watkins, W.S., Zenger, R., Fraley, A.E., Krakowiak, P.A., Carpenter, K.D., Soodyall, H., Jenkins, T. and Rogers, A.R. (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. **Am J Hum Genet**, 57, 523-538.
- Jorde, L.B. y Bamshad, M. (2000) Questioning evidence for recombination in human mitochondrial DNA. **Science**, 288, 1931.
- Kauppi, L., Sajantila, A. y Jeffreys, A.J. (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. **Hum Mol Genet**, 12, 33-40.
- Khrapko, K., Nekhaeva, E., Kraytsberg, Y. y Kunz, W. (2003) Clonal expansions of mitochondrial genomes: implications for in vivo mutational spectra. **Mutat Res**, 522, 13-19.
- Khrapko, K., Ebralidse, K. y Kraytsberg, Y. (2004) Where and when do somatic mtDNA mutations occur? **Ann N Y Acad Sci**, 1019, 240-244.
- Kimura, M. (1983) **The neutral theory of molecular evolution**. Cambridge University Press, Cambridge.
- Kivisild, T., Tolk, H.V., Parik, J., Wang, Y., Papiha, S.S., Bandelt, H.-J. y Villems, R. (2002) The emerging limbs and twigs of the East Asian mtDNA tree. **Mol Biol Evol**, 19, 1737-1751.
- Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., Parik, J., Geberhiwot, T., Usanga, E. y Villems, R. (2004) Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. **Am J Hum Genet**, 75, 752-770.
- Kong, Q.P., Yao, Y.G., Sun, C., Bandelt, H.-J., Zhu, C.L. y Zhang, Y.P. (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. **Am J Hum Genet**, 73, 671-676.
- Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J. y Khrapko, K. (2004) Recombination of human mitochondrial DNA. **Science**, 304, 981.

Bibliografía

- Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M. y Paabo, S. (1997) Neandertal DNA sequences and the origin of modern humans. **Cell**, 90, 19-30.
- Krings, M., Geisert, H., Schmitz, R.W., Krainitzki, H. y Paabo, S. (1999a) DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. **Proc Natl Acad Sci USA**, 96, 5581-5585.
- Krings, M., Salem, A.E., Bauer, K., Geisert, H., Malek, A.K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G. et al. (1999b) mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? **Am J Hum Genet**, 64, 1166-1176.
- Kurland, C.G. y Andersson, S.G. (2000) Origin and evolution of the mitochondrial proteome. **Microbiol Mol Biol Rev**, 64, 786-820.
- Lakshmiathy, U. y Campbell, C. (1999a) Double strand break rejoining by mammalian mitochondrial extracts. **Nucleic Acids Res**, 27, 1198-1204.
- Lakshmiathy, U. y Campbell, C. (1999b) The human DNA ligase III gene encodes nuclear and mitochondrial proteins. **Mol Cell Biol**, 19, 3869-3876.
- Landsteiner, K. (1901) Über agglutinationserscheinungen normalen menschlichen blutes. **Wien Klin Wochenschr**, 14, 1132-1134.
- Lang, B.F., Gray, M.W. y Burger, G. (1999) Mitochondrial genome evolution and the origin of eukaryotes. **Annu Rev Genet**, 33, 351-397.
- Lasker, G.W. (1954) Human evolution in contemporary communities. **Southwestern Journal of Anthropology**, 10, 353-365.
- Lee, H.Y., Chung, U., Yoo, J.E., Park, M.J. y Shin, K.J. (2004) Quantitative and qualitative profiling of mitochondrial DNA length heteroplasmy. **Electrophoresis**, 25, 28-34.
- Legros, F., Malka, F., Frachon, P., Lombes, A. y Rojo, M. (2004) Organization and dynamics of human mitochondrial DNA. **J Cell Sci**, 117, 2653-2662.
- Lewis, J.H., Walters, D., Didisheim, P. y Merchant, W.R. (1958) Application of continuous flow electrophoresis to the study of the blood coagulation proteins and the fibrinolytic enzyme system. I. Normal human materials. **J Clin Invest**, 37, 1323-1331.
- Lewontin R.C. (1972) The apportionment of human diversity. **In Evolutionary Biology** 6. T. Dobzhansky, M. K. Hecht y W.C. Steere (Eds.). Appleton-Century-Crofts, New York, pp. 381-398.
- Lewontin, R.C. (1984) **La Diversidad Humana**. Prensa Científica, S. A., Barcelona, España.
- Li, W.-H. (1997) **Molecular Evolution**. Sinauer Associates, Inc., Publishers, Massachusetts, USA.
- Lima, M. (1990) Consanguinidade aparente na população da Candelária. **Antropologia Portuguesa**, 8, 155-164.
- Lima, M. (1991) **Contribuição para o estudo biodemográfico da freguesia de Rabo de Peixe**. Pruebas APCC, Universidad de las Azores, Ponta Delgada, Portugal.

Bibliografia

- Lima, M. y Soares, A. (1991/1992) Niveis de consanguinidade e endogamia na população das Furnas (São Miguel / Açores). *Antropologia Portuguesa*, 9-10, 33-42.
- Lima, M. (1992/1993) Biodemografia das populações açorianas. *Arquipélago, Série Ciências Sociais*, 7-8, 99-114.
- Lima, M. (1996) *Doença de Machado-Joseph nos Açores: Estudo epidemiológico, biodemográfico e genético*. Tesis doctoral, Universidad de las Azores, Ponta Delgada, Portugal.
- Lima, M., Mayer, F., Coutinho P. y Abade, A. (1997) Prevalence, geographical distribution and genealogical investigation of Machado-Joseph disease in the islands of the Azores (Portugal). *Hum Biol*, 69, 383-391.
- Lima, M. (1998) *Demografia Histórica e Dinâmica Bio-Cultural da Ilha das Flores (séculos XVII a XIX) – Praxis/PCSH/C/DEM/56/96*. Universidad de las Azores, Ponta Delgada, Portugal.
- Lima, M., Mayer, F., Coutinho P. y Abade, A. (1998) The origin of Machado-Joseph disease gene in the islands of the Azores (Portugal): a genealogical study. *Hum Biol*, 70, 1011-1023.
- Lima, M. (1999) O contributo da Demografia Histórica para os estudos em Antropologia Genética. *Insulana*, LV, 201-210.
- Lima, M., Santos, C., Bettencourt, C., Silva, C. y Montiel, R. (2003) The African influence in the settlement of the Azores Islands: An interdisciplinary study using Biodemographic and Genetic data. Poster presentado en *Portugalia Genética 6th edition: Human Haploid Genomes*. OPorto, 20-22 Marzo.
- Lima, P. (1997) *Ilha Graciosa - história*. Retrieved October 8, 2001 from the World Wide Web: <http://www.geocities.com/TheTropics/2140/graciosp.html#historia>
- Loogväli, E.L., Roostalu, U., Malyarchuk, B.A., Derenko, M.V., Kivisild, T., Metspalu, E., Tambets, K., Reidla, M., Tolk, H.V., Parik, J. et al. (2004) Disuniting Uniformity: A Pied Cladistic Canvas of mtDNA Haplogroup H in Eurasia. *Mol Biol Evol*, 21, 2012-2021.
- Lum, J.K. y Cann, R.L. (2000) mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. *Am J Phys Anthropol*, 113, 151-168.
- Lundstrom, R., Tavaré, S. y Ward, R.H. (1992) Estimating substitution rates from molecular data using the coalescent. *Proc Natl Acad Sci USA*, 89, 5961-5965.
- Lutz, S., Weisser, H.J., Heizmann, J. y Pollak, S. (2000) Mitochondrial heteroplasmy among maternally related individuals. *Int J Legal Med*, 113, 155-161.
- Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., Flores, C. and Cabrera, V.M. (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet*, 2, 13.
- Maca-Meyer, N., González, A.M., Pestano, J., Flores, C., Larruga, J.M. y Cabrera, V.M. (2003) Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet*, 4, 15.
- Macaulay, V.A., Richards, M.B., Forster, P., Bendall, K.E., Watson, E., Sykes, B. y Bandelt, H.-J. (1997) mtDNA mutation rates - no need to panic. *Am J Hum Genet*, 61, 983-990.

Bibliografia

- Macaulay, V., Richards, M. y Sykes, B. (1999a) Mitochondrial DNA recombination-no need to panic. ***Proc R Soc Lond B Biol Sci***, 266, 2037-2039.
- Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B. y Torroni, A. (1999b) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. ***Am J Hum Genet***, 64, 232-249.
- Madeira, A. (1999) ***População e Emigração nos Açores (1766-1820)***. Patrimonia, Cascais, Portugal.
- Malhi, R.S., Schultz, B.A. y Smith, D.G. (2001) Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. ***Hum Biol***, 73, 17-55.
- Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Wozniak, M. y Miscicka-Sliwka, D. (2002) Mitochondrial DNA variability in Poles and Russians. ***Ann Hum Genet***, 66, 261-283.
- Malyarchuk, B.A., Grzybowski, T., Derenko, M.V., Czarny, J., Drobnic, K. y Miscicka-Sliwka, D. (2003) Mitochondrial DNA variability in Bosnians and Slovenians. ***Ann Hum Genet***, 67, 412-425.
- Marchington, D.R., Macaulay, V., Hartshorne, G.M., Barlow, D. y Poulton, J. (1998) Evidence from human oocytes for a genetic bottleneck in an mtDNA disease. ***Am J Hum Genet***, 63, 769-775.
- Margulis, L. (1971) The origin of plant and animal cells. ***Am Sci***, 59, 230-235.
- Marth, G., Schuler, G., Yeh, R., Davenport, R., Agarwala, R., Church, D., Wheelan, S., Baker, J., Ward, M., Kholodov, M. et al. (2003) Sequence variations in the public human genome data reflect a bottlenecked population history. ***Proc Natl Acad Sci USA***, 100, 376-381.
- Martin, W. y Muller, M. (1998) The hydrogen hypothesis for the first eukaryote. ***Nature***, 392, 37-41.
- Mateu, E., Comas, D., Calafell, F., Perez-Lezaun, A., Abade, A. y Bertranpetit, J. (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tome, Gulf of Guinea. ***Ann Hum Genet***, 61, 507-518.
- Matos, A. (1987) Do contributo Algarvio no povoamento da Madeira e dos Açores *In Actas das I Jornadas de História Medieval do Algarve e Andaluzia*. Câmara Municipal de Loulé, Loulé.
- Matos, A. (1989a) ***Povoamento e Colonização dos Açores in Portugal no Mundo***. Publicações Alfa, Lisboa.
- Matos, A. (1989b) Origem e reminiscência dos povoadores das ilhas atlânticas *In Actas do Congresso Internacional: Bartolomeu Dias e a sua época*, Volume III. Comissão nacional para as Comemorações dos Descobrimentos Portugueses, Universidade do Porto, Porto.
- Matos, P. (2004) Endogamia e estratégias familiares campesinas numa freguesia da periferia açoriana, A Ribeira Seca da ilha de São Jorge (Açores): 1800-1910. Comunicação presentada en el ***VII Congreso de la Asociación de Demografía Histórica***. Available at the World Wide Web: <http://www.ugr.es/~adeh/comunicaciones.htm>
- Mendonça, L. (1996) ***História dos Açores – Visão geral (sécs. XV-XIX)***. Centro de Apoio Tecnológico à Educação, Ponta Delgada, Azores.

Bibliografía

- Menozzi, P., Piazza, A. y Cavalli-Sforza, L. (1978) Synthetic maps of human gene frequencies in Europeans. **Science**, 201, 786-792.
- Merelim, P. (1966) **Os Hebraicos na Ilha Terceira**. Separata da Atlântida, Angra do Heroísmo, Açores, Portugal.
- Merriweather, D.A. y Kaestle, F.A. (1999) Mitochondrial recombination? (continued). **Science**, 285, 837.
- Mesa, N.R., Mondragon, M.C., Soto, I.D., Parra, M.V., Duque, C., Ortiz-Barrientos, D., Garcia, L.F., Velez, I.D., Bravo, M.L., Munera, J.G. et al. (2000) Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: pre- and post-Columbian patterns of gene flow in South America. **Am J Hum Genet**, 67, 1277-1286.
- Meyer, S., Weiss, G. y von Haeseler, A. (1999) Pattern of nucleotide substitutions and rate heterogeneity in the hypervariable regions I and II of human mtDNA. **Genetics**, 152,1103-1110.
- Meyer, S. y von Haeseler, A. (2003) Identifying site-specific substitution rates. **Mol Biol Evol**, 20, 182-189.
- Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D. et al. (2003) Natural selection shaped regional mtDNA variation in humans. **Proc Natl Acad Sci USA**, 100, 171-176.
- Monnat, R. J. Jr. y Loeb, L. A. (1985) Nucleotide sequence preservation of human mitochondrial DNA. **Proc Natl Acad Sci USA**, 82, 2895-2899.
- Monnat, R. J. Jr. y Reay, D. T. (1986) Nucleotide sequence identity of mitochondrial DNA from different human tissues. **Gene**, 43, 205-211.
- Montiel, R. (2001) *Estudio diacrónico de la variabilidad del DNA mitocondrial en población Catalana*. Tesis Doctoral, Universidad Autónoma de Barcelona, Barcelona, España. Available at the World Wide Web: <http://www.tdcat.cesca.es/TDCat-0726101-095837>
- Montiel, R., Bettencourt, C., Silva, C., Santos, C., Prata, M. J. y Lima, M. (2005) Analysis of Y-chromosome variability and its comparison with mtDNA variability reveals different demographic histories between islands in the Azores archipelago (Portugal). **Ann Hum Genet**, 69, 1-10.
- Morelli, L., Grosso, M.G., Vona, G., Varesi, L., Torroni, A. y Francalacci, P. (2000) Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. **Hum Biol**, 72, 585-595.
- Morley, J.M., Bark, J.E., Evans, C.E., Perry, J.G., Hewitt, C.A. y Tully, G. (1999) Validation of mitochondrial DNA minisequencing for forensic casework. **Int J Legal Med**, 112, 241-248.
- Morris, A.A. y Lightowlers, R.N. (2000) Can paternal mtDNA be inherited? **Lancet**, 355, 1290-1291.
- Morton, N. E. (1969) Human population structure. **Annu Rev Genet**, 3, 53-74.
- Mota, P. J. G. (1983) Consanguinidade aparente em S. Vicente da Beira. **Antropologia Portuguesa**, 1, 5-32.
- Mountain, J.L., Hebert, J.M., Bhattacharyya, S., Underhill, P.A., Ottolenghi, C., Gadgil, M. y Cavalli-Sforza, L.L. (1995) Demographic history of India and mtDNA-sequence diversity. **Am J Hum Genet**, 56, 979-992.

Bibliografía

- Mourant, A.E. (1954) *The Distribution of Human Blood Groups*. Blackwell Scientific, Oxford, UK.
- Mourant, A.E., Kopec, A.C. y Domaniewska-Sobczak, K. (1976) *The Distribution of the Human Blood Groups and Other Polymorphisms*. Oxford Univ. Press, London, UK.
- Muller, M. y Martin, W. (1999) The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. *Bioessays*, 21, 377-381.
- Nei, M. y Roychoudhury, A. (1974) Genic variation within and between the three major races of Man, Caucasoids, Negroids and Mongoloids. *Am J Hum Genet*, 26, 421-443.
- Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.
- NEPS - Núcleo de Estudos de População e Sociedade (2003) *Genealogias: Açores*. Available at the World Wide Web: <http://www.neps.ics.uminho.pt/acoresh.htm>
- Ngo, K.Y., Vergnaud, G., Johnsson, C., Lucotte, G. y Weissenbach, J. (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. *Am J Hum Genet*, 38, 407-418.
- Nielsen, R. y Yang, Z. (2003) Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol Biol Evol*, 20, 1231-1239.
- Noda, R., Kim, C.G., Takenaka, O., Ferrell, R.E., Tanoue, T., Hayasaka, I., Ueda, S., Ishida, T. y Saitou, N. (2001) Mitochondrial 16S rRNA sequence diversity of hominoids. *J Hered*, 92, 490-496.
- Pääbo S (1996) Mutational hot spots in the mitochondrial microcosm. *Am J Hum Genet*, 59, 493-496.
- Pacheco, P.R., Branco, C.C., Cabral, R., Costa, S., Araújo, A.L., Peixoto, B.R., Mendonça, P. y Mota-Vieira, L. (2005) The Y-chromosome heritage of the Azores Islands population. *Ann Hum Genet*, 69, 11-19.
- Palanichamy, M.G., Sun, C., Agrawal, S., Bandelt, H.-J., Kong, Q.P., Khan, F., Wang, C.Y., Chaudhuri, T.K., Palla, V. y Zhang, Y.P. (2004) Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet*, 75, 966-978.
- Parson, W., Parsons, T.J., Scheithauer, R. y Holland, M.M. (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Legal Med*, 111, 124-132.
- Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P. et al. (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet*, 15, 363-368.
- Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M. y Santachiara-Benerecetti, A.S. (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet*, 62, 420-434.
- Pauling, L., Itano, H. A., Singer, S. J. y Wells, I. C. (1949) Sickle-cell anemia, a molecular disease. *Science*, 110, 543-548.

Bibliografía

- Penny, D., Steel, M., Waddell, P.J. y Hendy, M.D. (1995) Improved analyses of human mtDNA sequences support a recent African origin for Homo sapiens. *Mol Biol Evol*, 12, 863-882.
- Pereira, L., Prata, M.J. y Amorim, A. (2000) Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet*, 64, 491-506.
- Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M.J. y Amorim, A. (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet*, 65, 439-458.
- Pesole, G., Sbisa, E., Preparata, G. y Saccone, C. (1992) The evolution of the mitochondrial D-loop region and the origin of modern man. *Mol Biol Evol*, 9, 587-598.
- Pinto, F., Gonzalez, A.M., Hernandez, M., Larruga, J.M. y Cabrera, V.M. (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet*, 60, 321-330.
- Pires, J. (1983) Ensaio Histórico: Povoamento do Faial In Rosa, J. (Ed.) *Em louvor do V Centenário do Povoamento da Ilha do Faial 1468-69 – 1968-69*. Horta, Açores.
- Pires, L. (1999) *A Origem da população dos Açores com base em estudos de DNA mitocondrial*. Trabajo realizado para obtención de la Licenciatura en Biología. Universidad de Coimbra, Coimbra.
- Plaza, S., Calafell, F., Helal, A., Bouzerna, N., Lefranc, G., Bertranpetit, J. y Comas, D. (2003) Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann Hum Genet*, 67, 312-328.
- Poulton, J., Macaulay, V. y Marchington, D.R. (1998) Mitochondrial genetics '98 is the bottleneck cracked? *Am J Hum Genet*, 62, 752-757.
- Prata, M. J., Gusmão, L. Mota, P., Silva D. y Amorim, A. (1998) Genetic profiling of S. Miguel Island (Azores). *Prog Foren Genet*, 7, 294-296.
- Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K. y Santachiara-Benerecetti, A.S. (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet*, 23, 437-441.
- Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S. et al. (2004) Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet*, 74, 827-845.
- Quintans, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M.V. y Carracedo, A. (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int*, 140, 251-257.
- Race, R. R. y Sanger, R. (1975) *Blood Groups in Man*. Blackwell Scientific, Oxford, UK.
- Rand, D.M. (2001) The units of selection on mitochondrial DNA. *Ann Rev Ecol Syst*, 32, 415-448.

Bibliografia

- Rando, J.C., Pinto, F., Gonzalez, A.M., Hernandez, M., Larruga, J.M., Cabrera, V.M. y Bandelt, H.-J. (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.
- Rando, J.C., Cabrera, V.M., Larruga, J.M., Hernandez, M., Gonzalez, A.M., Pinto, F. y Bandelt, H.-J. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet*, 63, 413-428.
- Rannala, B., Bertorelle, G. (2001) Using linked markers to infer the age of a mutation. *Hum Mutat*, 18, 87-100.
- Reich, D. E., Schaffner, S. F., Daly, M. J., McVean, G., Mullikin, J. C., Higgins, J. M., Richter, D. J., Lander, E. S. y Altshuler, D. (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet*, 32: 135-142.
- Reidla, M., Kivisild, T., Metspalu, E., Kaldma, K., Tambets, K., Tolk, H.V., Parik, J., Loogvali, E.L., Derenko, M., Malyarchuk, B. et al. (2003) Origin and diffusion of mtDNA haplogroup X. *Am J Hum Genet*, 73, 1178-1190.
- Relethford, J. (1988) Estimation of kinship and genetic distance from surnames. *Hum Biol*, 60, 475-492.
- Relethford, J. (2001) *Genetics and the search for modern human origins*. Wiley-Liss, New York, USA.
- Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H.-J. y Sykes, B. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet*, 59, 185-203.
- Richards, M., Macaulay, V., Bandelt, H.-J. y Sykes, B. C. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet*, 62, 241-260.
- Richards, M., Macaulay, V., Hickey, E., Veja, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T. et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet*, 67, 1251-1276.
- Richards, M. y Macaulay, V. (2001) The mitochondrial gene tree comes of age. *Am J Hum Genet*, 68, 1315-1320.
- Richards, M., Macaulay, V., Torroni, A. y Bandelt, H.-J. (2002) In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet*, 71, 1168-1174.
- Richards, M., Rengo, C., Cruciani, F., Gratrix, F., Wilson, J.F., Scozzari, R., Macaulay, V. y Torroni, A. (2003) Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am J Hum Genet*, 72, 1058-64.
- Richter, C., Park, J.W. y Ames, B.N. (1988) Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci USA*, 85, 6465-6467.
- Riley, C. (1991) Uma fronteira Ocidental. As Flores e o Corvo no contexto das rotas Atlânticas. *Boletim do Núcleo Cultural da Horta*. XI: 81-117.

Bibliografía

- Robin, E.D. y Wong, R. (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol*, 136, 507-513.
- Rodas, C., Gelvez, N. y Keyeux, G. (2003) Mitochondrial DNA studies show asymmetrical Amerindian admixture in Afro-Colombian and Mestizo populations. *Hum Biol*, 75, 13-30.
- Romualdi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M. A. y Barbujani, G. (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res*, 12: 602-612.
- Rosa, A., Brehm, A., Kivisild, T., Metspalu, E. y Villems, R. (2004) MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. *Ann Hum Genet*, 68, 340-352.
- Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A. y Feldman, M.W. (2002) Genetic structure of human populations. *Science*, 298, 2381-2385.
- Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V., Wallace, D.C. (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*, 303, 223-226.
- Saccone, C. (1994) The evolution of mitochondria DNA. *Curr Opin Genet Dev*, 4, 875-881.
- Saccone, C., Gissi, C., Lanave, C., Larizza, A., Pesole, G. y Reyes, A. (2000) Evolution of the mitochondrial genetic system: an overview. *Gene*, 261, 153-159.
- 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., Richards, M., De la Fe, T., Lareu, M.V., Sobrino, B., Sanchez-Diz, P., Macaulay, V. y Carracedo, A. (2002) The making of the African mtDNA landscape. *Am J Hum Genet*, 71, 1082-1111.
- Salas, A., Richards, M., Lareu, M.V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., y Carracedo, A. (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet*, 74, 454-465.
- Sanger, R. y Race, R.R. (1949) The combination of blood groups in a sample of 250 people. *Ann Eugen*, 15, 77-90.
- Santachiara-Benerecetti, A.S., Scozzari, R., Semino, O., Torroni, A., Brega, A. y Wallace, D.C. (1988) Mitochondrial DNA polymorphisms in Italy. II. Molecular analysis of new and rare morphs from Sardinia and Rome. *Ann Hum Genet*, 52, 39-56.
- Santos, C. y Lima, M. (2000) Trends in mortality in the parish of Santa Cruz (Island of Flores, Azores, Portugal), 1676-1876. Poster presentado en el **12th Congress of the European Anthropological Association - Millennial Perspectives: Past, Present and Future**. Cambridge, England, 8 – 11 September.
- Santos, M., Prata, M.J., Amorim, A., Almeida, U. y Santos, M. (1992) Ilha de S. Miguel: caracterização genética preliminar. *XXVII Jornadas Luso-Espanholas de Genética*, Badajoz, España. [Abst].

Bibliografia

- Santos, M. y Amorim, A. (1994) Population studies of genetic markers in Portugal - a comparison between S. Miguel (Azores) and Northern Portugal. **3rd International Workshop on Machado-Joseph disease**, S. Miguel, Portugal. [Abst].
- Santos, M., Amorim, A., Prata, M.J., Almeida, V., Santos, M. y Brandão, C. (1994) Population genetics of São Miguel Island. **3rd International Workshop on Machado-Joseph disease**, S. Miguel, Portugal. [Abst].
- Schwartz, M. y Vissing, J. (2002) Paternal inheritance of mitochondrial DNA. **N Engl J Med**, 347, 576-580.
- Schwartz, M. y Vissing, J. (2004) No evidence for paternal inheritance of mtDNA in patients with sporadic mtDNA mutations. **J Neurol Sci**, 218, 99-101.
- Scozzari, R., Cruciani, F., Santolamazza, P., Sellitto, D., Cole, D.E., Rubin, L.A., Labuda, D., Marini, E., Succa, V., Vona, G. et al. (1997) mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. **Am J Hum Genet**, 60, 241-244.
- Scozzari, R., Torroni, A., Semino, O., Sirugo, G., Brega, A. y Santachiara-Benerecetti, A.S. (1988) Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. **Am J Hum Genet**, 43, 534-544.
- Serre, D., Langaney, A., Chech, M., Teschler-Nicola, M., Paunovic, M., Menecier, P., Hofreiter, M., Possnert, G. y Paabo, S. (2004) No evidence of Neandertal mtDNA contribution to early modern humans. **PLoS Biol**, 2, E57.
- Shen, P., Lavi, T., Kivisild, T., Chou, V., Sengun, D., Gefel, D., Shpirer, I., Woolf, E., Hillel, J., Feldman, M.W. et al. (2004) Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. **Hum Mutat**, 24, 248-260.
- Sigurðardóttir, S., Helgason, A., Gulcher, J.R., Stefansson, K. y Donnelly, P. (2000) The mutation rate in the human mtDNA control region. **Am J Hum Genet**, 66, 1599-1609.
- Silva, C., Castro-Oliveira, P., Montiel, R. y Lima, M. Population genetics of 5 STR loci in S. Miguel Island. *Antropologia Portuguesa (in press)*.
- Silva, W.A., Jr., Bonatto, S.L., Holanda, A.J., Ribeiro-Dos-Santos, A.K., Paixao, B.M., Goldman, G.H., Abe-Sandes, K., Rodriguez-Delfin, L., Barbosa, M., Paco-Larson, M.L. et al. (2002) Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. **Am J Hum Genet**, 71, 187-192.
- Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. y Barbujani, G. (2000) Geographic patterns of mtDNA diversity in Europe. **Am J Hum Genet**, 66, 262-278.
- Smith, D.G., Malhi, R.S., Eshleman, J., Lorenz, J.G. y Kaestle, F.A. (1999) Distribution of mtDNA haplogroup X among Native North Americans. **Am J Phys Anthropol**, 110, 271-284.
- Smith, M., Abade, A. y Cunha, E. (1992) Genetic structure of the Azores: marriage and inbreeding in Flores. **Ann Hum Biol**, 19, 595-601.
- Sokal, R.R., Harding, R.M. y Oden, N.L. (1989) Spatial patterns of human gene frequencies in Europe. **Am J Phys Anthropol**, 80, 267-294.

Bibliografia

- Sokal, R.R. (1991) Ancient movement patterns determine modern genetic variances in Europe. *Hum Biol*, 63, 589-606.
- Soodyall, H., Jenkins, T., Mukherjee, A., du Toit, E., Roberts, D.F. y Stoneking, M. (1997) The founding mitochondrial DNA lineages of Tristan da Cunha Islanders. *Am J Phys Anthropol*, 104, 157-166.
- Spuhler, J. (1988) Evolution of mitochondrial DNA in monkeys, apes and humans. *Yearb Phys Anthropol*, 31, 15-48.
- SREA – Serviço Regional de Estatística dos Açores. (2001a) *Situação geográfica dos territórios portugueses: Continente, Madeira e Açores*. Retrieved October 8, 2001 from the World Wide Web: <http://www.ine.pt/srea/informacao/series/dados.asp?ID=3>
- SREA – Serviço Regional de Estatística dos Açores (2001b) *Área total, nº de freguesias, área média das freguesias e densidade populacional, por concelho*. Retrieved October 8, 2001 from the World Wide Web: <http://www.ine.pt/srea/informacao/series/dados.asp?ID=4>
- Stephens, J.C., Schneider, J.A., Tanguay, D.A., Choi, J., Acharya, T., Stanley, S.E., Jiang, R., Messer, C.J., Chew, A., Han, J.H. et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science*, 293, 489-493.
- Stevanovitch, A., Gilles, A., Bouzaid, E., Kefi, R., Paris, F., Gayraud, R.P., Spadoni, J.L., El-Chenawi, F. y Beraud-Colomb, E. (2004) Mitochondrial DNA sequence diversity in a sedentary population from Egypt. *Ann Hum Genet*, 68, 23-39.
- Stone, A.C. y Stoneking, M. (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am J Hum Genet*, 62, 1153-1170.
- Stoneking, M., Bhatia, K. y Wilson, A.C. (1986) Rate of sequence divergence estimated from restriction maps of mitochondrial DNAs from Papua New Guinea. *Cold Spring Harb Symp Quant Biol*, 51 Pt 1, 433-439.
- Stoneking M., Sherry, S.T., Redd, A.J. y Vigilant, L. (1992) New approaches to dating suggest a recent age for the human mtDNA ancestor. *Philos Trans R Soc Lond B Biol Sci*, 337, 167-175.
- Stoneking, M. (1993) DNA and recent human evolution. *Evol Anthropol*, 2, 60-73.
- Stoneking, M. (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet*, 67, 1029-1032.
- Sunyaev, S.R., Lathe, W.C., Ramensky, V.E. y Bork, P. (2000) SNP frequencies in human genes an excess of rare alleles and differing modes of selection. *Trends Genet*, 16, 335-337.
- Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C. y Schatten, G. (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod*, 63, 582-590.
- Sykes, B. y Irven, C. (2000) Surnames and the Y chromosome. *Am J Hum Genet*, 66, 1417-1419.

Bibliografía

- Tamura, K. y Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. ***Mol Biol Evol***, 10, 512-526.
- Tanaka, M., Cabrera, V.M., Gonzalez, A.M., Larruga, J.M., Takeyasu, T., Fuku, N., Guo, L.J., Hirose, R., Fujita, Y., Kurata, M. et al. (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. ***Genome Res***, 14, 1832-1850.
- Taylor, R.W., McDonnell, M.T., Blakely, E.L., Chinnery, P.F., Taylor, G.A., Howell, N., Zeviani, M., Briem, E., Carrara, F. y Turnbull, D.M. (2003) Genotypes from patients indicate no paternal mitochondrial DNA contribution. ***Ann Neurol***, 54, 521-524.
- Templeton, A.R. (1993) The Eve Hypotheses - a Genetic Critique and Reanalysis. ***Am Anthropol***, 95, 51-72.
- Thomas, M.G., Skorecki, K., Ben-Ami, H., Parfitt, T., Bradman, N. y Goldstein, D.B. (1998) Origins of Old Testament priests. ***Nature***, 394, 138-140.
- Thyagarajan, B., Padua, R.A. y Campbell, C. (1996) Mammalian mitochondria possess homologous DNA recombination activity. ***J Biol Chem***, 271, 27536-27543.
- Torrioni, A., Schurr, T.G., Yang, C.C., Szathmary, E.J., Williams, R.C., Schanfield, M.S., Troup, G.A., Knowler, W.C., Lawrence, D.N., Weiss, K.M. et al. (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. ***Genetics***, 130, 153-162.
- Torrioni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M. y Wallace, D.C. (1993a) Asian affinities and continental radiation of the four founding Native American mtDNAs. ***Am J Hum Genet***, 53, 563-590.
- Torrioni, A., Sukernik, R.I., Schurr, T.G., Starikorskaya, Y.B., Cabell, M.F., Crawford, M.H., Comuzzie, A.G. y Wallace, D.C. (1993b) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. ***Am J Hum Genet***, 53, 591-608.
- Torrioni, A., Chen, Y.S., Semino, O., Santachiara-Beneceretti, A.S., Scott, C.R., Lott, M.T., Winter, M. y Wallace, D.C. (1994a) mtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. ***Am J Hum Genet***, 54, 303-318.
- Torrioni, A., Lott, M.T., Cabell, M.F., Chen, Y.S., Lavergne, L. y Wallace, D.C. (1994b) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. ***Am J Hum Genet***, 55, 760-776.
- Torrioni, A., Miller, J.A., Moore, L.G., Zamudio, S., Zhuang, J., Droma, T. y Wallace, D.C. (1994c) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. ***Am J Phys Anthropol***, 93, 189-199.
- Torrioni, A., Neel, J.V., Barrantes, R., Schurr, T.G. y Wallace, D.C. (1994d) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. ***Proc Natl Acad Sci USA***, 91, 1158-1162.

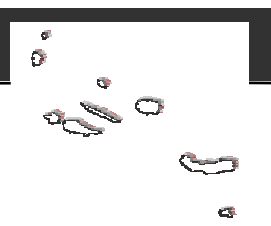
Bibliografía

- Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L. y Wallace, D.C. (1996) Classification of European mtDNAs from an analysis of three European populations. **Genetics**, 144, 1835-1850.
- Torroni, A., D'Urbano, L., Rengo, C., Scozzari, R., Sbracia, M., Manna, C., Cavazzini, C. y Sellitto, D. (1998a) Intracytoplasmic injection of spermatozoa does not appear to alter the mode of mitochondrial DNA inheritance. **Hum Reprod**, 13, 1747-1749.
- Torroni, A., Bandelt, H.-J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.L., Bonne-Tamir, B. et al. (1998b) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. **Am J Hum Genet**, 62, 1137-1152.
- Torroni, A., Richards, M., Macaulay, V., Forster, P., Villems, R., Norby, S., Savontaus, M.L., Huoponen, K., Scozzari, R. y Bandelt, H.-J. (2000) mtDNA haplogroups and frequency patterns in Europe. **Am J Hum Genet**, 66, 1173-7.
- Torroni, A., Rengo, C., Guida, V., Cruciani, F., Sellitto, D., Coppa, A., Calderon, F.L., Simionati, B., Valle, G., Richards, M. et al. (2001a) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? **Am J Hum Genet**, 69, 1348-1356.
- Torroni, A., Bandelt, H.-J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E. et al. (2001b) A signal, from human mtDNA, of postglacial recolonization in Europe. **Am J Hum Genet**, 69, 844-852.
- Tully, L.A., Parsons, T.J., Steighner, R.J., Holland, M.M., Marino, M.A. y Prenger, V.L. (2000) A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. **Am J Hum Genet**, 67, 432-443.
- Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L. y Oefner, P.J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. **Genome Res**, 7, 996-1005.
- Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P. et al. (2000) Y chromosome sequence variation and the history of human populations. **Nat Genet**, 26, 358-361.
- Velosa, R.G., Fernandes, A.T., Brehm, A. (2002) Genetic profile of the Açores Archipelago population using the new PowerPlex 16 system kit. **Forensic Sci Int**, 129, 68-71.
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. y Wilson, A. C. (1991) African populations and the evolution of human mitochondrial DNA. **Science**, 253, 1503-1507.
- Vigilant, L. y Bradley, B.J. (2004) Genetic variation in gorillas. **Am J Primatol**, 64, 161-172.
- von Heijne, G. (1986) Why mitochondria need a genome. **FEBS Lett**, 198, 1-4.
- Wakeley, J. (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. **J Mol Evol**, 37, 613-623.

Bibliografía

- Walberg, M.W. y Clayton, D.A. (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucl Acids Res*, 9, 5411-5421.
- Wallace, D. (1994) Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci USA*, 91, 8739-8746.
- Wallace, D. (1995) 1994 William Allan Address: Mitochondrial DNA variation in Human Evolution, Degenerative Disease, and Aging. *Am J Hum Genet*, 57, 201-223.
- Wallace, D., Stugard, C., Murdock, D., Schurr, T. y Brown, M.D. (1997) Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. *Proc Natl Acad Sci USA*, 94, 14900-14905.
- Wallace, D. (1999) Mitochondrial disease in man and mouse. *Science*, 283, 1482-1488.
- Wang, N., Akey, J.M., Zhang, K., Chakraborty, R. y Jin, L. (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am J Hum Genet*, 71, 1227-1234.
- Ward, R.H., Frazier, B.L., Dew-Jager, K. y Paabo, S. (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA*, 88, 8720-8724.
- Watson, E., Forster, P., Richards, M. y Bandelt, H.-J. (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet*, 61, 691-704.
- Wilson, A.C., Cann, R.L., Carr, S.M., Matthew, G., Gyllenstein, U.B., Helm-Bychowski, K.M., Higuchi, G., Palumbi, S.R., Prager, E.M., Sage, R.D. y Stoneking, M. (1985) *Biol J Linnean Soc*, 26, 375-400.
- Yasuda, N. y Morton, N.E. (1967) Studies on human population structure. In *Proceedings of the 3rd International Congress of Human Genetics*. J. F. Crow y J. V. Neel (Eds.). Johns Hopkins, Baltimore, pp. 249-265.
- Yasuda, N. y Furusho, T. (1971) Random and nonrandom inbreeding revealed from isonymy study. I. Small cities of Japan. *Am J Hum Genet*, 23, 303-316.
- Yasuda, N., Cavalli-Sforza, L.L., Skolnick, M. y Moroni, A. (1974) The evolution of surnames: an analysis of their distribution and extinction. *Theor Popul Biol*, 5, 123-142.
- Zeigler, G., Guglielmino, C.R., Siri, E., Moroni, A. y Cavalli-Sforza, L.L. (1983) Surnames as neutral alleles: observations in Sardinia. *Hum Biol*, 55, 357-365.
- Zeigler, G., Lisa, A., Fiorani, O., Magri, C., Quintana-Murci, L., Semino, O. y Santachiara-Benerecetti, A.S. (2003) From surnames to the history of Y chromosomes: the Sardinian population as a paradigm. *Eur J Hum Genet*, 11, 802-807.
- Zischler, H., Geisert, H., von Haeseler, A. y Paabo, S. (1995) A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature*, 378, 489-492.

VIII. APÉNDICE 1



VIII. APÉNDICE 1

En el apéndice 1 se presenta la información sobre la composición poblacional, por sexo y grupos de edad, así como el índice de masculinidad de los diferentes grupos islas del archipiélago de las Azores, desde 1864 hasta el año 2001.

VIII.1. COMPOSICIÓN POBLACIONAL

VIII.1.1. Grupo Oriental

Fig. I – Estructura poblacional por sexo y grupos de edad del grupo Oriental:

- I a) año 1864;
- I b) año 1890
- I c) año 1911
- I d) año 1930
- I e) año 1950
- I f) año 1970
- I g) año 2001

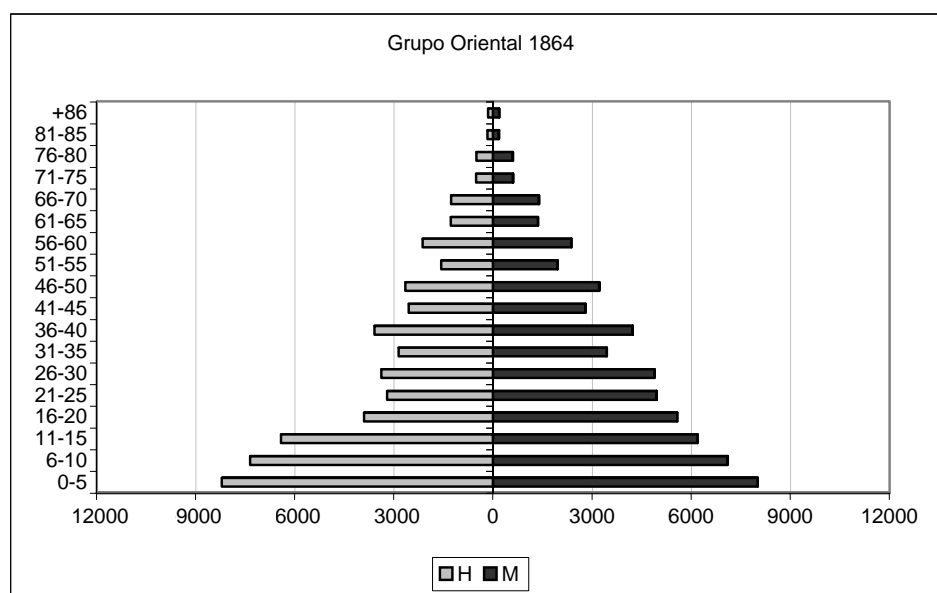


Fig. Ia

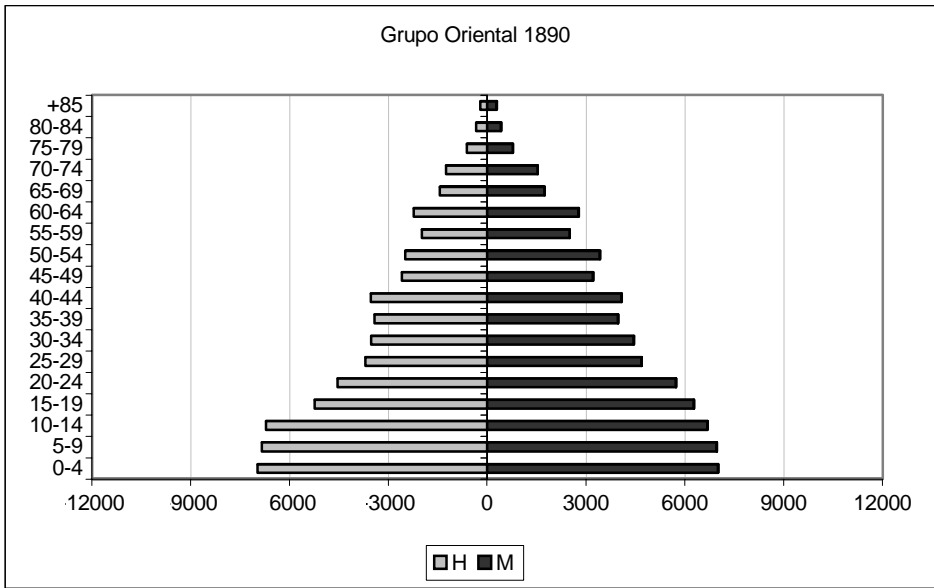


Fig. Ib

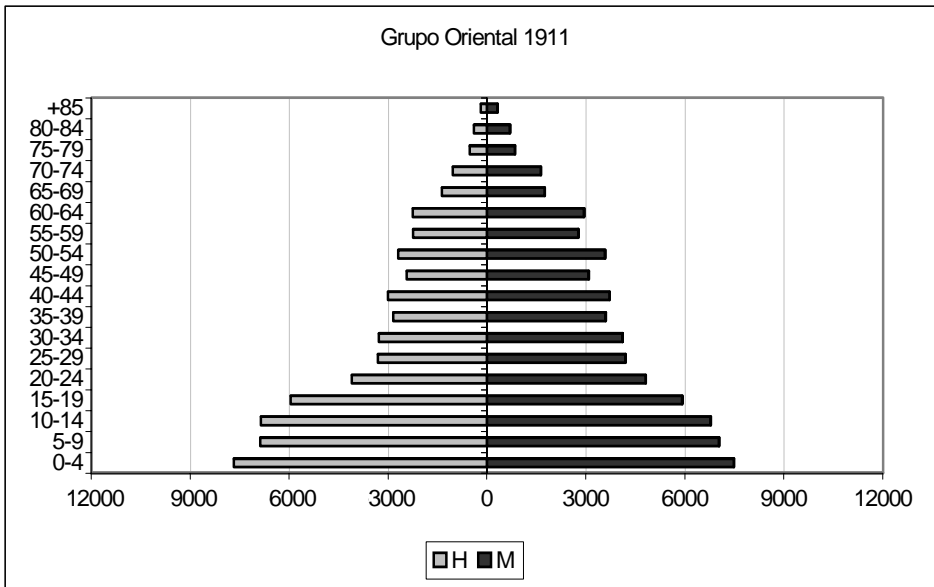


Fig. Ic

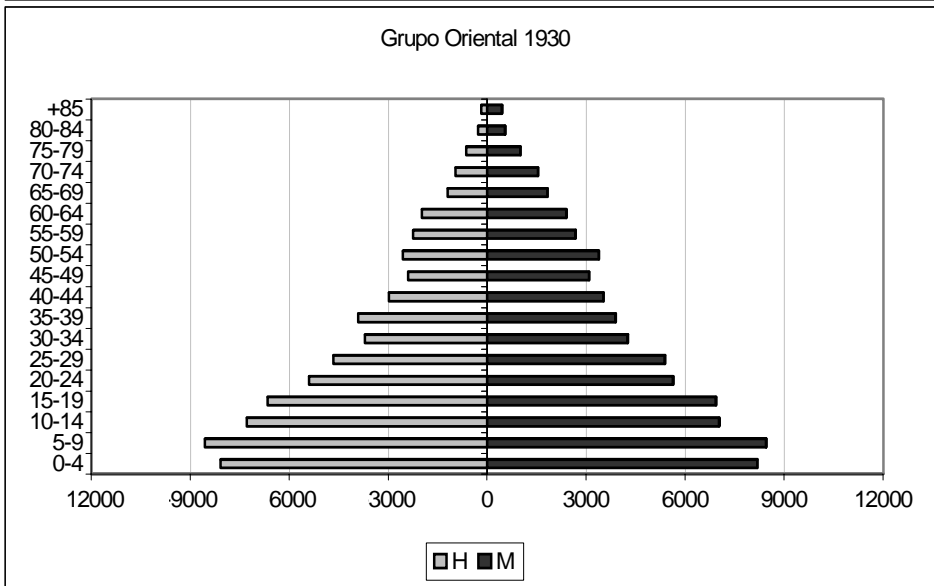


Fig. Id

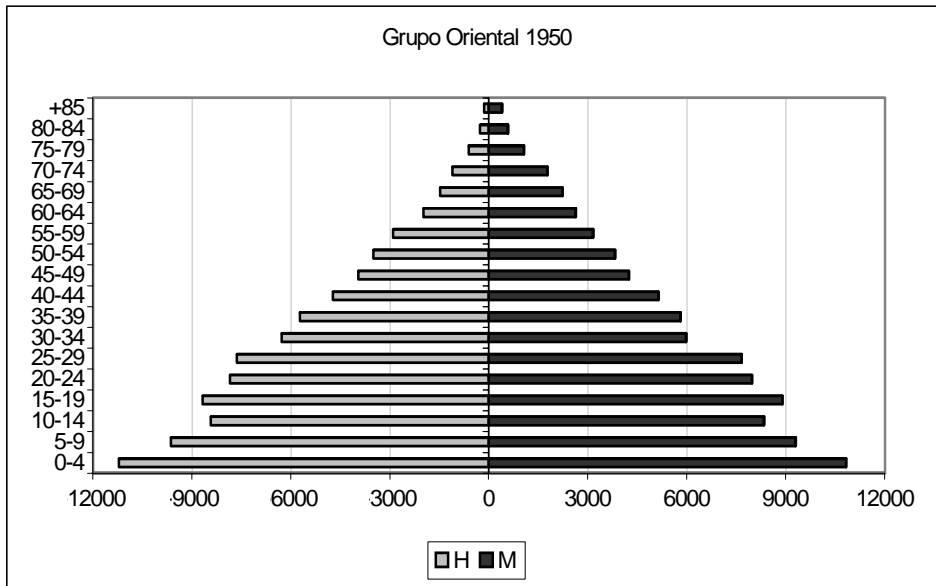


Fig. Ie

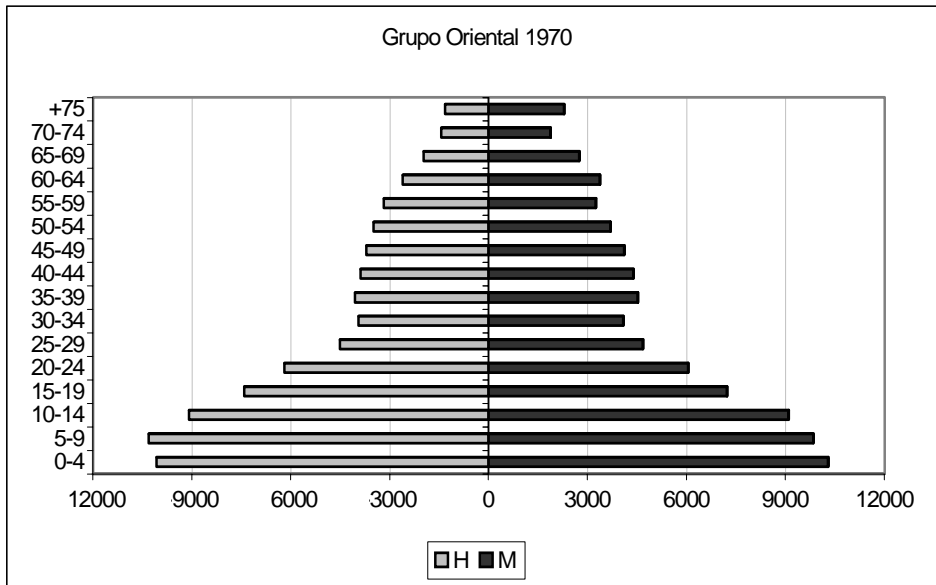


Fig. If

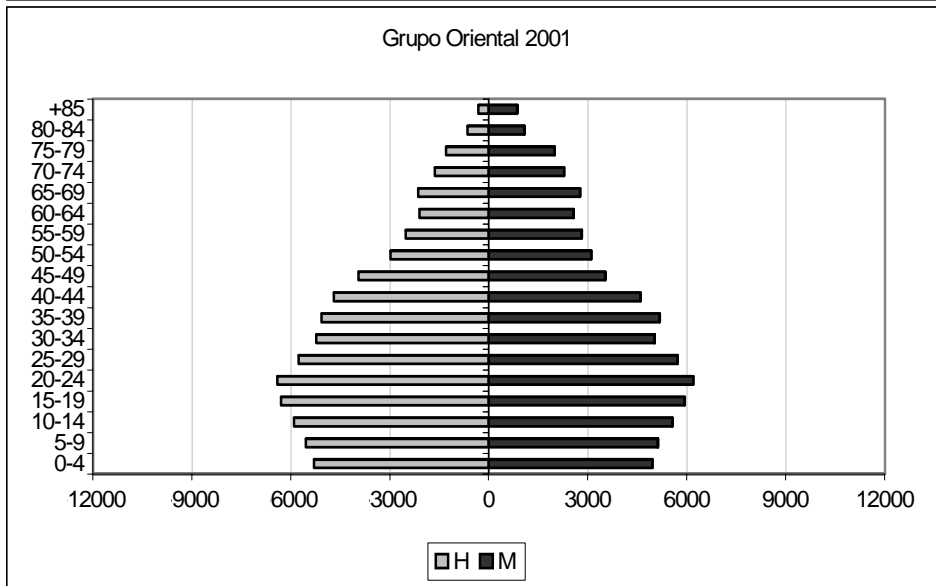


Fig. Ig

VIII.1.2. Grupo Central

Fig. II – Estructura poblacional por sexo y grupos de edad del grupo Central:

- II a) año 1864;
- II b) año 1890
- II c) año 1911
- II d) año 1930
- II e) año 1950
- II f) año 1970
- II g) año 2001

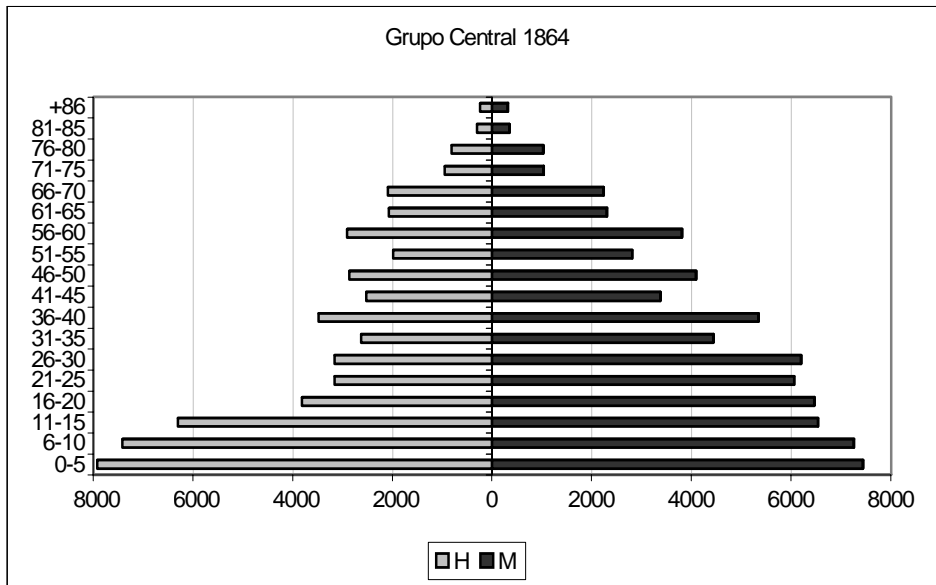


Fig. IIa

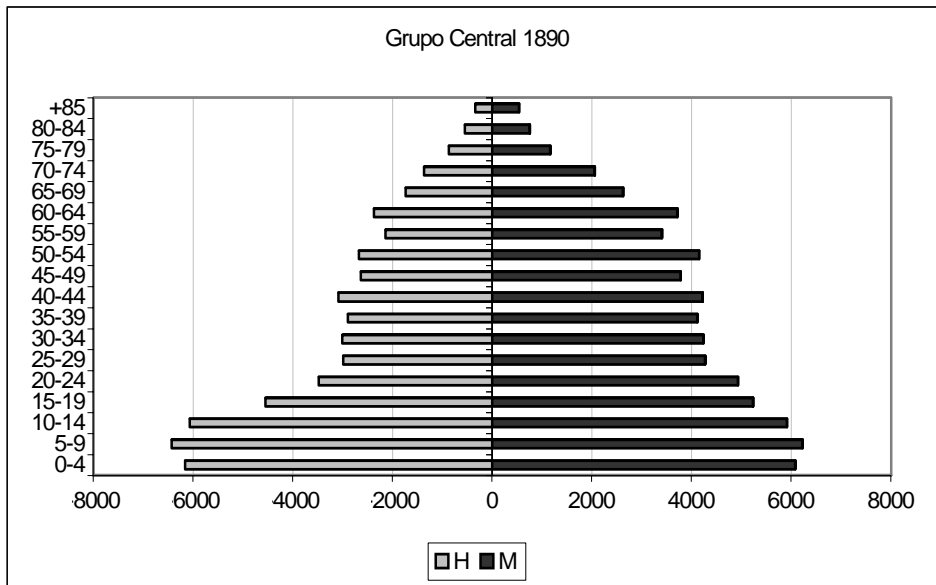


Fig. IIb

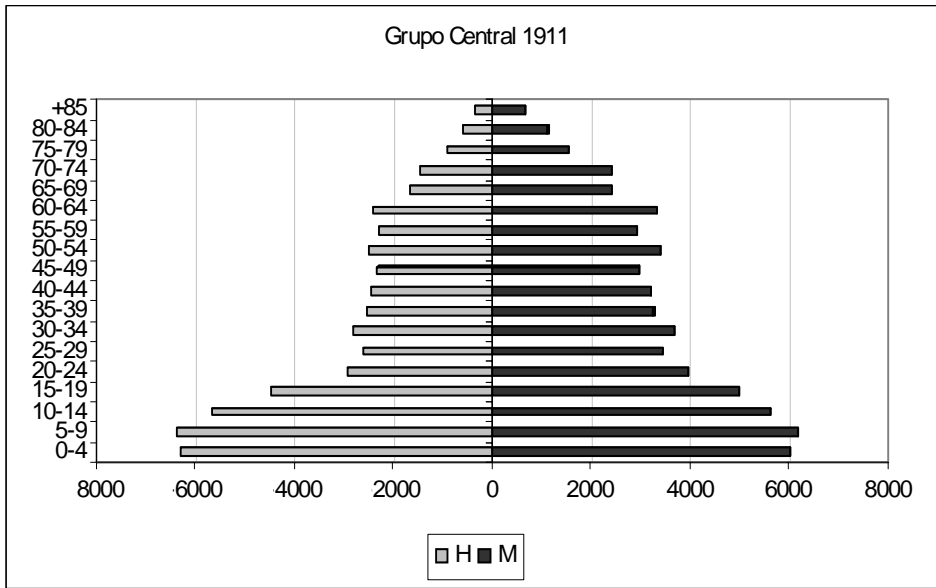


Fig. IIc

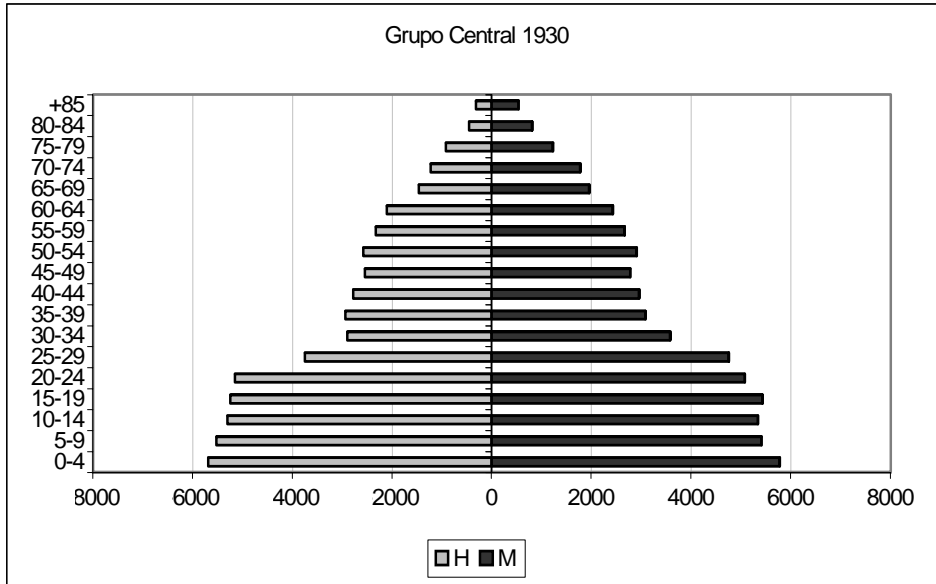


Fig. II d

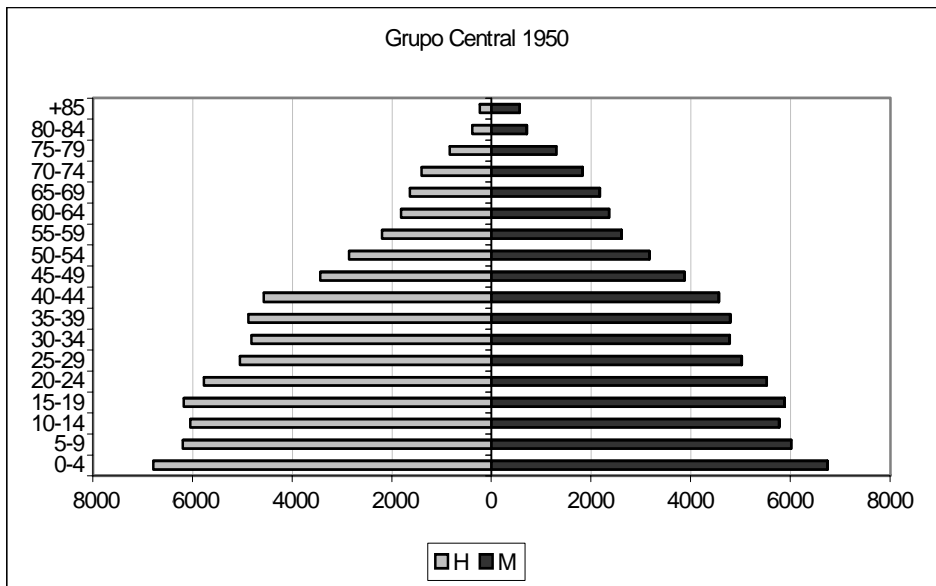


Fig. II e

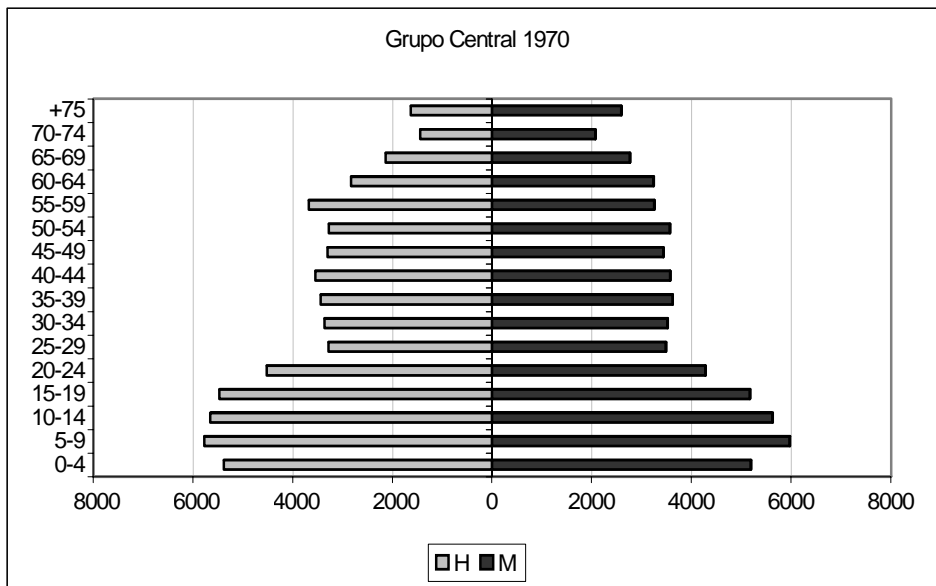


Fig. If

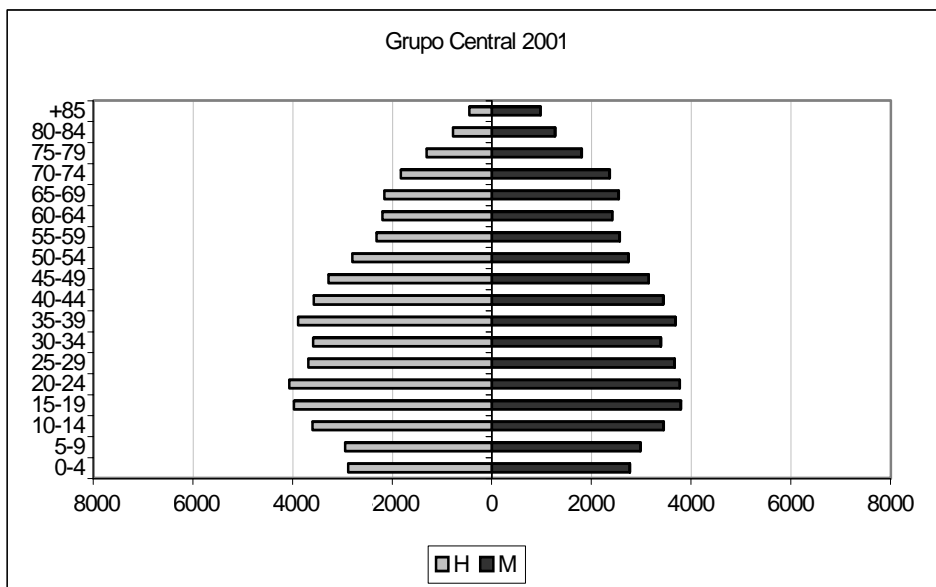


Fig. Iig

VIII.1.3. Grupo Occidental

Fig. III – Estructura poblacional por sexo y grupos de edad del grupo Occidental:

- III a) año 1864;
- III b) año 1890
- III c) año 1911
- III d) año 1930
- III e) año 1950
- III f) año 1970
- III g) año 2001

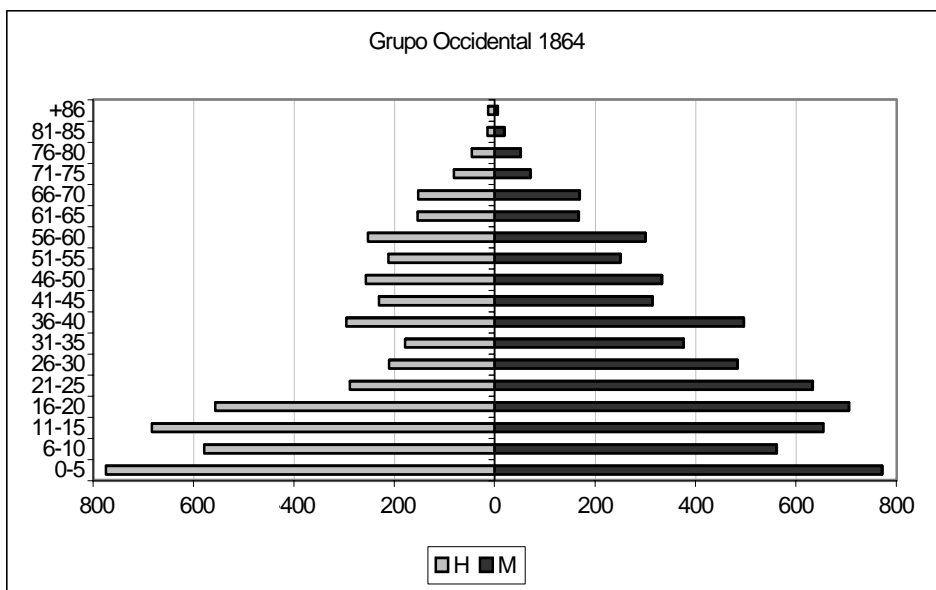


Fig. IIIa

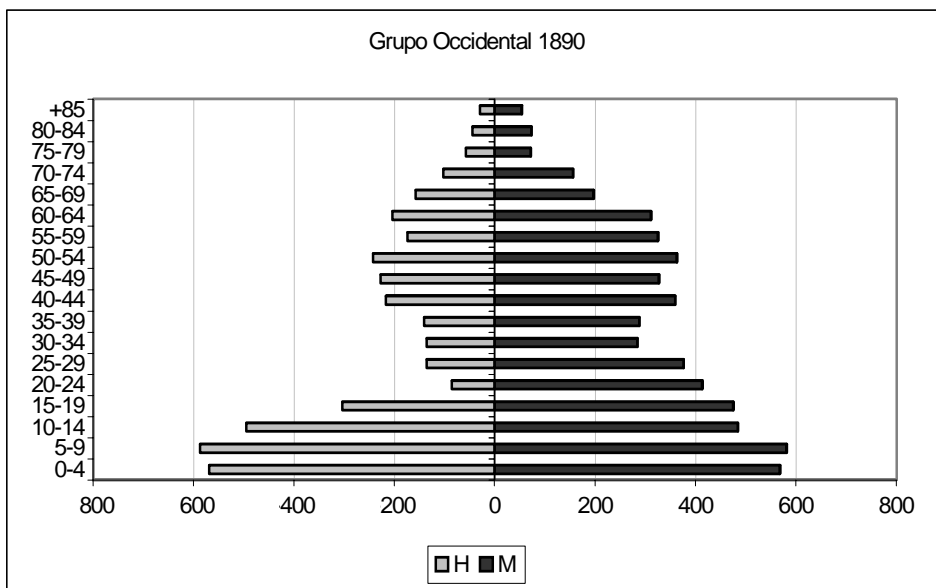


Fig. IIIb

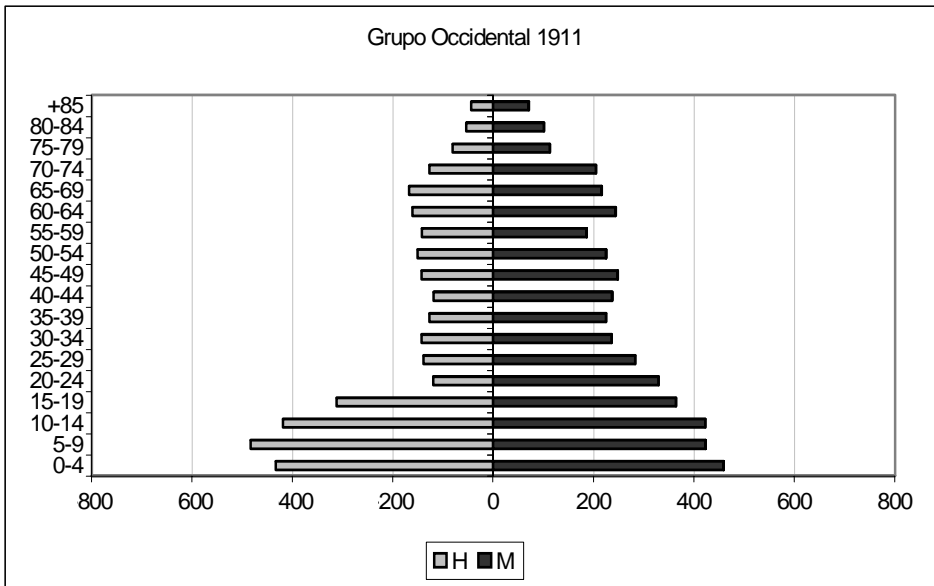


Fig. IIIc

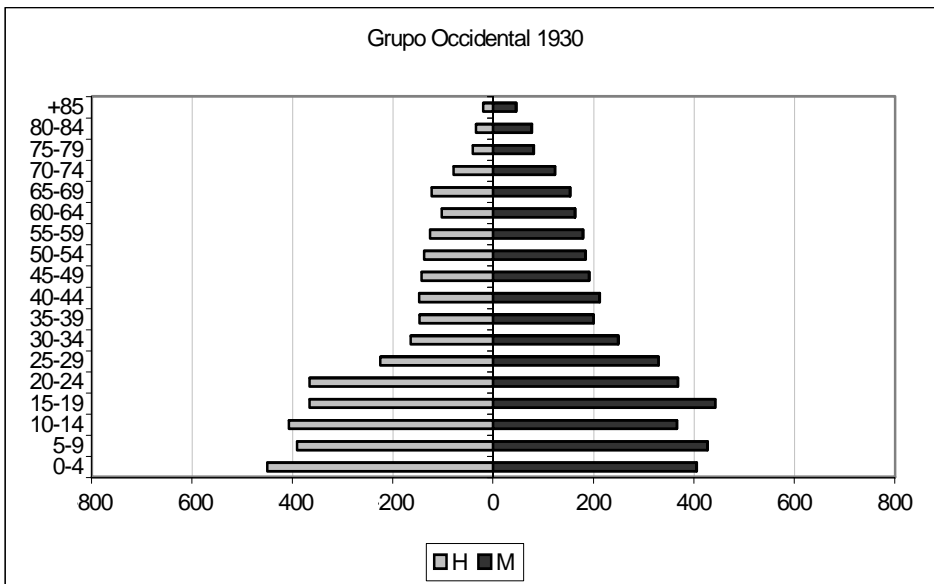


Fig. III d

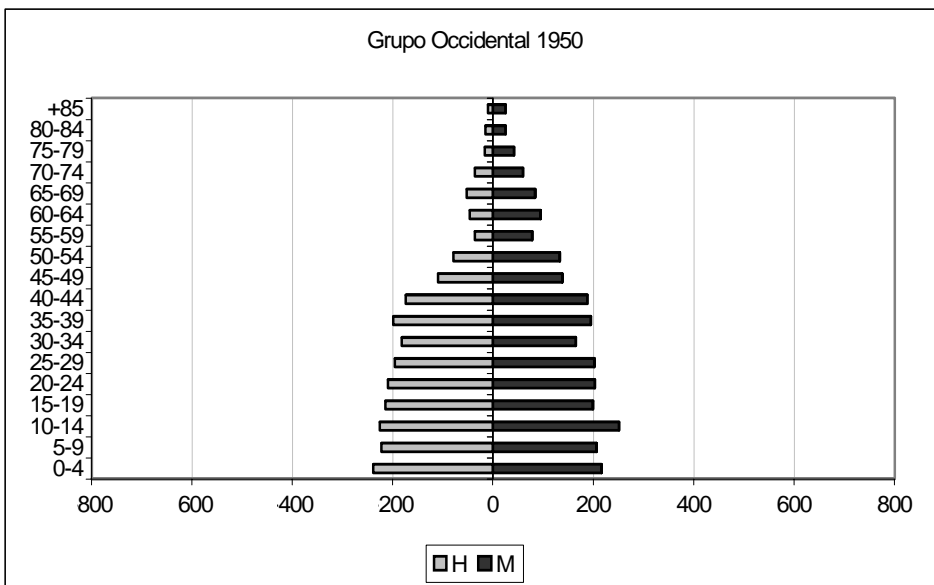


Fig. III e

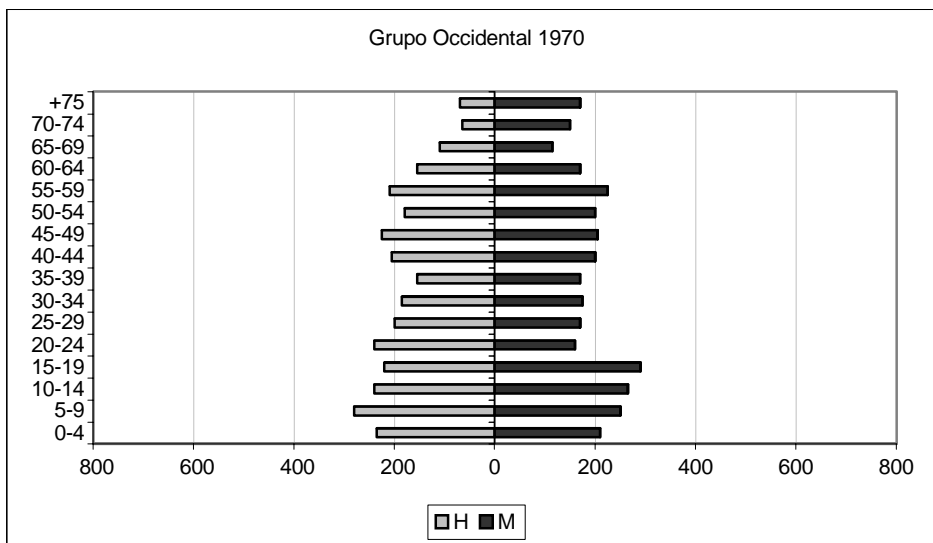


Fig. IIIf

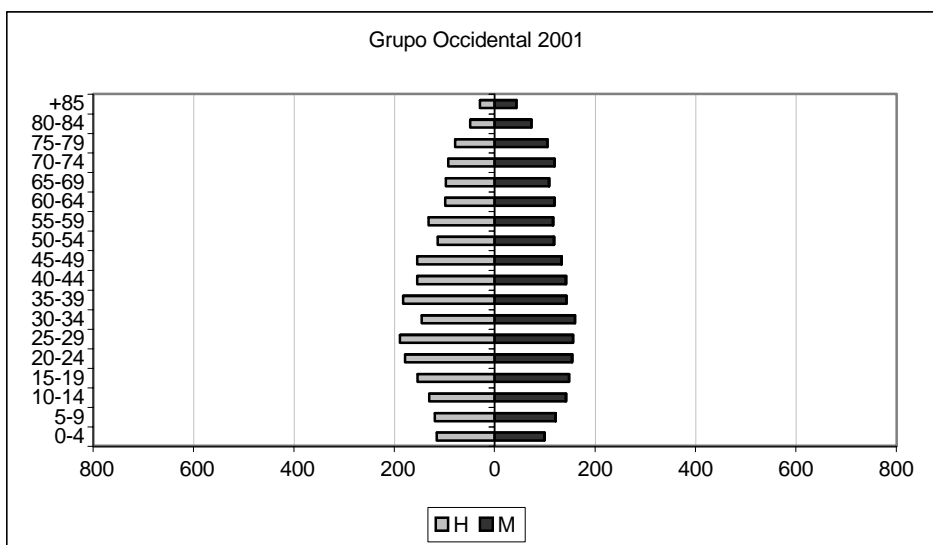


Fig. IIIg

VIII.2. ÍNDICE DE MASCULINIDAD

VIII.2.1. Índice de Masculinidad por grupos de edad

Fig. IV – Índice de Masculinidad por grupos de edad:

IV a) año 1864;

IV b) año 1890

IV c) año 1911

IV d) año 1930

IV e) año 1950

IV f) año 1970

IV g) año 2001

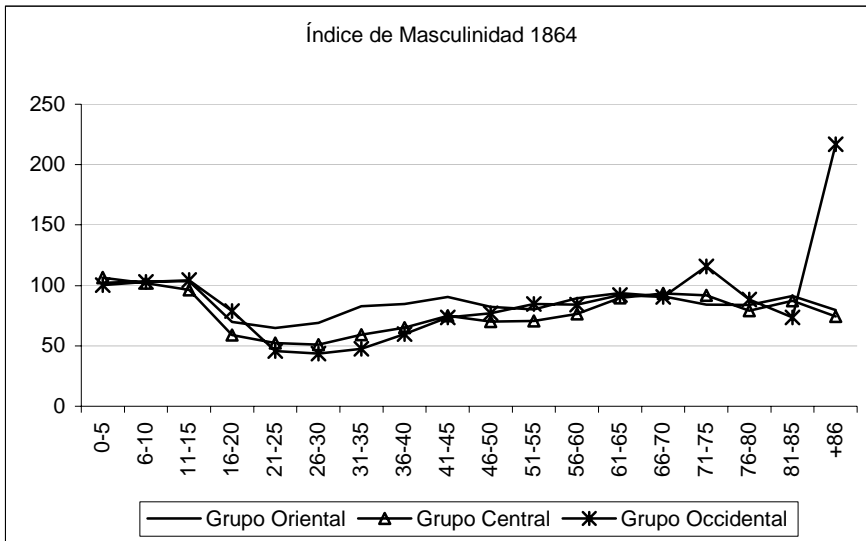


Fig. IVa

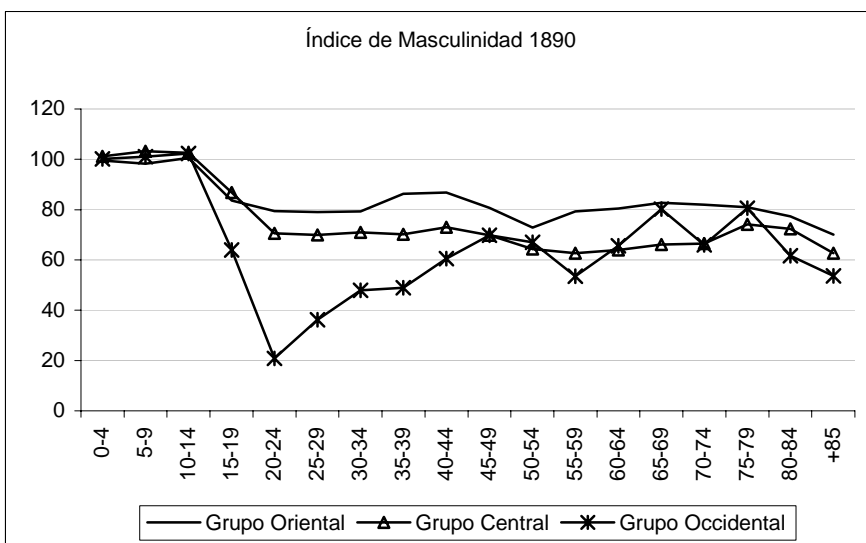


Fig. IVb

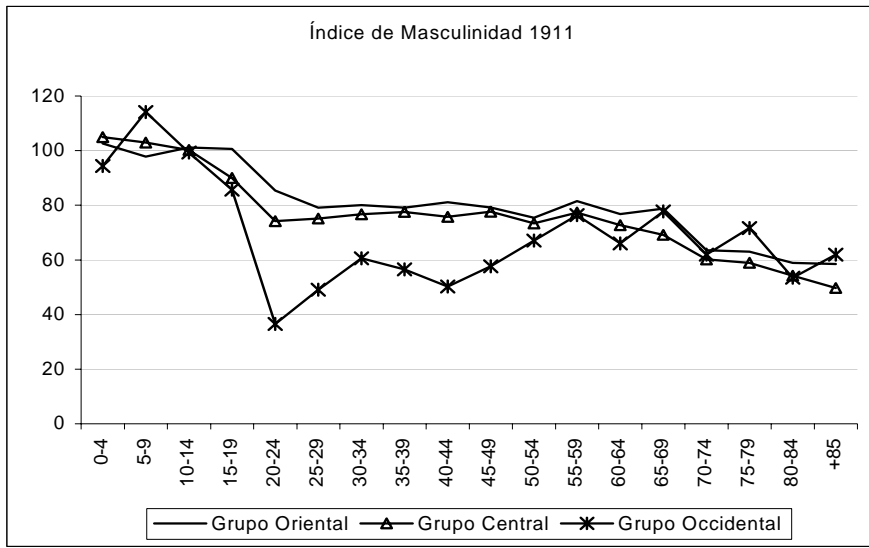


Fig. IVc

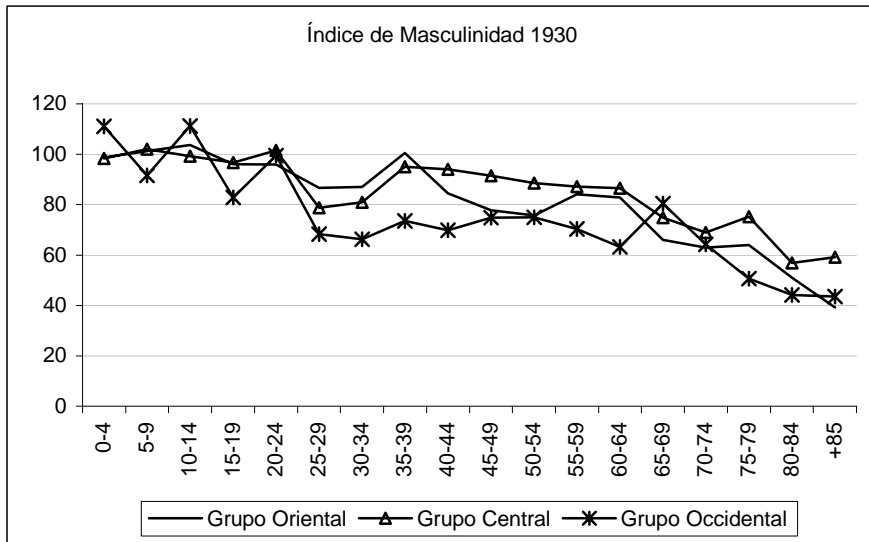


Fig. IVd

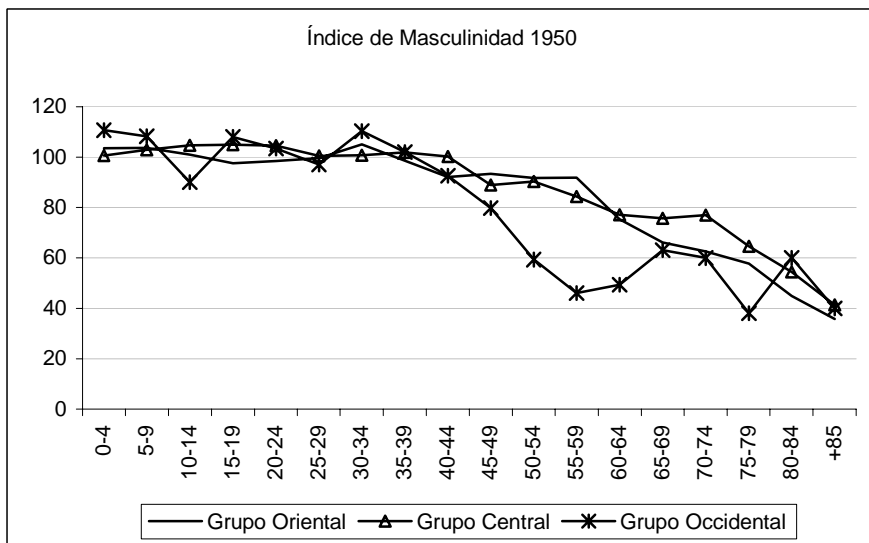


Fig. IVe

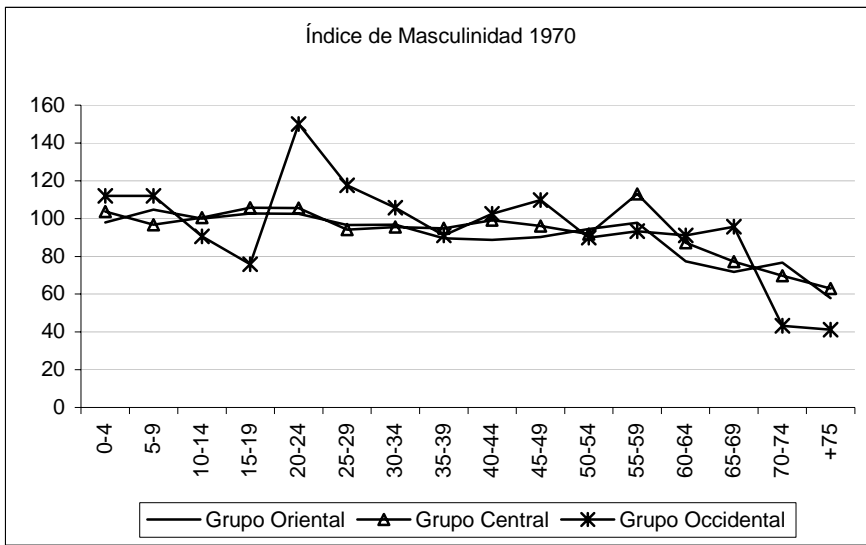


Fig. IVf

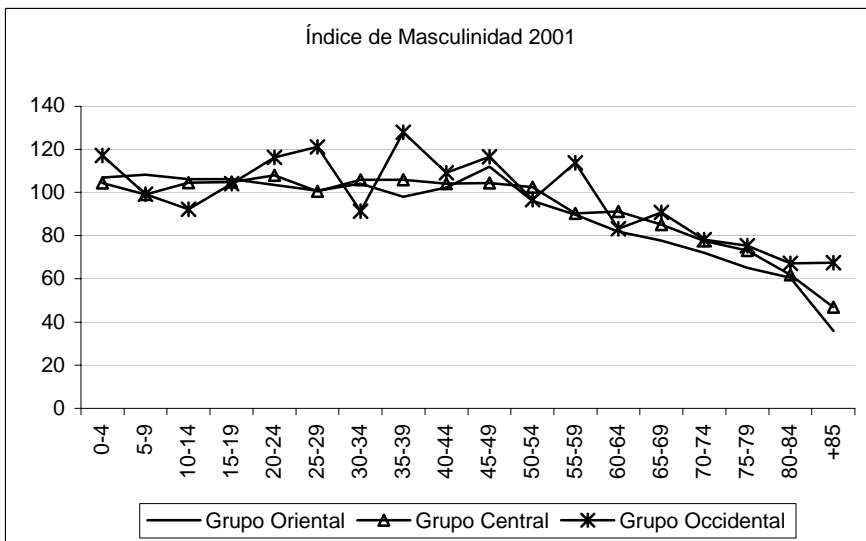


Fig. IVg

VIII.2.2. Evolución Temporal del Índice de Masculinidad

Fig. V – Evolución temporal del índice de masculinidad, para el intervalo de edades de los 15 a los 49 años, en los tres grupos de islas.

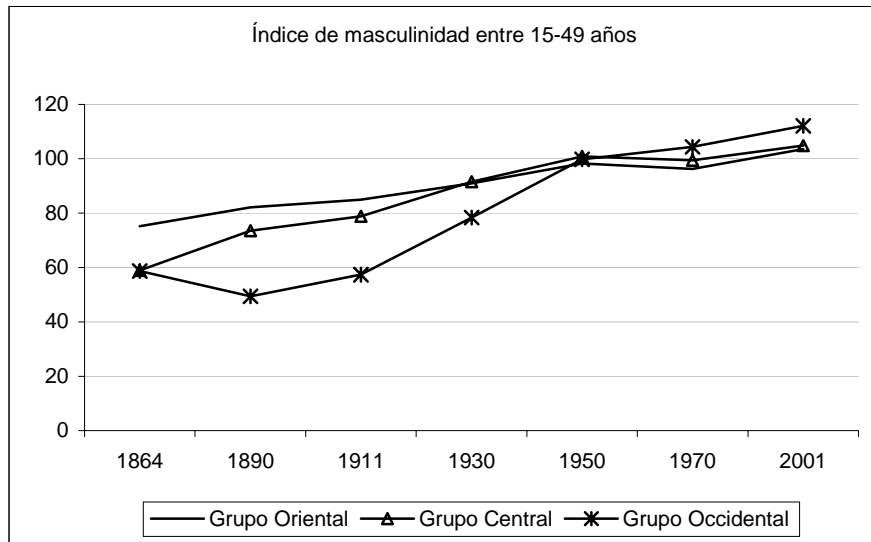


Fig. V